Statistical models for the location of lightning-caused wildfire ignitions

Nicholas Read

ORCID 0000-0002-8540-464X

Submitted in total fulfillment of the requirements of
the degree of Doctor of Philosophy

May 2018

School of Mathematics & Statistics

The University of Melbourne

Abstract

Lightning-caused wildfire is a significant concern for fire management agencies worldwide. Unlike other ignition sources, lightning fires often occur in remote and inaccessible locations making detection and suppression particularly challenging. Furthermore, individual lightning storms result in a large number of fires clustered in space and time which can overwhelm suppression efforts. Victoria, Australia, is one of the most fire prone environments in the world and the increased frequency of large-scale landscape fires over the last decade is of particular concern to local wildfire management authorities.

This thesis is concerned with modeling lightning-caused wildfire ignition locations in Victoria. Such models could be used for predicting daily lightning-caused ignition likelihood as well as simulating realistic point patterns for use in fire spread models for risk analyses.

The first half of this thesis looks at regression models. We review methods for the model selection, validation, approximation and interpretation of generalised additive models. A review of performance metrics, such as the AUC, shows the difficulties and subtleties involved in evaluating the predictive performance of models.

We apply this theory to construct a non-linear logistic regression model for lightning-caused wildfires in Victoria. The model operates on a daily time scale, with a spatial resolution of 20 km and uses covariate data including fuel moisture indices, vegetation type, a lightning potential index and weather. We develop a simple method to deconstruct model output into contributions from each of the individual covariates, allowing predictions to be explained in terms of the weather conditions driving them. Using these ideas, we discuss ranking the relative 'importance' of covariates in the model, leading to an approximating model with similar performance to the full model.

The second half of this thesis looks at point process models for lightning-caused ignitions. We introduce general theory for point processes, focusing

on the inhomogeneous Poisson process, cluster processes and replicated point patterns. The K-function is a useful summary function for describing the spatial correlation point patterns and for fitting models. We present a method for pooling multiple estimates of the K-function, such as those that arise when using replicated point patterns, intended to reduce bias.

We fit an inhomogeneous Poisson process model as well as a Thomas and Cauchy cluster process model to the Victorian lightning-caused ignition data set. The cluster process models prove to have significantly better fit than the Poisson process model, but still struggle to reproduce the complex behaviour of the physical process.

Declaration

This is to certify that:

i. the thesis comprises only my original work towards the PhD except where indicated in the Preface;

ii. due acknowledgment has been made in the text to all other material used; and,

iii. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed

Nicholas Read

Preface

This thesis is based soley on research I carried out at the School of Mathematics and Statistics, The University of Melbourne during my PhD candidature.

The bulk of the work in Chapter 5 has been published in *A lightning-caused wildfire ignition forecasting model for operational use* (Read, Duff, and Taylor, 2018). I was responsible for 90% of this work. In particular, Tom Duff advised me on constructing the simplified vegetation layers described in Section 2.1 and helped frame the work within the wildfire management community.

I am grateful to the staff at the Department of Environments, Land, Water and Planning, and in particular Andrew Mellor and Andy Ackland, who took the time to meet with me. It was very helpful to talk to end users about desirable features of models. These conversations motivated the development of the covariate contribution anomaly as a method of explaining model forecasts (Section 3.6.2).

I worked on the weighted pooled K-function (Section 7.4) with Adrian Baddeley, who had the initial idea after I found some issues with the standard estimator on the data set used in this study. We intend to publish the work surrounding the weighted pooled K-function in the paper Moradi et al. (n.d.). This paper also includes work on normalising terms for the inhomogeneous K-function (Section 6.6.3); this work was that of my coauthors.

Most of this thesis was read by Daryl Daley, who is extremely knowledgeable about point processes. Daryl provided line edits as well as suggesting some structural changes.

I thank the Bushfire and Natural Hazards Cooperative Research Centre for providing funding through the project *Probability of Fire Ignition and Escalation* and the Australian Research Council (ARC) for providing funding through the Laureate Fellowship FL130100039.

Acknowledgments

I would like to thank my supervisor, Peter Taylor, for his support throughout my lengthy stay at The University of Melbourne. His gentle encouragement during my undergraduate degree is partly responsible for my being here today. I thank him for encouraging me to travel and to work closely with others.

I thank Tom Duff for being so interested in my work and for his assistance on our paper, which would not have been possible without his insights.

I thank Adrian Baddeley for teaching me ways to think simply and clearly about point processes. I have never learnt as rapidly as I did during the month I spent working with him in Perth. Adrian was incredibly generous with his time and patiently answered many of my questions.

I thank Daryl Daley, for slogging through early drafts of my thesis and for providing valuable advice on how to write clearly and precisely.

I thank Peter Braunsteins for thanking me in his acknowledgements, but also for the good conversation shared over lunch.

Lastly, I would like to thank my parents, John, Annette and Tim, my sisters, Katherine and Sarah, and my partner, Adalya, for their ongoing support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Lightning-caused wildfire. General approaches to modelling wildfire ignition locations. An outline of the thesis.

1.1 Lightning-caused wildfire ignition

Lightning-caused wildfire, which we call *lightning fire*, is a significant concern for fire management agencies worldwide. Lightning is an important source of wildfire ignitions, with some studies attributing upwards of 40% of recorded ignitions to lightning (K. Anderson, 2002; Hall and Brown, 2006). Unlike other ignition sources, lightning fires often occur in remote and inaccessible locations making detection and suppression particularly challenging (Flannigan and Wotton, 1991; Todd and Kourtz, 1992). Furthermore, individual lightning storms result in a large number of fires clustered in space and time which can easily overwhelm suppression efforts (Podur, Martell, and Csillag, 2003). Due to these factors, lightning fires burn disproportionately large amounts of land. For example, in Ontario, Canada, roughly 40% of wildfires are lightning-caused yet they go on to burn 70% (Flannigan et al., 1991) to 81% of the total area (Wotton and Martell, 2005).

In Victoria, Australia, lightning fire was responsible for 70% of land burnt during the period from 1973 to 2014. The 2003 Alpine Fires saw a cold front creating lightning storms that started 87 fires in Victoria, which went on to

burn approximately 1.1 million ha (Stephenson, 2010). The same thunderstorms caused further fires in New South Wales and the Australian Capital Territory, which burnt a further 760,000 ha. Similarly, the 2006–07 Great Divide Fires saw 70 fires caused by a single thunderstorm, burning a cumulative 1.1 million ha. These large-scale landscape fires pose serious risks to life, property, agriculture and infrastructure. They also disrupt ecosystem services (Gill, Stephens, and Cary, 2013) such as water supply and water quality (Feikema, Sherwin, and Lane, 2013).

Lightning ignition can be understood as the product of three processes: (i) lightning strike occurrence, (ii) fire ignition given a lightning strike, and (iii) ignition survival until detection (K. Anderson, 2002). These processes are complex, involving vegetation, fuel moisture and weather conditions conducive to both fire growth and the occurrence of lightning. While there are strong correlations between 'fire weather' and lightning-caused wildfire ignition there is still much unknown about the likelihood of ignition on a given day. *Dry lightning*, defined as a lightning strike with less than 2.54 mm of accompanying precipitation (Rorig and Ferguson, 1999), is of particular interest to land managers. Dry lightning occurrence has been linked with the 850–500 hPa temperature lapse and 850 hPa dewpoint depression (Dowdy, 2015; Dowdy and Mills, 2009; Rorig et al., 1999). There is also a link to the Convective Available Potential Energy (CAPE) index, which is related to lightning occurrence more broadly. Work has looked at the relationship between elevation, vegetation and lightning strike density (Dissing and Verbyla, 2003; Kilinc and Beringer, 2007). Elevation can cause forced convection, resulting in lightning strikes, and topographic effects play an important role in defining vegetation type.

Given the occurrence of a lightning strike, fuel characteristics and fuel moisture determine if an ignition will occur and survive until detection; fuel must be present, it must be in a condition where it can ignite and there must

be a suitable degree of connectivity between flammable fuels to sustain a fire (Duff, Keane, et al., 2017). The ability of the fuel to combust is the most dynamic since it is strongly defined by the fuel moisture, which in turn is strongly defined by the moisture in the environment (Matthews, 2014). The Keetch–Byram Drought Index (KBDI) represents the moisture deficit in the top 200 mm soil layer and is calculated using the maximum air temperature, total rainfall over the past 24 hours and yesterday's KBDI (Keetch and Byram, 1968). As such, it reflects both daily- and long-term conditions. The drought factor (DF) represents the proportion of fine fuel available to burn in the event of a fire and is calculated from the KBDI and the precipitation (Griffiths, 1999). Both the KBDI and the DF play an important role in calculating the McArthur Forest Fire Danger Index (FFDI), which is the primary numerical tool used to communicate bushfire risk to the Australian public. It was originally developed to describe fire behaviour such as the rate of forward spread, flame height and spotting distance (McArthur, 1967), but it has been shown to be related to wildfire ignition (Bradstock et al., 2009).

Finally, aspects of the weather such as temperature, relative humidity, wind speed and precipitation affect the lightning ignition process. Beyond their direct involvement in calculating fuel moisture indices, they are related to lightning formation, and ignition survival and detection. For example, previous studies have found a negative relationship between wind speed and lightning ignition, postulating that this could be due to the wind dispersing smoke plumes and reducing detection likelihood (Wotton and Martell, 2005).

1.2 Models for lightning-caused wildfire ignitions

From the outset, lightning fire occurrence poses several challenges to the modeller. It is a rare event, influenced by large numbers of environmental factors, with a high degree of stochasticity caused by the "highly variable numbers of cloud-to-ground lightning strikes accompanied by very spotty rainfall" which

in turn results in a highly variable number of ignitions (Wotton and Martell, 2005).

Modelling work has been motivated by the need to forecast ignitions to assist fire managers in resource allocation and suppression efforts (Martell, Bevilacqua, and Stocks, 1989; Martell, Otukol, and Stocks, 1987; Plucinski et al., 2014; Preisler, Brillinger, et al., 2004; Preisler, Chen, et al., 2008; Preisler and Westerling, 2007; Wotton and Martell, 2005) as well as to understand the influence of fire indices on wildfire ignition (Peng, Schoenberg, and Woods, 2005; Schoenberg, Pompa, and Chang, 2008; Xu and Schoenberg, 2011), or how the length of time between burning can affect ignition likelihood (Penman, Bradstock, and Price, 2013), or the effects of climate change on ignition likelihood (Liu et al., 2012; Woolford, Cao, et al., 2010; Woolford, Dean, et al., 2014; Wotton, Martell, and Logan, 2003).

Models have often focused on either human- or lightning-caused ignition but the approaches are very similar. Early work looked at the probability of a day with one or more fires, known as a *fire day* (Haines et al., 1983; Martell, Otukol, et al., 1987; Plucinski et al., 2014). Others divided a large study region into smaller subregions and modelled the number of fires per day in each of the subregions (Plucinski et al., 2014). Explicitly spatial models have been developed, producing gridded maps of estimated daily ignition likelihood (Guo et al., 2016; Magnussen and Taylor, 2012; Preisler, Brillinger, et al., 2004; Woolford, Bellhouse, et al., 2011). Alternatively, some have modelled the probability of fire given a lightning strike has occurred, isolating the conditions suitable for ignition survival and detection and ignoring those conducive to lightning occurrence (Wotton and Martell, 2005).

A diverse range of models have been used for wildfire ignitions. The bulk of the literature is made up of *regression* models (Chapter 3), which use gridded or regional data to estimate the likelihood of the presence of ignition or the numbers of ignitions. These are well understood and relatively transparent

models with a wealth of supporting theory. Alternative models include *point process* models (Chapter 6), which use point data for ignitions and either gridded or continuous covariate data to fit a model. These models are usually used to estimate maps of the mean intensity of ignitions, although they can also be used to estimate probabilities of the number of points in arbitrary regions. Similarly to regression models, the theory is well developed and transparent, although it is less familiar to most researchers which probably explains the dominance of regression models in the literature. Finally, some work has been done using machine learning models (Oliveira et al., 2012; Ouyang, Han, and Tong, 2016; Vasconcelos et al., 2001; Vasilakos, Kalabokidis, Hatzopoulos, Kallos, et al., 2007; Vasilakos, Kalabokidis, Hatzopoulos, and Matsinos, 2008). Ouyang et al. (2016) and Vasconcelos et al. (2001) both found machine learning algorithms to have better predictive power than regression models for modelling wildfire ignitions, however both use crude, non-linear regression models making this an unfair comparison.

There are some important practical considerations when modelling wildfire ignitions, irrespective of the choice of model. As discussed in Section 1.1, wildfire ignition is a complex physical process and there are a large number of covariates that could be influential. This makes model selection, and in particular, the problem of which covariates to include in a model, an important component of modelling wildfire ignition location. Along with issues of parsimony, the need to obtain data for a large number of covariates before making predictions can make models unwieldy for operational use. Another practical consideration, which came to our attention during personal communication with firefighting authorities, is the need to explain model output in terms of the environmental conditions driving it. This was seen as important because it allows authorities to integrate the model output with their intuitive understanding of lightning ignition likelihood and gives them the ability to 'sanity check' the model forecasts.

1.3 Thesis outline

This thesis is interested in modelling lightning-caused wildfire igntions and is split in two halves; the first half explores the use of regression models and the second half explores the use of point process models. Each half is broken into chapters explaining relevant theory and finishing with a chapter that applies the theory to model lightning ignition locations. We are interested in understanding the practical outcomes these models can provide and spend a lot of time validating models in order to understand the differences between them.

Chapter 2 introduces the data and gives a preliminary analysis of the ignition location data set. In Chapter 3 we give an introduction to regression models with a focus on model selection, validation and interpretation. Performance classifiers are often used to measure the accuracy of regression models and in Chapter 4 we discuss surrounding theory and give some critiques and recommendations. In Chapter 5 we apply the regression theory to model lightning ignitions, using a logistic regression model. The second half of the thesis begins with Chapters 6 and 7, which introduce the theory of point processes including inference and validation. This theory is then applied to model lightning ignitions in Chapter 8.

Chapter 2

Data and preliminary analysis

Introducing the data. A preliminary analysis of the wildfire ignitions data set.

2.1 Data

Victoria is a state in southeastern Australia with an area of 227,000 km^2 and a Mediterranean type climate (Duff, Chong, et al., 2014) with dry summers and wet winters. Figure 2.1 shows a simple map of Victoria, with the capital city, Melbourne, on the southern coast. Victoria shares its northern border with the state of New South Wales and its western border with the state of South Australia.

We used the combined Victorian Department of Environment, Land, Water and Planning (DELWP) / Country Fire Authority (CFA) Bushfire Ignitions data set for ignition point locations and their associated causes. This data set is the combination of four different fire records, maintained by the two firefighting agencies in Victoria:

1. DELWP bushfire ignitions point database (1972–2014);

2. DELWP planned burn ignitions point database (2001–2014);

3. DELWP bushfire perimeters polygon databse (1903–2014); and,

Figure 2.1: *A map of Victoria (VIC), New South Wales (NSW), South Australia (SA) and the Australian Capital Territory (ACT), Australia. Produced using* ggmap *(Kahle and Wickham, 2013).*

4. CFA Fire Incident Reporting System (1997–2014).

The CFA is responsible for recording ignitions across a significant proportion of Victoria and so the combined DELWP/CFA bushfire ignitions data set should be considered reasonably complete only for the years following 1997. The data set has recently been quality checked under the DELWP project Bushfire Spatial Data Models and Ignition Data (Kilinc, 2017) and we restricted our study to ignition points that were considered 'good quality' by this analysis.

Ignition data on lightning-caused wildfire may be prone to errors from *sleeper* ignitions, those that remain undetected for days before escalating when conditions become more favourable for fire spread (K. Anderson, 2002; Otway et al., 2008). Although we restricted our study to 'good quality' records where there is a high degree of certainty of their cause, these records may still underestimate the true occurrence of lightning ignitions as sleeper fires could be misattributed to other causes. Furthermore, it may cause errors in the recorded time of ignition. In Australian systems, there has been limited investigation into the occurrence of sleeper fires. Evidence suggests that while such fires occur, the majority become apparent soon after ignition (Dowdy and

Mills, 2009). Australian sclerophyllous systems typically have shallow soils and rarely develop the deep duff layers (Costermans, 2009) where smouldering ignitions are usually sustained (K. Anderson, 2002). The lack of these fuels may limit the occurrence of sleeper fires relative to other localities, although there is no empirical evidence to support this.

To characterise the vegetation we used the Victorian Fuel Layer, provided to us by DELWP in June 2015, which is used operationally for fire simulation modelling (Paterson and Chong, 2011). This data set has 37 categories of vegetation type but, to avoid overfitting, we combined these into broader categories (wet forest, dry forest, grassland, Mallee spinifex, heath, plantation and residential) that we expect to have similar fire behaviour (Read et al., 2018). Information on the combined fuel layers can be found in Table 2.1 and they are plotted in Figure 2.2.

We sourced the maximum temperature, 3 pm vapour pressure and the daily precipitation totals from the Australian Water Availability Project (Raupach, Briggs, et al., 2009; Raupach, Haverd, et al., 2015). The 3 pm relative humidity was calculated from the 3 pm vapour pressure as

$$RH = 100 \left(\frac{VP}{VP_s} \right),$$

where VP is the 3 pm vapour pressure and VP_s is the saturated vapour pressure, both in hPa. The saturated vapour pressure was calculated as

$$VP_s = 6.11 \exp \left(\frac{17.27T}{T + 237.3} \right),$$

where T is the maximum daily temperature in degrees Celsius (Abtew and Melesse, 2013, equation 5.1).

The DF, KBDI, wind speed and FFDI were sourced from the Fire Weather Climatology for Victoria data set (Brown et al., 2016). The version we were provided with included data for the years 1972 to 2012 inclusive. Both the wind speed and FFDI were hourly measurements; we calculated the daily average wind speed and the daily maximum FFDI. Preliminary analysis showed

that the average FFDI was a poor indicator of wildfire ignitions due to large variations between day and night values.

We acquired the 500 hPa and 850 hPa temperature, 850 hPa relative humidity and the CAPE index from the from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim archive (Dee et al., 2011). The temperature and relative humidity data are 12 pm analyses (observed) values. The CAPE index data are 12 h forecast values using 0 am analyses values and so they represent the conditions at 12 pm. We manually calculated the 850–500 hPa temperature lapse, which is simply the difference between the temperatures at the the two atmospheric levels. The 850 hPa dewpoint depression was calculated using the R package *weathermetrics* 1.2.2 (G. B. Anderson, Bell, and Peng, 2013) and is the difference between the temperature, and the temperature at which water vapour condenses, both at the atmospheric level of 850 hPa.

The elevation was taken from the Vicmap Elevation DTM 20 m (Department of Environment, Land, Water & Planning, 2008).

Information on all the covariate data sets considered in this thesis, such as the spatial resolution, can be found in Table 2.2.

Simplified fuel type	**Behaviour**	**Victorian Fuel Layer values**
No fuel and uncommon	These fuel types constitute less than 3% of the area of Victoria and are ignored by the model.	0, 1, 2, 3
Grasslands	Natural grasslands, pastures and crops, possibly with scattered trees. The majority of the fine fuel consists of grasses.	4, 9, 16, 20, 29, 37, 43, 46
Dry forest	Woodlands or forests with tree cover of over 50%. A high cover and diversity of herbs and grasses in the ground layer.	5, 6, 7, 8, 18, 19, 22
Wet forest	Dense, tall eucalypt overstorey, typically over 30 m tall with scattered understorey trees. Typically has mesic understorey of broad-leaved shrubs and ferns.	11, 12, 13, 14, 15, 97
Heathland	Dense shrublands, up to 3 m in height, supporting a diverse array of narrow or ericoid-leaved shrubs.	10, 17, 21, 23, 24, 25
Mallee spinifex	Low woodland, up to 7 m tall, consisting of multi-stemmed Eucalyptus trees. Occurs on infertile siliceous sands on minor dunes and undulating sand plains. Has some understory species, but discontinuous surface fuels.	26, 27, 28
Plantation	Monoculture tree plantations for timber or paper pulp production. Possibly Eucalyptus or exotic conifers. Has a high density overstorey and minimal understorey vegetation.	98, 99
Residential	Residential environments consisting of highly variable fragmented fuels, including gardens, sports fields and grasslands. Has a high density of structures and fuel continuity is broken up by the presence of roads.	47

Table 2.1: The simplified fuel types. The Victorian Fuel Layer values are the corresponding fuel layers in the original data set.

Data set name	**Spatial resolution**	**Covariate**	**Units**
Climate Maps - Australian Water Availability Project	$0.05° \approx 5$ km	Maximum temperature	$C°$
		3 pm relative humidity	hPa
		Precipitation	mm
Fire Weather Climatology for Victoria	4 km	Drought Factor (DF)	
		Keetch-Byram Moisture Code (KBDI)	
		Average wind speed	kn
		Maximum Forest Fire Danger Index (FFDI)	
European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim archive	$0.7° \approx 70$ km	850–500 hPa temperature lapse	$C°$
		850 hPa dewpoint depression	$C°$
		Convective Available Potential Energy (CAPE) index	
Vicmap Elevation DTM 20m	20 m	Elevation	m
Victorian Fuel Layer	30 m	Wet forest	
		Grassland	
		Mallee Spinifex	
		Heath	
		Plantation	
		Residential	

Table 2.2: Information on the covariates considered in this study.

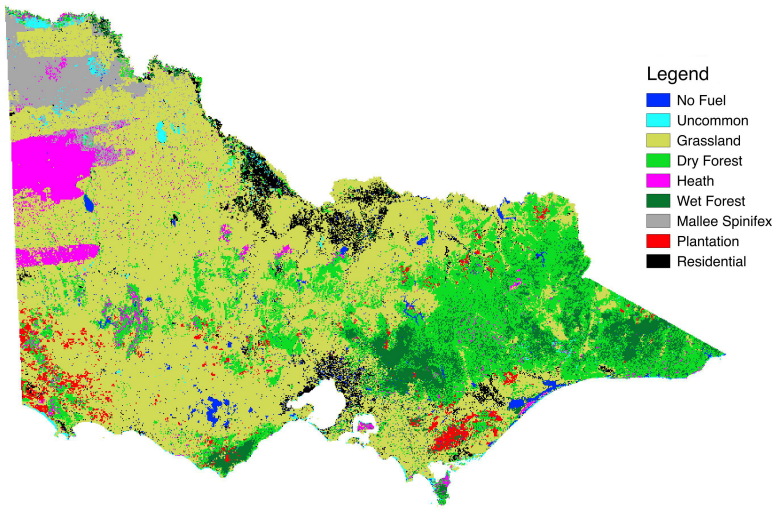


Figure 2.2: *The Victorian Fuel Layer after being reduced to 9 categories.*

2.2 Preliminary analysis of wildfire ignition data

The availability of data limited the modelling period to the years 1997–2012. Figure 2.3 shows the total number of ignitions by month and the total final area burnt by ignitions per month, stratified by ignition cause. Unsurprisingly, there is significant seasonal behaviour in the number and severity of ignitions. All models used in this thesis are *stationary*, depending on time only through the values of the covariates. Motivated by this as well as a desire to reduce the volume of data, we restricted our model to the summer months of October through April, which we denote as the *fire season*.

Figure 2.3 shows that the fire season captures the ignition process of interest to land managers; 82.1% of all ignitions and 98.6% of lightning ignitions occurred during a fire season. Ignitions that occurred during a fire season were responsible for 98.6% of the total area burnt, which jumps to 99.9% if restricted to lightning ignitions. The few ignitions that fell outside a fire season were relatively inconsequential in terms of risk management. Very little is lost by restricting our study to the fire season period of October–April.

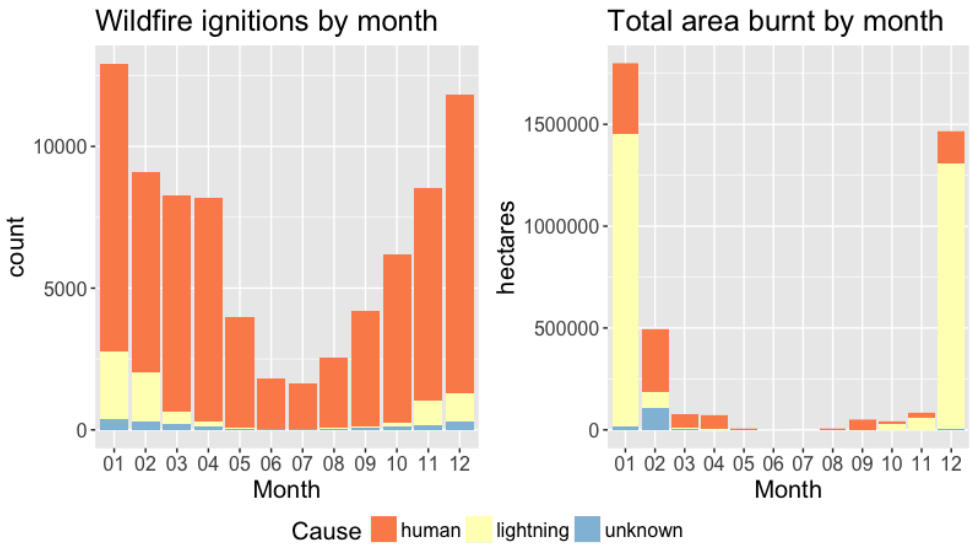


Figure 2.3: *Calculated using all ignitions between 1997 and 2012. The total area burnt by month is the total area burnt by all fires whose ignition occurred during said month.*

Most wildfire ignitions are a result of human activity: humans were responsible for 89.2% of all ignitions, lightning was responsible for 8.56% and 2.29% were recorded as having unkown cause. However, lightning fires were generally much more severe: human-caused fires were responsible for only 25.4% of the total area burnt by wildfires in the study period, whilst lightning-caused fires were responsible for 71.3% and fires with unknown cause were responsible for 3.28%.

As visualised in Figure 2.4, the total number of ignitions, and particularly lightning ignitions, varied significantly between fire seasons. The 2010–2011 fire season had anomalously few ignitions. This fire season corresponded with one of the strongest La Niña events on record, with very high rainfall and low temperatures across Australia. The effect was predicted to be so strong that in October 2010 the Bureau of Meteorology briefed firefighting agencies that there would be a decreased bushfire potential for the 2010–2011 bushfire season (Meteorology, 2012).

Focusing on lightning ignitions, Figure 2.5(a) shows high variability in the

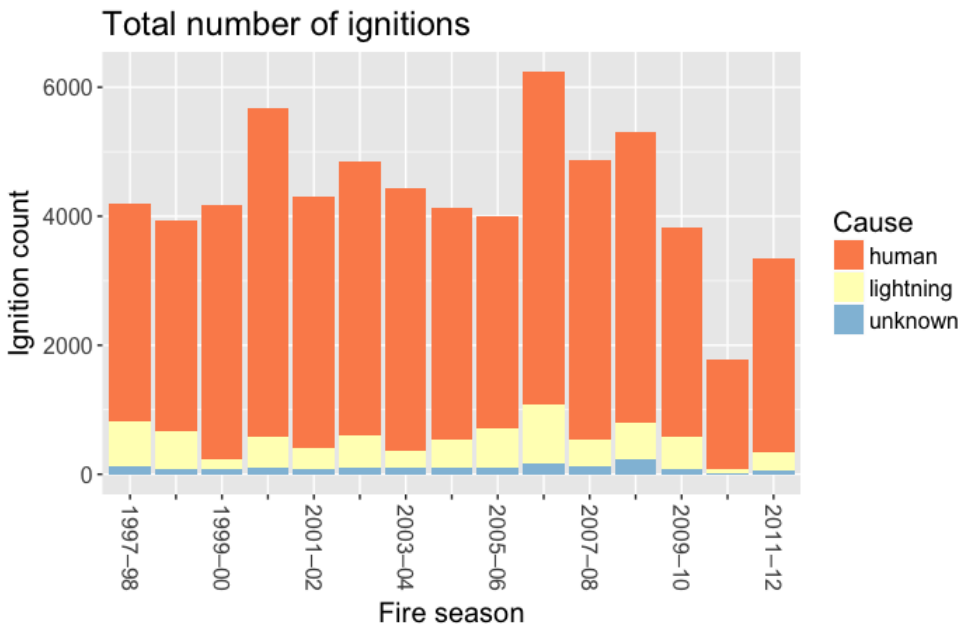


Figure 2.4

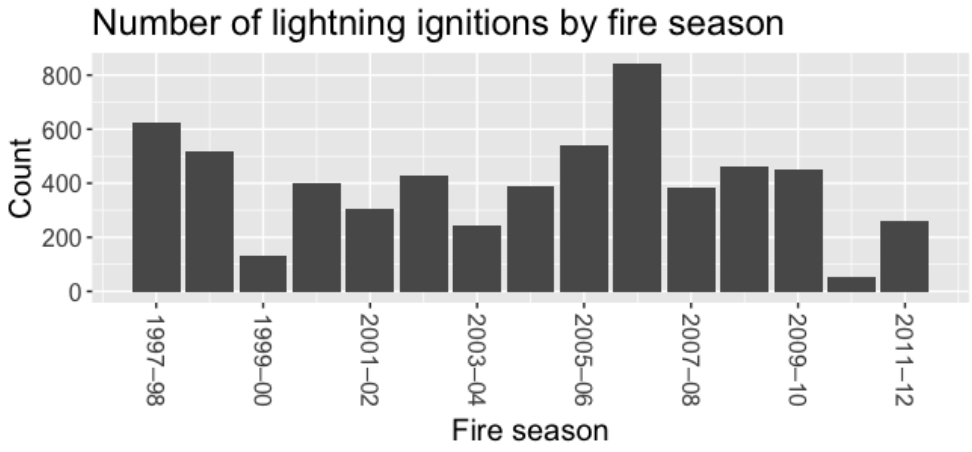
number of lightning ignitions per fire season, with a maximum of 924 lightning ignitions, attained by the 2006–2007 fire season, and a minimum of 52 lightning ignitions, attained by the 2010–2011 fire season. Even ignoring the La Niña affected 2010–2011 fire season, the minimum number of lightning-ignitions was 145, occurring during the 1999–2000 fire season.

Large scale convective storm events can result in a large number of ignitions on a single day and these events were responsible for a significant proportion of the variability in the total number of lightning ignitions between fire season. Figure 2.5(b) shows the maximum number of lightning ignitions on a single day by fire season, which which reflects the worst single day fire event from each fire season. These events were responsible for a significant proportion of the total number of ignitions in a fire season. For example, the 1997–1998 fire season saw 130 lightning fires in a single day which were responsible for 21% of the total lightning ignitions for that fire season. The scale of these single day events suggests a significant clustering effect between lightning ignitions.

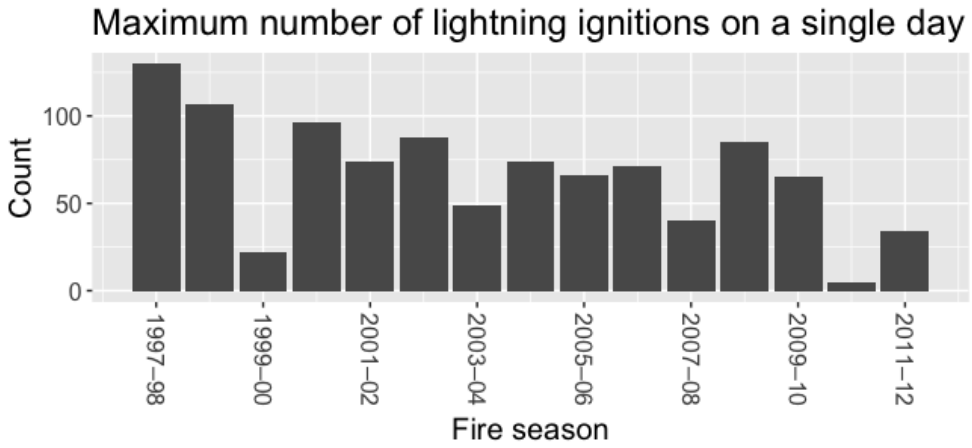
Figure 2.5(c) shows the number of *lightning-fire days*, defined as days with one or more lightning ignitions. Unsurprisingly, we see that fire seasons with a large number of lightning ignitions also had a large number of lightning-fire days. However the variability in the number of lightning-fire days between the fire seasons is less stark than the variablility in the total number of lightning ignitions. The extremely inactive fire season of 2010–2011 saw only 52 lightning ignitions yet still had 33 lightning-fire days, whilst the extremely active fire season of 2006–2007 saw 924 lightning-ignitions across 80 lightning-fire days. Again, this reflects the clustering behaviour of lightning ignitions.

Figure 2.6 shows a histogram of the number of lightning ignitions on a day given that there is at least one lightning ignition. This shows that whilst most lightning-fire days see very few ignitions, extreme observations of over a hundred ignitions play a significant role.

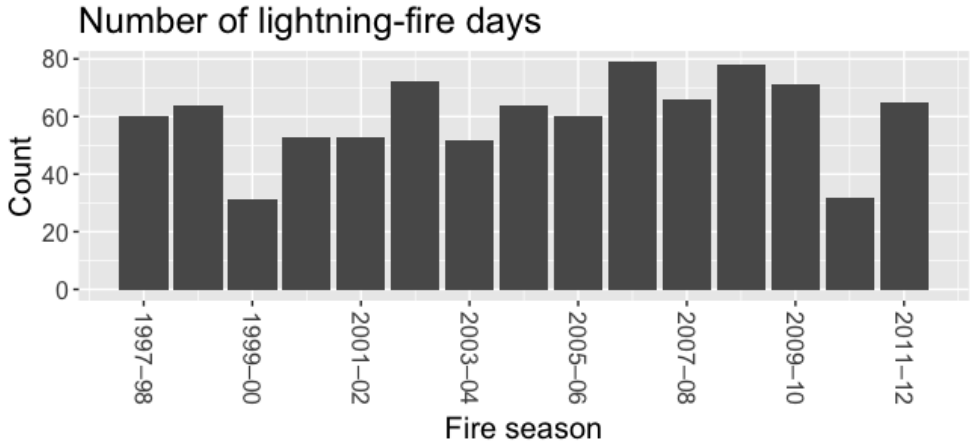
Preliminary analysis of the lightning ignition point data showed that most the lightning ignition activity occurs during the fire season, supporting the focus on these periods. The analysis showed that the lightning ignition process is highly stochastic, with large variability in the number of ignitions between fire seasons. This variability often came down to single day events, where landscape scale convective storms produced large numbers of ignitions.



(a)



(b)



(c)

Figure 2.5: *The lightning ignition data stratified by fire season (October–April inclusive)*

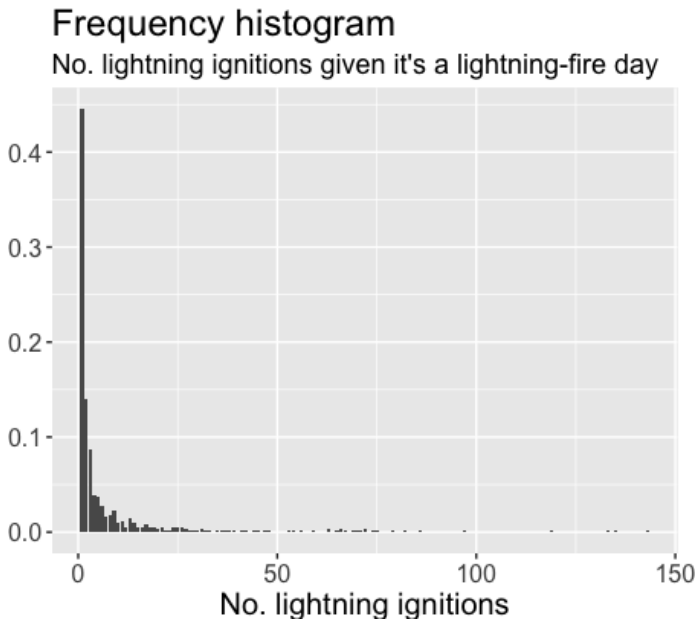


Figure 2.6: *A frequency histogram for the number of lightning ignitions given there is at least one lightning ignition. The distribution appears heavy tailed, with the bulk of the distribution clustered close to 0.*

Chapter 3

Regression models

Regression models, in particular Generalised Linear Models and Generalised Additive Models. Model selection using Purposeful Selection. Reducing model complexity by approximating models. Validation techniques for logistic regression. Interpreting fitted models in terms of the influence of covariates. Deconstructing model output in terms of contributions from the individual covariates.

3.1 Basic theory

In this section we introduce the basic structure of regression models, discussing generalised linear models as well as their non-linear generalisation. We pay particular attention to logistic and Poisson regression models and their spatio-temporal constructions since these form the backbone of the wildfire ignition modelling literature. We follow Harrell (2015) and Wood (2017) for general regression theory and Hosmer, Lemeshow, and Sturdivant (2013) for specifics of logistic regression.

3.1.1 Generalised linear models

Generalised linear models (GLMs) are flexible mechanisms for modelling relationships between variables. The variable of interest, $\mathbf{Y}$, known as the *response variable*, is understood through its relationship to a collection of *covariates*

(or *predictor* or *independent* variables), denoted $Z_1, Z_2, \ldots, Z_d$. The response and predictor variables can be either continuous or discrete or a mixture of both. Harrell (2015) defines a GLM by two objects:

1. $C(Y \mid \mathbf{Z})$, a property of the distribution of Y conditioned on the value of the covariates, $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)$; and,

2. the *link function*, $g(\cdot)$, which relates the covariates to the response variable by the relationship

$$g\left(C(Y \mid \mathbf{Z})\right) = \beta_0 + \sum_{j=1}^{d} \beta_j Z_j, \tag{3.1}$$

   where $\beta_0, \beta_1, \ldots, \beta_d$ are real-valued coefficients, known as the *regression coefficients*. We refer to the right-hand side of (3.1) as the *predictor term*.

The predictor term is a linear sum of the covariates, hence the name GLM. *Simple* linear regression is an example of a GLM, where $C(Y \mid \mathbf{Z}) = \mathbb{E}(Y \mid \mathbf{Z})$ and the link function is the identity function, giving the well known relationship $\mathbb{E}(Y \mid \mathbf{Z}) = \beta_0 + \sum_{j=1}^{d} \beta_j Z_j$.

A standard setup for a regression model is to have repeated observations of the response and predictor variables, perhaps corresponding to different experimental units or different individuals in a population. We denote Y_i as the i^{th} observation of the response variable and $\mathbf{Z}(i) = (Z_1(i), Z_2(i), \ldots, Z_d(i))$ as the i^{th} observation of the vector of covariates.

Regression models can also used for modelling spatial processes if the values of the response and predictor variables are available on a *regular grid*, which is an equally spaced grid, so that each cell is a square. In this setting the i^{th} observations, Y_i and $\mathbf{Z}(i)$, correspond to the i^{th} grid cell in the regular grid. For convenience, denote $i = 1$ for the first grid cell and $i = n_{\text{cells}}$ for the final grid cell. The regularity of the grid ensures a meaningful comparison between the observations, since these are often tied to the area of the cells.

This can be extended to spatio-temporal processes, where we have repeated observations of a spatial process. The first n_{cells} values of the index i correspond to the grid cells from the first time unit. The next n_{cells} values of the index i correspond to the grid cells from the second time unit and so on. For simplicity we refer to the time dimension in terms of days and denote n_{days} as the total number of time units. The spatio-temporal process results in $n_{\text{cells}}n_{\text{days}}$ observations of the response and predictor variables.

Regression models are used to predict the outcome of the response variable for new observations of the covariates, and for hypothesis testing in a variety of settings (Harrell, 2015), but we focus on regression models for binary and count response variables. In particular, we look at logistic regression for binary response variables and Poisson regression for count response variables.

Logistic regression

Many problems give rise to binary response data, where Y takes value 0 or 1; for example, the presence or absence of a species (Elith and Leathwick, 2009), bank credit scoring (Hand, 2005) or wildfire. Logistic regression is the most commonly used regression model for binary response data (Hosmer et al., 2013).

A binary response variable is modelled by a Bernoulli distribution, which is governed by a single parameter, the probability of $Y = 1$. This is also equal to the mean of Y. Because of this, regression models for binary data explore the relationship between the conditional probability $\pi(\mathbf{z}) := \mathbb{P}(Y = 1 \mid \mathbf{Z} = \mathbf{z})$ and the covariates. In the previous notation, $C(Y \mid \mathbf{Z}) = \pi(\mathbf{z})$.

Logistic regression is defined by the link function

$$g(x) = \log\left(\frac{x}{1-x}\right),$$

known as the *logit* function. A simple rearrangement of (3.1) for logistic

regression yields

$$\pi(\mathbf{z}) = \frac{\exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j z_j\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j z_j\right)} = \frac{1}{1 + \exp\left(-\beta_0 - \sum_{j=1}^{d} \beta_j z_j\right)} \tag{3.2}$$

From (3.2) we see that covariates taking any real value will result in $\pi(\mathbf{z}) \in (0, 1)$, allowing $\pi(\mathbf{z})$ to be interpreted as a probability. Alternative link functions exist for binary response variables, such as the inverse Normal distribution function used in Probit regression.

Inference for logistic regression is straightforward using maximum likelihood estimation. If we denote $\pi_{i,\theta} := \pi(\mathbf{z}(i))$ then the contribution to the likelihood function from the i^{th} observation is simply $\pi_{i,\theta}$ if $y_i = 1$ and $1 - \pi_{i,\theta}$ if $y_i = 0$, leading to the likelihood function

$$L(\theta) = \prod_{i=1}^{n} \pi_{i,\theta}^{y_i} \left(1 - \pi_{i,\theta}\right)^{1-y_i},$$

where $\theta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_d)$ is a vector of real coefficients and n is the total number of observations. The log-likelihood is given by

$$l(\theta) = \sum_{i=1}^{n} y_i \log(\pi_{i,\theta}) + (1 - y_i) \log(1 - \pi_{i,\theta}), \tag{3.3}$$

which can easily be numerically maximised.

Poisson regression

Poisson regression is a common method for modelling count data. Conditional on the value of the covariates, the response variable has a Poisson distribution with mean parameter

$$\lambda(\mathbf{z}) = \exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j z_j\right),$$

where $\beta_0, \beta_1, \ldots, \beta_d$ are real-valued coefficients. This corresponds to $C(Y \mid \mathbf{Z}) = \mathbb{E}(Y \mid \mathbf{Z})$ and link function $g(x) = \log(x)$. The likelihood function for Poisson regression is

$$L(\theta) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\lambda_\theta(\mathbf{z}(i))} (\lambda_\theta(\mathbf{z}(i)))^{y_i}}{y_i!},$$

which yields the log-likelihood function

$$l(\theta) \propto \sum_{i=1}^{n_{\text{cells}}} y_i \log(\lambda_\theta(\mathbf{z}_i)) - \lambda_\theta(\mathbf{z}_i). \tag{3.4}$$

Similarly to logistic regression, this is easily numerically maximised.

3.1.2 Generalised additive models

Generalised Additive models (GAMs) are a generalisation of GLMs that allow for non-linear relationships between the quantity $C(Y \mid \mathbf{Z})$ and the predictor variables. They follow the same setup as GLMs except that the predictor term is replaced by a sum of functions of the covariates. Hence (3.1) becomes

$$g\left(C(Y \mid \mathbf{Z})\right) = \beta_0 + \sum_{j=1}^{d} f_j(Z_j), \tag{3.5}$$

where β_0 is a real-valued constant and the $f_j(\cdot)$ are scaling functions. The constant β_0 is not necessary, since it can be incorporated into the scaling functions, but it can simplify the expression of many GAMs. There are countless ways of constructing the scaling functions and we discuss two of the more popular methods, *fractional polynomials* and *splines*. Later we use fractional polynomials to scale covariate effects in our application to wildfire ignition modelling.

Fractional polynomials

Early attempts to deal with non-linearity in regression models were ad hoc and used polynomials, usually quadratics, to reflect the non-linearity. Royston and Altman (1994) introduced fractional polynomials as a unified framework for scaling covariates using polynomials. A fractional polynomial of degree m is a function

$$f(z) = \beta_0 + \sum_{r=1}^{m} \beta_r z^{(p_r)},$$

where $\beta_0, \beta_1, \ldots, \beta_m \in \mathbb{R}$ and $p_1, p_2, \ldots, p_m \in \mathbb{R}$ with $p_1 < p_2 < \cdots < p_m$. The round bracket notation for the powers denotes the Box–Tidwell transformation:

$$z^{(p_r)} = \begin{cases} z^{p_r} & \text{if } p_r \neq 0, \\ \log(z) & \text{if } p_r = 0. \end{cases}$$

The definition of a fractional polynomial was extended to the case where one or more of the powers coincide. When $m = 2$ and $p_1 = p_2$ the fractional polynomial is defined as

$$f(z) = \beta_0 + \beta_1 z^{(p_1)} + \beta_2 z^{(p_1)} \log(z),$$

where $\log(z)$ is the natural logarithm. Details for when $m > 2$ can be found in Royston and Altman (1994).

Royston and Altman (1994) proposed to search for the best fitting fractional polynomial from a restricted set of possible powers. They found that fractional polynomials with $m = 2$ and powers drawn from the set

$$\mathcal{P} = \{-2, -2, -0.5, 0, 0.5, 1, 2, \ldots, \max(3, m)\} \tag{3.6}$$

were "sufficiently rich to cover many practical cases adequately".

In the case of multivariable regression, where there is more than a single covariate, the individual fractional polynomial terms are reparameterised by pooling the constants. That is, each fractional polynomial becomes

$$f(z) = \sum_{r=1}^{m} \beta_r z^{(p_r)}$$

and there is a single shared constant coefficient β_0. This prevents identifiability issues.

Royston and Sauerbrei (2008) has a wealth of information on regression modelling with fractional polynomials.

Splines

Splines offer an alternative method of scaling the effect of the covariates on $C(Y \mid \mathbf{Z})$. The overarching idea is to use piecewise polynomials to model non-linearity. The simplest spline is a piecewise linear spline (Harrell, 2015). For a set of *knots*, $\{k_1, k_2, \ldots, k_m\}$, a piecewise linear spline is a function

$$f(z) = \beta_0 + \beta_1 z + \sum_{r=2}^{m} \beta_r (z - k_r)_+,$$

where $\beta_0, \beta_1, \ldots, \beta_m$ are real-valued coefficients and

$$(z)_+ = \begin{cases} z & \text{if } z \geq 0, \\ 0 & \text{if } z < 0. \end{cases}$$

This is a continuous function and can approximate many non-linear relationships, but it struggles to reproduce relationships that have a high degree of curvature. To overcome this problem, polynomials of a higher degree can be used in the spline. Cubic splines are perhaps the most popular splines and use piecewise cubic polynomials. They are often constructed to be smooth functions by forcing the derivatives of the piecewise polynomials to agree at the knot locations. See Harrell (2015, Section 2.4.6) for details on choosing the number and location of knots.

Both splines and fractional polynomials can be used to scale the effect of covariates on $C(Y \mid \mathbf{Z})$ and comparisons have found negligible difference between their predictive performance (Binder, Sauerbrei, and Royston, 2013). Fractional polynomials produce simpler models in the sense that the final predictor term is a linear sum of 'nice' functions, whilst the final predictor term for spline models involves m knots and $m + 1$ polynomials. This simplicity of fractional polynomials can be useful in operational contexts where it can be important for models to be easily transported and deployed.

3.1.3 Equivalence of logistic and Poisson regression for low probability events

We have discussed how logistic and Poisson regression can be used to model binary and count response data respectively. An observation that becomes important when modelling wildfire ignitions is that, for low probability events, logistic and Poisson regression are approximately equivalent. Intuitively, this happens because the probability of $Y > 1$ in the Poisson regression model becomes so small that Y is essentially a binary response variable. However it is interesting to note that the relationship is very strong; not only do the probabilities approximate each other, but so too do the predictor terms. This relationship has been noted in the context of approximating point processes in Baddeley, Rubak, and Turner (2015, Section 9.9.3).

The relationship can be seen by equating the probabilies $\mathbb{P}(Y = 0)$ under both the Poisson and logistic regression models. Suppose we have a fitted logistic regression GAM with predictor term $F_l(\mathbf{z})$ and a fitted Poisson regression GAM with predictor term $F_p(\mathbf{z})$. Then the probability of $Y = 0$ is

$$\mathbb{P}(Y = 0 \mid \mathbf{Z} = \mathbf{z}) = \frac{1}{1 + \exp\big(F_l(\mathbf{z})\big)}$$

for the logistic regression model and

$$\mathbb{P}(Y = 0 \mid \mathbf{Z} = \mathbf{z}) = \exp\Big(-\exp\big(F_p(\mathbf{z})\big)\Big)$$

for the Poisson regression model. Equating these two expressions and rearranging yields

$$F_p(\mathbf{z}) = \log\left(\log\Big(1 + \exp\big(F_l(\mathbf{z})\big)\Big)\right). \tag{3.7}$$

Using the approximation $\log(1+x) \approx x$ for small x it follows that $F_p(\mathbf{z}) \approx F_l(\mathbf{z})$ for small values of $\exp(F_l(\mathbf{z}))$. This corresponds to small values of $\mathbb{P}(Y = 1 \mid \mathbf{Z} = \mathbf{z})$ under both the logistic and Poisson regression models.

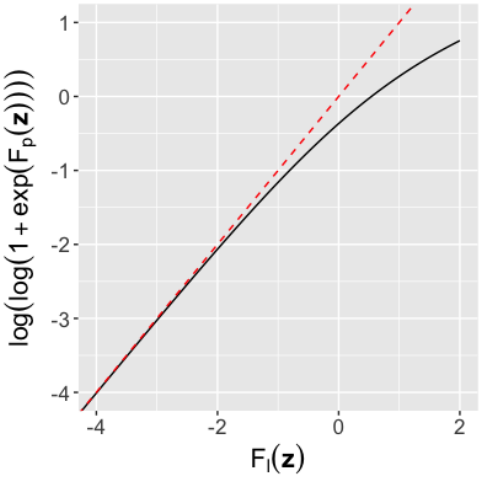


Figure 3.1: *The solid line shows the relationship* (3.7) *and the red dashed line corresponds with the identity function. The approximation* $F_l(\mathbf{z}) \approx F_p(\mathbf{z})$ *is very good when the predictor terms take values less than -2.*

As shown in Figure 3.1, this approximation is very good for values of $F_l(\mathbf{z})$ less than -2. A predictor term with a value of -2 corresponds to $\mathbb{P}(Y > 0) = 0.119$ and 0.123 for logistic and Poisson regression respectively.

We also need to consider the relationship between logistic and Poisson regression for $Y > 0$. For Poisson regression with a predictor term taking value 0.123, the probability $\mathbb{P}(Y > 1)$ is negligible at 0.007 and so $\mathbb{P}(Y > 0 \mid \mathbf{Z} = \mathbf{z}) \approx 1 - \mathbb{P}(Y = 0 \mid \mathbf{Z} = \mathbf{z})$ for Poisson regression with equality for logistic regression. This is the same situation as above.

This relationship implies that for low probability events, the logistic and Poisson regression models should approximately coincide: both the model probabilities and the actual form of the predictor term, such as coefficients and scaling functions, should be similar. In many modelling contexts, the probability of a non-zero response is well below 0.1 and this relationship arises.

3.2 Model selection

In Section 3.1 we discussed the structure of regression models as though the choice and scaling of covariates was given. In this section we discuss issues of model selection: how to choose covariates to include in the model and when and how to scale them. These two issues are closely connected and some methods of model selection deal with both issues simultaneously.

The guiding principle behind all model selection methods is that of parsimony: choose the simplest model that is an adequate representation of the data (Burnham and Anderson, 2002). Simple models guard against over fitting and can also be valuable in practical settings where more complex models can be difficult to interpret or where it is impractical to collect data for a large number of covariates.

A modeller usually has some prior understanding of which covariates may affect the response variable. This understanding might come from expert opinion or previous studies, but should not come from inspecting data that is being used in the study. We call this collection of covariates the *candidate covariates* and a key component of model selection is choosing which of these covariates to include in the final model. In Section 3.2.1 we discuss an approach to model selection known as *purposeful selection*, which we employ when developing our model for lightning ignitions in Chapter 5.

3.2.1 Model selection: purposeful selection

Hosmer et al. (2013, Section 4.2) advocated purposeful selection as a model selection process for logistic regression, although the basic principles of the approach apply more generally. Indeed, others recommend similar approaches to model selection for more general regression models (Harrell, 2015, Section 4.12 for example). Purposeful selection requires the modeller to inspect and judge the model at various stages of development. We begin with the like-

lihood ratio test and Wald statistics, which are the building blocks of the purposeful selection approach.

Likelihood ratio tests

Likelihood ratio tests are powerful tools in statistics and can be used for hypothesis tests involving *nested* regression models. We introduce the general likelihood ratio test before discussing its implications. Suppose that $x_1, x_2, \ldots, x_n$ are observations from the probability density (or mass) function $f(x \mid \boldsymbol{\theta})$ (satisfying regularity conditions, Casella and Berger, 2002, Miscellanea 10.6.2), where $\boldsymbol{\theta}$ is a vector of parameters. Suppose that $R(\cdot)$ is a vector-valued function such that $R(\boldsymbol{\theta}) = \mathbf{0}$ imposes r restrictions on the parameter vector $\boldsymbol{\theta}$. Then consider the composite hypothesis H_0: $R(\boldsymbol{\theta}) = \mathbf{0}$ against H_1: $R(\boldsymbol{\theta}) \neq \mathbf{0}$. The *likelihood ratio* test statistic is given by

$$LR = -2 \log \left(\frac{\mathcal{L}(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0})}{\mathcal{L}(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1})} \right),$$

where $\mathcal{L}$ is the likelihood function, $\widehat{\boldsymbol{\theta}}_{H_1}$ is the MLE and $\widehat{\boldsymbol{\theta}}_{H_0}$ is the value satisfying $R(\widehat{\boldsymbol{\theta}}_{H_0}) = \mathbf{0}$ that maximises the likelihood function (the restricted MLE). Then as $n \to \infty$,

$$LR \to \chi^2_r,$$

the chi-squared distribution with r degrees of freedom (Wood, 2017, Appendix A.5).

Now suppose we have two GLMs with d_1 and d_2 covariates respectively such that the covariates included in model 2 are a subset of the covariates included in model 1. Model 2 is *nested* in model 1 since model 2 can be recovered from model 1 by choosing coefficients equal to 0 for all the covariates in model 1 that are not in model 2. For nested regression models a likelihood ratio test can be performed to test H_0: the data comes from model 2 (the smaller model) against H_1: the data comes from model 1 (the larger model). The numerator and denominator of the likelihood ratio test statistic are calculated using the

coefficients from model 2 and model 1 respectively; finding evidence against the null hypothesis that model 2 is adequate if the LR is sufficiently large (as determined by a suitable quantile of the chi-squared distribution with $d_1 - d_2$ degrees of freedom).

Wald statistics

When regression models are fitted using maximum likelihood, asymptotic properties can be used to find the approximate distribution of the regression coefficients (Wood, 2017, Section 3.1.3). For $\widehat{\boldsymbol{\beta}}$, the MLE vector of regression coefficients, the large sample approximate distribution is

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (X^T \widehat{V} X)^{-1}),$$

where X is the *design matrix*, the $n \times (d+1)$ matrix with rows $(1, \mathbf{z}(i)), i = 1, 2, \ldots, n$; and $\widehat{V}$ is the $n \times n$ diagonal matrix with diagonal entries $\hat{\pi}_i(1-\hat{\pi}_i)$, where $\hat{\pi}_i$ is the i^{th} predicted probability (Hosmer et al., 2013, Section 2.3). The univariate Wald statistics are defined as

$$W_j = \frac{\widehat{\beta}_j}{\widehat{\text{SE}}(\widehat{\beta}_j)},$$

where $\widehat{\beta}_j$ is the MLE of the j^{th} regression coefficient and $\widehat{\text{SE}}(\widehat{\beta}_j)$ is the estimated standard error, given by the square root of the j^{th} diagonal entry of $(X^T \widehat{V} X)^{-1}$. If the true value of β_j is 0 then the univariate Wald statistic is approximately distributed according to a standard Normal distribution. This can be used to perform an approximate hypothesis test H_0: $\beta_j = 0$ against H_1: $\beta_j \neq 0$, rejecting H_0 if the value of $|W_j|$ exceeds a suitably chosen quantile of the standard Normal distribution.

Method

Here we give the steps of purposeful selection as outlined in Hosmer et al. (2013, Section 4.2). All significance levels are assumed to be 0.05 unless otherwise specified. We also assume that all covariates are continuous, since this

is the case for our intended application. See Hosmer et al. (2013) for details on how to deal with discrete or categorical covariates.

1. A univariate analysis of every covariate: fit a univariate regression model to each of the covariates. Compare the univariate and null model (model with no covariates) using a likelihood ratio test with significance level of 0.25. The generous significance level is recommended to capture variables which may show univariate insignificance, but have a confounding effect when included in a multivariate model.

2. Multivariate analysis: fit a multivariable model with all the covariates selected by the univariate analysis. Fit a second, smaller, multivariable model by removing any covariates with insignificant Wald statistics in the first multivariate model. Use a likelihood ratio test to compare the smaller model with the full model.

3. Multivariate analysis: compare values of the fitted coefficients between the full and reduced multivariable model. If any of the coefficients change by more than 20% it suggests that one or more of the removed covariates has an important interaction with the other covariates in the model. These covariates should be reintroduced to the model. Repeat steps 2 and 3 until all important covariates seem to be in the model. In this situation, it is recommended that only a few covariates be deleted each iteration.

4. Multivariate analysis: one at a time, add each covariate removed in step 1 to the model. Keep them in the model if their Wald statistic is significant. This is done to ensure that we don't miss any covariates that have important interaction effects, despite not being significant on their own. The model at the end of this step determines which covariates are included in the final model and is known as the *preliminary main effects model*.

5. Scaling: we assess the assumption that the logit-likelihood is linearly related to the covariates and choose scaling functions where appropriate. We discuss how to do this below. The model at the end of this step is known as the *main effects model*.

6. Interaction effects: there may be significant pairwise interactions between covariates in the model. This means that the effect of changes in a covariate is not constant over varying values of another covariate. Whether or not to include interaction terms "should be based on statistical as well as practical considerations" (Hosmer et al., 2013). Begin by compiling a list of pairs of covariates that could have a possible interaction effect from a practical perspective. Interaction terms are created by taking the arithmetic product of the covariates. Analyse each interaction term by constructing a model that includes all the covariates in the main effects model along with the selected interaction term. Check the significance of these models using a likelihood ratio test. Following this, create a model which includes all of the interaction terms deemed significant by the univariate analysis along with the covariates in the main effects model. Follow step 2 to produce a reduced model, but only allow the removal of interaction terms.

7. Validation: perform model validation and goodness-of-fit test, as discussed in Section 3.4 below.

We now provide extra detail on Step 5: scaling covariates. Scaling covariates typically increases the number of parameters in the model and so unnecessary scaling should be avoided. To begin, Nadaraya–Watson smooths can be inspected to determine if the relationship between the covariate and $C(Y \mid \mathbf{Z})$ departs from linearity in a meaningful way. A Nadaraya–Watson smooth of a function f is a non-parametric estimate formed by kernel smoothing the $(x, f(x))$ observations (Hastie, Tibshirani, and Friedman, 2009, Sec-

tion 2.8.2). In the case of logistic regression, a Nadaraya–Watson smooth can be calculated for each covariate–response pair and plotted on the logit scale. Any covariate whose plot shows a notable departure from linearity should be a candidate for scaling.

Hosmer et al. (2013) recommended several simple ad hoc approaches to scaling covariates along with a more methodological framework using fractional polynomials. They find that fractional polynomials of degree $m = 2$ and with powers drawn from the list (3.6) are sufficient for most practical applications. Let $\mathcal{S}$ be the set of covariates that are selected for scaling. The process of scaling is iterative, cycling through the covariates in $\mathcal{S}$ one by one and choosing suitable scaling functions. At each step, the model includes all covariates in the preliminary main effects model with those not in $\mathcal{S}$ remaining with a linear term.

To begin, the first covariate in $\mathcal{S}$, say z_j, is scaled. First, fit all possible models that include a degree 1 fractional polynomial term for z_j. This results in 8 models, one for each of the powers in $\mathcal{P}$. From this choose the best model, determined as the model that maximises the value of the likelihood function. Next, fit all possible models which include a degree 2 fractional polynomial term for z_j – resulting in 36 models – and choose the model that maximises the likelihood function. Finally, determine whether the scaled models offer a statistically significant improvement to the previous model. This can be done using a series of approximate likelihood ratio tests. First compare the best degree 2 model with the linear model; if the degree 2 model is not significant select the linear model, else compare the degree 2 model to the best degree 1 model; if the degree 2 model is still significant, reject the degree 1 model; else use the degree 1 model.

We repeat this process for the second covariate in $\mathcal{S}$ except we fix the fractional polynomial term for z_j chosen in the first cycle of the process. This process is repeated, cycling through each covariate in $\mathcal{S}$ until a full cycle is

made without any of the fractional polynomial terms changing. The resulting model is called the main effects model.

Royston and Altman (1994) argued that each term in a fractional polynomial contributes approximately 2 degrees of freedom to the likelihood ratio test. This means that the likelihood ratio that compares the best degree 2 fractional polynomial with the best degree 1 fractional polynomial is distributed approximately chi-squared with 2 degrees of freedom and the likelihood ratio that compares the best degree 1 fractional polynomial model with the unscaled (linear) model is distributed approximately chi-squared with 1 degree of freedom.

3.3 Model approximation

In some sense, traditional model selection techniques select covariates that have a detectable influence on the quantity being modelled. For data sets with a large number of observations it becomes possible for very subtle effects to be detected, often resulting in models that include a large number of covariates. Large models can be cumbersome to use; collecting and maintaining large databases of covariates may be too costly for operational use. In these situations it may be desirable to approximate the full model with a simpler model with fewer covariates. The full model will usually be the most accurate, but it may be possible to produce a reduced model with comparable performance by removing the relatively uninfluential covariates.

Harrell (2015) discussed a simple way to do this for linear regression models by using the weighted average of the predictions, as follows. Suppose we need to remove a particular covariate from the model. If the covariate is categorical then calculate the predicted value for every possible value of the missing covariate and take the average weighted by the prior probabilities of the category as our final prediction. In theory we could repeat this process until we had a sufficiently simple model, but this approach becomes unwieldy

as more covariates are removed.

An alternative approach that can deal with more complex models is to build a new, simpler model by regressing on the predictor terms from the full model. That is, rather than modelling the quantity of interest, $C(Y \mid \mathbf{Z})$, build a model for the predictor term from the full model, $\sum_j f_j(z_j)$, regressing on the original covariates $\mathbf{z}$. Harrell (2015) suggested that this can be done using regression trees (Breiman et al., 1984) or linear regression. The modeller can adjust the depth of the regression tree to find a simpler model which still has an acceptable degree of accuracy. Interestingly, these methods have been used to try and decode "black boxes" such as neural networks by constructing approximating models that yield more readily to interpretation (Harrell, 2015).

Measuring the accuracy of an approximating model can be difficult. Harrell (2015) recommended measuring the agreement between the predictor terms of the full and reduced models using the R^2 value. This approach is simple and easily understood, but it does not directly connect the approximating model to the quantity of interest, $C(Y \mid \mathbf{Z})$. An alternative approach is to measure the predictive performance of the full model and compare it with the predictive performance of the reduced model. Chapter 4 discusses performance measures for binary models such as logistic regression. This second approach allows the performance of the approximating model to be evaluated in reference to the quantity of interest, however measuring predictive performance is somewhat fraught (Section 4.5 below).

In general one expects the full model to be more accurate than the approximating model. A particularly important situation to consider is when a covariate mostly has little effect on the predictor term, but occasionally is highly influential. This covariate is relatively likely to be removed during model approximation since it has little effect on the bulk of observations. The approximating model may perform well on the bulk of the data, but can be

inaccurate for those few observations where the covariate is highly influential. The user should be aware that although an approximating model may seem to have good accuracy, it can perform poorly on particular subsets of the data.

In the next section we propose a very simple ad hoc method for identifying and removing relatively uninfluential covariates from a GAM model. The technique is not dissimilar to the approach to linear regression models (Harrell, 2015), but allows us to deal with more complex non-linear models.

3.3.1 Covariate contributions

We present an approach for finding covariates that are relatively uninfluential in the model. We do this by working on the scale of the predictor term, looking for covariates whose effect on the predictor term is relatively close to constant. Given a regression model, denote the contribution of the j^{th} covariate to the predictor term as the j^{th} *covariate contribution*

$$CC_j = f_j(Z_j), \tag{3.8}$$

where f_j is the scaling function corresponding to the j^{th} covariate. Note that CC_j incorporates effects from the regression coefficients, scaling functions and the inherent randomness of the covariate. The predictor term is the sum of the covariate contributions and the regression constant, if there is one.

We are interested in thinking of the covariate contributions as random variables themselves and considering their influence on the predictor term. If the link function is monotonic (which is true for all common regression models, such as Normal, exponential, gamma, Poisson, negative binomial and binomial), the predictor term is directly related to the quantity of interest, $C(Y \mid \mathbf{Z})$, in the sense that an increase (or decrease) in the predictor term will result in an increase (or decrease) in $C(Y \mid \mathbf{Z})$.

To identify relatively uninfluential covariates we might consider the variance of each CC_j. If $CC_{j'}$ has small variance relative to the other CC_j then

$CC_{j'}$ behaves more like a constant and can be factored into the regression constant. We propose approximating the full model by replacing these relatively uninfluential CC_js with their mean. A series of approximating models can be built by replacing the CC_js with their means in increasing order of variance. The modeller can then choose the smallest model that still provides suitable performance.

3.4 Model validation for logistic regression

Model selection and fitting procedures will always produce a model, so the next step in a modeller's journey should be to assess how faithfully the model reflects the data. This is not a judgement of the predictive power of a model. It is possible to have a well-fitting model with no predictive power, a Bernoulli random variable for a coin toss for example. Goodness-of-fit tests are formal hypothesis tests for assessing whether the data could reasonably be believed to come from the fitted model. These give some indication of model fit, but if a model fails a test it does not show why it failed. For this purpose, diagnostics are investigated. These give an indication of which observations are responsible for poor model fit and allow the modeller to understand where and why the model is failing.

The specifics of goodness-of-fit tests and diagnostics differ for different regression models and we focus solely on logistic regression.

3.4.1 The Hosmer–Lemeshow goodness-of-fit test

The *Hosmer–Lemeshow test* is a modified chi-squared test for logistic regression models, where the groups in the test are based on the estimated probabilities. After sorting the estimated probabilities into increasing order, there are two methods for creating the groups. The first splits the ordered estimated probabilities such that the observations are split equally across all the groups.

The second method uses fixed cut points; if there are g groups the cut points would be defined by $k/g, k = 1, 2, \ldots, g-1$. Then the Hosmer–Lemeshow test statistic is given by

$$\widehat{C} = \sum_{k=1}^{g} \left[\frac{(o_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(o_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right]$$

where

$$o_{1k} = \sum_{i=1}^{c_k} y_i, \quad o_{0k} = c_k - \sum_{i=1}^{c_k} y_i,$$
$$\hat{e}_{1k} = \sum_{i=1}^{c_k} \hat{\pi}_i, \quad \hat{e}_{0k} = c_k - \sum_{i=1}^{c_k} \hat{\pi}_i$$

and c_k is the number of observations in the k^{th} group.

Through a simulation study, the distribution of $\widehat{C}$ was found to be approximately chi-squared with $g-2$ degrees of freedom, and that the first grouping method, using percentiles, is preferable to fixed cut points (Hosmer et al., 2013, Section 5.2.2). They recommend that $g = 10$ groups is sufficient in most cases.

In the case of large data sets, goodness-of-fit tests such as the Hosmer–Lemeshow test can be overly sensitive, recognising a lack of fit from very slight deviations from the model. As Johnson and Wichern (2007, page 187) noted, "very large samples invariably produce statistically significant lack of fit. Yet the departure from the specified distribution may be very small and technically unimportant to the inferential conclusions".

3.4.2 Diagnostics for logistic regression

Diagnostics for logistic regression mirror those for linear regression. The *Pearson residual* for logistic regression is given by

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}};$$

assuming that the model is correctly specified and that there is no error in the estimate $\hat{\pi}_i$, r_i is the observed value y_i, standardised to have zero mean

and unit variance. The *Pearson chi-squared statistic* is $X^2 = \sum_{i=1}^{n} r_i^2$. The *deviance residual* is

$$d(y_i, \hat{\pi}_i) = \begin{cases} -\sqrt{2|\log(1-\hat{\pi}_i)|} & \text{if } y_i = 0 \\ \sqrt{2|\log(\hat{\pi}_i)|} & \text{if } y_i = 1 \end{cases}$$

and the *deviance* is $D = \sum_{i=1}^{n} d(y_i, \hat{\pi}_i)^2$ (Hosmer et al., 2013, Section 5.3). Where the Pearson chi-squared statistic represents the distance between the observed data and the model predictions, the deviance statistic represents the agreement between the log-likelihood functions of the fitted model and the saturated model.

In linear regression, the *hat matrix* is the matrix H satisfying $\widehat{\boldsymbol{y}} = H\boldsymbol{y}$. The diagonal entries of H, denoted h_i, are known as the *leverage values* and they reflect the influence of the i^{th} observation on the i^{th} prediction $\widehat{y}_i$. That is, the influence of the i^{th} observation on its own prediction. If the i^{th} leverage value is large relative to other leverage values, it indicates that the i^{th} observation is having a large impact on model fit.

Pregibon (1981) used linear approximation methods to find a matrix for logistic regression that is equivalent to the hat matrix from linear regression:

$$H = \widehat{V}^{1/2} X (X' \widehat{V} X)^{-1} X' \widehat{V}^{1/2},$$

where X is the design matrix, $\widehat{V}$ is an $n \times n$ diagonal matrix with diagonal entries

$$\hat{v}_i = \hat{\pi}_i \left(1 - \hat{\pi}_i\right).$$

The leverage values are of interest themselves and also play a key role in constructing other diagnostics. Another consequence of this approximation is a further standardisation of the Pearson residual (Hosmer et al., 2013, Section 5.3), namely the *standardised Pearson residuals*,

$$r_{s,i} = \frac{r_i}{\sqrt{1 - h_i}}.$$

Many diagnostics explore model fit by comparing the model to a reduced model constructed by deleting the i^{th} observation and refitting the model. Large differences between the full and reduced model would suggest that model fits poorly to the i^{th} point, or that the i^{th} point has a large influence on general model fit. Pregibon (1981) found linear approximations for the change in the Pearson chi-squared statistic,

$$\Delta X_i^2 = r_{s,i}^2,$$

the change in the deviance statistic,

$$\Delta D_i = \frac{d_i^2}{1 - h_i},$$

and *Cook's distance*

$$\Delta\widehat{\boldsymbol{\beta}}_i = \frac{r_{s,i}^2 h_i}{1 - h_i}.$$

Cook's distance represents the standardised difference between the regression coefficients of the full and reduced models (Cook, 1977; Hosmer et al., 2013; Pregibon, 1981).

Note that these linear approximations are necessary because otherwise it would be computationally infeasible to calculate these residuals. For example, we could calculate the change in the Pearson chi-squared statistic by calculating $X^2 - X_{(-i)}^2$ for each i, where $X_{(-i)}^2$ is the Pearson chi-squared statistic for the model with the i^{th} observation removed. However, this would require us to fit one model for each observation, which would be infeasible for moderate or large sample sizes.

Interpretation of residuals for logistic regression can be difficult. Hosmer et al. (2013) recommended looking for observations whose "diagnostic statistics fall well away from the rest of the values" for two or more of the diagnostics of interest. Generally, large value of $\Delta\mathbf{X}_i^2$ or ΔD_i reflects a large discrepancy between the model and the i^{th} observation, whilst a large value of $\Delta\widehat{\boldsymbol{\beta}}_i$ or h_i reflects a disproportionate influence of the i^{th} observation on model fit. Hos-

mer et al. (2013) recommended inspecting plots of (i) h_i against $\hat{\pi}_i$, (ii) ΔX_i^2 against $\hat{\pi}_i$, (iii) ΔD_i against $\hat{\pi}_i$, and (iv) $\Delta\widehat{\boldsymbol{\beta}}_i$ against $\hat{\pi}_i$.

3.5 Interpreting the model

Meaningful interpretation of a fitted model is important when the model is intended for operational use. To have trust in a model, users need to understand how it will behave in various scenarios. This allows users to compare the model to their intuitive understanding of the physical process and allows them to 'sanity check' the model. They are more likely to trust a model if the general underlying principles are consistent with their understanding of the process.

Interpreting regression models is easiest with the assumption that the link function is monotonic, which we assume is true since it holds for all common regression models. For GLMs, if the link function is monotonically increasing (decreasing) then a positive value of a regression coefficient indicates that the corresponding covariate is positively (negatively) related to the quantity of interest, $C(Y \mid \mathbf{Z})$. For linear regression, when the link function is the identity function, regression coefficients are equal to the increase in $C(Y \mid \mathbf{Z})$ corresponding to a unit increase in the covariate, given that all the other covariates are fixed (or more generally, the contribution to the predictor term from the other covariates remains constant). Interpreting regression coefficients for other models is much more difficult. For example, Hosmer et al. (2013, Chapter 3) explain how to interpret the regression coefficients in a logistic regression GLM in terms of increases in the log-odds. In Harrell (2015, Section 2.3) the author discusses this interpretation more generally, in terms of changes in the quantity $g(C(Y \mid \mathbf{Z}))$.

The interpretation of GAMs is even more difficult since we have functions of the covariates and they are non-linearly related to the quantity of interest. Under the assumption that the link function is monotonic, it is easiest to anal-

yse the effect of the covariates on the predictor term $g(C(Y \mid \mathbf{Z}))$ rather than $C(Y \mid \mathbf{Z})$. Plots of the scaling function f_j give some insight into the influence of z_j in the model, however they fail to show the effect of any interaction terms, if present. We recommend including vertical lines at the 0.025 and 0.975 quantiles of the covariate to highlight the section of the scaling function that sees most use. Plots of each of the covariates z_j against the predictor term holding all other covariates constant are known as 'partial effects plots' and also reflect interaction terms (Harrell, 2015, Section 5.1.1). In the presence of interaction terms the partial effects plot can be repeated for different levels of interacting covariates.

We propose another approach to interpreting a fitted GAM model using covariate contributions (Section 3.3.1), which we defined as $CC_j = f_j(Z_j)$ and view as a random variables. Following the logic outlined in Section 3.3.1, we can consider the variance of each of the CC_j as a rough guide to the relative influence of each of the covariates on the model. Further insight can be gained by looking at estimates of the probability densities of the CC_j. The strength of analysing the CC_j over other techniques such as partial effects plots is that it incorporates the randomness of the covariate itself. In the presence of interaction, multiple CC_j density plots can be made by conditioning on the value of interacting covariates.

3.6 Forecasts

Producing forecasts from regression models is simple enough: calculate the predicted value of $C(Y \mid \mathbf{Z})$ using new observations of the covariates. However in many practical settings, for example those involving weather, we have forecast rather than observed values of the covariates. The discrepancy between the predicted and observed values of the covariates introduces another source of error into the model forecasts. In this situation the best practice is to use historic forecast data to fit and evaluate the model, however it is

often difficult to acquire such data. Data repositories tend to keep only the observed data.

In Section 3.5 we discussed how it is important for models to be explainable if users are to build trust in them. In that section we discussed how to understand the general behaviour of a model. However, it can also be important to understand a particular forecast. In Sections 3.6.1 and 3.6.2 we discuss how individual forecasts can be deconstructed into contributions from each of the covariates, allowing users to relate model output to their own understanding.

3.6.1 Deconstructing forecasts - GLM

It is often recommended that the covariates be centered, by deducting their mean, prior to modelling. This gives the regression constant β_0 the interpretation of being the value of the predictor term when all the covariates are equal to their means. The quantity $g^{-1}(\beta_0)$ is the predicted outcome if all covariates are equal to their means and has the somewhat fraught interpretation as being the forecast outcome in 'normal' conditions. Of course, all covariates being equal to their respective means is not 'normal' – it ignores any dependence between the covariates and the mean need not be an observed value of the covariate – but in many practical settings this can be a useful interpretation.

If we assume the link function is monotonic then a consequence of model building with centered covariates is that we can deconstruct a given forecast into contributions from each of the covariates. Suppose that $\mathbf{z}^*(i)$ is an observation of the centered covariates and we have used our 'centered' model to forecast $C(Y \mid \mathbf{Z}^* = \mathbf{z}^*(i))$. If we think of all covariates being equal to their mean as 'normal' conditions, then for a given forecast the value of $\beta_j z_j^*(i)$ indicates the influence of the j^{th} covariate. A large positive (negative) value of $\beta_j z_j^*(i)$ suggests that the j^{th} covariate has a large positive (negative) influence on the predicted value of $C(Y \mid \mathbf{Z}^* = \mathbf{z}^*(i))$ when compared with 'normal' conditions. What constitutes a large or small value of $\beta_j z_j^*(i)$ is determined

by the relative values of the other covariates. Those with the largest (smallest) value of $\beta_j z_j^*(i)$ have the largest (smallest) influence on the given forecast when compared with 'normal' conditions.

In the presence of interaction terms it is not possible to make statements about the influence of one covariate without discussing the others. Instead the relative influence of the interactive covariates can understood by considering the combined term such as $\beta_1 z_1^*(i) + \beta_2 z_2^*(i) + \beta_{1,2} z_1^*(i) z_2^*(i)$, which captures the influence of two covariates simultaneously.

3.6.2 Deconstructing forecasts - GAM

We propose a method for deconstructing model forecasts from GAMs into contributions from each of the covariates by considering the value of centered covariate contributions (Section 3.3.1). We call the departure from the mean the covariate contribution anomaly,

$$CCA_j(i) = CC_j(i) - \mathbb{E}CC_j,$$

where $CC_j(i) = f_j(z_j(i))$ is the j^{th} covariate contribution evaluated for the i^{th} observation. The predictor term can be expressed in terms of the CCA as

$$\beta_0 + \sum_j f_j(z_j(i)) = \beta_0^* + \sum_{j=1}^{d} CCA_j(i),$$

where $\beta_0^* = \beta_0 - \sum_{j=1}^{d} \mathbb{E}CC_j$. The coefficient β_0^* has a similar interpretation to the regression coefficient from GLMs with centralised covariates.

The CCA has a similar interpretation to $\beta_j z_j^*(i)$ from the GLM case: a large positive (negative) value of $CCA_j(i)$ suggests that the j^{th} covariate has a large positive (negative) influence on the predicted value of $C(Y \mid \mathbf{Z}^* = \mathbf{z}^*(i))$ when compared with 'normal' conditions. Here, 'normal' conditions are with respect to the value of the CC not the covariate itself. In the presence of interaction terms we must again consider the joint contributions.

It's worth noting that although we have discussed using the mean to centralise the CC, any suitable constant will work. Possible alternatives are using the median or mode of the CC, or using the value of the scaling function f_j evaluated at the mean, median or mode of the j^{th} covariate. The best choice of centralisation quantity will be determined by the setting and the distribution of the covariate.

Chapter 4

Performance measures for binary response models

Classifiers and their link to logistic regression. The AUC, H measure, Kolmogorov–Smirnov statistic and Hedges' g. Issues with interpreting performance measures. Recommendations for measuring model performance.

4.1 Classifiers

For models that are intended to be used for prediction it is important to be able to judge the predictive accuracy of the model. Unless there is a very specific application where there exists a clearly defined performance metric, it can be difficult to assess 'general' performance. In this chapter we discuss approaches to measuring the performance of binary response models, such as logistic regression (Section 3.1.1). As we show, interpretations of these measures do not always link back to a meaningful quantity in terms of the underlying probabilities. Despite this, many continue to use them uncritically to draw conclusions about the value of models.

Many performance measures used for evaluating binary response models were initially designed to assess binary classifiers. A *classifier* is a function of the covariates that returns either a 0 (negative response) or a 1 (positive

response). More generally, a classifier is a pair $(s(\mathbf{z}), t)$ where $s(\mathbf{z})$ is the *score function*, which is monotonically related to the probability of a $y = 1$ response, and $t \in \mathbb{R}$ is the *threshold*. An observation of the covariates $\mathbf{z}$ is classified (predicted) as a positive response if $s(\mathbf{z}) \geq t$ and a negative response if $s(\mathbf{z}) < t$. In logistic regression, for example, the score function could be either the forecast probability $\pi(\mathbf{z})$ or the predictor term $\beta_0 + \sum_{j=1}^{p} f_j(z_j) = \text{logit}(\pi(\mathbf{z}))$. Logistic regression is a classifier when paired with a threshold. Binary response models return the probability of a positive response, whereas binary classifiers are decision rules, converting the output of the response model into a positive or negative response.

A *classifier performance measure* returns a numerical value that reflects the ability of a classifier to discriminate meaningfully between positive and negative responses. There are numerous performance measures with different interpretations and properties and we cover some of these in the following sections. However they all share a common property: they are summary statistics endeavoring to describe a complex object by a single number. A single performance measure cannot reflect all aspects of the accuracy of a classifier. For this reason it is important to understand the specific behaviour of any performance measures used in a study.

Performance measures are not a substitute for model validation and they do not reflect how well the model fits the data. For example, the coin toss classifier (randomly predict a positive response, with probability 0.5) will fit data from a coin toss very well and yet offers abysmal predictive performance. Hosmer et al. (2013) noted that the "abililty of the fitted model to discriminate between the two outcomes is more a function of the difference between the groups... we can have well fitting models that discriminate poorly, just as we could have models with poor fit that discriminate well". After we have performed model validation and are satisfied with the fit of the model, performance measures indicate the ability of the model to discriminate between

the classes.

The *sensitivity* (or true positive rate) of a classifier at threshold value t is the proportion of $y = 1$ cases that are correctly identified as $y = 1$ cases. That is, the proportion of $y = 1$ cases where $s(\mathbf{z}) > t$. Similarly, the *specificity* (or true negative rate) at threshold value t is the proportion of $y = 0$ cases that are correctly identified as $y = 0$ cases. That is, the proportion of $y = 0$ cases where $s(\mathbf{z}) \leq t$. If $G_0(t)$ is the distribution function of the score of a $y = 0$ observation and $G_1(t)$ is the distribution function of the score of a $y = 1$ observation then for a threshold with value t the sensitivity is $1 - G_1(t)$ and the specificity is $G_0(t)$.

Generally, the sensitivity and specificity are inversely related. For any classifier and as $t \to \infty$, every observation will eventually be classified as a $y = 0$ observation. This results in the sensitivity approaching 1 and the specificity approaching 0. Likewise if the threshold $t \to -\infty$, every observation will eventually be classified as a $y = 1$ observation and the sensitivity will approach 0 and the specificity will approach 1. The choice of threshold value is often informed by an acceptable level of sensitivity and specificity.

In many practical settings there is a need to assess the quality of a classifier without knowing the value of the threshold. For example, a bank might be interested in using a classifier to identify customers that are likely to default on their loan. Suppose that the model is developed centrally, but the thresholds are chosen to be specific to each of the branches of the bank to satisfy their individual constraints. The modeller needs to choose the 'best' model prior to knowing the threshold. Another reason to avoid prespecified thresholds is that they are usually highly subjective, chosen based on the problem at hand. These subjective decisions can make it impossible to compare models meaningfully. Performance measures attempt to measure the 'general' ability of a classifier to discriminate between positive and negative responses, without specifying a choice of threshold. In this chapter we discuss several popular

performance measures, highlighting some issues with them, before giving some recommendations for their use.

4.2 The AUC

The integral of the *Receiver Operating Characteristic* (ROC) curve, known as the area under the curve (AUC or AUROC), is the most popular measure of discrimination for binary classification models. Indeed, Hand and Anagnostopoulos (2013) found that the AUC was used in over 6000 papers in 2012 alone and noted the largely unrecorded prevalence of the AUC in commercial applications.

The ROC curve is a plot of the sensitivity against $1 -$ specificity, which is the curve defined by

$$\{(1 - G_1(t), 1 - G_0(t)) \mid t \in \mathbb{R}\}. \tag{4.1}$$

Since it is defined in terms of distribution functions, the ROC curve inhabits the space $[0, 1] \times [0, 1]$. A perfect classifier that correctly classifies every case for all threshold values will have an ROC curve equal to 1 everywhere in the $[0, 1]$ interval. A classifier that is equivalent to tossing a coin will have an ROC curve coresponding to the diagonal $\{(x, x) \mid x \in [0, 1]\}$. An example of a typical ROC curve can be found in figure 4.1.

Although it is possible for the ROC curve to go below the diagonal, this case is generally ignored since the corresponding classifier can easily be improved. The values of t where the ROC curve is below the diagonal correspond with the classifier reliably misidentifying the $y = 0$ observations as $y = 1$ observations. A simple transformation of the classifier will correct this and produce a new classifier whose ROC curve is above the diagonal. Due to this, we can think of the ROC curve as being bounded by 1 from above and the diagonal $\{(x, x) \mid x \in [0, 1]\}$ from below.

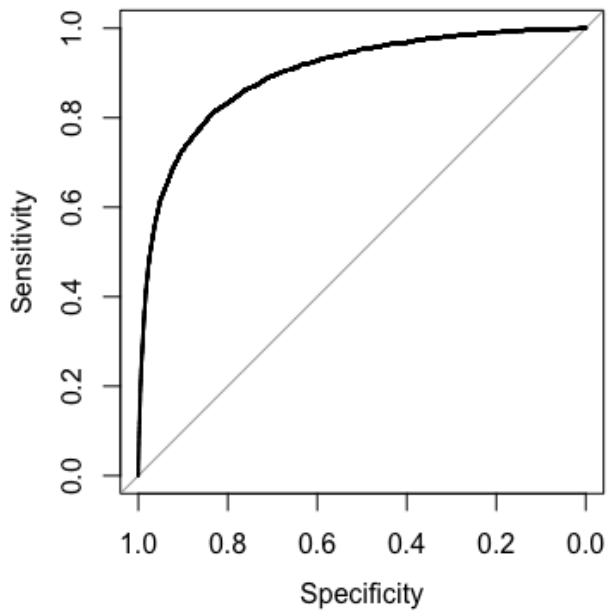


Figure 4.1: *An example of the reciever operating characteristic curve*

The area under the ROC curve is the integral of the ROC curve. Following from the above discussion, it takes values between 0.5, corresponding to the coin toss classifier, and 1, corresponding to the perfect classifier. The AUC is given by the integral of (4.1):

$$\begin{aligned} AUC &= \int_0^1 \left[1 - G_1(G_0^{-1}(1-s))\right] \mathrm{d}s \\ &= 1 - \int_{-\infty}^{\infty} G_1(u) g_0(u)\, \mathrm{d}u \qquad (4.2) \\ &= \int_{-\infty}^{\infty} (1 - G_1(u)) g_0(u)\, \mathrm{d}u, \qquad (4.3) \end{aligned}$$

where $g_1(u)$ is the density of $G_1(u)$, assuming it exists. Equation (4.2) follows from the change of variable $u = G_1^{-1}(s)$. An intuitive interpretation of the AUC is suggested by (4.3): it is the probability that a $y = 0$ observation selected uniformly at random will have a lower score than a $y = 1$ observation selected uniformly at random. An alternative interpretation is that the AUC

is the weighted average of the sensitivity, taking each value of the specificity as equally likely.

The AUC has drawn increased criticism in the past decade (Hand, 2009; Hand and Anagnostopoulos, 2013; Lobo, Jiménez-Valverde, and Real, 2008). In particular, the AUC has been criticised for: (i) treating false positives and false negatives equally, and (ii) summarising classifier performance over regions of ROC space that are not of practical interest. In most practical settings it is unreasonable to assume that false positive and false negatives have equal consequence. Hand (2009) discussed this in detail and proposed an alternative performance measure that considers different misclassification costs (the H measure, see below). The second criticism, that the AUC summarises classifier performance over irrelevant regions of ROC space, is addressed by the partial AUC (Dodd and Pepe, 2003), which is the integral of the ROC curve over a restricted domain,

$$pROC = \int_{s_0}^{s_1} \left[1 - G_1(G_0^{-1}(1-s))\right] \mathrm{d}s,$$

where (s_0, s_1) is the range of values of the specificity that are deemed 'reasonable'. The choice of (s_0, s_1) is arbitrary and problem specific, removing some of the appeal of the AUC as an objective measure. Despite these problems, the AUC remains the most popular measure of classifier performance.

4.3 The H measure

The H measure, first proposed by Hand (2009), was born out of a criticism of the AUC that arises in the context of misclassification costs. The arguments against the AUC were streamlined in Hand and Anagnostopoulos (2013), which showed that the AUC is related to the weighted average of the proportion of correctly classified cases. That is, they showed that the proportion of correctly classified cases, $p_0(1-G_0(t))+p_1G_1(t)$ where $p_0 = \mathbb{P}(Y=0)$,

$p_1 = \mathbb{P}(Y = 1)$, is related to the AUC by the weighted expectation

$$\int_{-\infty}^{\infty} \left[p_0(1 - G_0(t)) + p_1 G_1(t)\right] m(t)\, \mathrm{d}t = p_0^2/2 + p_1^2/2 + 2p_1 p_0 AUC, \quad (4.4)$$

where $m(t) = p_0 g_0(t) + p_1 g_1(t)$. They claimed that this makes the AUC "incoherent" in the sense that the weighting used to calculate the proportion of correctly classified cases depends on the score densities g_0, g_1 which are a property of the classifier. Hand and Anagnostopoulos (2013) wrote that "using the area under the ROC curve is equivalent to evaluating different classifiers using different metrics" which exposes a possibly fatal flaw in the AUC.

They sought to fix this problem with the H measure, essentially by changing the weighting function $m(t)$ to be independent of the score densities. Following from the arguments in Hand (2009), the definition is given in terms of minimum misclassification costs. Suppose that $c = c_0/c_1$ is the cost ratio of misclassifying a $y = 0$ response to misclassifying a $y = 1$ response and assume $c \in (0, 1)$. Then the loss function for threshold value t is proportional to

$$L(c; t) = cp_0(1 - G_0(t)) + (1 - c)p_1 G_1(t). \quad (4.5)$$

For a given cost ratio, there is a unique threshold t that minimises the loss,

$$T_c = \operatorname*{argmin}_t L(c; t).$$

Instead of using a fixed cost ratio, a weighting function $w(c)$ is used to integrate over a range of possible cost ratios and we arrive at the weighted expected minimum loss,

$$\int_0^1 L(c; T_c) w(c)\, \mathrm{d}c. \quad (4.6)$$

The AUC can be recovered from (4.6) by taking the weighting function (9) given in Hand (2009), which depends on the score distributions and is therefore incoherent. The H measure was constructed to avoid this incoherency by taking the weighting function $w(c)$ to be the density of a Beta distribution, which will be independent of the classifier score distributions. In Hand (2009)

the author proposed that a Beta$(2, 2)$ distribution be used, but this was revised in Hand and Anagnostopoulos (2012) to be Beta$(p_1 + 1, p_0 + 1)$ to better deal with highly imbalanced data ($p_0 << p_1$ or $p_1 >> p_0$).

It is important to note, as Flach, Hernandez-Orallo, and Ferro (2011) noted, that the original interpretations of the AUC in terms of the probability of a uniformly chosen $0, 1$ pair being correctly ranked, remains coherent. Further, Hand and Anagnostopoulos (2013) showed that the AUC can be coherent when external constraints inform the choice of threshold (or equivalently, the cost ratio). An example of this is when the threshold is chosen such that a certain number of cases are flagged as $y = 1$ responses. This might be the case in resource limited settings where the user can only afford to inspect 100 cases and wants to use the model to nominate the 100 most likely objects.

A key feature of the H measure that separates it from the AUC, or indeed many other performance measures, is that it depends on the class probabilities p_0 and p_1. This property is linked with the issues Hand has with the AUC since both the average proportion of misclassified cases (4.4) and the average minimum misclassification cost (4.5) involve the class probabilities. Intuitively, this makes sense. As $p_1 \to 0$ there will be so few positive responses that the net cost of misclassifying these cases will be negligible and focus will shift entirely to correctly classifying the negative responses. The (potentially appealing) interpretation of the AUC as the probability that a uniformly randomly chosen $0, 1$ pair will be correctly ranked is impossible to keep if we demand that it also depend on the class probabilities.

The H measure is clearly designed to be used for assessing the performance of classifiers in conjunction with a threshold. In this context, Hand and coauthors have written numerous papers showing the incoherence of the AUC and have proposed a coherent alternative. However, if a binary response model is used to calculate probabilities and these are not converted to decisions using a threshold, the AUC can still offer a valid measure of discrimination (although

it comes with its own problems, discussed in section 4.2).

4.4 Measures of separation of the score distributions

An alternative approach to assessing the ability of a binary response model to discriminate between positive and negative responses is to look at the score distributions themselves. A good classifier will have score distributions that are in some sense well separated. On one extreme, a perfect classifier will have score distributions that are completely separate in the sense that the support of the two score distributions will not overlap. On the other extreme, the score distributions for a coin toss classifier would be equal. Of course, there are many ways to measure separation between distributions.

The *Kolmogorov–Smirnov* (KS) statistic between the two score distributions, $KS = \max_t |G_1(t) - G_0(t)|$, is a measurement of the separation between the distribution functions. The KS statistic is equivalent to the minimum proportion of cases misclassified if the cost ratio is fixed to be $c = p_1$ (Hand, 2012). The *overlap coefficient*, $OVL = \int \min\{g_0(t), g_1(t)\}\,\mathrm{d}t$ (Royston and Altman, 2010), is the area of the overlap of the two score densities and is equivalent to the KS statistic through the transformation $KS = 1 - OVL$ (Hand, 2012).

The *Tjur coefficient of discrimination*, defined as $D_{\mathrm{Tjur}} = \overline{\pi}_1 - \overline{\pi}_0$, where $\overline{\pi}_i$ is the mean model probability of observations with $y = i$. Assuming that $y = 1$ observations are generally scored higher than $y = 0$ observations, $0 \leq D_{\mathrm{Tjur}} \leq 1$. Tjur (2009) proves an asymptotic relationship between D_{Tjur} and the coefficient of determination for linear regression. Unfortunately there is a simple example where the Tjur coefficient of discrimination will perform poorly. Figure 4.2 shows two classifiers where the predicted probabilities for the $y = 0$ and $y = 1$ observations are uniform on different supports with means

0.4 and 0.6 respectively. The classifier in Figure 4.2(a) has g_0 with support $[0.35, 0.45]$ and g_1 with support $[0.55, 0.65]$ and is a perfect classifier, since the supports do not overlap. The classifier in Figure 4.2(b) has g_0 with support $[0.2, 0.6]$ and g_1 with support $[0.4, 0.8]$, showing that it is not a perfect classifier. Clearly both classifiers have $D_{\text{Tjur}} = 0.2$. This demonstrates a serious flaw in the Tjur coefficient of discrimination. In the next section we discuss an alternative that avoids this problem.

This problem is avoided by measures of effect size. *Hedges' g* is one such measure,

$$g = \frac{\overline{g_1} - \overline{g_0}}{s}$$

where $\overline{g_i}$ is the mean score of observations with $y = i$ and s is the pooled standard deviation of the two score distributions ($s^2 = ((n_0 - 1)s_0^2 + (n_1 - 1)s_1^2)/(n_0 + n_1 - 2)$) (Royston and Altman, 2010). Another popular choice of effect size is *Cohen's d*, which is related to Hedges' g by a scaling factor. These measures avoid the problem with the Tjur coefficient of discrimination we have outlined by scaling the difference between the classes by an estimate of the variance.

In general, all of these measures look at the degree of separation between the score distributions and it can be helpful to plot the score densities g_0 and g_1 (Hosmer et al., 2013; Royston and Altman, 2010). For logistic regression Tjur (2009) recommends inspecting plots of the conditional densities for the probabilities, $\mathbb{P}(\pi(\mathbf{Z}) \approx x \mid Y = i)$ for $i = 0, 1$, however we find these difficult to interpret when probabilties are very small. Visual assessment, along with measurements such as those discussed in this section, can help the user gauge the ability for the model to discriminate between the two groups.

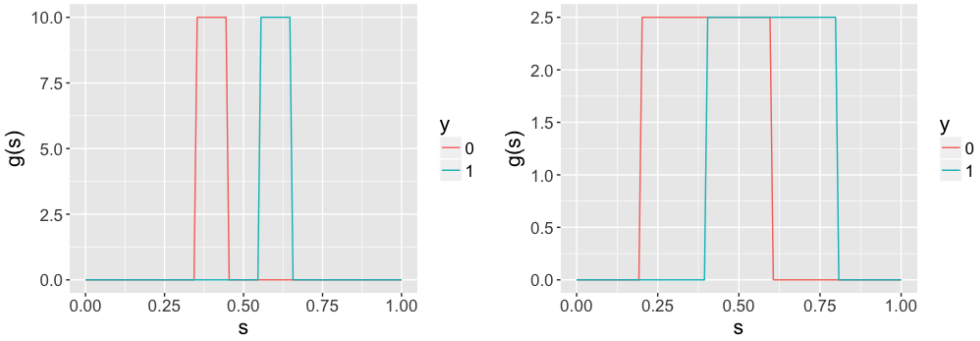


Figure 4.2: *Plots of score densities with equal D_{Tjur} values, however the left hand side shows a perfect classifier whilst the right hand side does not.*

4.5 Interpreting performance measures

Hand (1997) wrote that "classifier performance may be assessed for two basic reasons: to compare classifiers... or to determine an absolute measure of quality of performance". However both the relative and absolute interpretations of performance measures can be misleading.

Relative performance, comparing two classifiers against each other, can be a helpful component of model building for prediction. Different models will almost always report different values for performance measures and confidence intervals are needed for comparisons to be meaningful. These can usually be computed using bootstrapping methods.

There is also an idea of absolute performance in the literature. Guidelines are often given for what values of the AUC constitutes good performance. Hosmer et al. (2013) wrote that a value of the AUC betwen 0.8 and 0.9 shows excellent discrimination and a value about 0.9 shows outstanding discrimination. Harrell (2015) was more cautious and wrote that an AUC of greater than 0.8 indicates "some utility in predicting the responses of individual subjects". Yet as Royston and Altman (2010) noted, "judging by its presentation in the literature, interpretation of the c-index [AUC] is problematic". Indeed, as we

will see in the following sections, it can be very difficult to interpret performance measures in an absolute sense. In Sections 4.5.1 and 4.5.2 we look at the effect of the class probabilities p_0 and p_1 and the effect of 'obvious' 0s on the value of performance measures. This will demonstrate how the value of performance measures are highly dependent on the underlying data.

4.5.1 Influence of the class probabilities

Many performance measures, including the AUC, KS statistic and measures of effect size, are independent of the class probabilities p_0 and p_1. In the case of the aforementioned performance measures, this is because they depend only on the conditional score distributions G_0, G_1. This independence can lead to situations where a classifier is measured to have very good performance yet is useless in practice.

Suppose we have a classifier and have settled on a threshold t. In practice it is often important that the true positive rate is suitably high, or equivalently, that the $\mathbb{P}(Y = 1 \mid s(\mathbf{Z}) > t)$ is high. However,

$$\mathbb{P}(Y = 1 \mid s(\mathbf{Z}) > t) = \frac{p_1(1 - G_1(t))}{p_0(1 - G_0(t)) + p_1(1 - G_1(t))}$$

decreases to 0 as $p_0 \to 1$ and $p_1 \to 0$. So in this sense a classifier becomes less useful as $p_0 \to 1$, but performance measures such as the AUC, KS statistic and measures of effect size do not change with p_0 and hence do not reflect this change in performance. This issue has close parallels to the issues surrounding public screening for rare diseases, where the number of false positives can far outnumber the true positives rendering public screening useless even for tests with very low false positive rates.

The volume of false positives can be assessed by plotting the *precision* as the threshold value t varies. The precision is the ratio $TP/(TP + FP)$ where TP is the true positive rate and FP is the false positive rate. The quantity $1 - precision$ reflects how often the boy cries wolf. A model with low

precision may have high values of performance measures such as the AUC, KS statistic or Hedges' g, but forecasts from such a model will be rife with false positives and may be unusable. In settings with highly imbalanced classes we recommend inspecting plots of the the precision against threshold value alongside the score densities.

As discussed previously, the H measure is dependent on the class probability and so it may avoid the problem outlined in this section. However, it is not presently understood exactly how changes in the class probabilities affect the H measure. As such, using the H measure does not avoid the problem discussed in this section.

4.5.2 Influence of 'obvious' cases

In many practical situations, decisions are made about the extent of the data which is included in the modelling problem. For example, models for wildfire ignition may or may not include data from winter months or a species distribution model may or may not include spatial areas where the species is almost surely absent. There are good modelling reasons for choosing specific extents of the data included in a model, however these choices impact the value of performance measures. We illustrate this with an example.

Suppose a bank wishes to give credit scores to customers. In their world it turns out to be difficult to discriminate amongst employed people those who will default on their loans and those who will not. However unemployed people always default on their loan. The bank tasks their analyst with building a model for discriminating between credit worthy and unworthy customers. The modeller must choose whether to include or exclude the unemployed customers from their model. If the modeller builds a model only for employed customers then it will seem to perform poorly, but this is a result of the intrinsic difficulty of the problem. If the modeller builds a model for both the employed and unemployed customers then it will perform much better. As the proportion

of customers unemployed increases, the value of performance measures will also increase. The reason for this is that the first model tackles the 'real' problem, which happens to be difficult. The second model tackles a much easier problem, which is essentially distinguishing between the employed and unemployed customers.

We demonstrate this effect through simulations of a modified coin toss classifier. The classifier has 100 'real' observations which are difficult to discriminate between and the classifier flips a coin for these observations. We then added 'obvious' 0s to the data, which the classifier was able to perfectly separate from the 'real' observations. Figure 4.3 shows the value of various performance measures as the number of 'obvious' 0s increases. We see the 'obvious' 0s inflating the performance of the classifier.

This phenomenon is not a failure of the performance measures, they are working exactly as intended. The issue is that the choice of the extent of the data can influence the value of performance measures. We have demonstrated this by increasing the number of easily classified observations, but the concern applies more generally to any setting where a decision must be made about the extent of the data.

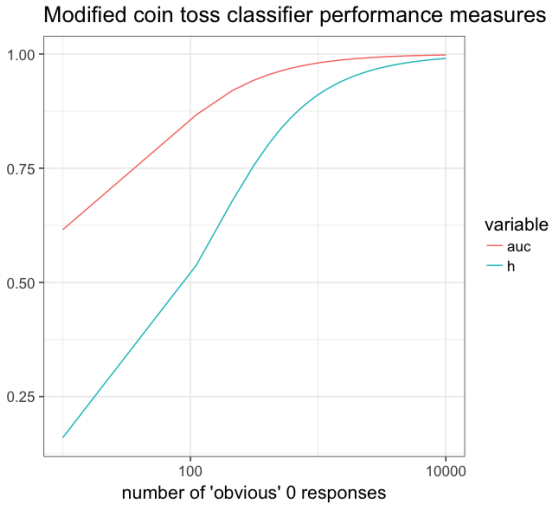


Figure 4.3: *The AUC and H measure of the modified coin toss classifier as the number of 'obvious' 0 responses increase. The classifier always correctly assigns the 'obvious' observations with a 0 response but then tosses a coin for the remaining observations. Hedges' g is not included because it is not bounded by 1, but it displays similar behaviour.*

4.6 Summary and recommendations

Binary response models have been used to calculate probabilities as well as to construct classifiers. Because of this overlap, performance measures intended for classifiers have been adopted to assess the ability for a binary response model to discriminate between responses in a probabilistic sense. The literature on performance measures can be difficult to navigate because of the subtly different objectives between classifiers and binary response models. This thesis is concerned with forecasting the probability of wildfire ignition and we do not believe that the model presented in Chapter 5 would produce a useful classifier.

For classifiers with a fixed threshold, performance is very clearly defined in terms of error rates and performance is easy to measure. However, for classifiers with no given threshold or for binary response models, it is difficult to interpret performance measures in an absolute sense, as demonstrated in Section 4.5. Performance measures such as the AUC and effect size (Section

4.4) in some way reflect the separation between the score distributions but give no indication of the precision $TP/(TP + FP)$ of the model, which is an important practical consideration. On the other hand, performance measures can be used to compare models using the same data set in a meaningful fashion, although some understanding of the variance of the performance measure needs to be considered.

When using performance classifiers for comparing relative performance of binary response models we recommend that the investigator does the following:

- calculates confidence intervals for performance measures;
- investigates multiple performance measures since they measure different aspects of performance;
- inspects plots of the score densities. We prefer these over plots of the probability densities since the later can be difficult to interpret for low probabilities; and,
- inspects plots of the precision against threshold value.

Investigators must be careful when drawing conclusions from performance measures. The correct interpretation of a performance measure must include an understanding of the data used to calculate the performance measure along with an understanding of what exactly is being measured.

Chapter 5

A logistic regression model for lightning-caused wildfire ignition

A review of regression models in the lightning ignition literature. An application of logistic regression to lightning ignitions. Detailed analysis of model output over several time periods. A discussion of the utility of the model along with practical considerations.

5.1 Introduction

Regression models are the most prevalent models in the wildfire ignition modelling literature. Tables 5.1, 5.2 and 5.3 detail the response variable, ignition cause, regression type, spatial resolution and location of study of many regression models found in the literature. Response variables are measured on a daily time scale, unless specified otherwise. Response variables denoted as 'large fire' are ignitions that result in fires whose final size is considered large; what constitutes large differs from one paper to another and can be a consequence of the local fire regime as well as the modelling objectives. In this section we discuss the regression models found in the literature; relevant

references can be found in the aforementioned tables.

In terms of spatial discrimination, models are either regional or gridded. Regional models look at ignitions in large irregularly shaped areas, such as states or national parks. Logistic regression is used to model the occurrence of a fire day as well as ignition presence in subregions (the presence of one or more ignitions on a day). It has also been used to model the occurrence of large fire days as well as large fire presence in subregions (ignitions that go on to be large fires). Poisson and negative-binomial regression have been used to model ignition counts, which can be used to calculate the probability of any number of ignitions rather than simply presence or absence of ignitions. Negative-binomial regression models are typically useful for over-dispersed count data, where the variance is too large for Poisson regression (Hilbe, 2011).

Regression models can also make use of spatial data on a regular grid (Section 3.1.1). Logistic regression models have been used to model ignition presence on scales varying from 1 km to 20 km. M. Rodrigues, Riva, and Fotheringham (2014) used logistic regression to model high/low ignition frequency, rather than ignition presence/absence. This was motivated by a misunderstanding of how regression models handle absences; namely that areas with no recorded ignitions are deemed to have zero probability of ignition by the model. Preisler, Chen, et al. (2008) and Preisler and Westerling (2007) coupled output from a logistic regression model for ignition presence with a logistic regression model for the probability of large fire given an ignition. Preisler, Westerling, et al. (2011) extended this by using extreme value theory to understand the distribution of large fire size. Collins, Price, and Penman (2015) and Oliveira et al. (2012) used linear regression with Normal errors to model the gridded annual ignition density on the log-scale.

Some studies restricted their focus to either human- or lightning-caused wildfire ignition. Models with restricted causes can be easier to interpret, since there is only a single physical process to consider, and may be more

accurate, since the relationships between the covariates and ignitions may differ between sources. Forecasts from multiple models for different ignition sources could easily be combined to produce net ignition likelihood forecasts if desired. Apart from the usual suspects, such as weather and fuel moisture covariates, models often include covariates such as population density and distance to the nearest road (Woolford, Bellhouse, et al., 2011) for human-caused wildfire ignition and lightning ground flash density (Collins et al., 2015) for lightning-caused wildfire ignition. Apart from the choice of covariates, there is essentially no difference in the modelling process between the causes for regression models.

Three more specialised techniques that have been used in the regression wildfire ignition literature are so called 'presence-only' regression, geographically weighted regression and mixed models. Presence-only regression uses the ignition points along with randomly generated 'pseudo absences' to estimate the relative likelihood, rather than absolute likelihood. This approach has been linked to point processes (Renner et al., 2015) and is essentially equivalent to using the coarse quadrature approximation to fit an inhomogeneous Poisson process, where the pseudo absences are quadrature points (Section 6.3.1). Geographically weighted (GW) regression has been used to allow the relationship between the response and predictor variables to vary across space (Martínez-Fernández, Chuvieco, and Koutsias, 2013; M. Rodrigues et al., 2014). This is done by replacing the regression coefficients by functions of space (Brunsdon, Fotheringham, and Charlton, 1998) and can result in a very large increase in the number of model parameters. Díaz-Avalos, Peterson, et al. (2001) used generalised linear mixed models (GLMM) (Wood, 2017) to model ignition presence, generalising the regression constant to be a random function of time and space. This can allow models to adapt to unobserved covariates that may be influencing the response variable.

A wide variety of approaches have been used for model selection. Guo

et al. (2016) considered 33 covariates and selected 11 for inclusion in the final model based on the significance of the Wald statistics (Section 3.2.1) of coefficients in intermediate models. Woolford, Bellhouse, et al. (2011) considered 16 covariates, choosing a model with 8 covariates based on an AIC and likelihood ratio test procedure. Magnussen et al. (2012) considered 70 covariates and settled on two models containing 15 and 19 covariates using stepwise regression (Hosmer et al., 2013).

In this chapter we apply a logistic regression model to gridded lightning ignitions data for the state of Victoria, Australia. We aim to produce a model that is suitable for both operational use and for improving long-term risk forecasts. At a minimum, this requires the model to produce daily, spatially explicit forecasts. We spend significant time interpreting the model in terms of the covariates driving it and demonstrate the model by analysing forecast likelihood maps for several periods, comparing them with observed ignitions.

	Response variable	Cause	Regression type	S.Res.	Location
Albert-Green et al. (2012)	fire day	lightning	logistic GAM	regional	Alberta and Ontario, Canada
Albertson et al. (2009)	fire day	combined	probit GLM	regional	English Peak District, UK
Bradstock et al. (2009)	large fire day	combined	bayesian logistic GLM	regional	Sydney, Australia
Collins, Price, and Penman (2015)	log annual ignition density	human, lightning	linear	regional	Victoria and NSW, Australia
Díaz-Avalos, Peterson, et al. (2001)	3 monthly ignition presence	lightning	logistic GLMM	2.5 km	Blue Mountains, Oregon, USA
Vega Garcia et al. (1995)	fire day	human	logistic GLM	regional	Alberta, Canada
Guo et al. (2016)	ignition presence	combined	logistic GLM	presence-only	southeast China
Haines et al. (1983)	fire day, ignition count given fire day, ignition count and large fire day	combined	weighted binomial	regional	northeast USA
Magnussen and Taylor (2012)	ignition presence, ignition count given ignition	lightning, human	logistic GLM, zero-truncated Poisson GLM	20 km	British Columbia, Canada
Mandallaz and Ye (1997)	ignition count	combined	Poisson GLM	regional	Europe

Table 5.1: See Section 5.1 text for details. S.Res. = spatial resolution.

	Response variable	Cause	Regression type	S.Res.	Location
Martell, Bevilacqua, and Stocks (1989)	fire day	human	logistic GLM	regional	north Canada
Martell, Otukol, and Stocks (1987)	fire day	human	logistic GLM	regional	Ontario, Canada
Martínez-Fernández, Chuvieco, and Koutsias (2013)	ignition presence, log annual ignition density	human	logistic GLM, GW regression, linear	regional	Spain
Mundo et al. (2013)	ignition presence	combined, lightning, human	logistic GLM	presence-only	Patagonia region, Argentina
Oliveira et al. (2012)	log annual ignition density	combined	linear	10 km	Europe
Penman, Bradstock, and Price (2013)	ignition presence	lightning, human	logistic GAM	presence-only	Sydney, Australia
Plucinski et al. (2014)	ignition count	human	negative-binomial GLM	regional	southwestern Western Australia
Preisler, Brillinger, et al. (2004)	ignition presence, large fire given presence	combined	logistic GAM	1 km	Oregon, USA
Preisler, Chen, et al. (2008)	monthly fire presence, large fire given presence	combined	logistic GAM	$1^\circ \approx$ 320 km	western USA

Table 5.2: See Section 5.1 text for details. S.Res. = spatial resolution.

	Response variable	Cause	Regression type	S.Res.	Location
Preisler and Westerling (2007)	monthly fire presence, large fire given presence	combined	logistic GAM	$1° \approx$ 320 km	western USA
Preisler, Westerling, et al. (2011)	monthly large fire presence	combined	logistic GAM	$0.125° \approx$ 40 km	California, USA
M. Rodrigues, Riva, and Fotheringham (2014)	high/low ignition presence	human	GW logistic GLM	1 km	Spain
Todd and Kourtz (1992)	ignition count	human	negative-binomial GLM	$0.25° \approx$ 220 km	Outaouais, Canada
Woolford, Bellhouse, et al. (2011)	ignition presence	human	logistic GAM	1 km	Ontario, Canada
Woolford, Cao, et al. (2010)	fire day	lightning	logistic GAM	regional	Ontario, Canada
Woolford, Dean, et al. (2014)	weekly ignition count	lightning	binomial GAM with mixture terms	regional	Ontario, Canada
Wotton, Martell, and Logan (2003)	ignition count	human	Poisson GLM	regional	eastern Canada
Wotton and Martell (2005)	ignition given lightning strike	lightning	logistic GLM	regional	Ontario, Canada
H. Zhang, Han, and Dai (2013)	ignition presence	regional	logistic GLM	1 km	northeastern China
Y. Zhang, Lim, and Sharples (2016)	ignition presence	regional	logistic GLM	1 km	southeastern Australia

Table 5.3: See Section 5.1 text for details. S.Res. = spatial resolution.

5.2 Data preparation

We split the data into training and validation data sets. The training data was constructed from the fire seasons (Section 2.2) from 1997–1998 to 2004–2005. The validation data was constructed from the fire season from 2005–2006 to 2011–2012.

The regression model in this chapter used the data described in Section 2.1, however some addition processing was required. To use a regression model with gridded data, all data must be projected onto the same regular grid. To this end we used the Geospatial Data Abstraction Library (GDAL) to project all covariate data (Table 2.2) onto the grid given by GDA94 VicGrid94 (EPSG:3111) with a spatial resolution of 20 km and extents $x_{\min} = 2120000, x_{\max} = 2940000, y_{\min} = 2260000$ and $y_{\max} = 2840000$. Choosing a finer spatial resolution would produce higher resolution forecasts, however it would increase the amount of data making inference a challenge and could decrease the accuracy of the model. We felt that a spatial resolution of 20 km was a good compromise between precision and accuracy whilst still being useful for operational decision making.

Using the wildfire ignitions data set, we produced daily grids with 1s in grid cells with one or more lightning ignitions and 0s elsewhere. We found 82% of 1 grid cells arose from exactly one ignition and 94% of 1 grid cells arose from two or fewer ignitions, suggesting that relatively little information is lost by doing this. Out of all grid cells, about 0.05% had more than a single ignition, which puts us safely in the realm of equivalence between logistic and Poisson regression approaches (Section 3.1.3).

The fuel covariates required slightly more work. The Victoria Fuel Layer consists of 37 layers with a spatial resolution of 30 meters, which we upscaled to a resolution of 1 km for computational reasons. We then combined these 37 layers into the 9 reduced fuel types discussed in Section 2.1 and for each of the nine reduced fuel types, the final 20 km resolution raster was calculated

to be the proportion of the 20 km cell covered by the reduced fuel type.

The final fuel data sets are rasters whose cells take values between 0 and 1 that represent the proportion of the grid cell covered by the fuel type. Since the sum of the fuel covariates is always 1, we removed the dry forest covariate to minimise issues of collinearity. Dry forest represents the largest proportion of ignitions by fuel type, so in some sense it can be thought of as the 'normal' condition and is factored into the constant term in the regression.

5.3 Method

We used a logistic regression GAM to model the presence of one or more lightning ignitions in a grid cell for a given day. Scaling of the covariates was done using fractional polynomials (Section 3.1.2) since they produce models with relatively compact analytic forms and can be easily implemented in various software environments. We didn't include interaction terms in the model, since we believed that not enough is understood about the lightning ignition process to determine a plausible set of interaction terms (Read et al., 2018).

For model selection we used the purposeful selection approach outlined in Section 3.2.1. We also built an approximating model to the final model using the covariate contributions method outlined in Section 3.3.1.

Using the validation data set described in Section 5.2, model fit was assessed with a Hosmer–Lemeshow test and by looking at diagnostic plots of fitted probabilities against the leverage, change in the Pearson chi-square statistic, change in deviance and Cook's distance (Section 3.4). We used the H measure, AUC and Hedges' g (Chapter 4) as indicators of model accuracy along with a plot of the precision against threshold. The aim of this study was to build a model suitable for operational use. Diagnostics and performance measures never fully capture the complexities of models and as such, we included model output and discussion from four periods of lightning-fire activity in the validation data set.

We performed all model building in R version 3.4.0, using the *stats* 3.2.4 package for fitting linear logistic regression models and the *mfp* 1.5.2 package (Ambler and Benner, 2015) for fitting fractional polynomial models. We performed the Hosmer and Lemeshow test using the *ResourceSelection* 0.3-2 package (Lele, Keim, and Solymos, 2017) and calculated the AUC and H measure with the *hmeasure* 1.0 package (Anagnostopoulos and Hand, 2012).

5.3.1 Deconstructing forecasts by covariate

To use the ideas discussed in Sections 3.6.1 and 3.6.2 for regression models with time and space dependent data, some extra care is needed to create a meaningful expected CC_j. Since we want the expected CC_j to represent the 'normal' conditions, it should reflect the local conditions rather than the global average. We define the j^{th} expected covariate contribution on the d^{th} day of the year at location u as

$$ECC_j(d,u) = \frac{1}{15n_y}\sum_{y'}\sum_{d'=d-7}^{d+7} CC_j(d',y',u),$$

where $CC_j(d',y',u)$ is the value of the j^{th} covariate contribution on day d' of year y' at location u, n_y is the number of years data used for model fitting and y' sums over all those years. This is an estimate of the expected value of the j^{th} covariate contribution over a fifteen-day window centred on day d. If the window is too small the ECC_j can be unduly influenced by a single bad fire day, whilst if the window is too large the ECC_j starts to reflect long term rather than local behaviour. The choice of fifteen days is arbitrary, but we believe it produces good smoothing without losing local behaviour. By using a local smoother, the ECC_j reacts to changes in ignition likelihood caused by geography and seasonal changes in the weather conditions.

For regression models with time and space dependent data, the covariate contribution anomaly of the j^{th} covariate on day d, year y at location u is

$$CCA_j(d,y,u) = CC_j(d,y,u) - EEC_j(d,u).$$

A positive value of $CCA_j(d, y, u)$ indicates that, for this time and location, the j^{th} covariate has an unusually positive effect on the probability of a lightning ignition. Likewise, a negative value of $CCA_j(d, y, u)$ signifies an unusually negative impact on the probability of lightning ignition. Spatial maps of each of the CCA_j can be produced for a given day, showing the relative influences of the covariates on the ignition likelihood.

5.3.2 Baseline logit-likelihood

Using a local average for ECC_j suggests that we should consider a local average equivalent of the centralised regression ceofficient, β_0^* (Sections 3.6.1 and 3.6.1). We define the baseline logit-likelihood on the d^{th} day of the year at location u as

$$BLL(d, u) = \beta_0 + \sum_j ECC_j(d, u), \tag{5.1}$$

where β_0 is the regression coefficient. The baseline-logit likelihood changes with time and location, allowing it to be interpreted as the expected likelihood at location u on the d^{th} day of the year, measured on the logit scale. If we denote the *total covariate contribution anomaly* by $TCCA(d, y, u) := \sum_j CC_j(d, y, u)$, then the ignition likelihood at location u on day d of year y can be expressed on the logit scale as

$$\text{logit}(\pi(d, y, u)) = BLL(d, u) + TCCA(d, y, u).$$

5.4 Results

Model selection resulted in a model including all candidate covariates (Table 2.2) except for the 850–500 hPa temperature lapse and plantation variables; these were dropped from the initial multivariate model in step 2 of purposeful selection (Section 3.2.1). This suggests that the information in these covariates was captured by other covariates in the model.

We inspected Nadaraya–Watson smooths (Section 3.2.1) between the individual covariates and the response variable on the logit scale (Appendix A). From these we chose to scale all covariates except the maximum temperature, wet forest, heath and residential variables. Some of the Nadaraya–Watson smooths corresponding with the vegetation layers are particularly rough. This is because these covariates do not change with time, producing a relatively small number of unique observations of the covariate. For example, the smooth of the residential covariates drops to 0 around the value of 0.47, however this is simply due to the lack of observations of the residential covariate at this value, rather than a real effect.

The terms in the final model are detailed in Table 5.4.

Covariate	first term	second term
Intercept	3.757	
Maximum temperature	$0.1012x$	
3 pm relative humidity	$5.052\log(x/100)$	$-17.38723\sqrt{x/100}$
Precipitation	$-0.09259(x+0.1)^{-1}$	$0.3369\sqrt{x+0.1}$
Wind speed	$-0.2542(x/10)^{-1}$	
DF	$0.9925x/10$	
KBDI	$4.056\sqrt{(x+0.1)/100}$	$-2.916(x+0.1)/100$
Maximum FFDI	$0.3590\log(x/10)$	$0.01749(x/10)^2$
CAPE index	$1.820\sqrt{(x+0.1)/100}$	$-0.2414(x+0.1)/100$
850 hPa Dewpoint depression	$-2.437(x+65.8)/100$	
Elevation	$2.298\sqrt{x/100}$	$-0.6946\sqrt{x/100}\log(x/100)$
Wet forest	$-0.4916x$	
Grassland	$-0.7623(x+0.1)^3$	
Heath	$-1.532x$	
Mallee spinifex	$0.9221((x+0.1)/0.1)^{-2}$	$-4.471((x+0.1)/0.1)^{-2}\log\big((x+0.1)/0.1\big)$
Residential	$-1.305x$	

Table 5.4: The final logistic regression model.

5.4.1 Diagnostics

Analysing plots of residuals proved to be challenging for such a large data set. Figure 5.1 contains the four plots discussed in Section 3.4.2. We looked for points which stood apart from the pack. From visual inspection we chose to treat points as large if they exceeded 0.004 for the leverage, 3000 for the change in Chi-square statistic, 16 for the change in deviance statistic or if they exceeded the curve $1.5 \times 10^{-3}/(x - 0.03)$ for Cook's distance. Observations that were large in two or more of the residuals were deemed suspicious and were investigated further.

Table 5.5 contains information on the 50 suspicious points, including the value of four of the most influential covariates (Section 5.4.2). There are five points with extremely large predicted probabilities ($\hat{\pi}_i > 0.1$), however there are 2136 predicted values with probabilities exceeding 0.1 so this does not show a widespread poor fit for large predicted probabilities. In general there is no obvious systemic pattern in the suspicious points (including in the covariates not listed in Table 5.5). However, there is some clear temporal dependence between the suspicious points, with many of them arriving within days of each other.

The model failed a Hosmer–Lemeshow goodness-of-fit test, reporting a p-value less than 10^{-7}. We believe that, due to the large size of the data set, very subtle deviations from the model assumptions are being detected (Johnson et al., 2007, page 187) and that this is not cause for concern.

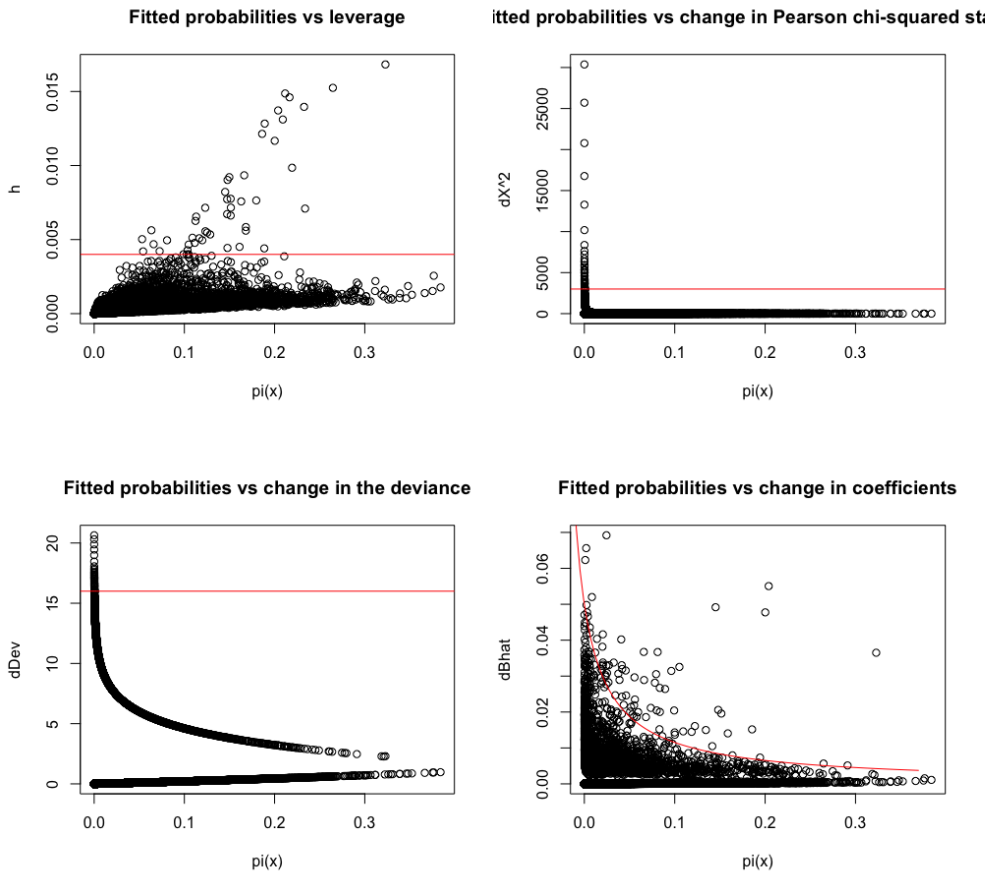


Figure 5.1: *Plots of the predicted probability against leverage, change in Pearson chi-squared statistic, change in the deviance and Cook's distance. Observations which were above the red curve on two or more of the residual plots were considered suspicious and investigated more closely.*

date	max. temp.	CAPE	max. FFDI	precipitation	$\hat{\pi}_i$
1997-10-27	19	4	2	0	0.0002
1997-11-27	21	1	2	0	0.0003
1997-11-27	21	0	2	1	0.0002
1997-12-04	22	0	4	0	0.0003
1998-10-17	39	784	116	0	0.2646
1998-10-17	39	687	112	0	0.1452
1998-10-17	38	786	113	0	0.3229
1998-10-19	18	3	9	1	0.0002
1998-12-26	23	0	27	1	0.0003
1998-12-26	23	0	29	2	0.0003
1998-12-27	25	0	20	0	0.0002
1999-01-14	23	0	2	0	0.0001
1999-01-25	24	42	1	0	0.0003
1999-04-23	19	0	5	0	0.0003
2000-01-07	21	0	1	0	0.0001
2000-01-12	23	4	3	0	0.0002
2000-11-04	22	16	5	0	0.0002
2001-04-26	22	3	5	0	0.0003
2001-04-27	19	4	2	0	0.0002
2001-04-29	21	0	6	0	0.0002
2001-12-02	36	170	110	1	0.2001
2001-12-02	36	169	116	1	0.2039
2001-12-03	20	18	4	1	0.0003
2002-04-28	17	0	1	0	0.0000
2002-10-28	14	0	5	0	0.0000
2002-11-04	21	0	15	1	0.0002
2002-11-04	21	0	14	0	0.0001
2002-11-04	21	0	14	0	0.0001
2002-12-07	24	0	17	0	0.0003
2003-02-03	24	0	6	0	0.0002
2003-04-01	19	16	2	0	0.0002
2003-11-25	19	0	4	0	0.0002
2003-12-05	24	0	3	22	0.0003
2003-12-06	22	0	1	0	0.0002
2003-12-07	24	0	2	0	0.0003
2003-12-24	28	0	4	0	0.0003
2004-10-09	16	0	1	0	0.0000
2004-10-16	13	1	1	6	0.0001
2004-10-19	24	0	8	0	0.0003
2004-10-29	20	0	3	0	0.0003
2004-11-23	19	0	3	0	0.0002
2004-12-07	19	6	0	14	0.0002
2004-12-09	19	62	0	10	0.0003
2004-12-13	24	101	1	0	0.0003
2004-12-25	28	0	24	0	0.0003
2005-01-21	21	0	0	28	0.0001
2005-01-30	29	0	17	0	0.0002
2005-02-05	24	0	4	4	0.0001
2005-03-03	22	0	2	0	0.0002

Table 5.5: Observations that had large values for two or more residuals. The shaded rows highlight residuals occurring on runs of days with no more than four days without a suspicious residual. The red highlights large values of the predicted probability.

5.4.2 Model interpretation

We interpret the model by understanding how the most influential covariates affect lightning ignition likelihood. Using the variance of the covariate contributions (Table 5.6) to judge relative importance, we found the CAPE index, precipitation, maximum temperature and maximum FFDI to be most influential. Estimates of the CC_j densities (Figure 5.2) for the covariate contributions of the CAPE index and precipitation have sharp peaks and heavy positive tails. This suggests that these covariates frequently exhibit little influence, but the large support of the CC_j suggests that they can be highly influential. The density estimate for the maximum temperature CC_j shows a relatively well spread distribution, suggesting that the maximum temperature is frequently highly influential. The maximum FFDI has a similar CC_j density estimate to the maximum temperature, but with longer tails. In particular, the length of the right-hand tail suggests that the maximum FFDI can be highly influential on extreme fire days. Plots of the CC_j densities for all covariates can be found in Appendix B.

Note that the spikes in both the CAPE index and precipitation plots may be misleading. Both the CAPE index and precipitation take value 0 with a high frequency, suggesting they would be better modelled by a mixed type distribution which has a discrete mass at 0 and a continuous part on the positive real numbers. Kernel density estimates assume that the distribution is continuous everywhere, which is incorrect, producing the spikes to the left side of the CC_j densities. However, these plots still give a good indication of how the CC_j behaves.

Plots of the scaling functions (Figure 5.3) show that the CAPE index attains its maximum positive influence on ignition likelihood at about 1400, whilst the precipitation, maximum temperature and maximum FFDI continue to have increasingly positive influences as they increase in value. The scaling functions for the CAPE index, maximum FFDI and precipitation all have

sharp drops towards 0, showing that that the zero/nonzero state of these covariates has a large influence on the predicted lightning ignition likelihood. Plots of the scaling functions for all covariates can be found in Appendix C.

Wet forest	0.013	KBDI	0.122
Residential	0.015	3pm RH	0.135
Wind speed	0.025	Mallee spinifex	0.187
850 hPa dewpoint depression	0.029	Maximum FFDI	0.251
Heath	0.052	Maximum temperature	0.425
DF	0.062	Precipitation	0.440
Elevation	0.070	Cape	0.545
Grassland	0.109		

Table 5.6: The variances of the CC_j. Relatively small CC_j suggest that the covariate is relatively uninfluential in the model.

Figure 5.4 shows the baseline logit-likelihood, (5.1), for the 1$^{\text{st}}$ of January, which is representative of the baseline logit-likelihood for other days of the year. In some sense the baseline logit-likelihood can be interpreted as the expected likelihood on the logit scale. The plot shows that the baseline logit-likelihood was lower than average in the Mallee area in the northwest and was above average in central and eastern Victoria. The baseline logit-likelihood at a set location varied by no more than 1.29 across the year. This change was caused by seasonal shifts in the expected values of the dynamic covariates across the year. This contrasts with the difference in logit-likelihood across space, which had a maximum difference of 3.55. This shows that changes in the static covariates (elevation, wet forest, grassland, heathland, Mallee spinifex and residential) are most influential in determining the baseline logit-likelihood of the model.

Evaluating the model with the validation data we recorded an H measure of 0.449 [0.436, 0.467], an AUC of 0.877 [0.871, 0.883] and Hedges' g value of 1.785 [1.735, 1.821], where the bracketed values are bootstrapped 95% confidence intervals. The value of the AUC and Hedges' g suggests good separation between the score distributions which is supported by Figure 5.5(a), showing

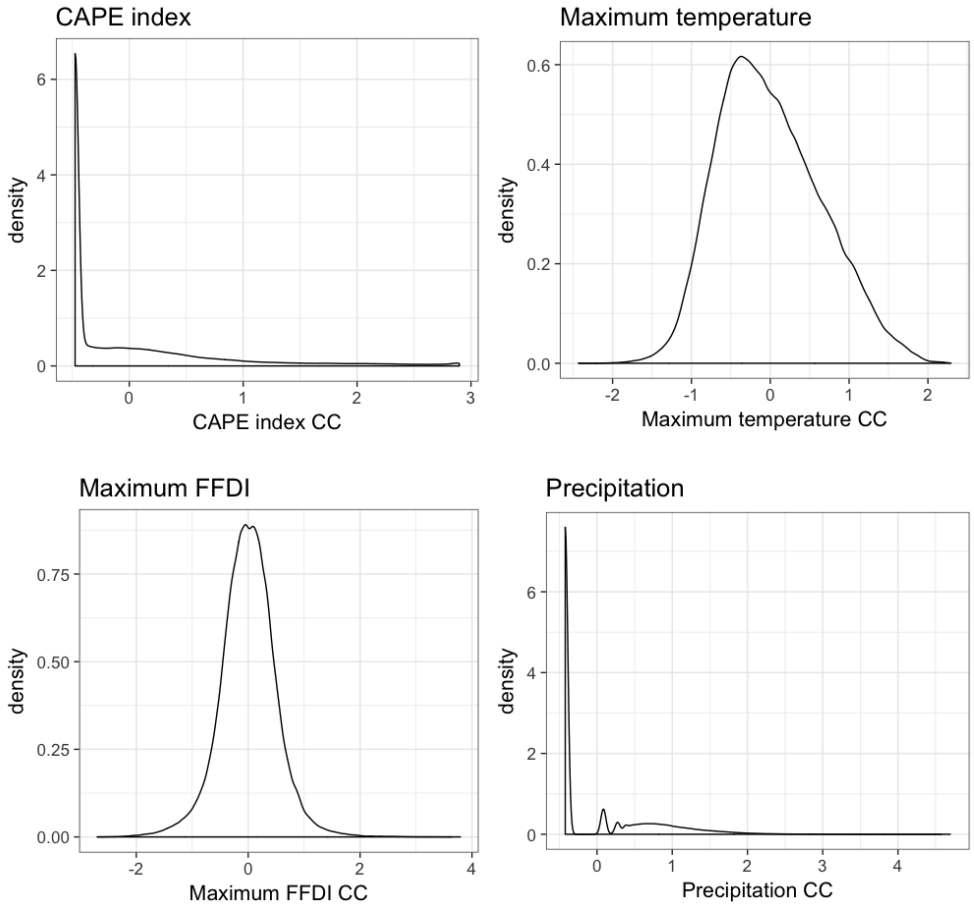


Figure 5.2: *Density estimates of the covariate contributions for the more influential covariates. The value of the covariate contributions have been centralised.*

the positive and negative score densities calculated with the validation data. Figure 5.6(a) shows the precision as the threshold varies. If we consider only those observations scored higher than a chosen threshold, the precision reflects the proportion of those observations that turn out to be positive responses. Although we have no intention of using the model as a classifier, this plot gives an indication of the volume of highly scored negative responses. The precision of the model is generally fairly low, which is to be expected given the general low likelihood of lightning ignition.

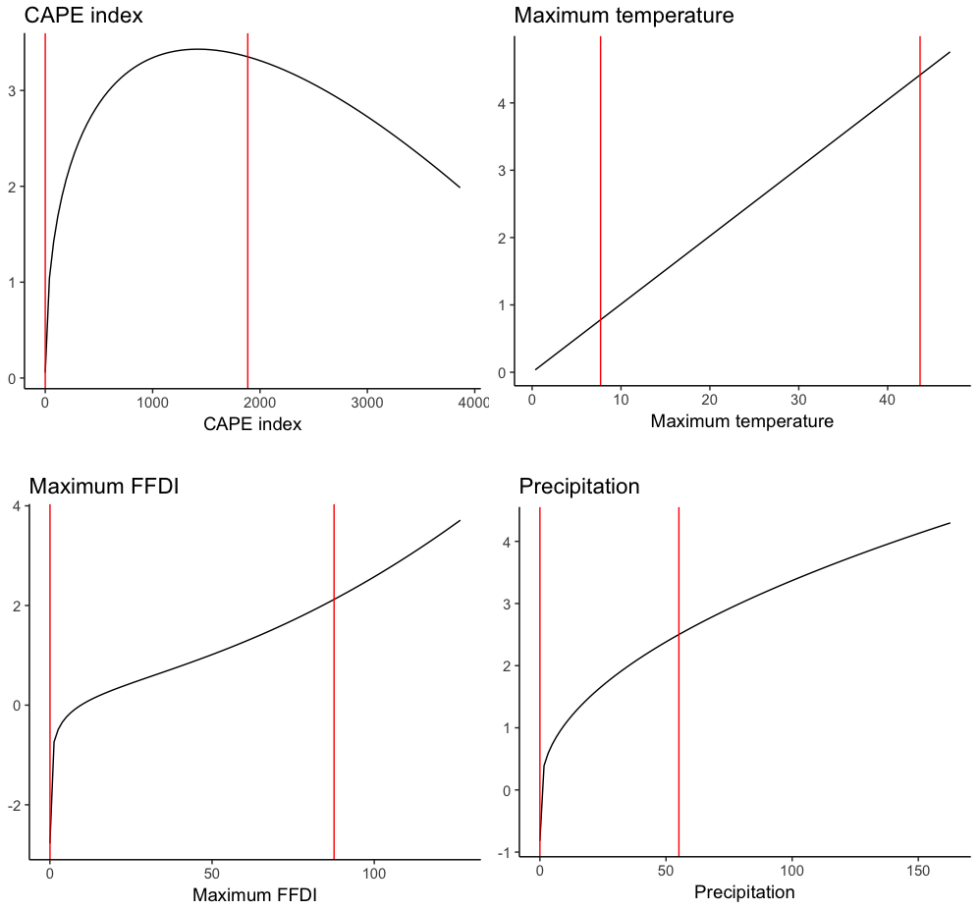


Figure 5.3: *The scaling functions for the most influential covariates included in the model. The red vertical lines show the 0.025 and 0.975 quantiles for the covariates.*

5.4.3 Approximating model

We created an approximating model by removing the relatively uninfluential covariates, as determined by the variance of the corresponding CC_j (Table 5.6). On this basis we removed the wet forest, residential, wind speed and 850 hPa dewpoint depression covariates and adjusted the regression coefficient to include the mean value of the respective CC_js. The approximating model consisted of the remaining terms in the full model with an adjusted regression coefficient of value 1.634. The approximating model was compared to the full model using performance measures (Table 5.7), calculated on the training

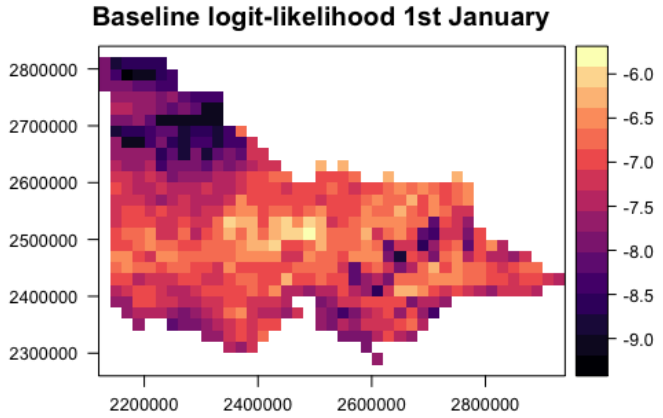


Figure 5.4: *The baseline logit-likelihood map for the first of January. Other days of the year have very similar baseline logit-likelihood maps.*

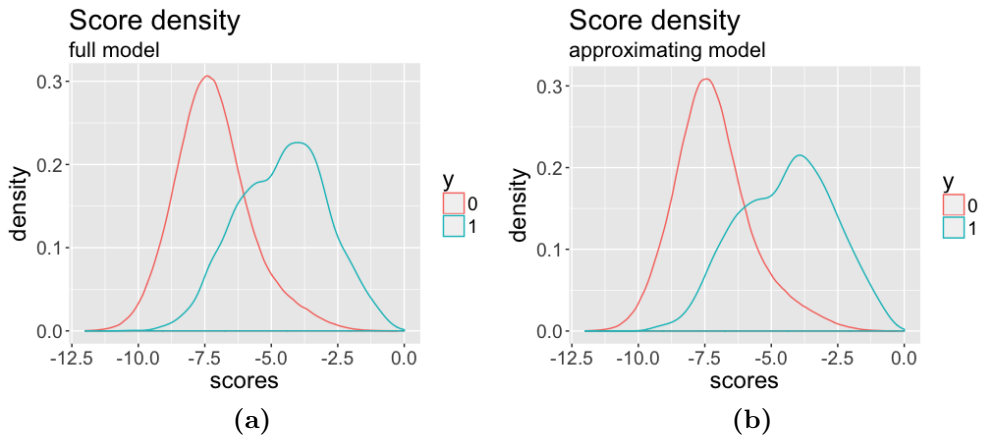
data set, since this is the only data available during model fitting.

Model	H	AUC	Hedges' g
full	0.522 [0.506, 0.538]	0.897 [0.892, 0.903]	2.483 [2.421, 2.538]
approximating	0.512	0.893	2.520

Table 5.7: The value of performance measures on the training data for the full and approximating model

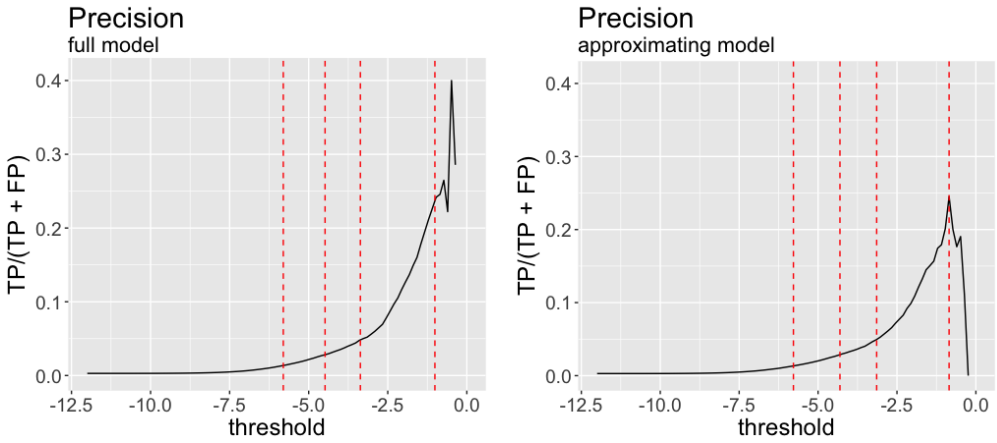
The approximating model had slightly lower values of the H measure and AUC and a slightly higher value of Hedges' g than the full model. However, the value of all performance measures fell within the respective 95% confidence intervals for the full model, suggesting that the change in performance was negligible. Furthermore, the score (logit probability) densities for the approximating model (Figure 5.5(b)) were visually similar to those for the full model (Figure 5.5(a)).

The R^2 value between the scores (the logit probability) of the full and approximating models was 0.959, showing good agreement between the two



(a) (b)

Figure 5.5: *Estimates of the score densities for the full and approximating model calculated using the validation data. Seperation between the densities reflects a good classifier. Here we have very similar distributions between the full and approximating models.*



(a) (b)

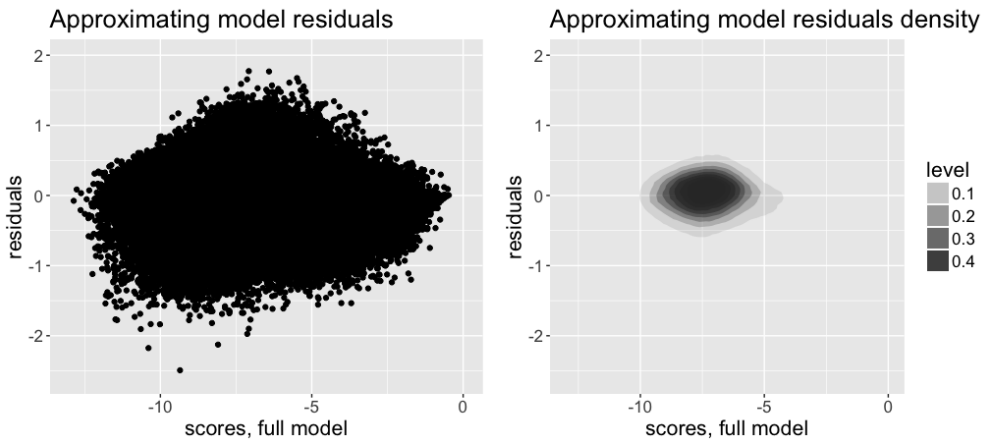
Figure 5.6: *The precision of the model as the threshold value changes. The red dashed lines show the 0.25, 0.5, 0.75 and 0.99 quantiles of the positive response score distribution. Unsurprisingly, the precision is fairly low for moderate values of the threshold.*

models. However we were concerned that there could still be important disagreements between the full and approximating models, prompting us to inspect a plot of the difference between the scores of the full and approximating models (Figure 5.7). The score differences are the residuals of the approximat-

ing model if we consider it as regressing on the scores of the full model. It can be difficult to interpret residual plots with so many points, but Figure 5.7(b) shows that most of the residuals are small. There are two interesting things to note about the residuals. First, they are smaller for large positive values of the full model scores. This tells us that the approximating model works relatively well during extreme fire weather conditions. Secondly, when considering large observations of the residuals, more emphasis should be given to the positive residuals than the negative residuals since these correspond with the approximating model underestimating the likelihood estimated by the full model. In the context of wildfire ignition modelling, these errors are more important than overpredicting risk (the negative residuals). With this in mind, the residuals show that the approximating model tends to perform well even in extreme conditions, with most of the large discrepancies occurring due to the approximating model overpredicting risk.

The plot of the precision as threshold varies (Figure 5.6(b)) shows a different story. The precision of the approximating model is slightly below that of the full model between the 0.75 and 0.99 quantiles of the positive response score distribution. However after the 0.99 quantile the precision of the approximating model drops sharply. This region represents the top 0.01% of scores of the validation data, although this still accounts for 102 observations.

Finally we assessed the performance of the approximating model on the validation data set, which gave an H measure of 0.446, an AUC of 0.875 and Hedges' g of 1.811. Again, they all fell within the 95% bootstrapped confidence intervals of the full model, indicating that the approximating model is performing well.



(a) (b)

Figure 5.7: *The residuals show that most of the discrepancy between the full and approximating models occurs in the bulk of the scores.*

5.5 Case studies

To understand the behaviour of the model better, we analysed model output from four periods chosen from the validation data. The four periods were chosen to reflect circumstances of interest to fire managers.

Black Saturday - February 2009. The period surrounding Black Saturday, February 7$^{\text{th}}$ 2009, resulted in the largest loss of life of any wildfire in Australia. Although the most devastating fires during this period were attributed to arson, there was still significant lightning-caused wildfire activity and the weather conditions during this period have formed the basis for what is considered extreme fire weather in Australia (Stephenson, 2010).

February 2010. The six-day period of the 9$^{\text{th}}$–11$^{\text{th}}$ February 2010 saw 152 lightning ignitions, putting it in the top 1% of six-day periods between 2005 and 2012 in terms of numbers of lightning ignitions. This period was constructed by selecting a day uniformly at random from those exceeding the 0.99 quantile of observed number of lightning ignitions.

January 2006. The six-day period of the 18^{th}–23^{rd} January 2006 included two days with 35 or more forecast $y = 1$ cells, reflecting very high forecast probabilities. This period was contructed by selecting a day uniformly at random from those whose predictions exceeded the 0.99 quantile of predicted probabilities

January 2011. The six-day period of the 20^{th}–25^{th} January 2011 was constructed from a day selected uniformly at random amongst those of moderate forecast lightning ignition risk. This was chosen to contrast the previous three examples, which were chosen to illustrate model performance under extreme conditions.

After randomly selecting a day, the surrounding six-day period was chosen subjectively to reflect the period of most practical interest. All model output was calculated from the full model.

5.5.1 Black Saturday - February 2009

Figure 5.8 shows the observed lightning ignitions superimposed on the forecast lightning ignition likelihood for the period of the 5^{th}–9^{th} February 2009, which includes Black Saturday on the 7^{th}. Below each forecast is information giving a simulated 95% confidence interval for the number of lightning ignition presence cells, along with the observed number of lightning ignition presence cells. The model correctly identified this period as generally high likelihood, but failed to correctly identify the precise spatial and temporal distribution of lightning ignitions.

The forecast likelihood for the 7^{th} was very high, with a 95% prediction interval on the number of presence cells of $[14, 31]$, which failed to match the observed 6 ignition presence cells in either quantity or spatial position. In fact, the forecast on the 7^{th} would better match the observed ignitions on the 8^{th}. The model underpredicted ignitions on the 8^{th} and 9^{th}, finding 95% prediction intervals of $[0, 6]$ and $[0, 2]$ ignition presence cells respectively rather than the observed 19 and 11 ignition presence cells.

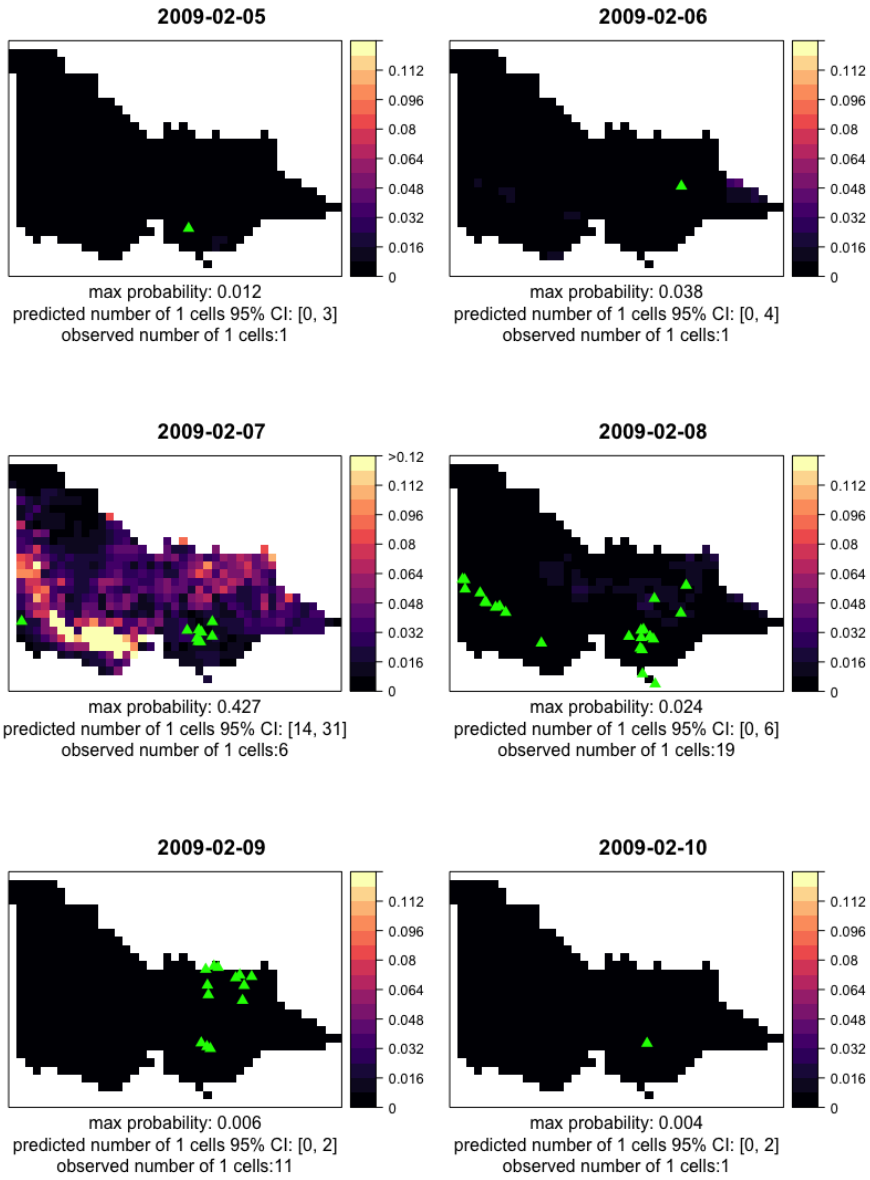


Figure 5.8: *Lightning ignition likelihood forecasts for the period surrounding Black Saturday, February 2009. The green triangles are the observed lightning ignitions. Note that the colour legend presents all likelihood values that exceed* 0.12 *as the same colour. This was done to prevent days with lesser likelihood appearing too dark and obscuring details.*

5.5.2 February 2010

The model performed exceptionally well for the six-day period of the 8$^{\text{th}}$–13$^{\text{th}}$ February 2010, which was chosen to reflect a period with a large number of observed lightning ignitions. Figure 5.10 shows the model capturing a range of behaviour shown by the observed ignitions. The model correctly classified the days of the 9$^{\text{th}}$ through to the 11$^{\text{th}}$ as high likelihood and the days of the 8$^{\text{th}}$, 12$^{\text{th}}$ and 13$^{\text{th}}$ as much lower likelihood. It did a good job of forecasting the spatial distribution of ignitions, correctly producing a band of high risk through the middle of the state.

Figure 5.9 shows the total covariate contribution anomaly for the 9$^{\text{th}}$ and 10$^{\text{th}}$. For both days the lightning ignition likelihood was higher than usual, with the exception of the far eastern tip of Victoria, which was slightly below average. From the 9$^{\text{th}}$ to the 10$^{\text{th}}$ there was a general increase in likelihood and a sharp increase in likelihood in central Victoria.

Figure 5.11 and 5.12 shows the corresponding covariate contribution anomalies. Focusing on the 9$^{\text{th}}$, the CAPE index was the primary driver of the high likelihood, with its CC_j well above average across the state. The maximum temperature and precipitation were also highly influential. The Precipitation CC_j was mostly below average but with high patches in the northeast. The other covariates were relatively less influential on the 9$^{\text{th}}$; the maximum FFDI had roughly the same influence as the sum of all the remaining dynamic covariates. The overlap of areas of high CCA_j for the CAPE index, maximum temperature and precipitation resulted in the region of elevated likelihood stretching from the southwest to the northeast of the lightning ignition likelihood map for the 9$^{\text{th}}$ seen in Figure 5.10.

The total covariate contribution anomaly increased from the 9$^{\text{th}}$ to the 10$^{\text{th}}$ and the covariate contribution anomalies for the 10$^{\text{th}}$ show that this was mostly driven by shifts in the precipitation. The maximum temperature CCA_j intensified its hold on the southwest and the CAPE index CCA_j continued

to be above average in the northwest, but a significant increase formed in the southwest. The overlapping areas of significantly above average CCA_j values for maximum temperature, CAPE index and precipitation in the south west gave rise to the band of high likelihood seen on the 10^{th} of Figure 5.10.

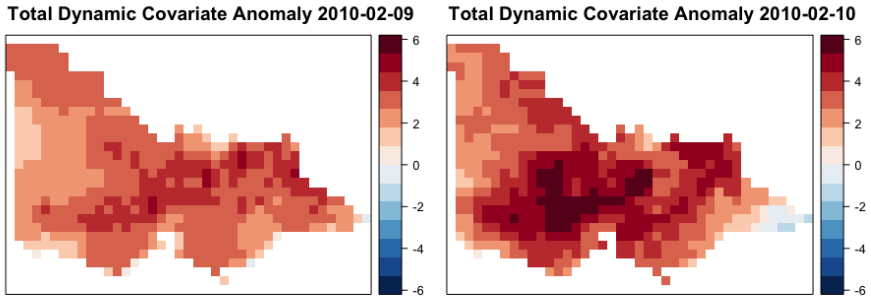


Figure 5.9: *Total covariate contribution anomaly for the 9^{th} and 10^{th} of February 2009. This represents the influence of dynamic covariates on the ignition likelihood.*

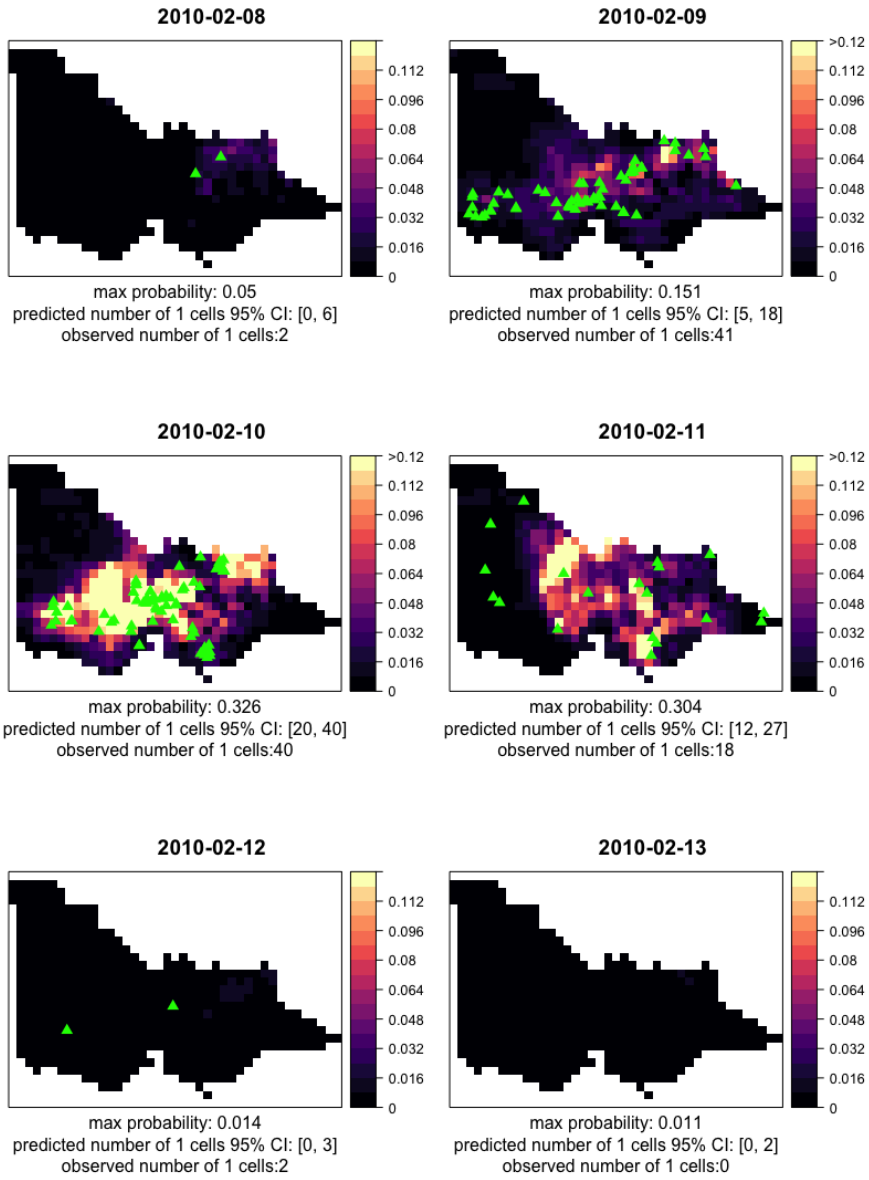


Figure 5.10: *Lightning ignition likelihood forecasts for a randomly chosen period of elevated lightning ignition activity. The green triangles are the observed lightning ignitions. Note that the colour legend presents all likelihood values that exceed* 0.12 *as the same colour. This was done to prevent days with lesser likelihood appearing too dark and obscuring details.*

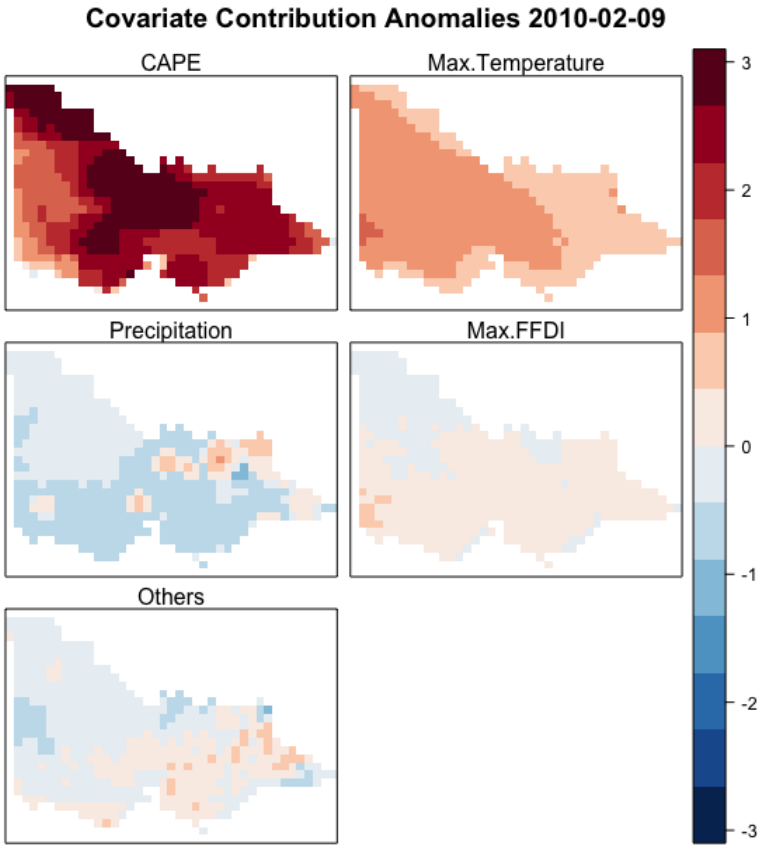


Figure 5.11: *The Covariate Contribution Anomalies (CCA_j) for the 9^{th} of February 2010. Positive (negative) values of the CCA_j suggest an above (below) average contribution to the lightning ignition likelihood. The relatively uninfluential dynamic covariates (wind speed, DF, KBDI, 850 hPa dewpoint depression and 3 pm relative humidity) have been grouped under 'others'.*

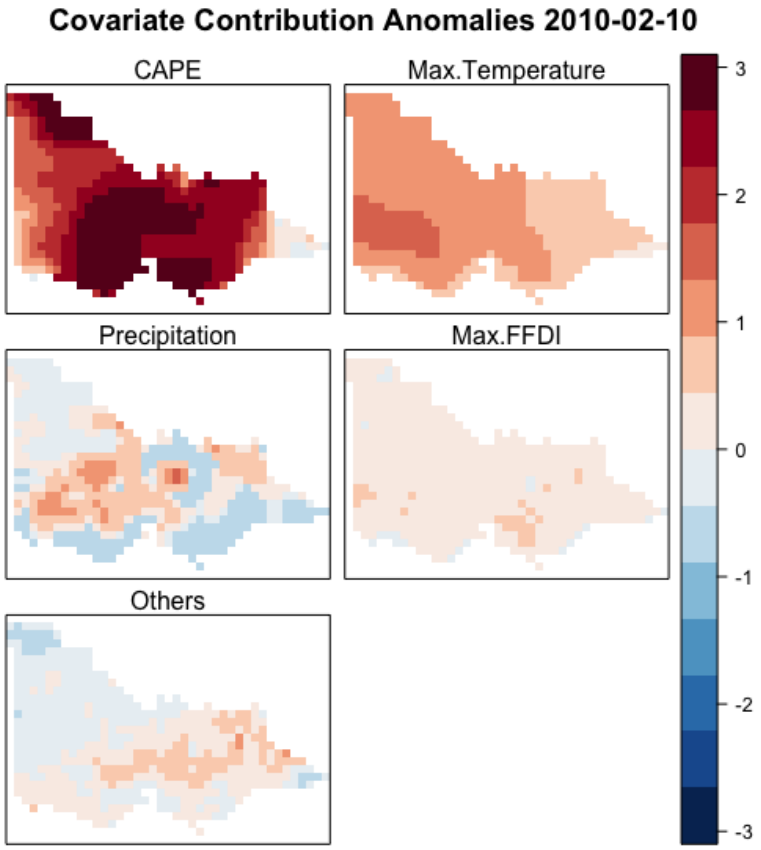


Figure 5.12: *The Covariate Contribution Anomalies (CCA_j) for the 10^{th} of February 2010. Positive (negative) values of the CCA_j suggest an above (below) average contribution to the lightning ignition likelihood. The relatively uninfluential dynamic covariates (wind speed, DF, KBDI, 850 hPa dewpoint depression and 3 pm relative humidity) have been grouped under 'others'.*

5.5.3 January 2006

The model performed very well for the randomly chosen period of high predicted likelihood. Figure 5.13 shows the predicted lightning ignition likelihood for the 18$^{\text{th}}$–23$^{\text{th}}$ January 2006. The model predicted extreme likelihood for the 20$^{\text{th}}$ and the 21$^{\text{st}}$ and both days saw a very large number of observed lightning ignitions, however it under predicted on the surrounding days of the 19$^{\text{th}}$ and 22$^{\text{nd}}$, which both saw a relatively large number of lightning ignitions. The model correctly predicted low likelihood on the 18$^{\text{th}}$ and 23$^{\text{rd}}$.

The spatial distribution of the predicted likelihood did a relatively good job of mirroring the observed lightning ignition patterns, even on the days when the model under predicted. On the 18$^{\text{th}}$, 19$^{\text{th}}$ and 22$^{\text{nd}}$ the model correctly identified the areas containing the observed ignitions. From the 20$^{\text{th}}$ to 21$^{\text{st}}$ the area of extreme likelihood shifted from southwest Victoria towards central Victoria, which was reflected in the observed lightning ignition pattern.

The total covariate contributions anomaly was significantly above zero across the state for the 19$^{\text{th}}$, 20$^{\text{th}}$ and 21$^{\text{st}}$ (Figure 5.14). The CCA_j in Figures 5.15, 5.16 and 5.17 shows that the CAPE index was again the primary driver of the likelihood, however the maximum temperature, precipitation and maximum FFDI were also highly influential. A band of elevated CAPE index CC_j moved from west to east across the state, whilst both the maximum temperature and precipitation CCA_j increased in west and central Victoria over the three days from the 19$^{\text{th}}$–21$^{\text{st}}$.

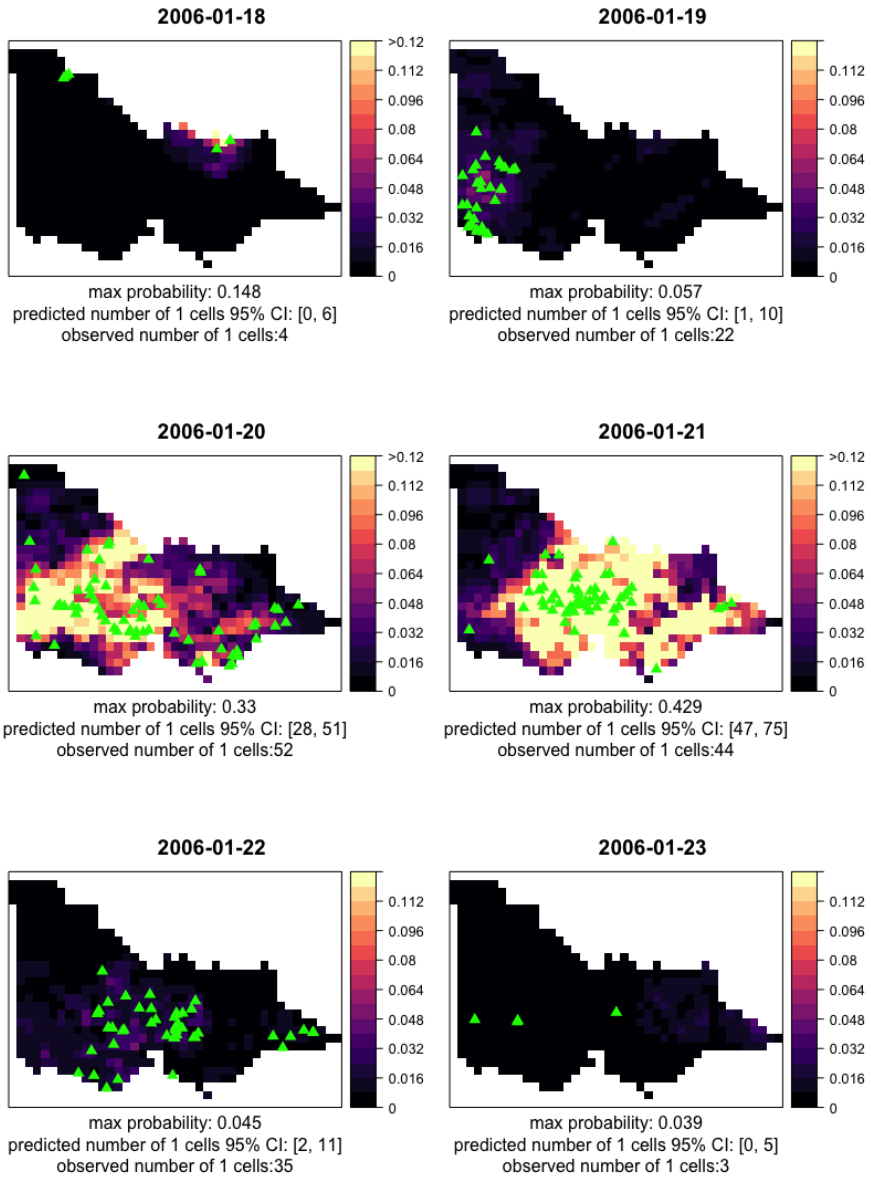


Figure 5.13: *Lightning ignition likelihood forecasts for a randomly chosen period with high predicted lightning ignition likelihood. The green triangles are the observed lightning ignitions. Note that the colour legend presents all likelihood values that exceed* 0.12 *as the same colour. This was done to prevent days with lesser likelihood appearing too dark and obscuring details.*

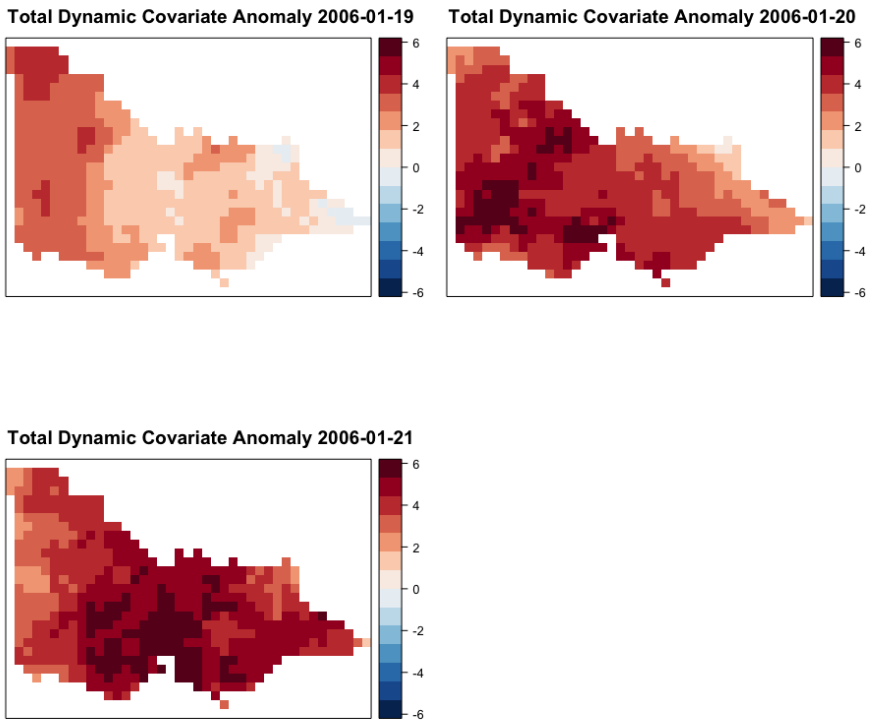


Figure 5.14: *The total covariate contribution anomaly shows the difference between the expected the observed influence of the covariates on the predicted likelihood. The ignition likelihood was significantly above average for this three day period, especially in west and central Victoria.*

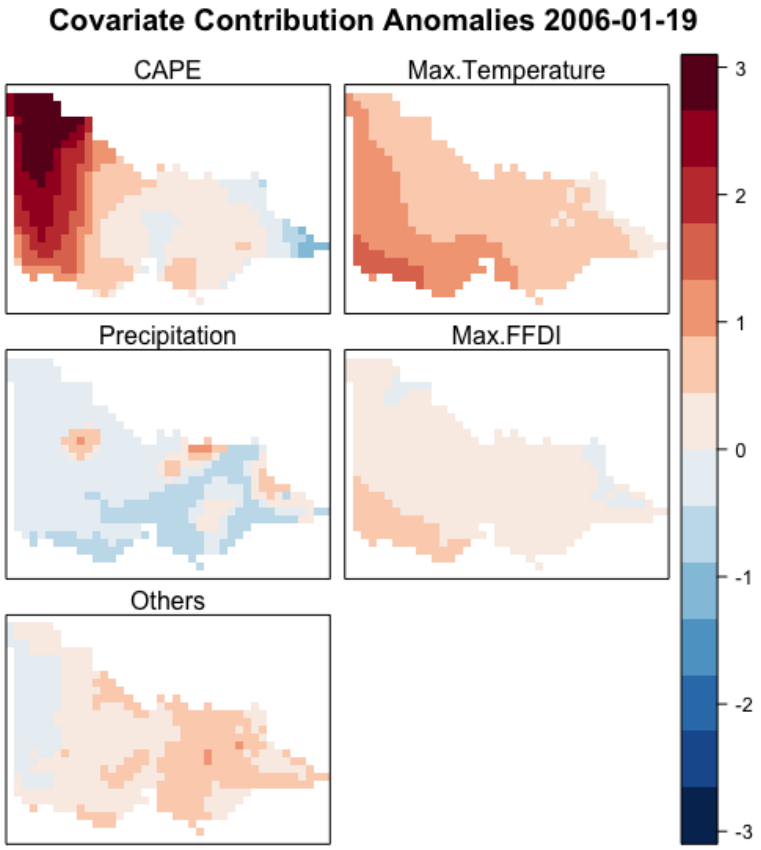


Figure 5.15: *The Covariate Contribution Anomalies (CCA_j) for the 19^{th} of January 2006. Positive (negative) values of the CCA_j suggest an above (below) average contribution to the lightning ignition likelihood. The relatively uninfluential dynamic covariates (wind speed, DF, KBDI, 850 hPa dewpoint depression and 3 pm relative humidity) have been grouped under 'others'.*

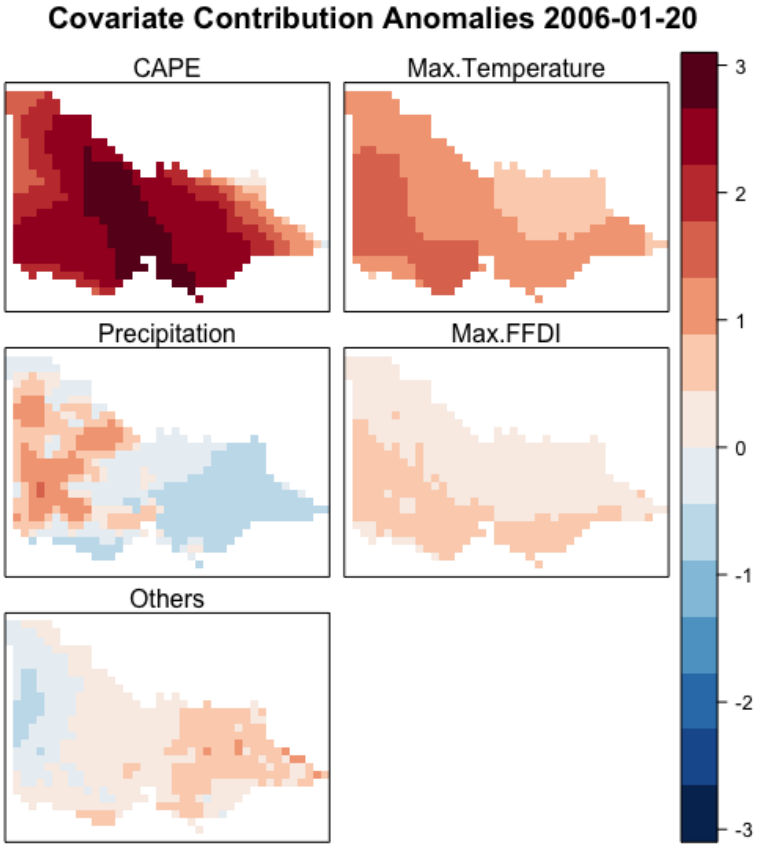


Figure 5.16: *The Covariate Contribution Anomalies (CCA_j) for the 20^{th} of January 2006. Positive (negative) values of the CCA_j suggest an above (below) average contribution to the lightning ignition likelihood. We have grouped the relatively uninfluential dynamic covariates (wind speed, DF, KBDI, 850 hPa dewpoint depression and 3 pm relative humidity) under 'others'.*

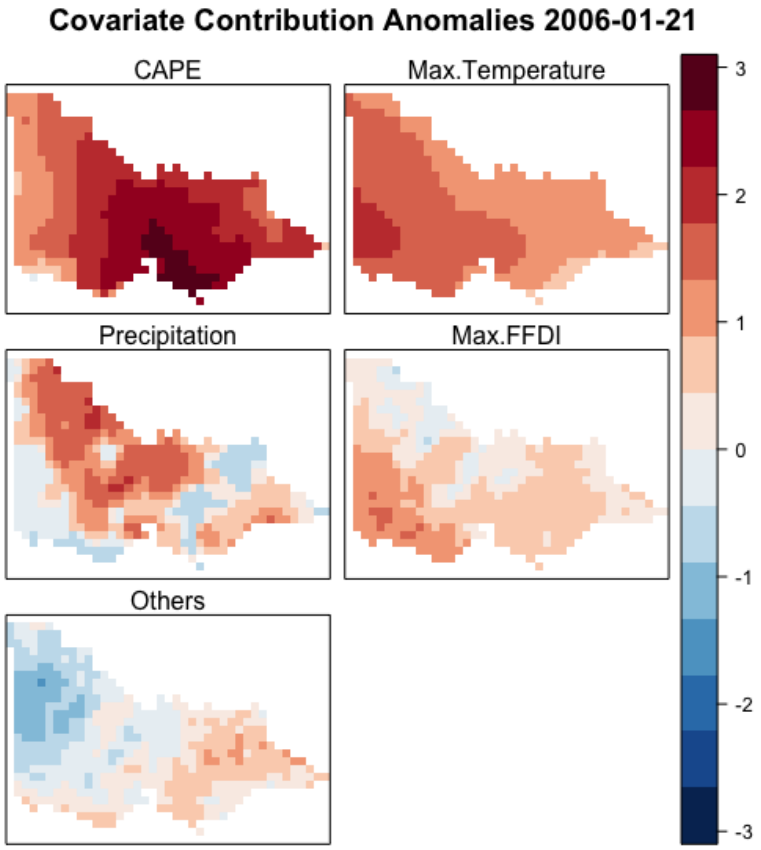


Figure 5.17: *The Covariate Contribution Anomalies (CCA_j) for the 21^{st} of January 2006. Positive (negative) values of the CCA_j suggest an above (below) average contribution to the lightning ignition likelihood. We have grouped the relatively uninfluential dynamic covariates (wind speed, DF, KBDI, 850 hPa dewpoint depression and 3 pm relative humidity) under 'others'.*

5.5.4 January 2011

Figure 5.18 shows ignition likelihood forecasts for the randomly chosen moderate period. The model generally performed well, showing generally low levels of ignition likelihood across the period. The likelihood increased for the days of the 21st–23rd, which coincided with the observed ignitions. The model also captured the observed spatial pattern, with the areas of elevated ignition likelihood concentrated in the east of the state.

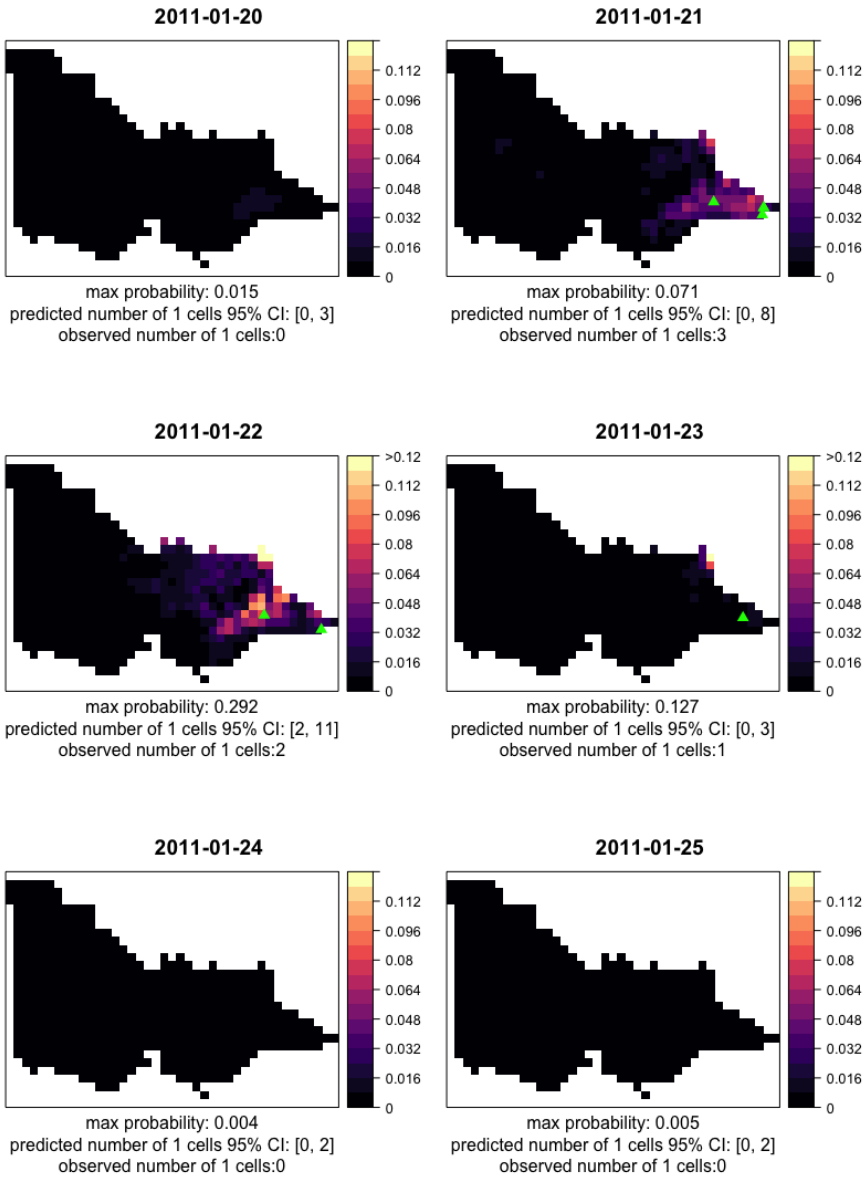


Figure 5.18: *Lightning ignition likelihood forecasts for a randomly chosen period with moderate observed lightning ignition. The green triangles are the observed lightning ignitions. Note that the colour legend presents all likelihood values that exceed 0.12 as the same colour. This was done to prevent days with lesser likelihood appearing too dark and obscuring details.*

5.6 Discussion

In this chapter we have presented a logistic regression GAM for forecasting lightning-caused ignitions for the purpose of decision making. The model produces daily, spatially-explicit forecasts using covariates that reflect both convective activity and fire behaviour. An analysis of diagnostics showed good model fit, which was confirmed through four case studies. The model performed well during randomly chosen periods of high observed lightning ignition as well as periods of high predicted likelihood, identifying temporal as well as spatial areas of elevated risk. Performance during the period of Black Saturday was less impressive, but models cannot be expected to perform well outside of the range of data used for model fitting — the fire weather conditions during this period were the most extreme on record. Nevertheless, during this period the model still demonstrated the ability to distinguish areas of high and low lightning ignition likelihood with sufficient accuracy to benefit operational use. The final case study showed the model performing well during a randomly chosen period with moderate conditions.

Due to the independence assumptions of regression models we see some systematic errors produced in the forecasts. The clustering nature of convective storms is passed down to lightning ignitions, which tend to cluster in both space and time. The model struggled to replicate this clustering behaviour, frequently resulting in the model failing to identify consecutive days of elevated lightning ignition likelihood. For example, in the January 2006 case study the model recognised two days as having extremely high ignition likelihood, yet the observed lightning ignitions were extreme for four consecutive days. The model provided a relative indication of ignitions; a similar issue has been raised in another study where the point was made that fire managers are comfortable with predictions that provide a reliable indication of the severity of the day and do not expect models to provide exact numbers (Plucinski et al., 2014). From an operational perspective, these outputs

are still valuable; the model recognises periods of extreme lightning ignition likelihood but underestimates their length.

Being able to identify areas of elevated lightning ignition likelihood at daily time scales can be of great assistance to fire management agencies. Identifying areas of high lightning ignition likelihood could allow fire management agencies to prepare by increasing surveillance efforts and relocating firefighting resources to ensure firefighting needs are met (Chow and Regan, 2011). Having an understanding of likely lightning patterns is also important for long-term landscape-scale risk assessment. These assessments are usually carried out with fire simulation systems that simulate the growth and spread of sets of hypothetical fires under different weather scenarios. Output from these models is used to identify vulnerable areas and opportunities for risk reduction. For a particular weather scenario, not all ignition locations are equally likely and by using a daily ignition likelihood model, the impacts of simulated fires can be weighted to provide a better indication of the true risk. Better estimates of ignition likelihood would improve estimates of risk at both short term scales, over the coming days (Duff, Chong, et al., 2014), and strategic scales, over years or decades (Department of Environment, Land, Water and Planning, 2015).

The approximating model was built using the CC method proposed in Section 3.3.1 and had four (27%) fewer covariates than the full model. Comparison of performance measures between the full and approximating model showed that the approximating model performed well, however we ultimately believe that this is not a good application of approximating models. Fire fighting agencies are very interested in understanding wildfire ignitions during extreme conditions and we are concerned that removing any significant covariates from the model may adversely affect performance during extreme conditions. Model approximation methods generally remove covariates that are uninfluential most of the time, but they may be highly influential in some

cases. Indeed, the precision of the approximating model dropped off for very high values of the predicted probability suggesting that the full model performed better for extreme observations. Nevertheless, we believe that this method of model approximation is simple and effective and could be useful when models are intended to be used on the bulk of the data rather than focused on the extremes.

We demonstrated a method of deconstructing model forecasts into components from the individual environmental conditions, so that fire-fighting agencies could more easily communicate and 'sanity check' the model output. The likelihood was split, on the logit scale, into a baseline and dynamic component. The baseline component was adjusted for each day of the year and captured the effects of seasonal changes in mean weather conditions along with the effect of time independent covariates. It can be understood as the expected log-odds of lightning ignition. The dynamic components came in the form of the covariate contribution anomaly, which captured the effects of anomalies in the dynamic environmental variables. Ultimately, this allows a given forecast to be understood relative to expected conditions for the time of year. This helps practitioners relate model output back to their own understanding of lightning ignition likelihood.

A lightning ignition likelihood model could be a powerful tool for assisting fire-fighting planning and decision making, but there are some important considerations when deploying the model. In this thesis we have used historical observed data, but in practice forecast data would need to be used for deployment, which would introduce some additional uncertainty into the model (this was explored in Preisler, Chen, et al., 2008 and Preisler and Westerling, 2007). For example, seven-day weather forecasts could be used to produce seven-day lightning ignition forecasts. Another consideration is the possibility of long-term shifts in the relationships between the covariates and ignition likelihood. This can be accounted for by periodically refitting the model, us-

ing a moving window of data from the preceding years. This approach sees the influence of old data on the model being removed once a certain amount of time has passed, making way for newer observations to impact the model. Annual refitting at the start of each fire season should be frequent enough and would avoid needless adjustments during the bulk of the annual fire activity.

The model as presented treats weather as a dynamic variable, but fuel as a static one. Improvements may be made by recognising that the properties of fuel change through time and that this may affect ignition potential. For example, after fire or timber harvest, forest structures are greatly altered, affecting the load, structure and moisture dynamics of fuels over timescales of years (Keane, 2015; McCaw, Neal, and Smith, 2002). This can affect the receptiveness of the fuelbed to ignitions. Fuel properties can also vary on shorter time scales, for example through the curing of annual grasses during dry seasons (Plucinski, 2013). While maintaining data sets of fuel condition would requires investment, the improvement to model predictions may warrant this.

Chapter 6

Point processes

The Poisson process; inference and validation. A connection between regression models and the Poisson process. Integrals with respect to point processes. Campbell's formulas. The K-function. The Neymann–Scott cluster process.

6.1 Introduction to point processes

The theory of point processes offers a flexible framework for studying point data, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, known as *point patterns*. Point processes have been used to study rainfall, the distribution of celestial bodies in the night sky, the spread of disease through a geographically diverse population and the distribution of plant species through a forest.

Point processes can be used to describe point data on very general spaces, but we restrict our conversation to *planar point processes*, $\mathbf{x} \subset \mathbb{R}^2$, since they can readily be used to model wildfire ignitions. Although point processes usually cover the entirety of $\mathbb{R}^2$, observed point patterns do not. We denote W as the bounded region of the observed point process, which can take arbitrary shape and is informed by the availability of data. Note that the observed region of $\mathbb{R}^2$ includes all observed points along with the observed empty space; the absence of points is just as important as the presence.

There are two approaches to analysing point patterns which we will call

the *random measure* and *random set* approaches. Both random measures and random sets necessitate a discussion about which sets are amenable to analysis. Matters are made much simpler by considering the *Borel sets* (Williams, 1991), which is essentially a collection of 'nice' sets so that measures are well behaved on them. Throughout this thesis we assume that all sets are Borel; almost all sets that would arise in applications are Borel.

To understand how random measures can be used to describe point processes we must first understand how a point pattern is equivalent to an integer-valued measure. If we have a point pattern on $\mathbb{R}^2$ we can define a function, $N(\cdot)$, on the subsets of $\mathbb{R}^2$ such that $N(B)$ is equal to the number of points that fall in B. This defines a measure since it is non-negative and since

$$N\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} N(B_i),$$

for any collection of disjoint regions $B_1, B_2, \ldots$. This process 'encodes' a point pattern as an integer-valued measure.

A point process can be described by a mechanism that randomly generates integer-valued measures. Although the theory of random measures is technical and rich, they can be understood as a generalisation of measures; where a measure of a set B returns a non-negative number, a random measure returns a non-negative random variable. In the context of point processes, $N(B)$ is the random variable describing the number of points in the region B. If ω is an element of the outcome space then the random measure evaluated at ω becomes the non-random measure $N(\,\cdot\,;\omega)$ that represents an observed point pattern. Hence a random measure is a mechanism that maps elements of the outcome space to point patterns. For a mathematically thorough exposition of random measures and how they pertain to point processes, see Daley and Vere-Jones (2003, 2008) and Kallenberg (1983).

Random set theory provides an alternative approach to describing point processes. In this context, a point process is described by a random set of

unordered points, $\mathbf{X}$, in the space of interest. Let $\mathrm{n}(\cdot)$ denote the cardinality function, so $\mathrm{n}(\mathbf{x} \cap B)$ is the number of points in the point pattern $\mathbf{x}$ which are also in the region B. Note that $\mathrm{n}(\mathbf{x} \cap W) = \mathrm{n}(\mathbf{x})$ since W is the region where we observe $\mathbf{X}$. Thus $\mathrm{n}(\mathbf{X} \cap B) = N(B)$, connecting the random measure construction of point processes to the random set theory construction. For a thorough exposition of random set theory and point processes, see (Baddeley, Rubak, et al., 2015; Gaetan and Guyon, 2009; Møller and Waagepetersen, 2003).

For all the results in this thesis, the random set and random measure approaches produce equivalent results and we can shift between the two notational styles with impunity. We generally use the random set notation because it allows for more intuitive expressions of integrals with respect to point processes, as we discuss in Section 6.5.

This chapter follows the approach of the excellent Baddeley, Rubak, et al. (2015) and provides the preliminary theory needed to understand the results in Chapters 7 and 8. We expand on some sections and pivot the emphasis to suit our ends, but the notation and approach are broadly consistent and we owe much to this clear and thorough text.

We finish this section with some basic definitions and theory. In Section 6.2 we introduce the *Poisson process*, which is a fundamental point process that plays a key role in general theory. We discuss inference and validation for the Poisson process in Sections 6.3 and 6.4, showing strong links between the Poisson process and traditional regression models. In Section 6.5 we discuss some key technical results for point processes that will inform the discussion of the spacing of points in Section 6.6. We finish the chapter by discussing cluster processes in Section 6.7; these point processes allow interaction effects between points and we use them to model lightning ignitions in Chapter 8.

6.1.1 Key definitions

Point processes can be exceedingly general and it is common to make some simplifying assumptions. A point process is *simple* if there are no coincident points (Daley et al., 2003, Definition 3.3.II). A point process is *locally finite point process* if the number of points in a bounded region is always finite, that is $\mathrm{n}(\mathbf{X} \cap B) < \infty$ for all bounded regions $B \subset \mathbb{R}^2$ (Baddeley, Rubak, et al., 2015, Section 5.2.2). A locally finite point process excludes the case where there is an 'explosion' of points, where an infinite number of points occurs in a finite interval. All processes discussed in this thesis are simple and locally finite and this greatly reduces the complexity of the surrounding theory.

There are two further definitions that play an important role in the theory of point processes.

Definition 1. (Baddeley, Rubak, et al., 2015, Section 5.6.3) A point process $\mathbf{X}$ on $\mathbb{R}^2$ is *stationary* if for all Borel sets $B_1, B_2, \ldots, B_k$ the joint distribution of

$$\big(\mathrm{n}(\mathbf{X} \cap \{B_1 + v\}), \mathrm{n}(\mathbf{X} \cap \{B_2 + v\}), \ldots, \mathrm{n}(\mathbf{X} \cap \{B_k + v\})\big)$$

is identical for all vectors $v \in \mathbb{R}^2$, where $\{B_i + v\} = \{u \in \mathbb{R}^2 \mid u - v \in B_i\}$.

Intuitively, stationarity means that the statistical properties of the point process are unaffected by translation; the points in a stationary point process have no preference for any spatial location.

Definition 2. (Baddeley, Rubak, et al., 2015, Section 5.6.3) Let $r_\theta(\cdot)$ be the operator that rotates space around the origin by the angle θ. A point process $\mathbf{X}$ is *isotropic* if for all Borel sets $B_1, B_2, \ldots, B_k$ the joint distribution of

$$\Big(\mathrm{n}\big(r_\theta(\mathbf{X} \cap B_1)\big), \mathrm{n}\big(r_\theta(\mathbf{X} \cap B_2)\big), \ldots, \mathrm{n}\big(r_\theta(\mathbf{X} \cap B_k)\big)\Big)$$

is identical for all angles θ.

The definition for isotropy implies that the origin is a unique point, but most point processes of interest are both stationary and isotropic. Combined,

they imply that the process is unaffected by translation and rotation around any point. Stationarity and isotropy are strong assumptions and they greatly simplify the theory. We will often discuss theory for the stationary case before generalising it to the non-stationary case.

6.1.2 Intensity

The *intensity* of a point process reflects the average density of points in the process and is a key quantity for characterising point processes. For a stationary point process the intensity reflects the intuitive idea of the density: the number of points per unit area.

Definition 3. For a stationary point process $\mathbf{X}$ the *intensity* is the constant $\lambda > 0$ such that

$$\mathbb{E}[\mathrm{n}(\mathbf{X} \cap B)] = \lambda |B|, \tag{6.1}$$

for all bounded sets $B \subset \mathbb{R}^2$, where $|B|$ is the Lebesgue measure of B.

For a non-stationary process we expect the mean number of points per unit area to vary over space and we must work with the *intensity function*.

Definition 4. For a point process $\mathbf{X}$, the *intensity function*, if it exists, is a function $\lambda : \mathbb{R}^2 \to [0, \infty)$ such that for all bounded sets $B \subset \mathbb{R}^2$,

$$\mathbb{E}[\mathrm{n}(B \cap \mathbf{X})] = \int_B \lambda(u) \, \mathrm{d}u.$$

If the process is simple then $\lambda(u)\delta$ is approximately the probability that there is a point in a small region around u with area δ.

In general there is no guarantee that the intensity function exists and instead we must rely on the *intensity (mean) measure*, $\Lambda(\cdot)$, which is the measure that satisfies

$$\mathbb{E}[\mathrm{n}(B \cap \mathbf{X})] = \Lambda(B) \tag{6.2}$$

for all bounded sets $B \subset \mathbb{R}^2$. For simplicity we assume that the intensity function always exists, which is not unreasonable in most applied settings.

The intensity function is an important summary statistic of a point process; plots of the intensity function can be used to identify areas with a high or low expected number of points. However, similarly to the mean of a random variable, it does not contain all the information about the process. It is possible for point processes to have identical intensity functions yet generate very different point patterns.

6.2 The Poisson process

The Poisson process is arguably the most natural point process. It appears in limiting theorems in a similar way to the Normal distribution in the Central Limit Theorem (Daley et al., 2008, Chapter 11) and forms the boundary between clustering and regularity (Section 6.6). In this section we discuss the definition and properties of the homogeneous (stationary) and inhomogeneous (non-stationary) Poisson processes.

6.2.1 The homogeneous Poisson process

In spatial statistics, the idea of *complete spatial randomness* requires some care. It is generally understood (Baddeley, Rubak, et al., 2015, Section 5.3, Diggle, 2013) to mean two properties:

Homogeneity The points have no preference for any spatial location; and,

Independence The numbers of points in disjoint regions are independent of each other.

Section 5.3.2 of Baddeley, Rubak, et al. (2015) showed how the Poisson process is constructed from these two properties.

Definition 5. (Baddeley, Rubak, et al., 2015, Section 5.3.2) A point process $\mathbf{X}$ on $\mathbb{R}^2$ is a *homogeneous Poisson process* with intensity $\lambda > 0$ if the following is true:

1. for any subset $B \subset \mathbb{R}^2$, the number of points in $\mathbf{X} \cap B$ has a Poisson distribution with mean $\lambda|B|$; and,

2. the random variables $\mathrm{n}(\mathbf{X} \cap B_1), \mathrm{n}(\mathbf{X} \cap B_2), \ldots, \mathrm{n}(\mathbf{X} \cap B_k)$ are independent for disjoint $B_1, B_2, \ldots, B_k$.

There are some important properties of the Poisson process that help us to understand its role in the theory (Baddeley, Rubak, et al., 2015, Section 5.3.3).

Stationarity The process is stationary since the distribution of the number of points in a region is dependent only on the area of the region and since disjoint regions are independent of each other.

Superimposing The superimposition of independent Poisson processes is another Poisson process. If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent Poisson processes with intensities $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively then the point process $\bigcup_i \mathbf{X}_i$ is a Poisson process with intensity $\sum_i \lambda_i$.

Thinning In some sense the reverse of superimpositioning, a Poisson process can be *thinned* to create another Poisson process with a reduced intensity. Thinning with probability $p > 0$ is the operation where each point in the process is independently discarded with probability $1 - p$. The remaining points form the thinned process. A Poisson process with intensity λ, when thinned with probability p, becomes a Poisson process with intensity λp.

Conditional distribution Conditional on the number of points in a region B, the location of the points is distributed independently and uniformly at random across B.

The conditional distribution property of the Poisson process immediately suggests way to simulate a Poisson distribution. Begin by determining the

total number of points in the window of interest, W, by simulating a Poisson variable with mean $\lambda|W|$. Then simulate the location of each of the points by uniformly sampling points from W. The conditioning property tells us that the resulting points form a Poisson process with intensity λ.

6.2.2 The inhomogeneous Poisson process

Generalising the Poisson process to allow the intensity to vary over space is natural and straightforward. The inhomogeneous Poisson process plays a fundamental role in the theory of point processes as well as being a useful model in its own right.

Definition 6. A point process $\mathbf{X}$ on $\mathbb{R}^2$ is an *inhomogeneous Poisson point process* with intensity function $\lambda : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ if it satisfies the following:

1. for any subset $B \subset \mathbb{R}^2$, the number of points in $\mathbf{X} \cap B$ has a Poisson distribution with mean $\int_B \lambda(u)\,\mathrm{d}u$; and,

2. the random variables $\mathrm{n}(\mathbf{X} \cap B_1), \mathrm{n}(\mathbf{X} \cap B_2), \ldots, \mathrm{n}(\mathbf{X} \cap B_k)$ are independent for disjoint $B_1, B_2, \ldots, B_k$.

The homogeneous Poisson process can be recovered from the inhomogeneous Poisson process by taking the intensity function to be a constant. As such, we use the name Poisson process to refer to the inhomogeneous Poisson process, which includes the case of constant intensity.

As in the homogeneous case, the superimposition of independent inhomogeneous Poisson processes results in a Poisson process. If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent Poisson processes with intensity functions $\lambda_1(u), \lambda_2(u), \ldots, \lambda_{\mathrm{n}}(u)$ then the superimposition $\bigcup_{i=1}^{n} \mathbf{X}_i$ is a Poisson process with intensity function $\lambda(u) = \sum_{i=1}^{n} \lambda_i(u)$.

The inhomogeneous Poisson process (with bounded intensity function) can also be understood as a thinned homogeneous Poisson process. Suppose that

$\lambda(u)$ is a function such that $\lambda(u) \leq M$ for all $u \in W$. Let $\mathbf{X}$ be a Poisson process with rate M. If each point x_i in $\mathbf{X}$ is thinned with probability $p(x_i) = \lambda(x_i)/M$ then the resulting process will be an inhomogeneous Poisson process with intensity function $\lambda(u)$. This characterisation suggests a simple way to sample from an inhomogeneous Poisson process: simulate a point pattern $\mathbf{x}$ from a Poisson process with constant intensity M such that $M \geq \lambda(u)$ for all $u \in W$, then for each point $x_i \in \mathbf{x}$, remove x_i with probability $1 - \lambda(u)/M$. The remaining points will form a realisation of an inhomogeneous Poisson process with intensity function $\lambda(u)$.

Covariate effects

An inhomogeneous Poisson processs can be used as a simple model for point patterns that are influenced by covariates. Typically this is done by assuming that, conditioned on the value of the covariates at every location, $\mathbf{X}$ is an inhomogeneous Poisson process with intensity function that is a log-linear sum of the covariates,

$$\lambda(u) = \exp\left(\alpha + \sum_{j=1}^{p} \beta_j z_j(u)\right), \tag{6.3}$$

where $\alpha, \beta_1, \ldots, \beta_p$ are constants and $z_j(u)$ is the value of the j^{th} covariate at location u. The log-linear form of the intensity function is convenient because it maps covariates that take any value in $\mathbb{R}$ to the positive real line and so allows (6.3) to be meaningfully interpreted as an intensity function.

The log-linear form can be understood as a series of thinning or 'modulating' steps. The log-linear intensity function can be expressed as the product

$$\lambda(u) = \exp(\alpha)\exp(\beta_1 z_1(u))\exp(\beta_2 z_2(u))\cdots\exp(\beta_p z_p(u)).$$

If $\exp(\beta_j z_j(u)) < 1$ for all j then the process is equivalent to a homogeneous Poisson process with intensity $\exp(\alpha)$ that undergoes a series of thinning steps with probabilities $\exp(\beta_j z_j(u))$. For general $\exp(\beta_j z_j(u))$, not necessarily less

than 1, we can think of it as a series of 'modulating' steps that thin or 'excite' the point process based on the value of the covariates.

6.3 Inference for the Poisson process

The definition of the intensity of a stationary point process (Definition 3) immediately suggests that the intensity of a homogeneous Poisson process (and other stationary point processes) can be estimated non-parametrically using the method-of-moment style estimator

$$\widehat{\lambda} = \frac{\mathrm{n}(\mathbf{x})}{|W|}. \tag{6.4}$$

Things are complicated by inhomogeneity, where both parametric and non-parametric methods of inference can be appropriate. Non-parametric methods are identical to kernel smoothing methods for point data and give an empirical estimate of the intensity function $\widehat{\lambda}(u)$. We focus on parametric model fitting using maximum likelihood estimation since this is traditionally the approach taken to incorporate covariate effects.

Defining the likelihood function for the spatial Poisson process, and spatial point processes more generally, requires some advanced theory about probability densities with respect to point processes. These are defined for point processes with (almost surely) finite numbers of points; the Poisson process is a finite point process when restricted to the observational window W. Roughly speaking, the density of $\mathbf{X}$, $f(\mathbf{x})$, is the probability of observing the point pattern $\mathbf{x}$ from the point process $\mathbf{X}$ divided by the probability of observing the same point pattern from a stationary Poisson process with unit intensity. Formally, the density of a point process $\mathbf{X}$ with respect to a Poisson process $\mathbf{Y}$ with unit intensity, if it exists, is a function $f : N_{lf} \to \mathbb{R}_{\geq 0}$ satisfying

$$\mathbb{E}[h(\mathbf{X})] = \mathbb{E}[h(\mathbf{Y})f(\mathbf{Y})]$$

for all functions $h : N_{lf} \to \mathbb{R}$, where N_{lf} as the set of all locally finite point patterns on $\mathbb{R}^2$ ($\mathbf{x} \in N_{lf}$ if $n(\mathbf{x} \cap B) < \infty$ for all bounded B) (Baddeley, Rubak,

et al., 2015, equation 13.62). The likelihood function is simply $\mathcal{L}(\theta) = f(\mathbf{x})$. For the inhomogeneous Poisson process, this is

$$\begin{aligned}\mathcal{L}(\theta) &= \left(\prod_{x_i \in \mathbf{x}} \lambda_\theta(x_i) \right) \exp\left(|W| - \int_W \lambda_\theta(u)\, \mathrm{d}u \right) \\ &\propto \left(\prod_{x_i \in \mathbf{x}} \lambda_\theta(x_i) \right) \exp\left(- \int_W \lambda_\theta(u)\, \mathrm{d}u \right) \end{aligned} \tag{6.5}$$

(Baddeley, Rubak, et al., 2015, equation 9.39). The log-likelihood function is

$$l(\theta) = \left(\sum_{x_i \in \mathbf{x}} \log \lambda_\theta(x_i) \right) - \int_W \lambda_\theta(u)\, \mathrm{d}u, \tag{6.6}$$

where the $|W|$ term has been omitted. For general parametric forms of the intensity function, the integral in the log-likelihood function prevents an analytic expression of the maximum likelihood estimate of θ. Instead we resort to numerical techniques, known as *numerical quadrature*, which is a generalisation of the idea of Riemann sum approximations to a curve. In Section 6.3.1 we discuss coarse quadrature approximations, which are commonly used to fit Poisson processes. We discuss alternative approaches, known as fine pixel approximations, in Section 6.3.2. Fine pixel approximations show a strong connection between Poisson process models and standard regression models.

Finally, note that the naive estimator (6.4) for the intensity of a stationary point process is also the maximum likelihood estimator for the homogeneous Poisson process. This can easily be seen by setting $\lambda_\theta(x_i) = \lambda$ in (6.6).

6.3.1 Coarse quadrature approximation

Similarly to Reimann approximations of integrals, course quadrature approximations use the value of the function at a series of points, known as *quadrature points*, to estimate the value of an integral. Unlike Reimann approximations, the points are not regularly spaced and are chosen so that the 'important' points are the focus.

The integral term in the log-likelihood function of the Poisson process (equation 6.6) can be approximated by

$$\int_W \lambda(u)\,\mathrm{d}u \approx \sum_{j=1}^{m} \lambda(u_j) w_j,$$

where $\{u_j\}_{j=1}^m$ are the quadrature points and w_j are weights summing to $|W|$. If $\{u_j\}$ is chosen to include all points in $\mathbf{x}$ as well as suitably chosen dummy points then the log-likelihood can be approximated by

$$l(\theta) \approx \sum_{j=1}^{m} \left(\mathbb{1}_{\mathbf{x}}(u_j) \log \lambda_\theta(u_j) - \lambda_\theta(u_j) w_j \right), \tag{6.7}$$

where $\mathbb{1}_{\mathbf{x}}(u_j) = 1$ if $u_j \in \mathbf{x}$ and 0 otherwise. As noted by Berman and Turner (1992), this can be rewritten as

$$l(\theta) \approx \sum_{j=1}^{m} \left(\frac{\mathbb{1}_{\mathbf{x}}(u_j)}{w_j} \log \lambda_\theta(u_j) - \lambda_\theta(u_j) \right) w_j. \tag{6.8}$$

Although the value of $\mathbb{1}_{\mathbf{x}}(u_j)/w_i$ is not necessarily integer, the form of (6.8) is the same as the weighted log-likelihood function for a series of independent Poisson random variables with mean $\lambda_\theta(u_j)$ (see Harrell, 2015, Section 9.9 for information on weighted log-likelihood functions). The form of the approximate log-likelihood in (6.8) allows standard GLM fitting software to estimate the parameters of a Poisson process model.

There are many methods for choosing the dummy quadrature points and weights, but the goal is always the same. Intuitively, the dummy quadrature points are chosen to form a rough grid-like structure across the study window W and the quadrature weights are the area associated with each of the 'grid' cells. A precise discussion of how quadrature schemes can be chosen is found in Section 9.8.1.2 of Baddeley, Rubak, et al. (2015).

6.3.2 Fine pixel approximations

By default, *spatstat* (the companion package to Baddeley, Rubak, et al., 2015) uses coarse quadrature approximation techniques to fit Poisson processes to

data. Fine pixel approximations are an alternative quadrature approximation and show an even stronger connection between regression models and the Poisson process. Unlike coarse quadrature approximation, fine pixel approximations use the value of the function on a regular grid to estimate the integral.

Pixel counts

Suppose that the domain W is split into a regular grid with cell centers $\{u_j\}_{j=1}^m$ and cell area a. Then the integral in the Poisson log-likelihood function (equation 6.6) can be approximated by

$$\int_W \lambda_\theta(u)\,\mathrm{d}u \approx \sum_{j=1}^m \lambda_\theta(u_j)a. \tag{6.9}$$

Furthermore, if the point x_i falls in the j^{th} grid cell, approximate the value of $\log \lambda_\theta(x_i)$ by $\log \lambda_\theta(u_j)$. By summing over grid cells instead of points the sum in the log-likelihood function (6.6) can be rewritten as

$$\sum_{x_i \in \mathbf{x}} \log \lambda_\theta(x_i) \approx \sum_{j=1}^m y_j \log \lambda_\theta(u_j), \tag{6.10}$$

where y_j is the number of points in the j^{th} grid cell. Putting approximations (6.9) and (6.10) together yields the approximation

$$l(\theta) \approx \sum_{j=1}^m y_j \log \lambda_\theta(u_j) - \lambda_\theta(u_j)a. \tag{6.11}$$

By adding and subtracting the terms $y_j \log a$, (6.11) becomes

$$l(\theta) \approx \sum_{j=1}^m y_j \log(\lambda_\theta(u_j)a) - \lambda_\theta(u_j)a - y_j \log(a) \tag{6.12}$$

$$\propto \sum_{j=1}^m y_j \log(\lambda_\theta(u_j)a) - \lambda_\theta(u_j)a, \tag{6.13}$$

where (6.13) is proportional to (6.12) since $y_j \log(a)$ is independent of θ. Equation (6.13) is the log-likelihood function for Poisson regression with mean

$\lambda_\theta(u_j)a$ (equation 3.4). Like coarse quadrature approximation, this allows inference techniques for regression models to be used for the Poisson process.

The relationship between Poisson regression and the Poisson process is not surprising. The numbers of points in disjoint regions in a Poisson process is independently distributed according to a Poisson distribution, a property shared by Poisson regression.

Pixel presence–absence

An alternative fine pixel approximation is linked with logistic regression. Similarly to the pixel counts approximation, divide the spatial region W into a regular grid with centers $\{u_j\}_{j=1}^m$ and grid cell area a. If the grid cells $\{B_j\}_{j=1}^m$ are sufficiently small such that there is no more than one point in each grid cell, then logistic regression with

$$\begin{aligned}\pi_j = \mathbb{P}(\mathrm{n}(\mathbf{x} \cap B_j) > 0) &\approx 1 - \exp(-\lambda_\theta(u_j)a) \\ &\approx 1 - (1 - \lambda_\theta(u_j)a) \\ &= \lambda_\theta(u_j)a\end{aligned}$$

is a suitable model for the data. This leads to the log-likelihood

$$l(\theta) = \left(\sum_{j=1}^m y_j \log(\lambda_\theta(u_j)a)\right) + \left(\sum_{j=1}^m (1 - y_j)\log(1 - \lambda_\theta(u_j)a)\right) \tag{6.14}$$

for the logistic regression model (equation 3.3). As the grid becomes increasingly fine, and hence the area $a \to 0$, the first sum in (6.14) becomes

$$\sum_{j=1}^m y_j \log(\lambda_\theta(u_j)a) \propto \sum_{j=1}^m y_j \log(\lambda_\theta(u_j)) \tag{6.15}$$

$$\to \sum_{x_i \in \mathbf{x}} \log(\lambda_\theta(x_i)). \tag{6.16}$$

Likewise, by noting that $\log(1+x) \approx x$ when $x \approx 0$, the second sum in (6.14) is approximated by

$$\sum_{j=1}^{m}(1-y_j)\log(1-\lambda_\theta(u_j)a) \approx \sum_{j=1}^{m}(1-y_j)(-\lambda_\theta(u_j)a) \tag{6.17}$$

$$= \sum_{j=1}^{m} -\lambda_\theta(u_j)a + \sum_{j=1}^{m} y_j\lambda_\theta(u_j)a \tag{6.18}$$

$$\to -\int_W \lambda_\theta(u)\,\mathrm{d}u. \tag{6.19}$$

The final line follows from noting that $\sum_{j=1}^{m} -\lambda_\theta(u_j)a$ is a Riemann approximation of $-\int_W \lambda_\theta(u)\,\mathrm{d}u$ and that $\sum_{j=1}^{m} y_j\lambda_\theta(u_j)a \to 0$, since the number of of $y_i = 1$ cells remains fixed as the parition becomes finer ($a \to 0$). Combining (6.16) and (6.19) gives the log-likelihood function of the Poisson process (equation 6.6). This approximation allows inference for logistic regression to be used for the Poisson process.

6.4 Validation for the Poisson process

Validation of a Poisson process model involves checking that the intensity function is appropriate and that the number of points in disjoint regions are independent of each other. Both residuals and hypothesis tests exist to investigate the fit of the intensity function. In this section we discuss residuals for interrogating the fit of the intensity function of a Poisson process. Hypothesis tests for the intensity function can be found in Baddeley, Rubak, et al. (2015, Section 10.4 and 10.5). Approaches for validating the independence assumption can be found in Baddeley, Rubak, et al. (2015, Section 11.6).

6.4.1 Residual measure

The *raw residual* measure of a point process is the difference between the observed and expected number of points and is given by the signed measure

$$\mathscr{R}(B) = \mathrm{n}(\mathbf{x} \cap B) - \int_B \widehat{\lambda}(u)\,\mathrm{d}u$$

(Baddeley, Turner, et al., 2005). It is commonly visualised as the observed points superimposed on a plot of the negative estimated intensity function. To make this easier to interpret, the residuals are often smoothed. The *smoothed residual field* is defined as the function

$$s(u) = \frac{1}{e(u)}\left[\sum_{x_i \in \mathbf{x}} k(u - x_i) - \int_W \widehat{\lambda}(v) k(u - v)\,\mathrm{d}v\right],$$

where $k(u)$ is a smoothing kernel and $e(u) := \int_W k(u - v)\,\mathrm{d}v$ is an edge correction (Baddeley, Rubak, et al., 2015, Section 6.5.1.1 for details on the edge correction, not to be confused with the edge corrections discussed in Section 6.6.4). The smoothed residual field is essentially the difference between a non-parametric estimate of the intensity function and a smoothed version of the fitted intensity function. Values of $s(u)$ should be close to 0 for a model that fits well.

6.4.2 Rescaled residuals

It can be difficult to interpret the values of the residual measure, since the variance of the residual measure is dependent on the intensity measure. These sorts of issues are commonly tackled by reweighting residuals to have constant variance. The *weighted residual measure* is

$$\mathscr{R}^{(w)}(B) = \sum_{x_i \in \mathbf{X} \cap B} w(x_i) - \int_B w(u)\widehat{\lambda}(u)\,\mathrm{d}u,$$

where $w(u)$ is a weight function (Baddeley, Turner, et al., 2005), possibly dependent on the estimated intensity function. If $\widehat{\lambda}(u) = \lambda(u)$, the mean of the residual measure takes value 0 on all sets (this results from a simple application of Campbell's formula, see Section 6.5 below). Good choices of the weight function can control the variance of the residual measure.

One such weighting is used in the *Pearson residual measure*,

$$\mathscr{R}^{(P)}(B) = \sum_{x_i \in \mathbf{x} \cap B} \widehat{\lambda}(x_i)^{-1/2} - \int_B \widehat{\lambda}(u)^{1/2} \, \mathrm{d}u.$$

To prevent the Pearson residual measure being unbounded, set the measure to take value 0 on sets with $\widehat{\lambda}(u) = 0$. The Pearson residual measure has unit variance, that is $\mathrm{Var}(\mathscr{R}^{(P)}(B)) = |B|$ for all bounded sets B (this results from an application of Campbell's Formula and the Extended Slivnyak–Mecke Theorem, Section 6.5 below).

Similarly to the raw residual measure, we can smooth the rescaled residuals to make visual interpretation more straightforward. The *smoothed Pearson residual field* is given by the function

$$s^{(\mathrm{P})}(v) = \sum_{x_i \in \mathbf{X} \cap B} \widehat{\lambda}(x_i)^{-1/2} k(v - x) - \int_B \widehat{\lambda}(u)^{1/2} k(v - u) \, \mathrm{d}u,$$

where $k(u)$ is the smoothing kernel. Although the Pearson measure is constructed to have unit variance, we expect the variance of the smoothed Pearson field to differ due to the dependence introduced by the smoothing kernel. Baddeley, Rubak, et al. (2015, Section 11.3.5) found that if the smoothing kernel $k(x)$ is the radially symmetric Gaussian density with variance σ^2, the pointwise variance of the smoothed Pearson residual field is $\mathrm{Var}(s^{(\mathrm{P})}(u)) \approx 1/(4\pi\sigma^2)$.

6.5 Integrals and Campbell's formula

Integrals with respect to point processes play an important role in concepts such as the K-function (Section 6.6.1) and in more advanced point processes, such as cluster processes (Section 6.7). The integral of a function $f : \mathbb{R}^2 \to \mathbb{R}$ with respect to a point process $\mathbf{X}$ is written as

$$\sum_{x_i \in \mathbf{X}} f(x_i),$$

the sum of the values of f at each point in the point process. In the notation of random measures, this is the stochastic integral with respect to the random

measure N,

$$\sum_{x_i \in \mathbf{X}} f(x) = \int_{\mathbb{R}^2} f(u)\, \mathrm{d}N(u).$$

One of the most useful sets of results in point process theory is Campbell's formula and its generalisations. These results allow us to calculate the expectation of an integral with respect to a point process by using the standard Lebesgue measure and the intensity function.

Theorem 1. *(Campbell's formula, equation 6.11, Baddeley, Rubak, et al., 2015) If* $\mathbf{X}$ *is a point process with intensity function* $\lambda(u)$ *then*

$$\mathbb{E}\left[\sum_{x_i \in \mathbf{X}} f(x)\right] = \int f(u)\lambda(u)\, \mathrm{d}u. \tag{6.20}$$

Section 6.5.1 discusses the theory needed to understand the extension of Campbell's formula, but it presents a significant increase in complexity. The extensions of Campbell's formula are presented in Section 6.5.2.

6.5.1 Palm probabilities

Conditioning plays an important role in elementary probability theory and is invaluable in the study of point processes. The distribution of the distance from a typical point to its nearest neighbouring point, that is the distance from u to the closest point $x_i \in \mathbf{X}$ given that $u \in \mathbf{X}$, is closely tied to ideas of clustering and regularity of point processes. However, it is not immediately obvious how to construct conditional probability in this context. In this section we discuss the Palm probability, $\mathbb{P}^u(A)$, which is essentially the probability of event A occurring conditioned on the event that $u \in \mathbf{X}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space of the point process. Then for any event $A \in \mathcal{F}$ and Borel set $B \subset \mathbb{R}^2$, define the measure $C(A, B) = \mathbb{E}[n(\mathbf{X} \cap B)\mathbb{1}_A]$, known as the *Campbell measure* (Baddeley, Bárány, and Schneider, 2007, Section 3.2, Daley et al., 2008, Section 13.1). If we

consider the event A as being fixed we can define a measure μ_A such that $\mu_A(B) = C(A, B)$. Then

$$\begin{aligned}\mu_A(B) &\leq \mathbb{E}[n(\mathbf{X} \cap B)] \\ &= \Lambda(B).\end{aligned}$$

This implies that μ_A is absolutely continuous with respect to Λ and so the Radon–Nikodym theorem (Daley et al., 2003, theorem A1.4.IV) ensures the existence of a function $f_A(u)$ such that

$$\mu_A(B) = \int_B f_A(u)\,\mathrm{d}\Lambda(u).$$

The function $f_A(u)$ is known as the Palm probability.

Definition 7. (Baddeley, Rubak, et al., 2015, equation 8.48) Suppose $\mathbf{X}$ is a point process with intensity function $\lambda(u)$. Suppose there exists a function $Q(u, A)$ such that, for any event A and any spatial region B,

$$\mathbb{E}[n(\mathbf{X} \cap B)\mathbb{1}_A] = \int_B Q(u, A)\lambda(u)\,\mathrm{d}u.$$

Then $Q(u, A) := \mathbb{P}^u(A)$ is the *Palm probability* of the event A given a point at u.

The Palm probability measures constitute a family of probability measures on Ω indexed by u that correspond to the point process $\mathbf{X}$ conditioned on $u \in \mathbf{X}$. The expectation with respect to the Palm probability measure $\mathbb{P}^u(\cdot)$ is well defined and is denoted by the operator $\mathbb{E}^u(\cdot)$.

The Palm probabilities are difficult to calculate for a general point process. However they are simple to calculate for the Poisson process, courtesy of the independence property; the Palm probability measure $\mathbb{P}^u(\cdot)$ corresponds with the probability measures of the process $\mathbf{X} \cup \{u\}$, which is the superimposition of the the Poisson process $\mathbf{X}$ and the non-random single point at location u.

6.5.2 Extensions of Campbell's formula

We can now state the generalisation of Campbell's formula, known as Campbell–Mecke's formula. Denote N_{lf} as the set of all locally finite point patterns on $\mathbb{R}^2$. So $\mathbf{x} \in N_{lf}$ if $n(\mathbf{x} \cap B) < \infty$ for all bounded B.

Theorem 2. *(Campbell–Mecke's formula, equation 8.50, Baddeley, Rubak, et al., 2015) If* $\mathbf{X}$ *is a point process with intensity function* $\lambda(u)$ *and* $f : W \times N_{lf} \to \mathbb{R}$ *is a function that is integrable with respect to the Campbell measure then*

$$\mathbb{E}\left[\sum_{x_i \in \mathbf{X}} f(x_i, \mathbf{X} \setminus \{x_i\})\right] = \int \mathbb{E}^u[f(u, \mathbf{X} \setminus \{u\})]\lambda(u)\,\mathrm{d}u.$$

If $\mathbf{X}$ is Poisson, Campbell–Mecke's formula can be restated without the Palm probabilties. This is known as the Slivnyak–Mecke theorem.

Theorem 3. *(Slivnyak–Mecke theorem, equation 3.7, Møller and Waagepetersen, 2003) If* $\mathbf{X}$ *is a Poisson process with intensity function* $\lambda(u)$ *and* $f : \mathbb{R}^2 \times N_{lf}$ *is a function, then*

$$\mathbb{E}\left[\sum_{x_i \in \mathbf{X}} f(x_i, \mathbf{X} \setminus \{x_i\})\right] = \int_{\mathbb{R}^2} \mathbb{E}\left[f(u, \mathbf{X})\right] \lambda(u)\,\mathrm{d}u.$$

Lastly, we state a useful extension of the Slivnyak–Mecke Theorem.

Theorem 4. *(Extended Slivnyak–Mecke theorem, equation 3.9, Møller and Waagepetersen, 2003) If* $\mathbf{X}$ *is a Poisson process with intensity function* $\lambda(u)$ *and* $f : (\mathbb{R}^2)^k \times N_{lf}$ *is a function, then*

$$\begin{aligned}&\mathbb{E}\left[\sum_{x_1 \neq x_2 \neq \ldots \neq x_k \in \mathbf{X}} f(x_1, x_2, \ldots, x_k, \mathbf{X} \setminus \{x_1, x_2, \ldots, x_k\})\right] \\ &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} \mathbb{E}\left[f(u_1, u_2, \ldots, u_k, \mathbf{X})\right]\left(\prod_{i=1}^{k}\lambda(u_i)\right)\mathrm{d}u_1\,\mathrm{d}u_2\ldots\mathrm{d}u_k.\end{aligned}$$

To see how to extend the Slivnyak–Mecke Theorem, consider the point process on $(\mathbb{R}^2)^k$ constructed by taking every k-tuple of unique points in

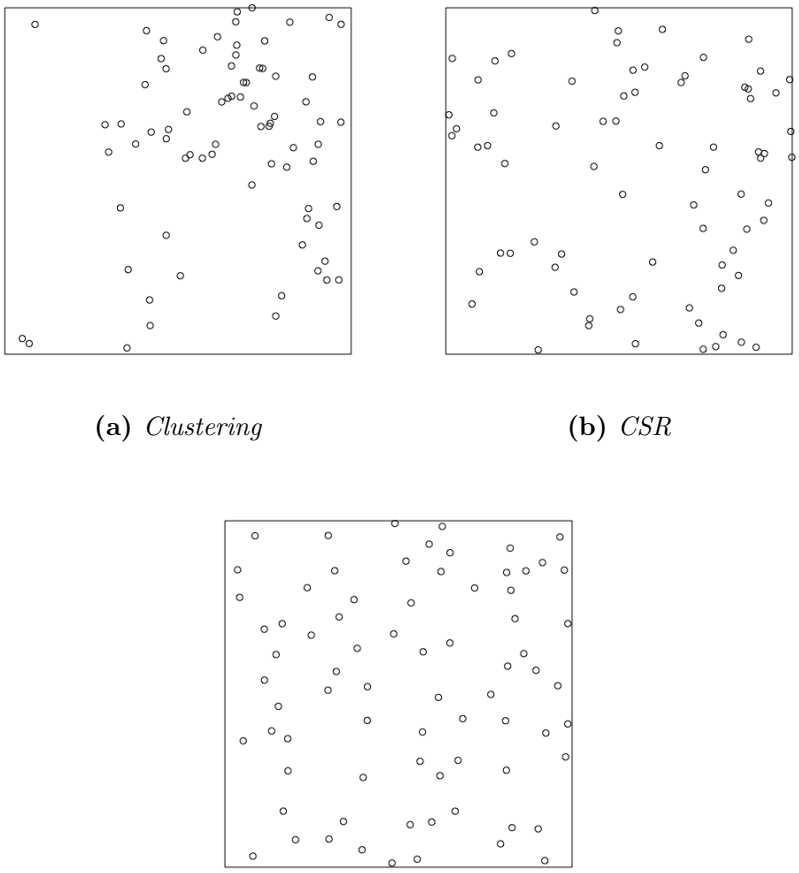
X. The Extended Slivnyak–Mecke Theorem comes from an application of Campbell–Mecke's formula to this extended point process. We present the Slivnyak–Mecke theorems here because they give more explicit expressions for the quantities we will calculate later.

6.6 Correlation

More complex point processes than the Poisson process often involve dependence between points. Similarly to random variables, one way to describe dependence is through concepts of correlation. In this section we discuss the second moment properties of point processes and how they link to the clustering and regularity of points. Intuitively, *clustering* is the tendency of points to be attracted to one another whilst *regularity* is the tendency of points to be repulsed by one another. Regularity is the regular spacing of points, or equivalently, the absence of clustering.

Figure 6.1 shows three point patterns, one exhibiting clustering, one exhibiting *complete spatial randomness* (CSR) and the last showing regularity. Although CSR (Figure 6.1(b)) has regions showing clustering behaviour and regions showing regularity behaviour, these arise purely by chance and do not represent an underlying interaction between the points. As such, the Poisson process is sometimes referred to as CSR and acts as the boundary between clustering and regularity. A point pattern is said to exhibit *clustering* if there is attraction between points in the process, resulting in point patterns with more clustering of points than would be expected for the Poisson process. A point pattern is said to exhibit *regularity* if points repel each other, resulting in point patterns with less clustering of points than would be expected for the Poisson process.

Notions of correlation for point processes are significantly complicated by non-stationarity. Intuitively, correlation is a measurement of the interaction between points. However, an inhomogeneous Poisson process with distinct



(a) *Clustering* **(b)** *CSR*

(c) *Regularity*

Figure 6.1: *Simulated point patterns exhibiting clustering, CSR and regularity. Each point pattern has 75 points simulated from (a) a Thomas cluster process (Section 6.7), (b) a homogeneous Poisson process and (c) a Hardcore process (Møller and Waagepetersen, 2003).*

areas of high and low intensity will result in point patterns with clusters, although the points are independent of each other.

For inhomogeneous point processes, clustering and regularity are understood after taking spatial variation in the intensity function into consideration. An inhomogeneous point process exhibits clustering if it shows a higher degree of clustering than an inhomogeneous Poisson process with the same intensity

function. Likewise, an inhomogeneous point process exhibits regularity if it shows a higher degree of regularity than an inhomogeneous Poisson process with the same intensity function.

In practice, it is impossible to determine from a single point pattern whether a point process is an inhomogeneous Poisson process or a non-Poisson point process. A cluster in a point pattern may be due to local variation in the intensity function, or it may be due to interactions between the points. In the absence of repeated observations of the point process (known as a replicated point pattern, see Chapter 7) we must use our understanding of the physical process to decide whether the point pattern is exhibiting interactions between the points or simply spatial variation in the intensity function.

As this discussion suggests, it is easiest to understand correlation for stationary point processes. We begin our discussion with the *homogeneous K-function*, which is a summary function of the correlation between points in a stationary point process, in Section 6.6.1. This theory can be generalised for non-stationary point processes; Section 6.6.2 covers the *inhomogeneous K-function*.

6.6.1 The homogeneous K-function

The K-function, also known as Ripley's K-function, is a summary function that describes the degree of clustering in a point process. It is independent of the intensity of the process, so meaningful comparisons can be made between processes with different intensities.

Definition 8. (Baddeley, Rubak, et al., 2015, Section 7.3.2) If $\mathbf{X}$ is a stationary point process with intensity $\lambda > 0$, the K-function is

$$K(r) = \frac{1}{\lambda}\mathbb{E}^u \left[\sum_{x_i \in \mathbf{X}} \mathbb{1}\left\{0 < \|u - x_i\| \leq r\right\} \right], \tag{6.21}$$

where $\|\cdot\|$ denotes the Euclidean distance.

The expectation in the definition of the K-function is the expectation with respect to the Palm probability (Section 6.5.1) and will be independent of u since $\mathbf{X}$ is stationary. Intuitively we think of the K-function as the expected number of points within distance r of a point u given that the point u is in $\mathbf{X}$.

For a homogeneous Poisson process the expectation in the K-function is simply the expected number of points in a circle of radius r and hence $K(r) = \pi r^2$. Since the Poisson process acts as the boundary between clustering and regularity, the curve πr^2 is used to distinguish between clustering and regularity behaviour.

The easiest K-functions to interpret are those where either $K(r) > \pi r^2$ or $K(r) < \pi r^2$ for all $r > 0$, as these cases represent clustering and regularity respectively. In practice, most K-functions tend to fall strictly above or below the πr^2 curve (Baddeley, Rubak, et al., 2015, Section 7.3.3). Interpreting a K-function that crosses the πr^2 function is difficult. For example, if $K(r) > 1$ for r less than some constant c, it can be tempting to conclude that the process shows clustering at spatial scales less than c, however this interpretation is incorrect since $K(r)$ is a cumulative function. In this case it can be simpler to work with the pair correlation function.

Definition 9. (Baddeley, Rubak, et al., 2015, Section 7.6.1) The *pair correlation function* for a stationary isotropic point process is

$$g(r) = \frac{K'(r)}{2\pi r}, \tag{6.22}$$

where $K'(r)$ is the derivative of the K-function.

If $g(r) > 1$ for some value r then points distance r apart from each other would occur more frequently than would be expected for a Poisson process with identical intensity. The scale of interaction of a process can defined in terms values of r where the pair correlation function is greater than or less than 1.

A common alternative to the K-function is the L-function, $L(r) = \sqrt{K(r)}/\pi$, which is defined such that the L-function of the Poisson process is the straight diagonal line, $L(r) = r$.

To derive the empirical K-function we first find an expression for the K-function in terms of $\mathbb{E}$ rather than $\mathbb{E}^u$ by considering that for any bounded region B,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{x_i \in X \cap B} \sum_{x_j \in X} \mathbb{1}\left\{0 < \|x_j - x_i\| \leq r\right\}\right] \\
&= \int_B \mathbb{E}^u \left[\sum_{x_j \in X} \mathbb{1}\left\{0 < \|x_j - u\| \leq r\right\}\right] \lambda \, \mathrm{d}u \qquad (6.23) \\
&= \lambda |B| \mathbb{E}^u \left[\sum_{x_j \in X} \mathbb{1}\left\{0 < \|x_j - u\| \leq r\right\}\right], \qquad (6.24)
\end{aligned}$$

where (6.23) follows from the Campbell–Mecke formula and (6.24) is because the value of $\mathbb{E}^u\left[\sum_{x_j \in X} \mathbb{1}\left\{0 < \|x_j - u\| \leq r\right\}\right]$ is independent of u since $\mathbf{X}$ is stationary. So for any choice of B, and in particular $B = W$, we can reformulate the K-function as

$$K(r) = \frac{1}{\lambda^2 |W|} \mathbb{E}\left[\sum_{x_i \in X \cap W} \sum_{x_j \in X} \mathbb{1}\left\{0 < \|x_j - x_i\| \leq r\right\}\right]. \qquad (6.25)$$

Finally, note that for a Poisson process

$$\mathbb{E}\left[n(\mathbf{X} \cap W)\left(n(\mathbf{X} \cap W) - 1\right)\right] = \lambda^2 |W|^2. \qquad (6.26)$$

Equations (6.25) and (6.26) motivate the *empirical K-function*,

$$\widehat{K}(r) = \frac{|W|}{n(\mathbf{x})(n(\mathbf{x}) - 1)} \sum_{x_i \in \mathbf{x}} \sum_{\substack{x_j \in \mathbf{x} \\ x_j \neq x_i}} \mathbb{1}\{\|x_i - x_j\| \leq r\} e(x_i, x_j), \qquad (6.27)$$

where $e(u, v)$ is an edge correction function. The edge correction addresses any error introduced by ignoring points outside the study window that may have an effect on the observed points. We discuss edge corrections in Section 6.6.4.

6.6.2 The inhomogeneous K-function and the general pair correlation function

To discuss the generalisation of the K-function for non-stationary point processes we must first discuss the second moment intensity function and the general pair correlation function.

Definition 10. (Baddeley, Rubak, et al., 2015, Definition 7.4) For a point process $\mathbf{X}$ and two disjoint, compact regions A, B, the *second moment intensity function*, if it exists, is the function $\lambda_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ satisfying

$$\mathbb{E}[\mathrm{n}(A \cap \mathbf{X}) n(B \cap \mathbf{X})] = \int_A \int_B \lambda_2(u, v) \, \mathrm{d}v \, \mathrm{d}u.$$

Roughly speaking, we can interpret $\lambda_2(u, v)\delta^2$ as the probability of having one point in a small region around u and another point in a small region around v where δ is the area of each of the small regions. For the Poisson process, $\lambda_2(u, v) = \lambda(u)\lambda(v)$, which is a consequence of the independence of disjoint regions. The second moment intensity is related to the mean of random sums over all pairs of distinct points,

$$\mathbb{E}\left[\sum_{x_i \in \mathbf{X}} \sum_{\substack{x_j \in \mathbf{X} \\ x_j \neq x_i}} f(x_i, x_j)\right] = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(u, v) \lambda_2(u, v) \, \mathrm{d}u \, \mathrm{d}v. \tag{6.28}$$

This follows from an application of Cambell's formula (equation 6.20) to the point process of all pairs of distinct points in $\mathbf{X} \times \mathbf{X}$. The general pair correlation function is defined by the relationship between the second and first moment intensity functions.

Definition 11. (Baddeley, Rubak, et al., 2015, equation 7.46) For a point process $\mathbf{X}$ with intensity function $\lambda(u)$ and second moment intensity function $\lambda_2(u, v)$, the *pair correlation function* is defined for distinct $u, v \in \mathbb{R}^2$ as

$$g_2(u, v) = \frac{\lambda_2(u, v)}{\lambda(u)\lambda(v)}. \tag{6.29}$$

Notice that the pair correlation function for a stationary isotropic point processes, equation (6.22), is a function of the radius r, whereas in general the pair correlation function is a function of two point locations u and v. For a stationary isotropic process, these two definitions are related by $g_2(u, v) = g(\|u - v\|)$.

The pair correlation function (6.29) for a Poisson process is equal to 1 for all distinct values of u and v. A process with $g_2(u, v)$ greater (less) than 1 exhibits positive (negative) correlation between points at positions u and v. This indicates that when a point is present at location u there is an increased (decreased) likelihood of a point being at position v.

The inhomogeneous K-function

In Baddeley, Møller, and Waagepetersen (2000) the authors proposed a K-function for correlation-stationary point processes, which are more general point processes than stationary point processes and includes the inhomogeneous Poisson process and cluster processes (Section 6.7).

Definition 12. (Baddeley, Rubak, et al., 2015, equation 7.47) A point process $\mathbf{X}$ is *correlation-stationary* if the pair correlation function is dependent only on the relative locations of the two points. That is, if there exists a function f such that

$$g_2(u, v) = f(v - u).$$

The requirement of correlation-stationarity is essentially the bare minimum needed for the idea of a K-function to make sense. Intuitively, $K(r)$ is the expected number of points within radius r of a 'typical' point in the process. Correlation-stationarity means that the (second moment) interaction between points depends only on the relative position of the points and not on their absolute location, so the idea of a typical point still makes sense.

The idea behind the inhomogeneous K-function is to weight the contribution of each point to the K-function so as to remove the effect of changes in the value of the intensity function.

Definition 13. (Baddeley, Møller, et al., 2000, equation 2) For a correlation-stationary point process $\mathbf{X}$ with intensity function $\lambda(u)$, the inhomogeneous K-function is

$$K_{\text{inhom}}(r) = \mathbb{E}^u \left[\sum_{x \in \mathbf{X}} \frac{\mathbb{1}\left\{0 < \|u - x\| \leq r\right\}}{\lambda(x)} \right],$$

where $a/0$ is interpreted to be 0 for all a.

The interpretation of $K_{\text{inhom}}(r)$ is identical to that of $K(r)$ since the inhomogeneous K-function for the Poisson process is

$$K_{inhom}(r) = \int \frac{1}{\lambda(v)} \mathbb{1}\left\{0 < \|u - v\| \leq r\right\} \lambda(v)\, \mathrm{d}v = \pi r^2.$$

If the process is stationary, the intensity function is a constant and the inhomogeneous K-function reduces to the homogeneous K-function.

To derive the empirical inhomogeneous K-function we take a similar approach to the homogeneous case. For any bounded region B

$$\begin{aligned}
&\mathbb{E}\left[\sum_{x_i \in X \cap B} \sum_{x_j \in X} \frac{\mathbb{1}\left\{0 < \|x_j - x_i\| \leq r\right\}}{\lambda(x_j)\lambda(x_i)} \right] \\
&= \int_B \mathbb{E}^u \left[\sum_{x_j \in X} \frac{\mathbb{1}\left\{0 < \|x_j - u\| \leq r\right\}}{\lambda(x_j)\lambda(u)} \right] \lambda(u)\, \mathrm{d}u \qquad (6.30) \\
&= |B| \mathbb{E}^u \left[\sum_{x_j \in X} \frac{\mathbb{1}\left\{0 < \|x_j - u\| \leq r\right\}}{\lambda(x_j)} \right], \qquad (6.31)
\end{aligned}$$

where (6.30) follows from the Campbell-Mecke formula and (6.31) is a consequence of being correlation-stationary. This suggests that

$$\widehat{K}^*_{inhom}(r) = \frac{1}{|W|} \sum_{x_i \in \mathbf{x}} \sum_{\substack{x_j \in \mathbf{x} \\ x_j \neq x_i}} \frac{\mathbb{1}\left\{\|x_i - x_j\| \leq r\right\}}{\widehat{\lambda}(x_i)\widehat{\lambda}(x_j)} e(x_i, x_j), \qquad (6.32)$$

known as the 'plug-in' estimator, is a suitable estimator for the inhomogeneous K-function. As before, $e(x_i, x_j)$ is an edge correction (Section 6.6.4).

Unfortunately this estimator is prone to errors caused by bias in the estimate of the intensity function, especially if there are points associated with small values of the estimated intensity function. This has caused it to perform poorly in some cases (Diggle, 2013; Diggle, Gómez-Rubio, et al., 2007; Gabriel and Diggle, 2009). In the Section 6.6.3 we discuss ways to normalise the estimate so that it is less sensitive to bias in the estimate of the intensity function.

Connecting the K-function and the pair correlation function

Definition 9 showed a strong connection between the pair correlation function and the K-function for stationary isotropic point processes. A similar relationship holds more generally and can be helpful when calculating the theoretical K-function for complex point processes. Relation (6.31) implies that, for any bounded region B,

$$K_{inhom}(r) = \frac{1}{|B|}\mathbb{E}\left[\sum_{x_i \in X \cap B} \sum_{x_j \in X} \frac{\mathbb{1}\left\{0 < \|x_j - x_i\| \leq r\right\}}{\lambda(x_i)\lambda(x_j)}\right] \tag{6.33}$$

$$= \frac{1}{|B|}\int_B \int_{B_r(u)} \frac{1}{\lambda(u)\lambda(v)} \lambda_2(u, v)\,\mathrm{d}v\,\mathrm{d}u \tag{6.34}$$

$$= \frac{1}{|B|}\int_B \int_{B_r(u)} g_2(u, v)\,\mathrm{d}v\,\mathrm{d}u \tag{6.35}$$

where $B_r(u)$ is a closed ball centred at u of radius r. Equation (6.34) follows from Campbell's formula applied to the process of all pairs of distinct points, see (6.28). If $g_2(u, v)$ is correlation-stationary (Definition 12), then $g_2(u, v)$ depends only on the distance $\|u - v\|$ and (6.35) can be expressed as

$$K_{\text{inhom}}(r) = \frac{1}{|B|}\int_B \int_{B_r(0)} g_2(0, v)\,\mathrm{d}v\,\mathrm{d}u$$

$$= \int_{B_r(0)} g_2(0, v)\,\mathrm{d}v \tag{6.36}$$

We can further simplify expression (6.36) if we assume that the pair correlation function is isotropic.

Definition 14. A function $f : \mathbb{R}^2 \to \mathbb{R}$ is *isotropic* if there exists another function $f^* : \mathbb{R} \to \mathbb{R}$ such that $f(u) = f^*(\sqrt{\|u\|})$ for all $u \in \mathbb{R}^2$.

A function is isotropic if it depends only on the distance from the origin. If the pair correlation function is isotropic it is common to reformulate it in terms of the distance from the origin.

Definition 15. The *isotropic pair correlation function*, if it exists, is the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfying

$$g_2(u, v) = g(\|v - u\|)$$

for all $u, v \in \mathbb{R}^2$.

Replacing the pair correlation function in (6.36) with the isotropic pair correlation function and rewriting the integral in terms of polar coordinates yields

$$K_{\text{inhom}}(r) = 2\pi \int_0^r sg(s)\,\mathrm{d}s. \tag{6.37}$$

For stationary point processes, relationship (6.37) implies Definition 9, the pair correlation function for stationary point processes.

6.6.3 The normalised inhomogeneous K-function

The 'plug-in' estimator for the inhomogeneous K-function, equation (6.32), can perform poorly due to bias in the estimated intensity function. One approach to reducing this error is to divide the empirical K-function by a normalising term that is similarly affected by the bias in the estimated intensity function. If chosen appropriately, the bias in the normalising term should 'cancel out' with the bias in the empirical K-function.

A simple improvement to the 'plug-in' estimator for the inhomogeneous K-function, equation (6.32), is suggested by the observation that

$$\mathbb{E}\left[\sum_{x_i \in \mathbf{X} \cap W} \frac{1}{\lambda(x_i)}\right] = |W|, \tag{6.38}$$

which follows from Campbell's formula (Section 6.5). Replacing $|W|$ in (6.32) with $\sum_{x_i \in \mathbf{X} \cap W} \left[\widehat{\lambda}(x_i)\right]^{-1}$ results in an estimator for the inhomogeneous K-function, which may have reduced error over the 'plug-in' estimator.

Baddeley, Rubak, et al. (2015) suggested estimators for the inhomogeneous K-function of the form

$$\widehat{K}_{inhom}(r) = \frac{1}{|W| D(\mathbf{x})} \sum_{x_i \in \mathbf{x}} \sum_{\substack{x_j \in \mathbf{x} \\ x_j \neq x_i}} \frac{\mathbb{1}\left\{\|x_i - x_j\| \leq r\right\}}{\widehat{\lambda}(x_i)\widehat{\lambda}(x_j)} e(x_i, x_j), \tag{6.39}$$

where $D(\mathbf{x})$ is a data-dependent normalising term chosen to counter the effects of bias in the estimate of the intensity function. The normalising term suggested by (6.38) is

$$D_1(\mathbf{x}) = \frac{1}{|W|} \sum_{x_i \in \mathbf{x}} \frac{1}{\widehat{\lambda}(x_i)}.$$

$D_1(\mathbf{x})$ is the default choice of normalising term used by *spatstat*.

An extreme example of bias would be if the intensity function was incorrectly estimated by a constant factor at every location, say $\widehat{\lambda}(u) = V\lambda(u)$. Then for all $r > 0$, the 'plug-in' estimator (6.32) would be incorrect by a factor of $1/V^2$ and the estimator (6.39) with normalising term $D_1(\mathbf{x})$ would be incorrect by a factor of $1/V$. An immediate solution to this is to take

$$D_2(\mathbf{x}) = \left[D_1(\mathbf{x})\right]^2 = \frac{1}{|W|^2} \left(\sum_{x_i \in \mathbf{x}} \frac{1}{\widehat{\lambda}(x_i)}\right)^2,$$

however this has some issues, namely that the empirical inhomogeneous K-function with normalising term $D_2(\mathbf{x})$ (and $D_1(\mathbf{x})$ for that matter) does not reduce to the empirical homogeneous K-function if we assume that the intensity function is a constant.

The unpublished work Moradi et al. (n.d.), proposes a better normalising term for the case when $\mathbf{X}$ is an inhomogeneous Poisson process,

$$D_{2-}(\mathbf{x}) = \frac{1}{|W|^2}\left(\sum_{x_i\in\mathbf{x}}\sum_{\substack{x_j\in\mathbf{x}\\ x_j\neq x_i}}\frac{1}{\widehat{\lambda}(x_i)\widehat{\lambda}(x_j)}\right). \tag{6.40}$$

This is motived by the observation that

$$\mathbb{E}\left[\sum_{x_i\in\mathbf{x}}\sum_{\substack{x_j\in\mathbf{x}\\ x_j\neq x_i}}\frac{1}{\lambda(x_i)\lambda(x_j)}\right] = \int_W\int_W \frac{1}{\lambda(u)\lambda(v)}\lambda_2(u,v)\,\mathrm{d}u\,\mathrm{d}v = |W|^2,$$

since the second moment intensity function of the Poisson process is $\lambda_2(u,v) = \lambda(u)\lambda(v)$. If we have the extreme case where the estimated intensity function is biased by a constant factor V then, for the Poisson process, the normalising term $D_{2-}(\mathbf{x})$ will be an unbiased estimator of $1/V^2$.

In the case when the intensity function is a constant and estimated by $\widehat{\lambda}$, the empirical inhomogeneous K-function with normalising term $D_{2-}(\mathbf{x})$ reduces to the homogeneous K-function. To see this, simply substitute

$$D(\mathbf{x}) = D_{2-}(\mathbf{x}) = \frac{1}{|W|^2}\left(\sum_{x_i\in\mathbf{x}}\sum_{\substack{x_j\in\mathbf{x}\\ x_j\neq x_i}}\frac{1}{\widehat{\lambda}^2}\right) = \frac{n(n-1)}{|W|^2\widehat{\lambda}^2},$$

into equation (6.39).

6.6.4 Edge corrections

The K-function describes the propensity of points to cluster or space themselves in a point process and reflects interactions between the points in the process. In practice we have knowledge about the point process only in the observation window W and must use this to estimate the K-function. Points that lie close to the boundary of W will have fewer neighbouring points than

those in the middle of W simply because we fail to observe the points outside W. Edge corrections attempt to reduce this bias by using fairly general assumptions about point processes to reweight points that are close to the boundary of W.

There are numerous methods of edge correction and in practice the choice of method should not overly influence the results. We discuss the *translation* edge correction to give a feel for how these corrections work, and refer the reader towards Baddeley, Rubak, et al. (2015, Section 7.4.2) for alternative methods and further details.

The contribution of a pair of points $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$ to the K-function is related to the likelihood of observing these points given the distance and orientation between them. If the point process is stationary then the likelihood of observing a pair of points (x_i, x_j) is proportional to the area of the part of W that admits the observation of such a pair. The extreme case is when we are interested in points which are spaced further apart than the maximum distance within W, which we have no chance of observing. We can write the subset of W that admits pairs of points such as (x_i, x_j) as $W \cap (W - (x_j - x_i))$. This leads to the translation edge correction

$$e(x_i, x_j) = \frac{|W|}{|W \cap (W - (x_j - x_i))|},$$

which reflects the proportion of W that admits a pair of points with the same orientation and seperating distance as (x_i, x_j).

6.7 Cluster point processes

Cluster processes are a class of point processes that use a two-stage mechanism to induce clustering between the points. At its most general, a cluster process is defined by a *parent* point process Φ, along with *offspring* processes $\{\mathbf{X}_c\}_{c \in \Phi}$ that cluster their points around their centres, c. The cluster process consists

of the superimposition of all the offspring processes,

$$\mathbf{X} = \bigcup_{c \in \Phi} \mathbf{X}_c.$$

These processes are exceedingly general and so we restrict our attention to Neymann–Scott cluster processes (NSCP) which are cluster processes satisfying the following:

1. the parent process, Φ, is a Poisson process;

2. the offspring processes, $\mathbf{X}_c$, are independent of each another;

3. the offspring processes, if shifted to the same location, are identically distributed; and,

4. the offspring within a cluster are independently and identically distributed.

Things are futher simplified by making further assumptions:

5. each cluster contains a Poisson number of points; and,

6. the distribution of points around the cluster centre is isotropic.

If the preceding six conditions are true, then the stationary NSCP can be described as follows. The parent process Φ is a homogeneous Poisson process with intensity κ. Each offspring process $\mathbf{X}_c$ is an inhomogeneous Poisson process with intensity function $\mu h(u - c)$ where $\mu > 0$ is the mean number of points in a cluster and $h(u)$ is an isotropic density function, known as the *spatial kernel*. The superimposition of all the offspring processes conditioned on the parent process, $\bigcup_{c \in \Phi} \mathbf{X}_c \,|\, \Phi$, is a Poisson process with intensity function

$$\nu(u) = \mu \sum_{c \in \Phi} h(u - c). \tag{6.41}$$

This follows directly from the superimposition property of the Poisson process (Section 6.2). The expected number of points in a region B can be found using Campbell–Mecke's formula,

$$\begin{aligned}
\mathbb{E}\left[\mathrm{n}(X \cap B)\right] &= \mathbb{E}\left[\sum_{c \in \Phi} \mathrm{n}(X_c \cap B)\right] \\
&= \int_{\mathbb{R}^2} \mathbb{E}\left[\mathrm{n}(X_c \cap B)\right] \kappa \, \mathrm{d}c \\
&= \kappa \int_{\mathbb{R}^2} \int_B \mu h(u - c) \, \mathrm{d}u \, \mathrm{d}c \\
&= \kappa\mu \int_B \int_{\mathbb{R}^2} h(u - c) \, \mathrm{d}c \, \mathrm{d}u \\
&= \kappa\mu |B|.
\end{aligned}$$

This shows that the intensity of a stationary NSCP is

$$\lambda = \kappa\mu. \tag{6.42}$$

As we discuss in Section 6.7.3, the Neyman–Scott cluster process has the attractive property that statistical inference can be performed without knowledge of the points in the parent process. This is important in practice, when the parent process is usually unobserved.

Cluster processes are part of a more general class of point processes known as *Cox processes* or *doubly stochastic Poisson processes*. They are a generalisation of the Poisson process, constructed by allowing the intensity measure (equation 6.2) to be a random measure. That is, suppose that Λ is a random measure on $\mathbb{R}^2$; a point process $\mathbf{X}$ on $\mathbb{R}^2$ is a Cox process with driving intensity measure Λ if $\mathbf{X} \,|\, \Lambda$ is a Poisson process with intensity measure Λ. The NSCP is a Cox process with driving intensity given by (6.41). It is because of this interpretation that the NSCP is sometimes refered to as the Shot-Noise Cox process.

6.7.1 The pair correlation function and K-function for the NSCP

The K-function for the NSCP is most easily calculated using the pair correlation function and the relationship (6.37). To calculate the pair correlation function we first need the second moment intensity function. For a NSCP $\mathbf{X}$ and disjoint regions $A, B \subset \mathbb{R}^2$,

$$\begin{aligned}
\mathbb{E}\left[n(\mathbf{X}\cap A)n(\mathbf{X}\cap B)\right] &= \mathbb{E}\left[\sum_{c_i\in\mathbf{\Phi}} n(\mathbf{X}_{c_i}\cap A)\sum_{c_j\in\mathbf{\Phi}} n(\mathbf{X}_{c_j}\cap B)\right] \\
&= \mathbb{E}\left[\sum_{c_i\in\mathbf{\Phi}}\sum_{\substack{c_j\in\mathbf{\Phi}\\ c_j\neq c_i}} n(\mathbf{X}_{c_i}\cap A)n(\mathbf{X}_{c_j}\cap B)\right] \qquad (6.43)\\
&\quad + \mathbb{E}\left[\sum_{c_i\in\mathbf{\Phi}} n(\mathbf{X}_{c_i}\cap A)n(\mathbf{X}_{c_i}\cap B)\right]. \qquad (6.44)
\end{aligned}$$

The expectation (6.43) can be calculated using the Slivnyak–Mecke Theorem,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{c_i\in\mathbf{\Phi}}\sum_{\substack{c_j\in\mathbf{\Phi}\\ c_j\neq c_i}} n(\mathbf{X}_{c_i}\cap A)n(\mathbf{X}_{c_j}\cap B)\right] \\
&\quad = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \mathbb{E}\left[n(\mathbf{X}_x\cap A)n(\mathbf{X}_y\cap B)\right]\kappa^2\,\mathrm{d}x\,\mathrm{d}y \\
&\quad = \kappa^2\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\left(\int_A \mu h(u-x)\,\mathrm{d}u\int_B \mu h(v-y)\,\mathrm{d}v\right)\mathrm{d}x\,\mathrm{d}y \\
&\quad = \kappa^2\mu^2|A||B|. \qquad (6.45)
\end{aligned}$$

The expectation (6.44) can be calculated using Campbell's formula,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{c_i\in\mathbf{\Phi}} n(\mathbf{X}_{c_i}\cap A)n(\mathbf{X}_{c_i}\cap B)\right] \\
&\quad= \int_{\mathbb{R}^2} \mathbb{E}\left[n(\mathbf{X}_x\cap A)n(\mathbf{X}_x\cap B)\right]\kappa\,\mathrm{d}x \\
&\quad= \kappa\int_{\mathbb{R}^2}\left(\int_A \mu h(u-x)\,\mathrm{d}u\int_B \mu h(v-x)\,\mathrm{d}v\right)\mathrm{d}x \\
&\quad= \kappa\mu^2\int_A\int_B\left(\int_{\mathbb{R}^2} h(u-x)h(v-x)\,\mathrm{d}x\right)\mathrm{d}v\,\mathrm{d}u. \qquad (6.46)
\end{aligned}$$

Then the second moment intensity can be found by combining (6.45) and (6.46) and shrinking the regions A and B to the points u and v,

$$\begin{aligned}
\lambda_2(u,v) &= \lim_{A\downarrow\{u\}}\lim_{B\downarrow\{v\}}\frac{\mathbb{E}\left[\mathrm{n}(\mathbf{X}\cap A)\mathrm{n}(\mathbf{X}\cap B)\right]}{|A||B|} \\
&= \kappa^2\mu^2\left(1+\frac{1}{\kappa}\int_{\mathbb{R}^2} h(u-x)h(v-x)\,\mathrm{d}x\right).
\end{aligned}$$

So the pair correlation function for the NSCP is

$$\begin{aligned}
g_2(u,v) &= \frac{\lambda_2(u,v)}{\lambda(u)\lambda(v)} \\
&= 1+\frac{1}{\kappa}\int_{\mathbb{R}^2} h(u-x)h(v-x)\,\mathrm{d}x. \qquad (6.47)
\end{aligned}$$

If $h(u)$ is isotropic, we can rewrite the integral in (6.47) as

$$\int_{\mathbb{R}^2} h(u-x)h(v-x)\,\mathrm{d}x = \int_{\mathbb{R}^2} h\left((u-v)-x\right)h(x)\,\mathrm{d}x,$$

which is the convolution of h with itself, evaluated at $u-v$. We can write

$$g_2(u,v) = 1+\frac{1}{\kappa}h*h(u-v), \qquad (6.48)$$

where $*$ is the convolution operator. This implies that the isotropic pair correlation function exists and is given by

$$g(r) = 1+\frac{1}{\kappa}h*h(r), \qquad (6.49)$$

where $h*h(r)$ is the convolution of h with itself evaluated at a point with distance r from the origin,

$$h*h(r) = \int_{\mathbb{R}^2} h((0,r)-u)h(u)\,\mathrm{d}u.$$

The K-function for a specific NSCP is typically calculated using relationship (6.49) to find the isotropic pair correlation function and then using relationship (6.37),

$$K(r) = 2\pi \int_0^r g(s)\,\mathrm{d}s.$$

This is used in Section 6.7.2 to calculate the K-function for some common NSCPs.

6.7.2 Specific NSCP

Different choices of spatial kernel $h(u)$ give rise to different NSCPs with different clustering behaviours. In this section we discuss a few of the commonly used NSCPs.

The Matérn cluster process

A simple NSCP is the Matérn cluster process, which takes a uniform density on a circle of radius R to be the spatial kernel.

$$h(u) = \frac{1}{2\pi R^2} \mathbb{1}\left(\|u\| \leq R\right)$$

The value of R controls the degree of clustering.

The Thomas cluster process

The Thomas cluster process is a NSCP where the spatial kernel is a radially symmetric bivariate Normal density,

$$h(u) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right),$$

where $\sigma > 0$ is called the *scale* parameter.

The pair correlation function for the Thomas cluster process can easily be calculated from equation (6.49). Since the Normal distribution is stable, the convolution $h * h$ a radially symmetric bivariate Normal density with marginal

variance $2\sigma^2$. So the isotropic pair correlation function is

$$g(r) = 1 + \frac{1}{4\pi\kappa\sigma^2} \exp\left(- \frac{r^2}{4\sigma^2} \right).$$

Using the isotropic pair correlation function and the relationship (6.37), the K-function of the Thomas cluster process is

$$K(r) = \pi r^2 + \frac{1 - \exp\left(-r^2/(4\sigma^2)\right)}{\kappa}.$$

The Cauchy cluster process

The Cauchy cluster process is a NSCP where the spatial kernel $h(u)$ is a radially symmetric bivariate Cauchy density,

$$h(u) = \frac{1}{2\pi\sigma^2} \left(1 + \frac{\|u\|^2}{\sigma^2} \right)^{-3/2},$$

where $\sigma > 0$ is called the *scale* parameter. This distribution is heavy-tailed, meaning that points can be very far from the center process that spawned them. This may be suitable for data which occasionally has points far from what appears to be their parent cluster, although this can be difficult to assess since the parent process is typically unobserved.

Since the Cauchy distribution is stable, the convolution of $h(u)$ with itself is a bivariate Cauchy density with scale parameter 2σ. Using (6.49), an expression for the isotropic pair correlation function is

$$g(r) = 1 + \frac{1}{4\pi\kappa\sigma^2} \left(1 + \frac{r^2}{2\sigma^2} \right)^{-3/2}.$$

Then, from (6.37), the K-function is

$$\begin{aligned}
K(r) &= 2\pi \int_0^r s g(s) \, \mathrm{d}s \\
&= \pi r^2 + \frac{1}{8\pi\kappa\sigma^2} \int_0^r s \left(1 + \frac{s^2}{4\sigma^2} \right)^{-3/2} \mathrm{d}s \\
&= \pi r^2 + \frac{1}{8\pi\kappa\sigma^2} \int_0^r -4\sigma^2 \frac{\mathrm{d}}{\mathrm{d}s} \left[\left(1 + \frac{s^2}{4\sigma^2} \right)^{-1/2} \right] \mathrm{d}s \\
&= \pi r^2 + \frac{1}{\kappa} \left(1 - \left(1 + \frac{r^2}{4\sigma^2} \right)^{-1/2} \right).
\end{aligned}$$

6.7.3 Inference for the NSCP

In general, likelihood methods of inference are not available for the NSCP since the likelihood function is intractable (Baddeley, Rubak, et al., 2015, Section 12.6.1). Instead, what is essentially a method-of-moments style approach known as *Minimum contrast estimation* is employed. The intensity (the first moment) is estimated using (6.4). Then the *cluster parameters*, κ and σ, are estimated by choosing the values that minimise the 'distance' between the model and empirical K-functions (the second moment). Formally, the *contrast* between the model and empirical K-functions is

$$D(\theta) = \int_a^b \left(K_\theta(s)^q - \widehat{K}(s)^q \right)^p \mathrm{d}s, \tag{6.50}$$

where $K_\theta(r)$ is the model K-function which depends on the parameters θ, and $\widehat{K}(r)$ is the empirical K-function. If $p = 2$ and $q = 1$ this is equivalent to minimising the mean squared error between the two functions. The default in *spatstat* is $p = 2$ and $q = 1/4$. The choice of the interval to compare the functions over, (a, b), depends on the data; however *spatstat* will attempt to choose suitable values for the user. Finally, the mean cluster size can be estimated by $\widehat{\mu} = \widehat{\lambda}/\widehat{\kappa}$, which follows from (6.42).

6.7.4 The inhomogeneous NSCP

There are several possible ways to define the inhomogeneous NSCP: the parent process Φ, the offspring processes $\{X_c\}$ or both can be inhomogeneous. Model fitting with an inhomogeneous parent process proves to be difficult and so usually one works with a homogeneous parent process and inhomogeneous offspring processes.

Again we have a parent process Φ which is a Poisson process with rate κ, but this time the offspring processes $\mathbf{X}_c$ are inhomogeneous Poisson processes with intensity functions

$$\lambda(u) = \mu(u)h(u - c),$$

where $\mu(u)$ is known as the *modulating intensity*. Conditioned on the parent process, the offspring process is a Poisson process with intensity function

$$\nu(u) = \mu(u) \sum_{c \in \mathbf{\Phi}} h(u - c).$$

The intensity of the inhomogenous NSCP is

$$\lambda(u) = \kappa \mu(u),$$

which follows from an application of Campbell's formula, similar to the homogeneous case (6.42).

Similarly to the Poisson process, an inhomogeneous NSCP (with bounded intensity function) is equivalent to a thinned homogeneous NSCP. Let $\mathbf{X}$ be an inhomogeneous NSCP with centre process intensity κ, spatial kernel $h(u)$ and (bounded) modulating intensity function $\mu(u)$. Let $\mathbf{Y}$ be a homogeneous NSCP with identical centre process intensity and spatial kernel and with mean cluster size M, where $M > \mu(u)$ for all $u \in W$. Then $\mathbf{X}$ has the same distribution as $\mathbf{Y}$ that has been thinned with probability $\mu(u)/M$.

This thinning construction of the inhomogeneous NSCP suggests a method of calculating the second moment intensity function (Definition 10), which is required to calculate the pair correlation function and inhomogeneous K-function. If $\mathbf{X}$ and $\mathbf{Y}$ are the inhomogeneous and homogeneous NSCPs described in the previous paragraph then we can rewrite integrals with respect to $\mathbf{X}$ to be with respect to $\mathbf{Y}$;

$$\sum_{x_i \in \mathbf{X}} f(x_i) = \sum_{y_i \in \mathbf{Y}} f(y_i) \mathbb{1}_{\mathbf{X}}(y_i), \tag{6.51}$$

where $\mathbb{1}_{\mathbf{X}}(y_i)$ is an indicator function taking value 1 if the point y_i survived the thinning process and is in the NSCP $\mathbf{X}$. Using relationship (6.51), it is relatively straightforward to show that the pair correlation function, and hence the K-function, of the inhomogeneous NSCP is identical to that of the homogeneous NSCP. This implies that the inhomogeneous NSCP is also correlation stationary.

Inference for the inhomogeneous NSCP is much the same as that for the homogeneous NSCP; minimum contrast estimation (Section 6.7.3) is used to estimate the cluster parameters. For the homogeneous NSCP, the non-parametric estimator (6.4) is used to estimate the intensity. For the inhomogeneous NSCP, either parametric or non-parametric estimators are available, but we focus on parametric estimators since we are interested in incorporating the influence of covariates on the intensity function.

Schoenberg (2005) showed that the maximum likelihood estimator of the intensity function of the Poisson process is an approximate estimator for the intensity function of certain non-Poisson point processes. This allows the machinery of Poisson process inference to be used for the inhomogeneous NSCP. It is common to assume that the modulating intensity is a log-linear sum of the covariates, which makes estimating the intensity function equivalent to fitting a Poisson process with a log-linear intensity function (Section 6.2.2).

This approach to inference shows a connection between a Poisson process and a NSCP which share the same intensity function. It is interesting to see how these processes compare in terms of the variance of the number of points in a bounded region B. An expression of this variance for spatial Cox processes, which includes the NSCP, can be found in Jalilian, Guan, and Waagepetersen (2013):

$$\operatorname{Var}\big[n(\mathbf{X} \cap B)\big] = \int_B \lambda(u)\,\mathrm{d}u + \int_B \int_B \lambda(u)\lambda(v)\big[g_2(u,v) - 1\big]\,\mathrm{d}u\,\mathrm{d}v, \quad (6.52)$$

where $g_2(u,v)$ is the pair correlation function. The first integral in (6.52), $\int_B \lambda(u)\,\mathrm{d}u$, is the variance of the number of points in B of a Poisson process with intensity function $\lambda(u)$. The second integral in (6.52) is a modifying effect on the variance caused by the interaction between the points. In the case of a NSCP, the pair correlation function is always greater than 1, resulting in a larger variance than a Poisson process with identical intensity function.

6.7.5 Validation for the NSCP

Present approaches to validating NSCP models are very limited. Residual measures (Section 6.4) of NSCPs are very difficult to analyse since there is dependence between the residual measure at nearby locations. Because of this, well fitting models can have residual measures that exceed two standard deviations over large regions. The degree of clustering can be validated by comparing the fitted and estimated K-functions. To make this comparison meaningful, confidence intervals must be placed around the estimated K-function. This can be done using bootstrap approaches (Baddeley, Rubak, et al., 2015, Sections 7.7.2 and 7.7.3).

Chapter 7

Replicated point patterns

Collections of independent point patterns. Spatio-temporal point processes. Inference with replicated point patterns. Ratio estimators and pooling ratio estimators. The pooled K-function. A new weighting for the pooled K-function. Model validation with replicated point patterns.

7.1 Introduction to replicated point patterns

Classically, inference for point processes is considered as though we have only a single observation to work with and in many applications, this is the case. Species distribution modelling, for example, is often forced to work with a single point pattern of recorded animal sightings. However some situations are closer to traditional statistics, where repeated independent observations of the same process are available. For example, Diggle, Lange, and Beneš (1991) looked at the spatial distribution of pyramidal neurons between normal, schizoaffective and schizophrenic persons at time of death by comparing multiple observations of brain tissue from each group. Point processes that evolve through time offer another potential source of repeated observations. For example, daily point patterns of wildfire ignitions can be considered as repeated observations of the same process.

A collection of independent point patterns is known as a *replicated point*

pattern (RPP) and we denote them as $\{\mathbf{X}_t\}_{t\in T}$, where T is an index set, which we denote as $T = \{1, 2, \ldots, n_{\text{days}}\}$ for convenience. The simplest case of a RPP is when each point pattern is identically and independently distributed, representing repeated observations of the same experiment. At its most general a RPP consists of a collection of conditionally independent, given the value of covariates, point patterns which are not identically distributed. For example, a covariate dependent inhomogeneous Poisson process can have intensity function

$$\lambda(u,t) = \exp\left\{\alpha + \sum_{j=1}^{d} \beta_j z_j(u,t)\right\},$$

where $z_j(u,t)$ is the value of the j^{th} covariate at location u on day t, giving rise to a collection of conditionally independent point patterns.

As we have mentioned above, point patterns generated by processes that evolve through time can be thought of as RPPs, but only if we are willing to assume independence between the observations. If this is an unreasonable assumption, an alternative approach is to consider the point patterns as coming from a point process on an expanded space. A *spatio-temporal point process* is a point process defined on the space $\mathbb{R}^2 \times T$, where T is the time dimension, usually $T = \mathbb{R}$ or $\mathbb{Z}$. The Poisson process is easily extended to this setting; the number of points in a space-time region $B \times [0,t]$ is Poisson distributed with mean $\int_0^t \int_B \lambda(u,t)\,\mathrm{d}u\,\mathrm{d}t$. Expanding the state space to include the time dimension allows for processes to have dependence in both space and time. Diggle (2013) provides a good introduction to the theory and application of spatio-temporal point processes.

The transition to RPPs has consequences for inference and validation of point process models. In Sections 7.2 and 7.3 we discuss how information from multiple point patterns can be combined to get estimates of model parameters or the K-function. Section 7.5 looks at validating models when replicated point patterns are available.

7.2 Inference with replicated point patterns

In Sections 6.3 and 6.7.3 we discussed inference for the Poisson process and the Neyman–Scott cluster process when there is a single point pattern available. Generalising those inference techniques to accommodate replicated point patterns is straight forward.

Homogeneous Poisson process The maximum likelihood estimator of the intensity of the homogeneous Poisson process becomes

$$\widehat{\lambda} = \frac{\sum_{t=1}^{n_{\text{days}}} n(\mathbf{X}_t \cap W)}{n_{\text{days}}|W|},$$

which is simply the average of the intensity estimates from each day.

Inhomogeneous Poisson process The likelihood function of the inhomogeneous Poisson process becomes

$$\mathcal{L}(\theta) = \prod_{t=1}^{n_{\text{days}}} \left(\prod_{x_i \in \mathbf{x}_t \cap W} \lambda_\theta(x_i, t) \right) \exp\left(- \int_W \lambda_\theta(u, t)\, \mathrm{d}u \right),$$

which can be maximised using the same approaches as before.

Cluster processes Inference for Neyman–Scott cluster process begins by estimating the intensity function using the method outlined above for the Poisson process. Cluster parameters are estimated using minimum contrast methods and requires an estimate of the K-function. As we will see in Section 7.3, estimating the K-function from RPPs requires some additional care.

7.3 Estimating the K-function from replicated point patterns

It is not immediately obvious how to estimate the K-function using replicated point patterns. One approach is to calculate an estimate of the K-function for

each observation in the RPP and then combine them into a single estimate. This is a common enough idea within statistics, known as *pooling*. An illustrative example of pooling is the case where we have two samples, $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$ from the same population and we wish to estimate the population mean. The sample means from the unpooled data sets, $\bar{x}$ and $\bar{y}$, can be combined as $(\bar{x}n + \bar{y}m)/(n+m)$ to calculate a pooled estimate for the mean.

More generally, suppose we are interested in estimating a quantity that can be expressed as a ratio $R = A/B$. A *ratio estimator* is an estimator of the form $\widehat{R} = \widehat{A}/\widehat{B}$, where $\widehat{A}$ is an estimator of A and $\widehat{B}$ is an estimator of B. If we have a collection of such estimators, $\widehat{R}_1 = \widehat{A}_1/\widehat{B}_1, \widehat{R}_2 = \widehat{A}_2/\widehat{B}_2, \ldots, \widehat{R}_n = \widehat{A}_n/\widehat{B}_n$, then a pooled estimator is given by

$$\overline{R} = \frac{\sum_{i=1}^n \widehat{A}_i}{\sum_{i=1}^n \widehat{B}_i} = \left(\sum_{i=1}^n \widehat{B}_i \right)^{-1} \sum_{i=1}^n \widehat{B}_i \widehat{R}_i \tag{7.1}$$

If the $(\widehat{A}_i, \widehat{B}_i)$ pairs are independent and identically distributed and they satisfy

$$\mathbb{E}(\widehat{A} \mid \widehat{B} = b) = bR$$

and

$$\mathrm{Var}(\widehat{A} \mid \widehat{B} = b) = cb,$$

where $c > 0$ is a constant then $\overline{R}$ is the minimum-variance linear unbiased estimator of R given $\widehat{B}_1, \widehat{B}_2, \ldots, \widehat{B}_n$ (Baddeley, Rubak, et al., 2015, Section 16.8.1). That is, given $\widehat{B}_1, \widehat{B}_2, \ldots, \widehat{B}_n$, the estimator $\overline{R}$ has minimum variance amongst all unbiased estimators that are linear combinations of $\widehat{A}_1, \widehat{A}_2, \ldots, \widehat{A}_n$.

Both the homogeneous and inhomogeneous K-functions can be expressed as ratio estimators. The empirical homogeneous K-function (6.21) can be expressed as

$$\widehat{K}(r) = \frac{\widehat{A}(r)}{\widehat{B}},$$

where

$$\widehat{A}(r) = |W| \sum_{x_i \in \mathbf{x}} \sum_{\substack{x_j \in \mathbf{x} \\ x_j \neq x_i}} \mathbb{1}\{\|x_i - x_j\| \leq r\} e(x_i, x_j) \tag{7.2}$$

and

$$\widehat{B} = n(\mathbf{x})(n(\mathbf{x}) - 1). \tag{7.3}$$

If we denote the t^{th} empirical homogeneous K-function as $\widehat{K}_t(r)$, then a pooled empirical homogeneous K-function follows directly from (7.1);

$$\widehat{K}(r) = c^{-1} \sum_{t=1}^{n_{\text{days}}} n(\mathbf{x}_t)(n(\mathbf{x}_t) - 1)\widehat{K}_t(r)$$

where $c = \sum_{t=1}^{n_{\text{days}}} n(\mathbf{x}_t)(n(\mathbf{x}_t) - 1)$ (Baddeley, Rubak, et al., 2015, Section 16.8.1).

Pooling empirical inhomogeneous K-functions is less straightforward since they are not identically distributed. The empirical inhomogeneous K-function (6.39) can be expressed as the ratio estimator

$$\widehat{K}_{\text{inhom}}(r) = \frac{\widehat{A}(r)}{\widehat{B}}, \tag{7.4}$$

where

$$\widehat{A}(r) = \frac{1}{|W|} \sum_{x_i \in \mathbf{x}} \sum_{\substack{x_j \in \mathbf{x} \\ x_j \neq x_i}} \frac{\mathbb{1}\left\{\|x_i - x_j\| \leq r\right\}}{\widehat{\lambda}(x_i)\widehat{\lambda}(x_j)} e(x_i, x_j) \tag{7.5}$$

and

$$\widehat{B} = D(\mathbf{x}). \tag{7.6}$$

For each point pattern in the RPP denote

$$\widehat{K}^{(t)}_{\text{inhom}}(r) = \frac{\widehat{A}_t(r)}{\widehat{B}_t},$$

where $\widehat{A}_t(r)$ and $\widehat{B}_t$ are the estimates of (7.5) and (7.6) coming from the t^{th} point pattern.

Despite the empirical inhomogeneous K-functions failing to be identically distributed, the currently accepted estimator for the pooled inhomogeneous

K-function is derived from (7.1),

$$\widehat{K}^{\text{pool}}_{\text{inhom}}(r) = \frac{\sum_{t=1}^{n_{\text{days}}} \widehat{A}_t(r)}{\sum_{t=1}^{n_{\text{days}}} \widehat{B}_t}, \tag{7.7}$$

(Baddeley, Rubak, et al., 2015, section 16.8.1).

However this estimator is overly influenced by bias in the estimate of the intensity function. The contribution of a pair of points to $\widehat{K}_{\text{inhom}}(r)$ is related to the reciprocal of the estimated intensity function at the locations of the points. Hence, points in regions where the estimated intensity function takes very small values have a large influence on $\widehat{K}_{\text{inhom}}(r)$. In practice, points that correspond to low values of the estimated intensity function can dominate the pooled estimator (7.7) resulting in jumps in the estimated function, even for large data sets. In the next section we discuss a new method for pooling estimates of the inhomogeneous K-function that reduces the variance over the naive approach.

7.4 The weighted pooled K-function

We propose a weighted pooled K-function which has a reduced variance over the naive pooled K-function. First, observe that the pooled estimator (7.1) can be expressed as

$$\overline{R} = \frac{\sum_{i=1}^{n} n^{-1} \widehat{A}_i}{\sum_{i=1}^{n} n^{-1} \widehat{B}_i},$$

which is the ratio of the means of the estimators. This suggests that we might be interested in estimating the numerator and denominator terms using weighted averages such as

$$\widehat{A} = \sum_{i=1}^{n} w_i \widehat{A}_i, \tag{7.8}$$

subject to $\sum_{i=1}^{n} w_i = 1$. Assuming that the $\widehat{A}_i$ are independent and unbiased, the choice of weights that minimises the variance of (7.8) are

$$w_i = \frac{\theta_i^{-1}}{\sum_{i=1}^{n} \theta_i^{-1}}, \tag{7.9}$$

where $\theta_i = \text{Var}(\widehat{A}_i)$ (Rubin and Weisberg, 1974). Note that this result is unchanged if all the θ_i are multiplied by the same constant factor, that is $\theta_i = c\text{Var}(\widehat{A}_i)$ for all i where $c > 0$ is a constant.

Consider the generalisation of (7.7),

$$\widehat{K}^{\text{pool}}_{\text{inhom}}(r) = \frac{\sum_{t=1}^{n_{\text{days}}} w_t^{-1} \widehat{A}_t(r)}{\sum_{t=1}^{n_{\text{days}}} w_t^{-1} \widehat{B}_t}, \tag{7.10}$$

where the w_t are weights. We propose that a good choice of weights is given by (7.9) with $\theta_t = \text{Var}(\xi_t)$ where

$$\xi_t = \sum_{x_i \in \mathbf{x}_t} \sum_{\substack{x_j \in \mathbf{x}_t \\ x_j \neq x_i}} \frac{1}{\lambda(x_i, t)\lambda(x_j, t)}. \tag{7.11}$$

We argue that these weights approximately minimise the variance of the numerator of (7.10) and if the normalising term is $D_{2-}(\mathbf{x})$ (equation 6.40), it also minimises the variance of the denominator.

To see this for the numerator, replace the estimated intensity function by the true intensity function and ignore the edge corrections. Then

$$\begin{aligned} \widehat{A}_t(r) &\approx \frac{1}{|W|} \sum_{x_i \in \mathbf{x}_t} \sum_{\substack{x_j \in \mathbf{x}_t \\ x_j \neq x_i}} \frac{\mathbb{1}\left\{\|x_i - x_j\| \leq r\right\}}{\lambda(x_i, t)\lambda(x_j, t)} \\ &\to |W|^{-1} \xi_t, \end{aligned}$$

as r becomes large enough that $\|x_i - x_j\| \leq r$ for all $x_i, x_j \in \mathbf{x}_t$. This assumes that W is bounded, which is a very mild assumption. So $\text{Var}[\widehat{A}_t(r)] \approx c\text{Var}(\xi_t)$, where c is a constant, showing that $\theta_t = \text{Var}(\xi_t)$ produces weights that approximately minimise the variance of the weighted sum (7.8).

The same holds for the denominator if the normalising term $D_{2-}(\mathbf{x})$ is used. This follows directly from the definition of $D_{2-}(\mathbf{x})$ (equation 6.40), assuming that estimated intensity function is equal to the true intensity function t; $D_{2-}(\mathbf{x}_t) = |W|^{-2}\xi_t$.

In the next section we cover the calculation of $\text{Var}(\xi_t)$ for the inhomogeneous Poisson process; it turns out to be very difficult for general point

processes. We believe that these weights may provide better performance even for non-Poisson processes. The weighted pooled estimator (7.10) is approximately ratio unbiased regardless of the choice of weights or whether $\mathbf{X}$ is Poisson. Because of this, using weights calculated for the Poisson process is unlikely to cause harm. In future work we will explore this by using simulated data to compare the weighted and unweighted pooled estimators of the K-function.

7.4.1 Calculation of weights

In this section we calculate the value of $\mathrm{Var}(\xi_t)$ for the inhomogeneous Poisson process; it is difficult to calculate for more general processes. To reduce the complexity of the notation, we supress the time index t and work with

$$\xi = \sum_{x_i \in \mathbf{X} \cap W} \sum_{\substack{x_j \in \mathbf{X} \cap W \\ x_j \neq x_i}} \frac{1}{\lambda(x_i)\lambda(x_j)} \tag{7.12}$$

where $\mathbf{X}$ is an inhomogeneous Poisson process. The mean of ξ is

$$\mathbb{E}(\xi) = |W|^2,$$

which follows from the extended Slivnyak–Mecke formula (Section 6.5.2). The second moment of ξ is

$$\mathbb{E}(\xi^2) = \mathbb{E}\left[\sum\sum_{x_i \neq x_j} \sum\sum_{x_{i'} \neq x_{j'}} \frac{1}{\lambda(x_i)\lambda(x_j)\lambda(x_{i'})\lambda(x_{j'})} \right],$$

where the sum is understood to be over all $x_i, x_j, x_{i'}, x_{j'} \in \mathbf{X} \cap W$.

To apply the Slivnyak–Mecke theorem we need the sum to be over all distinct pairs of points, so we split this sum into three cases. Either there are (i) two pairs of coincident points, (ii) one pair of coincident points or (iii) all points are distinct. Since $x_i \neq x_j$ and $x_{i'} \neq x_{j'}$ there are two ways for case (i)

to occur and four ways for case (ii) to occur. So

$$\begin{aligned}\mathbb{E}(\xi^2) = \mathbb{E}&\left[\sum\sum\sum\sum_{x_i\neq x_j\neq x_{i'}\neq x_{j'}} \frac{1}{\lambda(x_i)\lambda(x_j)\lambda(x_{i'})\lambda(x_{j'})}\right] \\ &+ 4\mathbb{E}\left[\sum\sum\sum_{x_i\neq x_j\neq x_i'} \frac{1}{\lambda(x_i)^2\lambda(x_j)\lambda(x_{i'})}\right] \\ &+ 2\mathbb{E}\left[\sum\sum_{x_i\neq x_j} \frac{1}{\lambda(x_i)^2\lambda(x_j)^2}\right].\end{aligned}$$

Then the extended Slivnyak–Mecke theorem yields

$$\mathbb{E}(\xi^2) = |W|^4 + 4|W|^2 \int_W \frac{1}{\lambda(u)}\,\mathrm{d}u + 2\left(\int_W \frac{1}{\lambda(u)}\,\mathrm{d}u\right)^2.$$

Putting the two moments together gives the variance of ξ as

$$\mathrm{Var}(\xi) = 4|W|^2 \int_W \frac{1}{\lambda(u)}\,\mathrm{d}u + 2\left(\int_W \frac{1}{\lambda(u)}\,\mathrm{d}u\right)^2.$$

Extending this result to non-Poisson processes presents significant challenges. Calculating the mean and second moment of ξ would require the Palm probabilities (Section 6.5.1), which can be very difficult to calculate for point processes that are non-Poisson.

7.5 Validation with replicated point patterns

Replicated point patterns provide new opportunities for model validation. If we think of the index t as time, we can reduce the point patterns in the RPP down to a purely spatial point process and a purely temporal process by considering $\bigcup_t \mathbf{X}_t$ and $\{n(\mathbf{X}_t \cap W)\}_{t\in T}$. Since the point patterns in a RPP are independent, the spatial and temporal processes can be easy to work with.

Since $\bigcup_t \mathbf{X}_t$ is a spatial point process, we can analyse its residual measure. Each point pattern in the RPP gives rise to a residual measure, which can be summed to give the residual measure of the spatial process $\bigcup_t \mathbf{X}_t$. For

the Poisson process, the summed residual measures will correspond with the residual measure of a Poisson process with intensity function equal to the sum of intensity functions from each of the days, $\lambda(u) = \sum_t \lambda(u,t)$. This can be analysed as usual, revealing possible misfits in the intensity function. As before, it is not obvious how to analyse residual measures for cluster processes, due to the dependence between spatial locations.

Thankfully we can analyse the temporal process $\{n(\mathbf{X}_t \cap W)\}_{t\in T}$ for both Poisson and cluster processes. By definition, the distribution of $n(\mathbf{X}_t \cap W)$ for the Poisson process is Poisson with mean parameter $\int_W \lambda(u,t)\,\mathrm{d}u$. Suitably chosen quantiles of the Poisson distribution can act as guidelines for large values of the observed temporal process. The distribution of $n(\mathbf{X}_t \cap W)$ for cluster processes is less obvious, but it is easy to estimate the quantiles using simulations. Days with observed values $n(\mathbf{x}_t)$ that fall outside suitably chosen quantiles may reflect poor fit and the model behaviour on these days should be investigated further.

Simulation offers further ad hoc methods for model validation. Simulated and observed RPPs can be compared using a suite of summary statistics, such as the number of days with one or more points. By including summary statistics that reflect different properties of the process, we can build up a picture of how well the model replicates the behaviour of the observed process. Suitable quantiles for the summary statistics can be estimated by quantiles from the simulated RPPs.

Choosing relevant summary statistics for this comparison is important and should be based on the objectives of the modelling project. They should be chosen to reflect properties of the model that are desirable for the given application. For example, if we want a model to forecast extreme fire days then we could consider the number of days with greater than twenty ignitions. By having a comprehensive suite of summary statistics we can see if and how a model fails to reflect the observed data.

Chapter 8

Point process models for lightning-caused wildfire ignition

A review of point process models in the lightning ignition literature. Methods and results for Poisson and cluster process models for lightning ignition. A return to the case studies. A discussion of the accuracy and practical value of the models.

8.1 Introduction

Point processes offer a natural framework within which to study wildfire ignition point data. Unlike regression models, the precise ignition location can be used and covariate data can be on different scales rather than having to adopt a common resolution. Like regression models, point processes can be used for inference or prediction. Inference can involve inspection of a fitted intensity function or an estimated K-function. Prediction can involve estimating the intensity as well as simulating point patterns under new conditions.

In terms of modelling wildfire ignitions, there is an important distinction between Poisson and non-Poisson point process models. In Section 6.3 we

discussed how inference for the inhomogeneous Poisson process is approximately equivalent to fitting a spatial logistic or Poisson regression model. Furthermore, in Section 3.1.3 we showed the equivalence of logistic and Poisson regression for events of low probability. Because of this we believe that any of a suitably constructed Poisson point process, logistic regression and Poisson regression model will yield similar modelling results, in the sense that the inferential conclusions and the predicted values from the models will be similar.

However, non-Poisson point processes offer a genuine difference to regression models. Regression models generally rely on the assumption of conditional independence between the observations: given the values of the covariates the response variables are independent. This is essentially the same assumption used for covariate dependent Poisson process models. However, non-Poisson point processes such as Cluster, Cox and Gibbs processes (Baddeley, Rubak, et al., 2015; Daley et al., 2008; Diggle, 2013) have interaction effects between the points. When modelling wildfire ignitions there are several reasons we might expect the assumption of conditional independence to be poor. Lightning storms can cause large numbers of lightning strikes clustered in time and space. Similary, for human-caused ignition, a single defective train or piece of machinery may cause multiple ignitions in the same region on the same day. These physical processes of ignition suggest that non-Poisson point processes may be more suitable for modelling wildfire ignition location than Poisson point processes.

8.1.1 Literature review

The use of point processes in the wildfire ignition literature can be roughly split into three groups: (i) Poisson process models, (ii) exploratory analysis with spatial correlation summary functions such as the K- and L-functions, and (iii) non-Poisson process models. As discussed above, Poisson process

models yield very similar results to suitable regression models. Many of the Poisson process papers have similar aims and conclusions to their regression model brethren (Brillinger, Preisler, and Benoit, 2003; Liu et al., 2012; Peng et al., 2005; Schoenberg et al., 2008; Xu et al., 2011; Yang, He, et al., 2007; Yang, Weisberg, et al., 2015), although they typically investigate spatial residuals much more rigorously. We focus our review on papers using non-Poisson process methods.

Work using K- and L-functions (Section 6.6) does not assume a model for the point patterns, but instead aims to detect the presence of interactions between the points of the process. Podur et al. (2003) analysed the L-function of lightning ignition point patterns for various years in Ontario, Canada, and found evidence of clustering. Similary, Wang and Anderson (2010) analysed K-functions of human- and lightning-caused wildfire ignition point patterns in Alberta, Canada, and found evidence of clustering for both causes. Wang et al. (2010) also looked at the cross-type K-function (Baddeley, Rubak, et al., 2015, Section 14.6.4) to study the interaction between the human- and lightning-caused ignition point locations, finding evidence of inhibition between the two types of points. Other exploratory work, Nichols et al. (2011), calculated summary point patterns known as prototypes, which are "analogous in many ways to the median of a collection of real numbers". They argued that these prototype point patterns represent the 'median' wildfire ignition point pattern.

Non-Poisson point processes can be used for both the prediction of points, and the exploration of interactions between them. The tutorial paper Turner (2008) used Gibbs point process models (Baddeley, Rubak, et al., 2015, Chapter 13) to analyse annual forest fire ignition point patterns for New Brunswick, Canada. Gibbs point processes are typically constructed using the *Papangelou conditional intensity function*, which can be roughly understood as the density function for the location of a point u conditional on the location of the remainder of the point process $\mathbf{x}\backslash\{u\}$. Gibbs point processes are generally

used for modelling inhibition between points (Baddeley, Rubak, et al., 2015, Section 13.1), although the Geyer model used in Turner (2008) can produce clustering.

Møller and Díaz-Avalos (2010) used an inhomogeneous spatio-temporal cluster process (Section 7.1) to model daily lightning-caused wildfire ignitions in the Blue Mountains, USA. The spatio-temporal cluster process $\mathbf{X}$ is defined on the space $\mathbb{R}^2 \times \mathbb{Z}$ and follows a similar construction to the cluster process defined in Section 6.7. The parent processes $\{\Phi_t\}_{t=-\infty}^{\infty}$ are (generally homogeneous) Poisson processes, the spatial kernel $h(u)$ is replaced by a bivariate density function $h(u,t)$ on $\mathbb{R}^2 \times \mathbb{Z}$ and the modulating intensity $\mu(u)$ is replaced by a bivariate function $\mu(u,t)$ on $\mathbb{R}^2 \times \mathbb{Z}$. Møller and Díaz-Avalos (2010) assumed that that the modulating intensity function was separable, $\mu(u,t) = \mu_1(u)\mu_2(t)$, and the spatio-temporal kernel was separable, $h(u,t) = h_1(u)h_2(t)$, where $h_1(u)$ is the radially symmetric bivariate Normal density function and $h_2(t)$ is a decreasing linear function on $\{0,1,\ldots,19\}$ such that it forms a probability mass function. These separability assumptions allowed the authors to find expressions for the K-functions of the spatial point process $\bigcup_t \mathbf{X}_t$ and the temporal point process $N = \{n(\mathbf{X}_t)\}_t$, where $\mathbf{X}_t$ is the point process corresponding to the points arriving at time t, $\mathbf{X}_t = \mathbf{X} \cap (\mathbb{R}^2 \times \{t\})$. These K-functions were used to estimate parameters using minimum contrast methods (Section 6.7.3).

Juan, Mateu, and Saez (2012) looked at annual ignition point locations for wildfires in Catalonia, Spain, split into four causes: natural, negligence, intentional and unkown. For their initial analysis they fitted Poisson process, Thomas cluster process (Section 6.7.2) and Gibbs process models for each year and each ignition cause. They extended this work by considering the years as observations from a replicated point pattern and fitting a single model per cause. To deal with the between-year variation they introduced random effect terms (Wood, 2017).

Serra et al. (2014) used the same data set as Juan et al. (2012), but fitted a log-Gaussian Cox process model (Baddeley, Rubak, et al., 2015, Section 12.2.4) to the annual ignition point locations. A log-Gaussian Cox process is a Cox process whose driving intensity function is a Gaussian field (an extension of Brownian motion to $\mathbb{R}^2$). Similarly to their previous work (Juan et al., 2012), random effect terms were included to deal with unexplained variation.

8.1.2 Relating point process models to wildfire management

A review of the point process literature as applied to the modelling of wildfire ignitions reveals a wealth of approaches that are often very complex. It can be helpful to consider these modelling approaches in terms of their modelling outcomes for wildfire management. In this section we discuss four management problems that models can assist with. We take regression models to be the default approach and discuss the potential improvements that non-Poisson point process models may provide.

Daily ignition forecasts for operational use

Models such as the logistic regression model presented in Chapter 5 can be used to produce daily estimates of ignition likelihood, or for point process models, the intensity function. These forecasts can assist fire managers to make decisions about resource allocation on a given day.

Typically the use of a non-Poisson point process model has no impact on the estimated intensity function. As we discuss in Section 6.7.4, the standard method for estimating the intensity function of a Neyman–Scott cluster process is identical to that for the Poisson process. The intensity function of the Gibbs process is not a simple function of the model parameters and must be approximated using a *Poisson-saddlepoint approximation* (Baddeley, Rubak, et al., 2015, Section 13.12.6). This approximation could introduce further errors into the estimation of the intensity function. Estimates of the intensity

function for a Poisson and non-Poisson point process model should be similar since the intensity is a first moment property of a point process. Non-Poisson point processes are concerned with interactions between the points, which are second (or higher) moment properties of the process.

Prediction intervals for the number of wildfire ignitions on a given day can also be an important component of model output for operational use. A Poisson process model can produce unrealistically narrow prediction intervals, giving a false impression of model certainty. If there is evidence of clustering or inhibition between the points, a non-Poisson point process model may produce more accurate prediction intervals. The variance of the number of points in a region for a cluster process will be larger than that for a Poisson process with identical intensity (Equation 6.52), resulting in wider prediction intervals.

Daily ignition forecasts for risk modelling

As discussed in Section 5.6, risk is often estimated using fire spread simulation models along with collections of initial ignition locations. Current approaches simulate each fire independently (Department of Environment, Land, Water and Planning, 2015), but future spread models may allow for interactions between multiple fires. If these models were to be developed, it would be important for the spatial patterns of ignition locations to reflect those likely to be experienced. Simulated point patterns from non-Poisson point process models could be used to initialise spread models for forecasting wildfire risk, and this could improve the accuracy of risk modelling. We have no insight into spread modelling and how achievable it is to model multiple fires simultaneously and so it may be the case that non-Poisson process models will offer little improvement to risk modelling in the foreseeable future.

Long term ignition forecasts

Point process models have been fitted to annual ignition patterns (Aragó et al., 2016; Díaz-Avalos, Juan, and Serra-Saurina, 2016; Juan et al., 2012; Serra et al., 2014; Turner, 2008), however these models may not be of much practical use since they do not link ignitions with the local weather conditions. The wildfire ignitions of most interest to management authorities are those occurring on extreme fire weather days when rapid escalation and spread is likely. In Victoria, single-day wildfire events have historically resulted in the largest loss of life and property (Department of Environment, Land, Water and Planning, 2015).

As we note in Section 5.6, fire spread models use weather conditions to simulate fire spread. Ignition models should reflect the daily weather conditions if they are to be used in conjunction with spread models. Models of annual ignition patterns could be used to study the clustering behaviour of ignitions, or the influence of covariates on the ignition process; however this faces similar criticism. A process may exhibit clustering on an annual time scale, but not on a daily time scale, making it difficult to interpret the results in terms of the fire events facing wildfire managers. For these reasons, we believe that models operating on a daily time scale give more insight.

Understanding ignition factors

Point process theory has been used to investigate the determinants of wildfire ignition. Point process models whose intensity is a function of covariates can be used to explore the impact of weather conditions or planning measures on ignition likelihood. Usually these studies look at the intensity function, whose estimate will be roughly equivalent for Poisson and non-Poisson point process models, as previously discussed.

Non-Poisson point processes allow the second (or higher) moment properties of wildfire ignition point patterns to be investigated. The K- and L-

functions have been used to explore the clustering behaviour of ignitions in both time and space. These papers often analyse annual ignition patterns making it difficult to interpret these functions in terms of the events of interest to wildfire managers; it would be more helpful to analyse these functions for daily point patterns. Nonetheless, non-Poisson point processes and the K- and L-functions are valuable tools for exploring the clustering or regularity behaviour of wildfire ignitions.

8.1.3 Chapter aims

This thesis is predominantly interested in using wildfire ignition models to assist operational decision making and risk analysis. In this chapter we compare cluster process and Poisson process models for daily lightning ignition point patterns. The Poisson process model is considered to be roughly equivalent to the logistic regression model discussed in Chapter 5 and we take this to be the baseline performance of an ignition location model. Given the above discussion on how these models can be used to assist wildfire managers, we are interested to see if the cluster process can improve the following:

1. the accuracy of the prediction interval for the number of wildfire ignitions on a given day; and,

2. the realism of the simulated ignition patterns for extreme fire weather days.

8.2 Method

We organised the data into point patterns for each day in the training period (Section 5.2, the 1997–1998 fire season until the 2004–2005 fire season) and validation period (Section 5.2, the 2005–2006 fire season until the 2011–2012 fire season), resulting in 1698 and 1486 point patterns respectively. These are the same time periods used in Chapter 5 for the logistic regression model.

Unlike regression models, point process models use the exact point locations data rather than converting them to presence–absence rasters.

Similarly, we used the same covariate data (Table 2.2) for the point process models as we did for the logistic regression model. Although we could have used the raw data, which would have avoided coercing the covariate data onto the shared regular grid, we used the same data as the regression model so that fair comparisons could be made between the models. In general, point process models do not need covariate data to share a common grid structure.

We considered three point process models for lightning-caused wildfire ignition: an inhomogeneous Poisson process model as well as two inhomogeneous cluster process models. The intensity function of the inhomogeneous Poisson process model has the log-linear form

$$\lambda(u,t) = \exp\left\{\sum_{j=1}^{d} f_j(z_j(u,t))\right\},$$

where $\{f_j\}$ are scaling functions.

We considered both a Thomas and a Cauchy cluster process model (Section 6.7.2), which we call the Thomas and Cauchy model respectively. Both cluster models share the following structure and associated notation:

- the center processes Φ_t are homogeneous Poisson processes with intensity $\kappa > 0$;
- the offspring processes $\mathbf{X}_{c,t}$ have intensity functions $\mu(u,t)h(u-c)$, where $\mu(u,t)$ is the modulating intensity and $h(u)$ is the spatial kernel; and,
- the modulating intensity has the form

  $$\mu(u,t) = \exp\left\{\sum_{j=1}^{d} f_j(z_j(u,t))\right\},$$

  where $\{f_j\}$ are scaling functions.

The cluster models differ in their spatial kernels, which are symmetric bivariate Normal and Cauchy densities for the Thomas and Cauchy models respectively.

To evaluate model fit we looked at both spatial and temporal diagnostics, although spatial residual measures are currently limited to the Poisson process. To assess how well the models reflect the clustering behaviour of the observed lightning ignition data we compared the empirical K-function with the model K-functions and calculated numerous summary statistics using simulated data. We returned to the case studies from Section 5.5 to compare the cluster process models with the logistic regression model. Finally, we examined how the cluster process models improve on the logistic regression model by evaluating the accuracy of prediction intervals for the number of fires on a day and by judging the realism of simulated point patterns.

8.2.1 Inference

Inference for the Poisson process is typically carried out using quadrature methods (Section 6.3); these approximate the likelihood function for the Poisson process and allow regression inference methods to be used. Some of these approaches are equivalent to fitting a logistic regression model (Section 6.3.2). This link suggests that model selection approaches for regression models could be co-opted to perform model selection for the intensity function of point processes.

Motivated by this, we scaled the covariates in the intensity function of our point process models using the scaling functions selected during model selection in Chapter 5. Expressing the intensity function in terms of fractional polynomials,

$$\lambda(u,t) = \exp\left\{\beta_0 + \sum_{j=1}^{d} \beta_{j1}(z_j(u,t))^{(p_{j1})} + \beta_{j2}(z_j(u,t))^{(p_{j2})}\right\},$$

we used Box–Tidwell transformations $(z_j(u,t))^{(p_{j1})}$ and $(z_j(u,t))^{(p_{j2})}$ identical to those in the logistic regression model (Table 5.4). With the fractional polynomial powers fixed, we used `kppm()` from *spatstat* to fit the coefficients, β_0

and $\beta_{j1}, \beta_{j2}, j = 1, 2, \ldots$. Allowing the coefficients to be fitted by `kppm()` gave the model some ability to adapt to the differences between the two processes.

We used K-functions and minimum contrast methods (Section 6.7.3) to estimate the cluster parameters of the Thomas and Cauchy processes. The empirical K-function was calculated for each point pattern and then combined to produce a pooled estimate of the K-function using the variance minimising weights proposed in Section 7.4. After the clustering parameters were estimated, the modulating intensity was estimated by $\widehat{\mu}(u,t) = \widehat{\lambda}(u,t)/\widehat{\kappa}$.

All work was carried out with the R package *spatstat* 1.55-0, which is the companion package to Baddeley, Rubak, et al. (2015).

8.3 Results

As a consequence of the model fitting process, the intensity function was identical for all three point process models. Table 8.2 details the terms in the log-intensity function.

Estimates of the cluster parameters can be found in Table 8.1. The Thomas model had larger estimated parent process intensity, $\widehat{\kappa}$, than the Cauchy model, corresponding to more frequent clusters of points. Figure 8.2 shows a radial cross section of the fitted spatial kernels. Compared to the Thomas model, the Cauchy model corresponds to more tightly clustered points, although the heavy tails allow small numbers of ignitions to fall far from their centre.

Figure 8.1 shows the weighted pooled empirical K-function, equation (7.10), as well as the unweighted version, equation (7.7), for comparison. The empirical K-function was notably above the πr^2 curve that signifies complete spatial randomness, suggesting a high degree of clustering in the process. In particular, the empirical K-function departs very rapidly from the πr^2 curve for very small values of r. Comparing the weighted and unweighted estimates

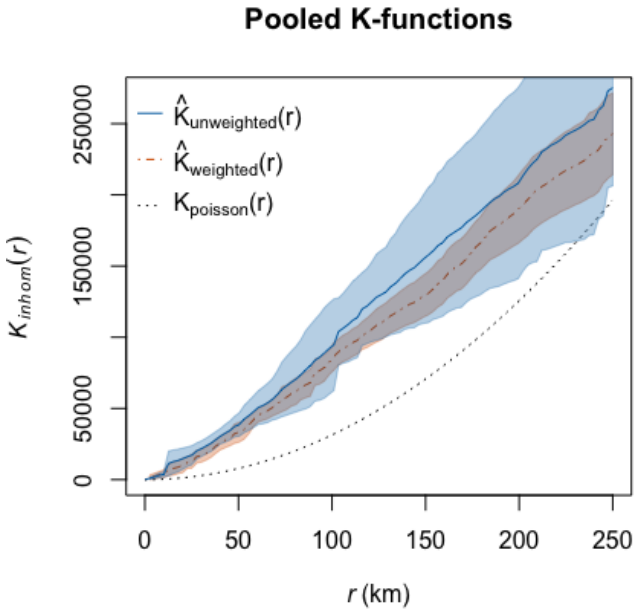


Figure 8.1: *The weighted and unweighted empirical K-functions with their associated 95% confidence intervals. The weighted K-function has a narrower pointwise confidence and was unaffected by the jump near $r = 0$.*

we see that the weighting smoothed the jump in the estimate near $r = 0$ and significantly reduced the width of the pointwise confidence intervals.

Parameter	$\widehat{\kappa}$	$\widehat{\sigma}$
Thomas	1.790×10^{-5}	27.41
Cauchy	1.339×10^{-5}	18.33

Table 8.1: Estimates of the cluster parameters.

Covariate	first term	second term
Intercept	-2.231	
Maximum temperature	$0.09708x$	
3 pm relative humidity	$4.790\log(x/100)$	$-17.03\sqrt{x/100}$
Precipitation	$-0.07866(x+0.1)^{-1}$	$0.07866\sqrt{x+0.1}$
Wind speed	$-0.2173(x/10)^{-1}$	
DF	$1.021x/10$	
KBDI	$4.364\sqrt{(x+0.1)/100}$	$-3.225(x+0.1)/100$
Maximum FFDI	$0.4131\log(x/10)$	$0.01420(x/10)^2$
CAPE index	$1.930\sqrt{(x+0.1)/100}$	$-0.2631(x+0.1)/100$
850 hPa Dewpoint depression	$-2.830(x+65.8)/100$	
Elevation	$2.017\sqrt{x/100}$	$-0.6426\sqrt{x/100}\log(x/100)$
Wet forest	$-0.6595x$	
Grassland	$-0.9642(x+0.1)^3$	
Heath	$-2.830x$	
Mallee spinifex	$1.205((x+0.1)/0.1)^{-2}$	$-3.726((x+0.1)/0.1)^{-2}\log\big((x+0.1)/0.1\big)$
Residential	$-1.714x$	

Table 8.2: The intensity function for the point process models is the exponential of the sum of all the terms in this table.

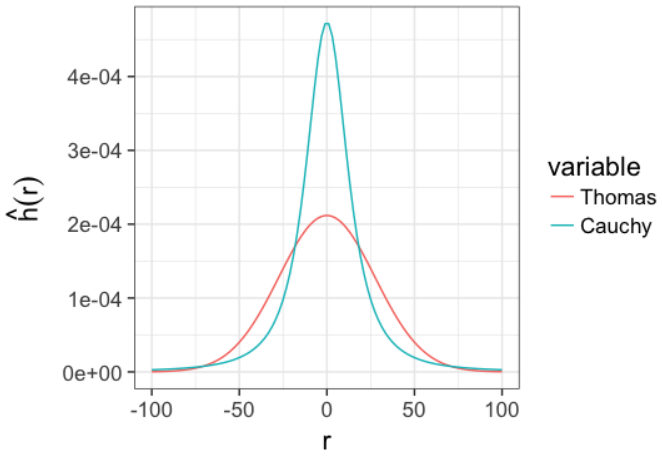


Figure 8.2: *A radial cross section of the estimated spatial kernels for the Thomas and Cauchy cluster processes. The Cauchy cluster process has tighter clustering around 0, but heavier tails than the Thomas process.*

8.3.1 Validation of spatial fit

Figure 8.3(a) shows the spatial residual measure (Section 6.4) of the Poisson model over the fitting period. The background colour map shows the negative of the sum of the intensity functions over all days. The point masses have positive charge and represent ignitions. Point masses with a charge greater than 1 are at locations with more than one recorded ignition, although these occurred on different days. Points should not coincide in a Poisson process (nor presumably in real life); coincident points are a result of data collection methods. Ignition point locations are often recorded from roads, rather than the actual site of ignition, which can cause records to have overlapping locations. The raw residual field shows that the Poisson process broadly predicts high intensity in a horizontal band across the middle of the state, which generally matches the ignition pattern.

The Pearson residual measure, Figure 8.3(b), shows the residuals after being reweighted to account for the natural variance of the Poisson process. To interpret this more easily, we calculated the smoothed Pearson residual field, as shown in Figure 8.3(c), using a Gaussian smoother with bandwidth 67.5 km,

the default chosen by *spatstat*. We felt that this smoother was too aggressive. Figure 8.3(d) shows the smoothed Pearson residual field calculated with a smaller bandwidth of 15 km. For an individual point pattern, the variance of the smoothed Pearson residual field (with smoothing kernel bandwidth 15 km) was 3.54×10^{-4}. Assuming that the point patterns that make up the RPP are independent, the variance of the smoothed Pearson residual field of $\bigcup_t \mathbf{X}_t$ is $3.54 \times 10^{-4} \times n_{\text{days}} = 0.600$. We added a contour to Figure 8.3(d) at twice the standard deviation about 0, namely ± 1.55.

The smoothed Pearson residual field showed generally poor fit for the intensity function for the Poisson model. Figure 8.3(c) shows that the model overpredicted likelihood in the west of the state and underpredicted likelihood in the east of the state. Figure 8.3(d) shows that within these regions of poor fit, the spatial distribution of the poor fitting residuals has no obvious pattern.

Figure 8.4 shows the model K-functions for the Thomas and Cauchy models, along with the empirical K-function and its 95% pointwise confidence intervals. The K-function of the Cauchy cluster process shows marginally better fit than that of the Thomas cluster process, managing to keep inside the pointwise confidence interval over the plotted range, $[0, 250]$.

8.3.2 Validation of temporal fit

To analyse the temporal fit we checked if the observed temporal process, $\{n(\mathbf{X}_t)\}_{t=1}^{n_{\text{days}}}$, fell within the model 95% prediction intervals. This is easy for the Poisson process since the distribution of the number of points on a given day is Poisson with mean parameter equal to the intensity measure of the state for that day, $\widehat{\Lambda}(W, t)$. For the cluster processes the distribution of the number of points in a region is not obvious and so we use simulated quantiles from the fitted models.

For the Poisson model, 92.2% of days and 77.5% of lightning-fire days fell between their 0.025 and 0.975 quantiles, suggesting a degree of poor fit. The

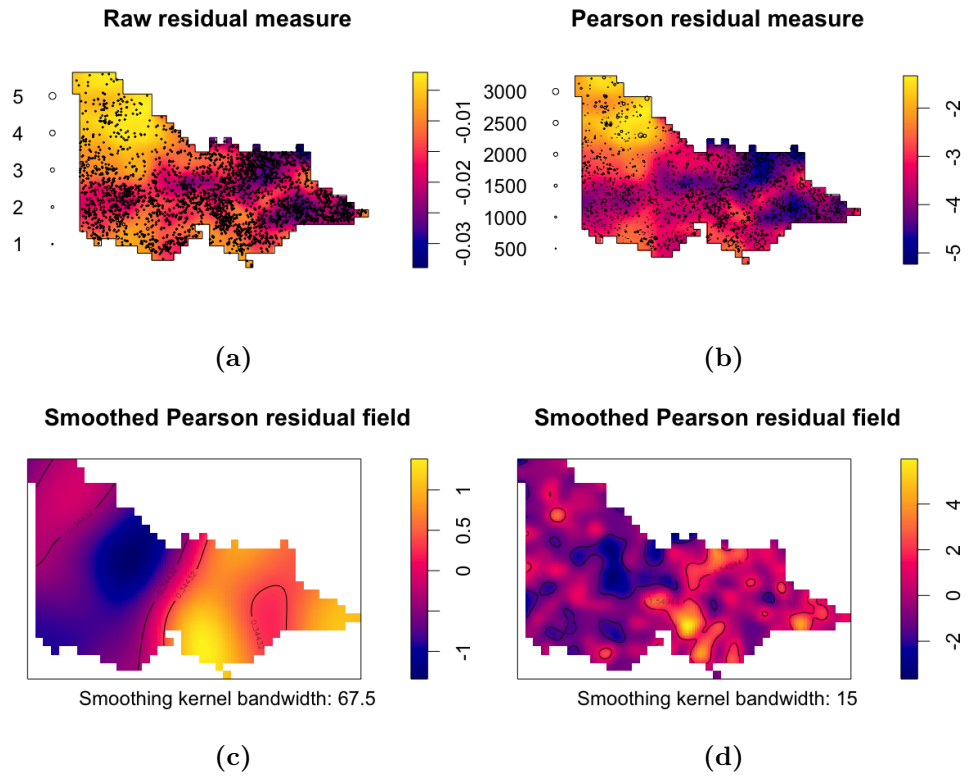


Figure 8.3: *Various residual measures for the inhomogeneous Poisson model. The smoothed field was calculated with a Gaussian smoother with the stated standard deviation (kernel bandwidth).*

Cauchy model performed much better, reporting 96.8% of days and 88.6% of lightning-fire days between their 0.025 and 0.975 quantiles. The performance of the Cauchy and Thomas cluster processes were almost identical and, for brevity's sake, we exclusively analyse the Cauchy process.

Closer inspection of the days with poor fit for the Cauchy model revealed exceptionally poor fit for days with very large numbers of ignitions. For example, the training data set included 29 days with 25 or more lightning ignitions and the Cauchy model showed poor fit for 15 of these (Table 8.3). The Cauchy cluster process underpredicted the number of ignition for all of these 15 poorly fitted days, in the sense that the 0.975 quantile for the number of ignitions was less than the observed number of ignitions. Often the underprediction was

spectacular with the 0.975 quantile sitting below half the observed number of ignitions.

Model performance on 2003-01-08 is worthy of further discussion, since this day marks one of the more significant fire events in recent Victorian history and since the model performed so abysmally. The Alpine and Canberra Fires were started by a series of lightning strikes on the 7^{th} and 8^{th} of January 2003 by a cold front moving over southeast Australia (Stephenson, 2010). In Victoria, the fires quickly escalated and merged, eventually burning 1.1 million ha over 59 days. The model found the period surrounding and including 2003-01-08 to have a very low likelihood of ignition. The reason for this is simple: the values of the covariates included in the model were all moderate over this period. However, Australia was enduring one of the worst droughts in recorded history (Worboys, 2003) and there is concern that the dryness indices (the Drought Factor and KBDI) were hitting their maximum values too easily, hiding the true level of dryness.

Date	No. Ignitions	$\widehat{\Lambda}(W_t)$	0.025 quantile	0.975 quantile
1997-11-26	130	27.0	4	76
1998-02-27	26	0.9	0	3
1998-12-26	51	6.2	0	18
1999-01-28	107	25.8	3	62
2001-01-04	96	32.0	6	68
2001-01-05	54	4.9	0	14
2002-02-15	74	20.0	1	45
2002-11-02	73	9.4	1	25
2002-12-03	37	3.1	0	10
2003-01-08	88	0.6	0	3
2003-02-27	26	4.3	0	15
2003-11-30	49	17.2	1	41
2004-01-04	26	2.3	0	8
2005-01-20	56	18.0	1	44
2005-01-26	74	19.8	2	43

Table 8.3: Days with at least 25 lightning ignitions and for which the observed number of ignitions exceeded the estimated 0.975 quantile of the fitted Cauchy cluster process.

8.3.3 Validation using simulated point patterns

For each day in the validation period we used the estimated intensity function for that day to simulate 100 point patterns from each of the three fitted processes. This resulted in 100 independent replicated point patterns for each process, simulating the lightning ignition process over the entire validation period. For each of the 100 simulation runs we calculated the total number of lightning ignitions, total number of lightning-fire days, the maximum number of ignitions on a single day and the mean number of lightning ignitions on a lightning-fire day. Table 8.4 shows the 95% prediction intervals for each of these statistics, estimated using the 0.025 and 0.975 sample quantiles.

In terms of the total number of ignitions, none of the models showed a lack of fit. This is unsurprising given that the intensity of the process is strongly connected to the total number of points. The two cluster processes had wider prediction intervals, reflecting a higher variance caused by the clustering of points (equation 6.52). Similarly, none of the models showed poor fit in terms of the maximum number of ignitions on a single day, although the prediction intervals for the two cluster processes were exceptionly large, possibly unreasonably so.

All models showed poor fit in terms of the number of lightning-fire days. A total of 451 lightning-fire days were observed during the validation period, whilst the Thomas cluster process performed best with a prediction interval of $[702, 767]$. This misfit was reflected in the mean number of ignitions on a lightning-fire day, which, for all models, was much lower than than the empirical mean of the observed data. For example, the 95% confidence interval for the mean number of ignitions on a lightning-fire day for the Thomas model was $[3.73, 4.50]$, whereas the observed empirical mean was 6.63.

We also calculated the value of the summary statistics in Table 8.4 for each fire season, seen in Figures 8.5 and 8.6. The model misfit over the whole validation period is repeated for each year. Furthermore, we see that

Statistic	Observed	Poisson	Thomas	Cauchy
Total no. ignitions	2991	[2924, 3114]	[2719, 3291]	[2784, 3364]
Total no. lightning-fire days	451	[778, 839]	[702, 767]	[717, 775]
Max no. ignitions on single day	85	[74.5, 107]	[63.0, 247]	[58.0, 241]
Mean no. ignitions on lightning-fire day	6.63	[3.61, 3.88]	[3.73, 4.50]	[3.69, 4.52]

Table 8.4: Comparisons of the observed and simulated data over the validation period. Simulated 95% prediction intervals are given for each of the processes.

the model failed to react to the significant variation in the lightning ignition process between years. For example, the 2006–2007 fire season saw a very large number of wildfire ignitions whilst the 2010–2011 fire season saw very few. Figure 8.5(a) shows that the models predicted moderate numbers of ignitions for the 2006–2007 fire season, roughly half the number of ignitions observed. The models predicted slightly below average numbers of ignitions for the 2010–2011 fire season, but failed to reflect the extremely low numbers observed. The models failure to react to annual variation in ignition counts could be due to missing covariates or it could be that the lightning ignition process is non-stationary. It is also worth noting that single day fire events can have a huge impact on the total number of ignitions in a fire season. A model that performs poorly on the most extreme day of the fire season may report terrible fit for that fire season, despite fitting well for all other days.

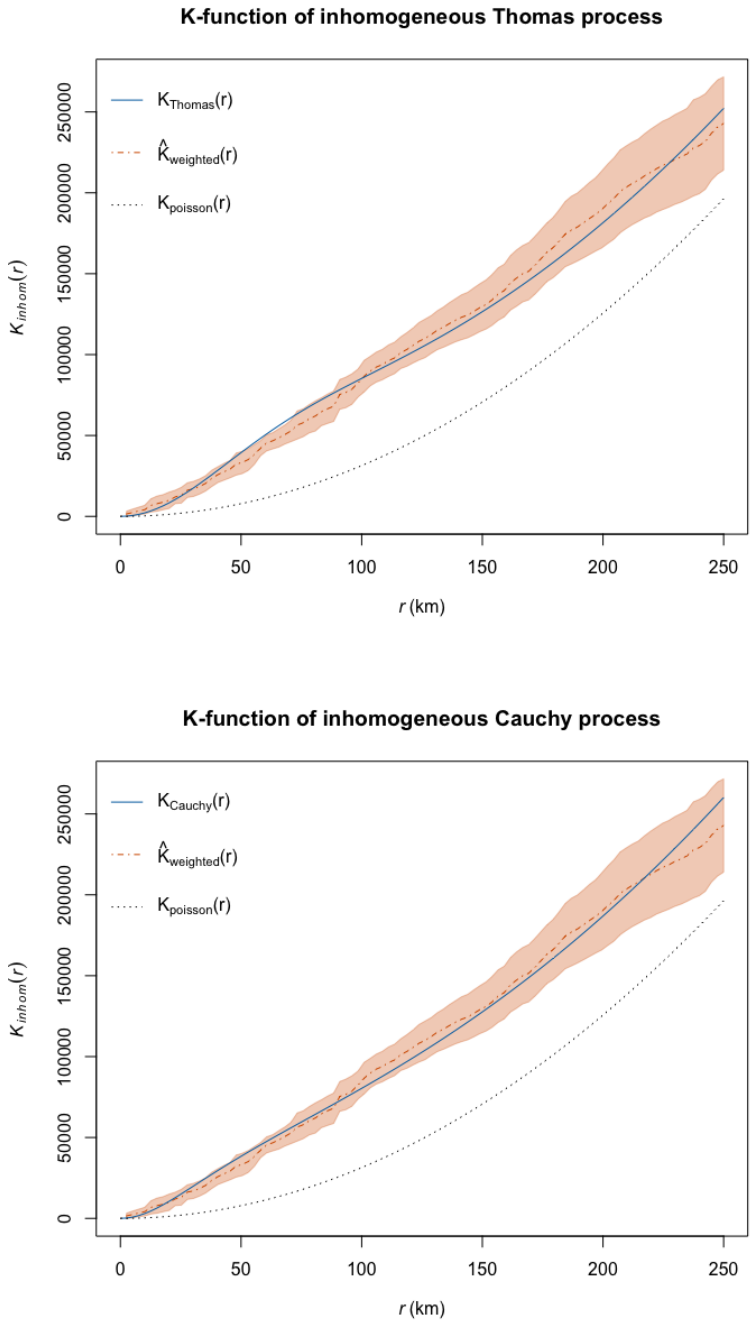
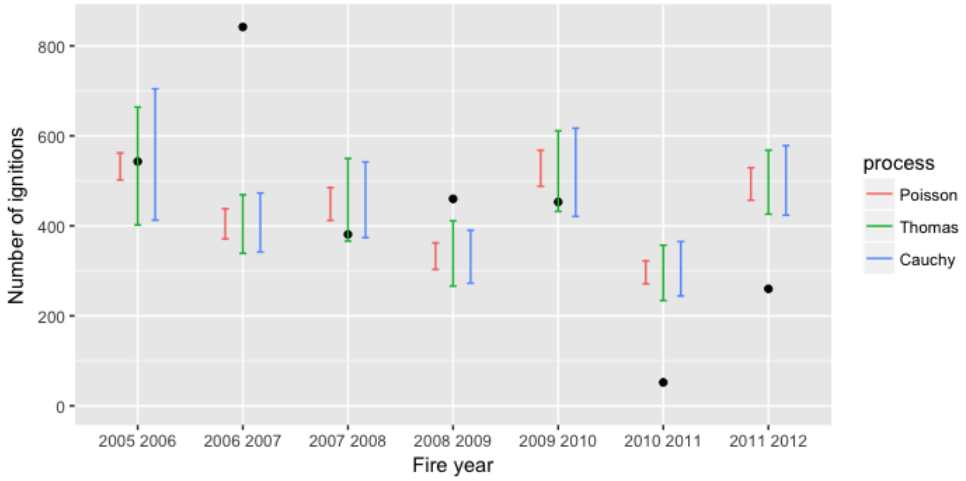
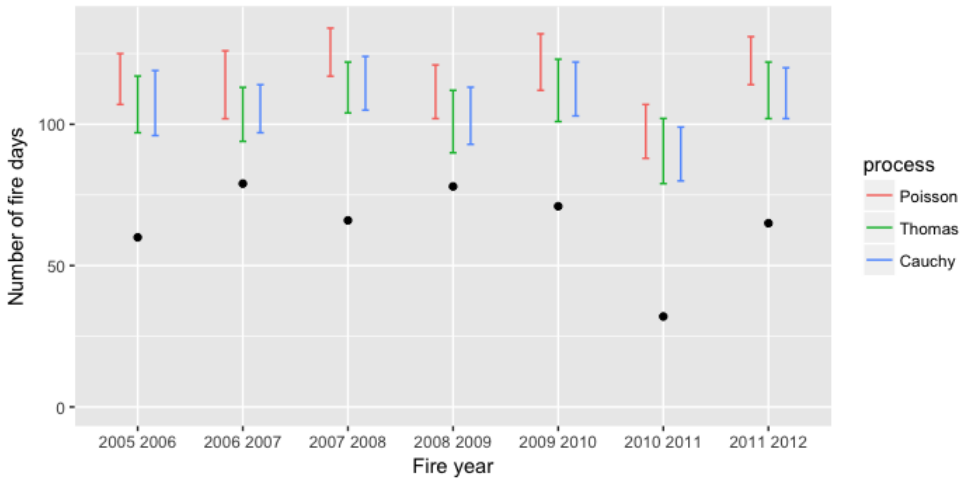


Figure 8.4: *The K-functions for the fitted models. The solid blue line is the empirical K-function and the dashed lines form the pointwise 95% confidence interval. The solid black line is the πr^2 curve and is the K-function for a Poisson process, representing complete spatial randomness.*

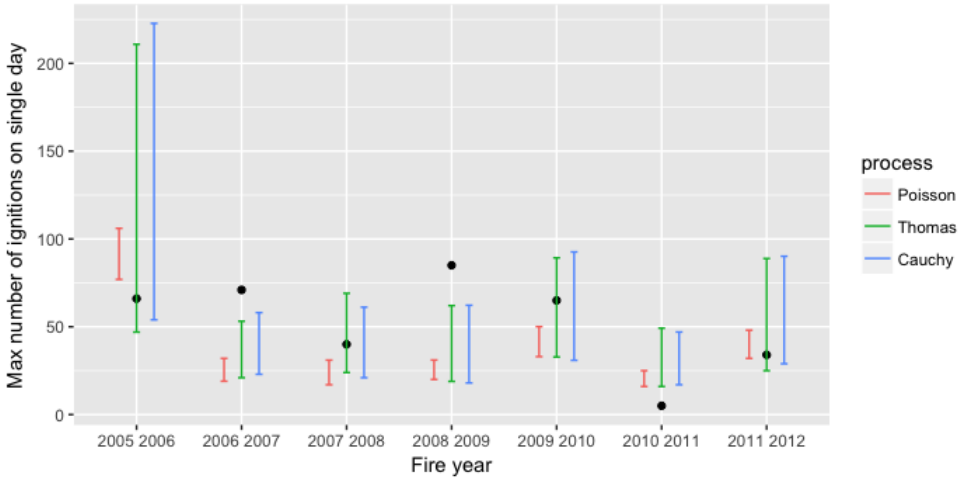


(a) *Total number of ignitions by fire season.*

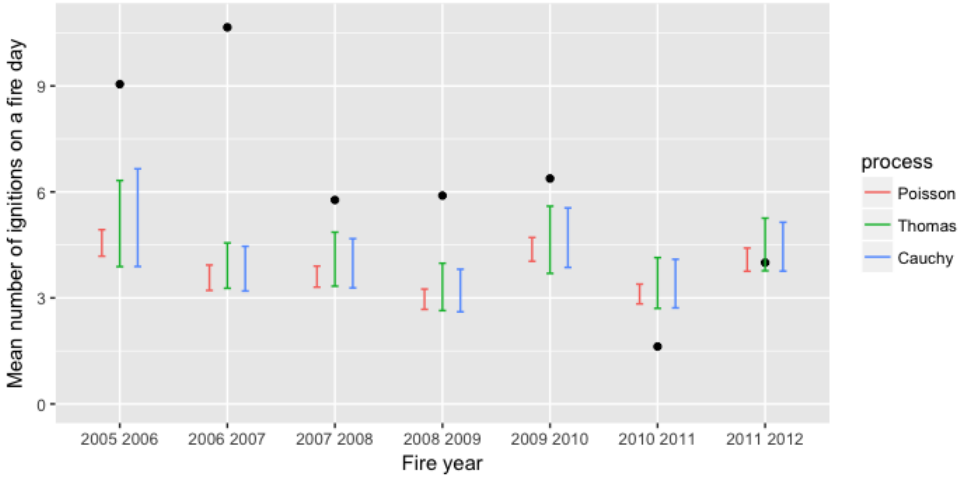


(b) *Total number of fire days by fire season*

Figure 8.5: *Plots comparing observed data with 95% confidence intervals from model simulated data.*



(a) *Maximum number of ignitions on a single day by fire season.*



(b) *Mean number of ignitions on a fire day by fire season.*

Figure 8.6: *Plots comparing observed data with 95% confidence intervals from model simulated data.*

8.4 Case studies

Similar to our analysis of the logistic regression model, we looked at model output from the fitted point process models to get a better sense of model performance. In order to compare the cluster process models with the logistic regression model, we returned to the same four periods chosen for analysis in Section 5.5. Although the intensity function of the point process models is of practical interest, it is visually almost identical to the likelihood function of the logistic regression model and so we focused entirely on the prediction intervals produced by the point process models.

Table 8.5 shows the observed number of ignitions together with the 95% prediction intervals for the Poisson process, Thomas and Cauchy models. Unsurprisingly, the cluster process models do not make up for a poor fitting intensity function. For example, on the 8^{th} of February 2009 the Poisson model produced a prediction interval of $[0, 6]$ whilst the Thomas and Cauchy models produced prediction intervals of $[0, 9]$ and $[0, 7]$ respectively. No model accurately reflected the 23 observed ignitions.

The cluster process models produced much wider prediction intervals. For 2009-02-07, 2010-02-10 and 2006-01-21, this made the difference between the observed number of ignitions falling inside or outside of the prediction interval. In general we believe that the wider prediction intervals are more realistic. The lower bound of the prediction intervals given by the cluster process models were often much closer to 0. Even on days with extreme fire weather, if there is no lightning then there can be no lightning-caused wildfire ignition, suggesting that accurate prediction intervals should often include 0. The prediction intervals for the Poisson model could have unrealistically large lower bounds, for example, 25, 41 and 74 for 2010-02-10, 2006-01-20 and 2006-01-21 respectively.

However the upper bound of the prediction intervals for the cluster process models may be unrealistically large. The Thomas and Cauchy cluster

Date	Observed	Poisson	Thomas	Cauchy
Black Saturday				
2009-02-05	1	$[0, 4]$	$[0, 4]$	$[0, 5]$
2009-02-06	1	$[0, 5]$	$[0, 7]$	$[0, 5]$
2009-02-07	9	$[15, 31]$	$[5, 58]$	$[4, 48]$
2009-02-08	23	$[0, 6]$	$[0, 9]$	$[0, 7]$
2009-02-09	13	$[0, 2]$	$[0, 2]$	$[0, 3]$
2009-02-10	1	$[0, 2]$	$[0, 3]$	$[0, 2]$
Elevated ignition activity				
2010-02-08	2	[0,6]	[0,10]	[0,9]
2010-02-09	58	[6,22]	[1,33]	[1,35]
2010-02-10	65	[25,51]	[1,98]	[5,105]
2010-02-11	19	[14,32]	[2,66]	[2,60]
2010-02-12	2	[0,3]	[0,4]	[0,3]
2010-02-13	0	[0,2]	[0,2]	[0,3]
High forecast ignition likelihood				
2006-01-18	5	[0,6]	[0,13]	[0,12]
2006-01-19	30	[1,13]	[0,14]	[0,21]
2006-01-20	66	[41,68]	[4,134]	[9,130]
2006-01-21	56	[74,107]	[6,230]	[10,241]
2006-01-22	43	[2,12]	[0,18]	[0,18]
2006-01-23	4	[0,5]	[0,8]	[0,9]
Moderate ignition activity				
2011-01-20	0	[0,4]	[0,4]	[0,4]
2011-01-21	4	[2,10]	[0,18]	[0,17]
2011-01-22	2	[2,13]	[0,20]	[0,26]
2011-01-23	1	[0,4]	[0,6]	[0,5]
2011-01-24	0	[0,2]	[0,2]	[0,1]
2011-01-25	0	[0,2]	[0,2]	[0,2]

Table 8.5: Observed ignitions and 95% prediction intervals for the four case study periods.

processes produced the prediction intervals $[6, 230]$ and $[10, 241]$ respectively for 2006-01-21, whereas the Poisson process produced a prediction interval of $[74, 107]$. An upper bound in excess of 200 may be physically unrealistic; the maximum number of lightning ignitions observed in a single day during the study period was 130.

8.5 Prediction intervals

To evaluate the accuracy of the prediction intervals we calculated the proportion of days in the validation period where the observed number of ignitions fell within the 95% prediction interval of the model. This is closely linked with the analysis of temporal diagnostics (Section 8.3.2), except that we now focus on the validation data. We interpret this quantity as the estimated coverage probability of the prediction interval. Note that the calculated prediction intervals are simply the 95% quantiles from simulated data and do not incorporate uncertainty in the parameter estimates. Table 8.6 shows the estimated coverage probability for each of the three models. The prediction intervals for the Poisson model were too narrow, reflected in the estimated coverage probability of 0.888 across all days in the validation period. The coverage probabilities for the Thomas and Cauchy models were 0.948 and 0.950 respectively, showing good accuracy.

However we were concerned that the reported accuracy of the prediction intervals was being inflated by days with no lightning fires. Many of these days are easy to predict, indeed a model which always predicts 0 ignitions would have a coverage probability of 0.697, which is the proportion of days with no fires. A more meaningful quantity is the estimated coverage probability for lightning-fire days. The estimated coverage probability for lightning-fire days was 0.716 for the Poisson model and 0.834 and 0.849 for the Thomas and Cauchy models respectively. As expected, restricting the calculation to lightning-fire days reduced the accuracy of the prediction intervals.

We also explored model performance during severe conditions. We looked at *observed severe days*, defined as days with 10 or more ignitions, as well as *predicted severe days*, defined as days where the intensity measure for the state was above 8.89. The choice of threshold for observed severe days was made based on practical considerations and led to 78 observed severe days in the validation data. The choice of threshold for predicted severe days was

made so that the number of predicted severe days in the validation data was also 78. Looking at observed severe days, we saw a further drop in accuracy for all models, with an estimated coverage probability of 0.244 for the Poisson process and 0.410 and 0.462 for the Thomas and Cauchy models respectively. On predicted severe days the cluster process models performed much better than the Poisson model, with coverage probabilties of 0.821 and 0.782 for the Thomas and Cauchy models respectively and 0.295 for the Poisson model.

In general, the Poisson model produced unrealistically narrow prediction intervals. The two cluster process models performed similarly to each other, and both were a marked improvement over the Poisson model. However the coverage probabilities for all of the models were low for severe observed days.

	Poisson	Thomas	Cauchy
All days	0.888	0.948	0.950
Lightning-fire days	0.716	0.834	0.849
Observed severe days	0.244	0.410	0.462
Predicted severe days	0.295	0.821	0.782

Table 8.6: The proportion of days in the validation data set where the modelled 95% prediction interval contains the observed number of ignitions. Observed severe days were those with 10 or more ignitions and predicted severe days were days where those were the intensity measure of the state was above 8.89.

8.6 Lineups

If ignition models are used in risk assessment then it is reasonable to expect that the observed ignition pattern could have been generated from the fitted model. One way to assess this is to produce a 'police lineup' consisting of the observed point pattern along with simulated point patterns from a model (Buja et al., 2009). We could conclude that the simulated point patterns are 'realistic' if a human struggles to accurately identify the observed point pattern. This sort of visual inspection test is helpful for noticing issues with

model fit that might be missed by specific summary statistics.

Risk analyses typically model fire spread using the 'worst' observed fire weather conditions. To keep in line with this we investigated severe days, as defined in Section 8.5, during the training period. There were 69 observed severe days and 69 predicted severe days, resulting in a total of 95 unique days selected for analysis.

To demonstrate the performance of the models, we present lineups for three days chosen uniformly at random from those selected as severe days. Figures 8.7 to 8.9 show lineups for the Poisson model, Figures 8.10 to 8.12 show lineups for the Thomas model and Figures 8.13 to 8.15 show lineups for the Cauchy model. We have intentially left the labelling ambiguous so that the reader may challenge themselves to identify the observed point pattern; their true identities are revealed in the captions.

Identifying observed point patterns from the Poisson process lineups was relatively easy. Only 30.5% of the days selected for analysis where such that the 95% prediction interval for the total number of ignitions included the observed number of ignitions. Typically the observed point pattern had notably more or few points than the simulated point pattern (Figure 8.7 for example). For days when the simulated point patterns had similar numbers of points, the observed point pattern could often be identified because it seemed to have a higher degree of clustering (Figures 8.8 and 8.9 for example).

It was more difficult to identify the observed point patterns from the cluster process lineups. Firstly, the cluster process models were more likely to generate point patterns with reasonable numbers of points; for the Thomas and Cauchy processes respectively, 73.7% and 70.5% of the days selected for analysis where such that the 95% prediction interval for the total number of ignitions included the observed number of ignitions. Both cluster process models produced convincing point patterns for days with moderate numbers of points (Figures 8.12 and 8.15 for example).

The spatial distribution of observed lightning ignitions on days with very large numbers of lightning ignitions was difficult to replicate. There was significant variation in the clustering dynamics between days. For example, the observed point pattern on 1998-01-14 had 30 points with relatively regular spacing. The Thomas model produced overly clustered point patterns for this day (Figure 8.10). The Cauchy model fared much better (Figure 8.13), thanks to the heavy tails of the spatial kernel. Contrast this with the observed point pattern on 2001-01-02, which showed a high degree of clustering. Figures 8.11 and 8.14 show both cluster process producing much more convincing point patterns during these conditions, however the observed point pattern can be easily identified by its highly non-symmetrical clusters. Overall, the heavy tails of the spatial kernel of the Cauchy model allowed it to produce more realistic point patterns on these extreme days, although it was far from reliable.

It is also worth noting that although it can be difficult to identify the observed point patterns, the cluster models would often produce clearly unrealistic point patterns (plot 12 from Figure 8.10 for example). Although it is a sign of underlying poor fit, these unrealistic point patterns could be discarded before use in risk analysis. Lastly, the Poisson model performed very poorly on days that had high estimated ignition likelihood but no observed ignitions. Whereas the cluster models could produce point patterns with no points, courtesy of the large variance of the total number of points (equation 6.52), the Poisson model would almost always produce point patterns with a nonzero number of points.

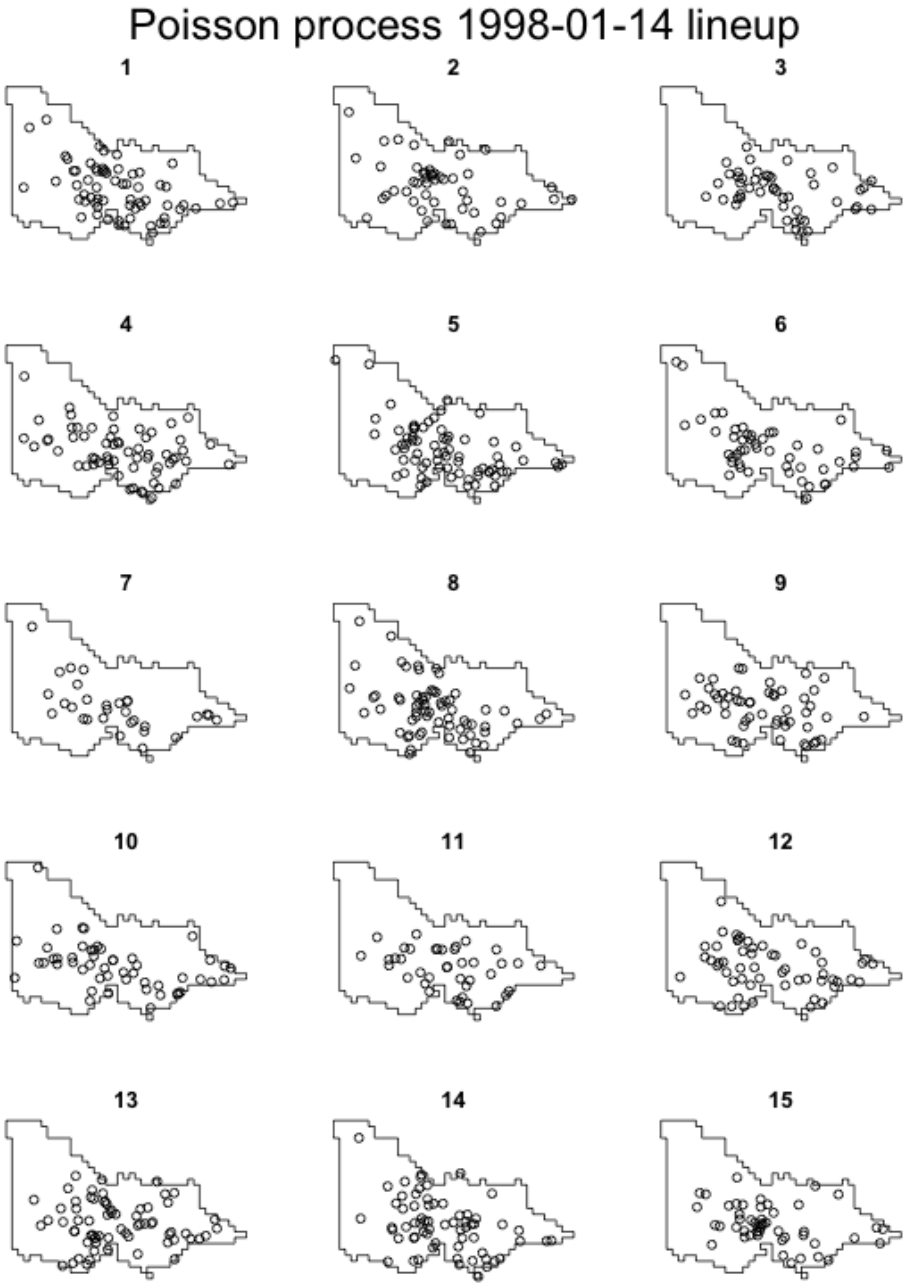


Figure 8.7: *A lineup of simulated point patterns from the Poisson model along with the observed point pattern (pattern 7, which has 30 points).*

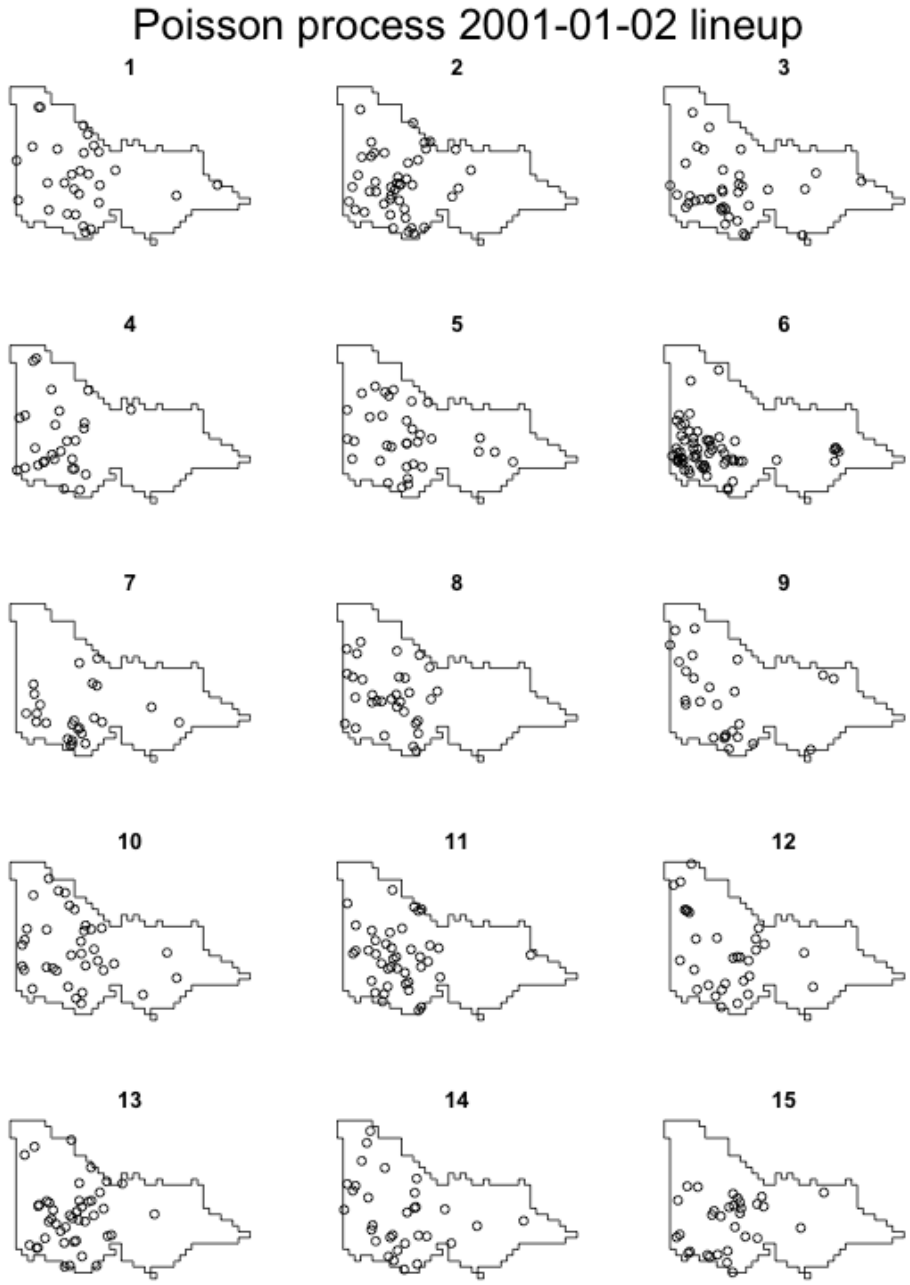


Figure 8.8: *A lineup of simulated point patterns from the Poisson model along with the observed point pattern (pattern 6, which has 63 points).*

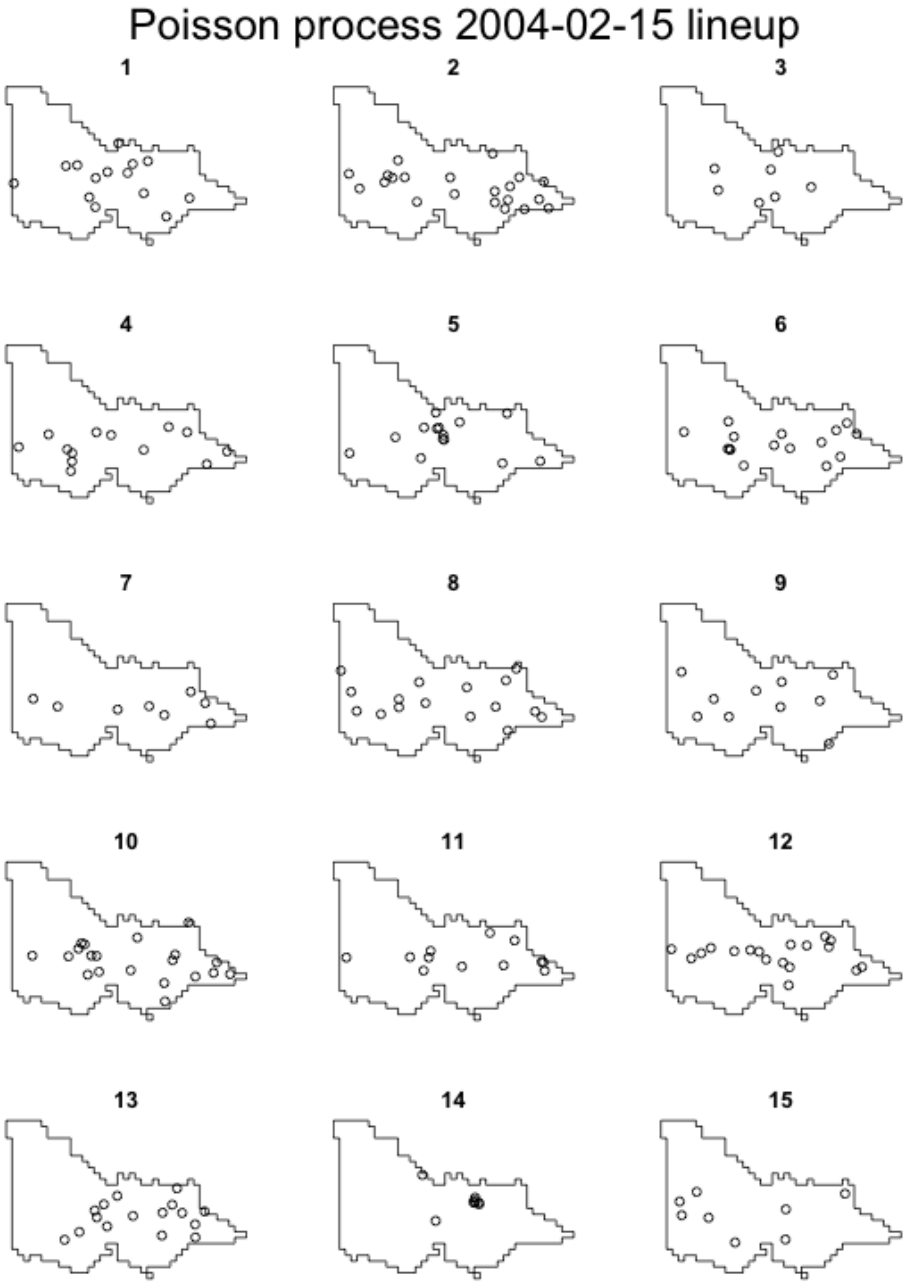


Figure 8.9: *A lineup of simulated point patterns from the Poisson model along with the observed point pattern (pattern 14, which has 8 points).*

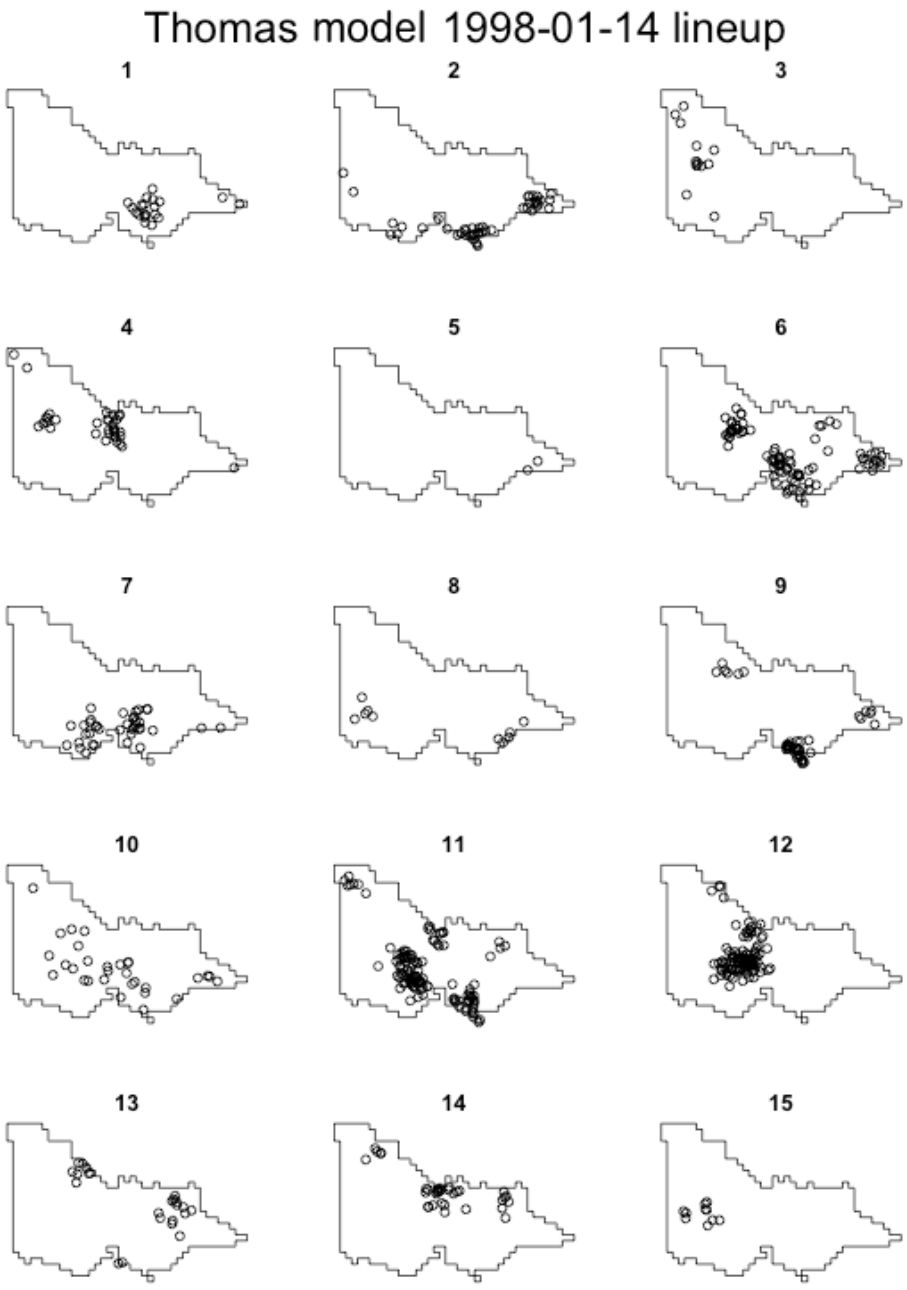


Figure 8.10: *A lineup of simulated point patterns from the Thomas model along with the observed point pattern (pattern 8, which has 30 points).*

Figure 8.11: *A lineup of simulated point patterns from the Thomas model along with the observed point pattern (pattern 3, which has 63 points).*

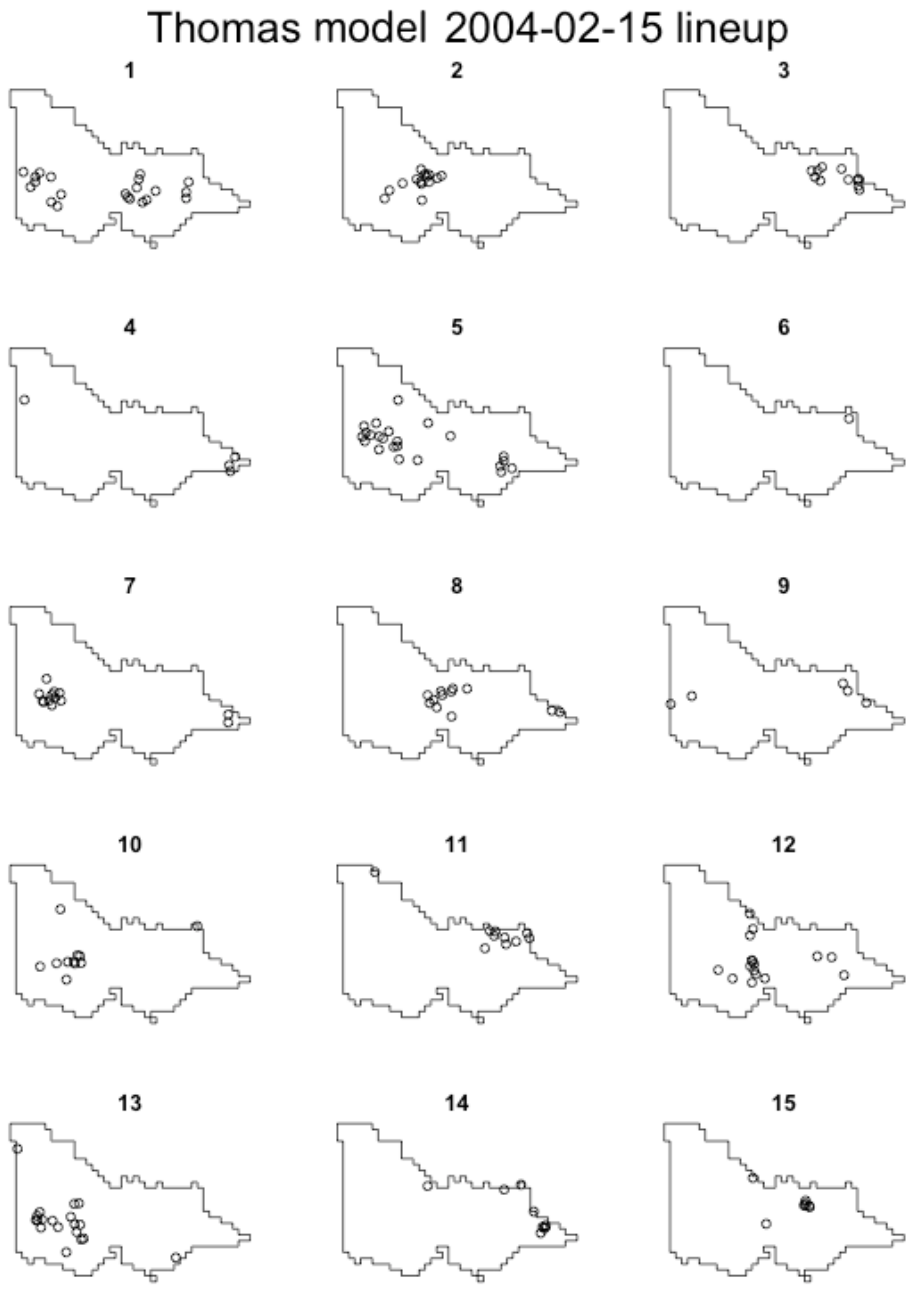


Figure 8.12: *A lineup of simulated point patterns from the Thomas model along with the observed point pattern (pattern 15, which has 8 points).*

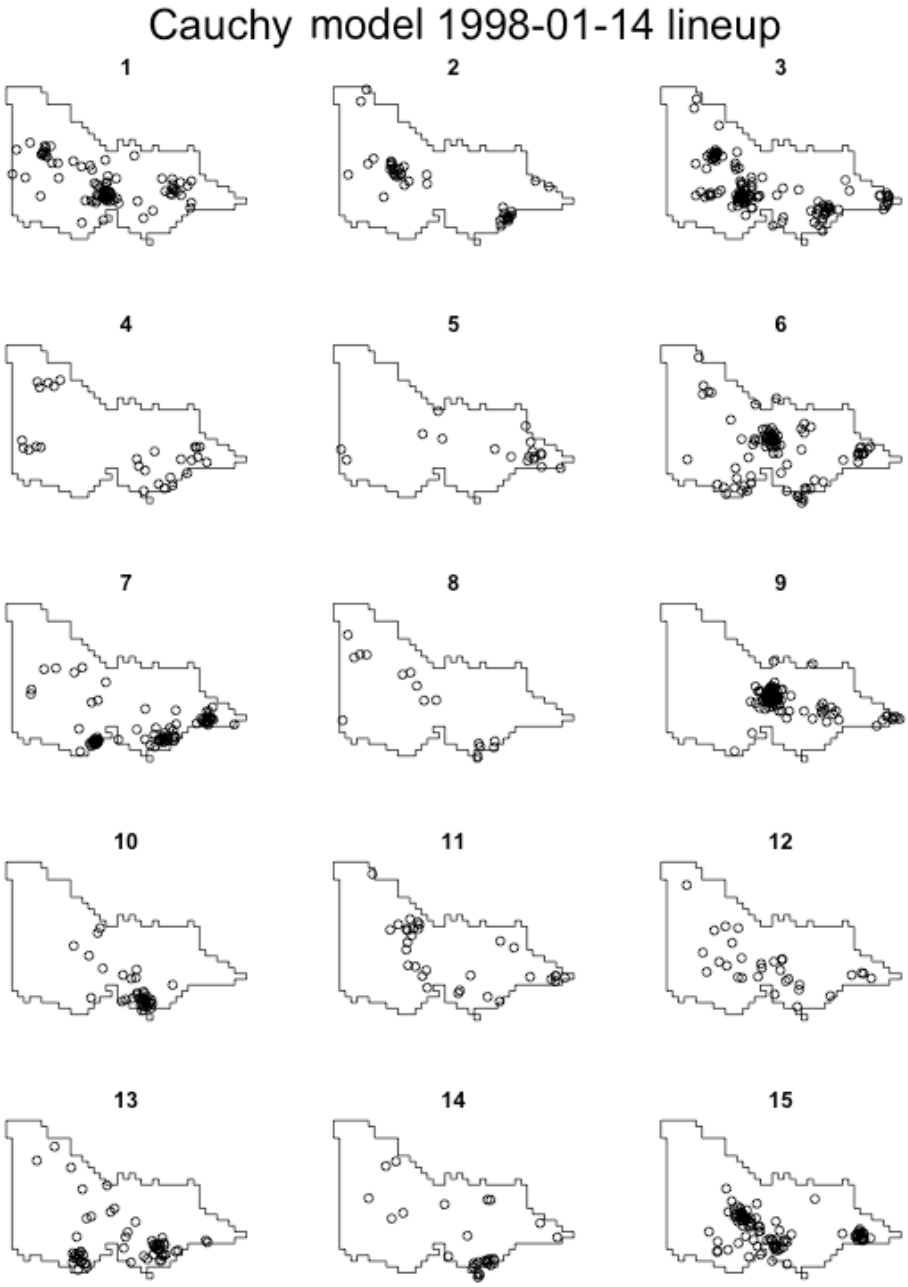


Figure 8.13: *A lineup of simulated point patterns from the Cauchy model along with the observed point pattern (pattern 12, which has 30 points).*

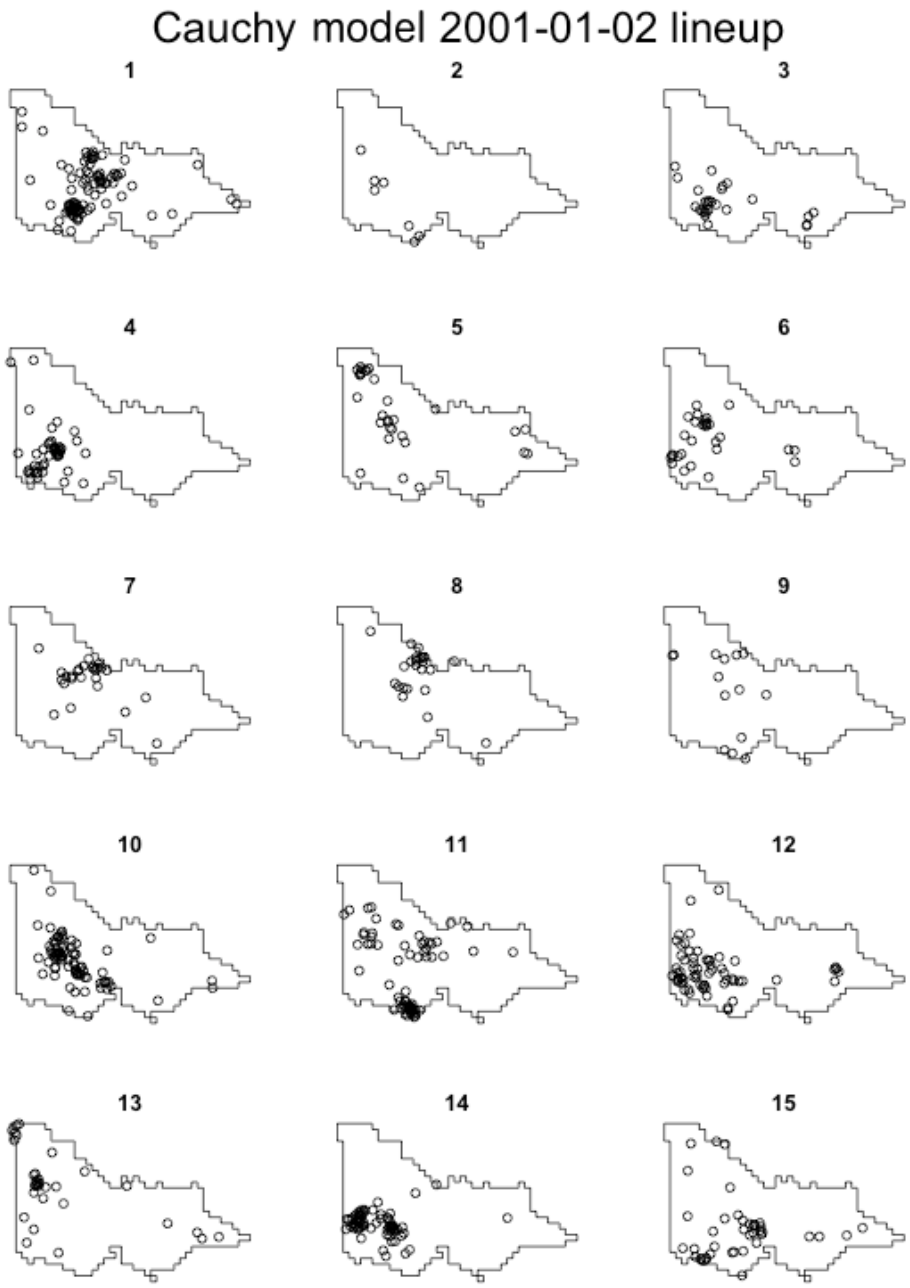


Figure 8.14: *A lineup of simulated point patterns from the Cauchy model along with the observed point pattern (pattern 12, which has 63 points).*

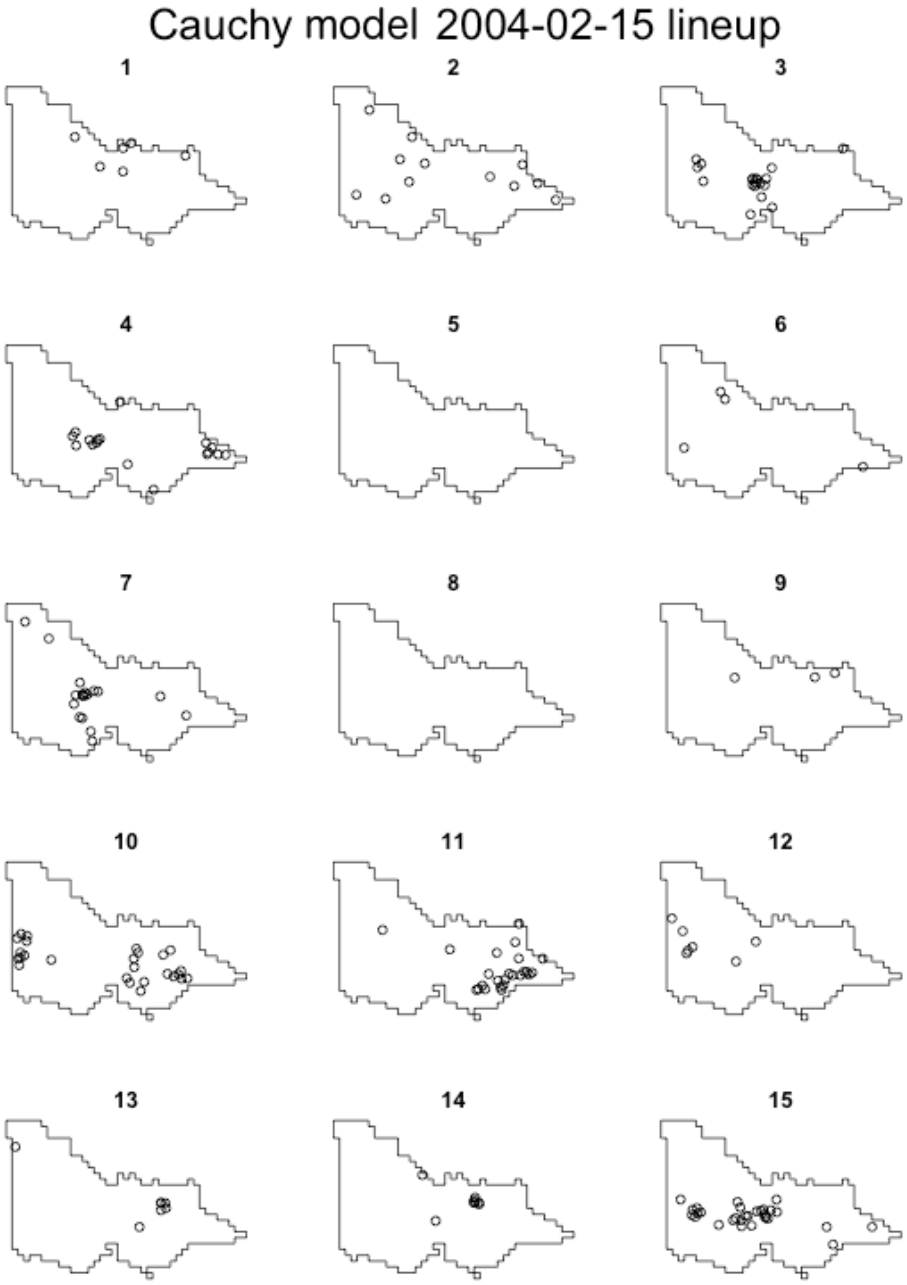


Figure 8.15: *A lineup of simulated point patterns from the Cauchy model along with the observed point pattern (pattern 14, which has 8 points).*

8.7 Discussion

This chapter explored the the use of Poisson and cluster process models for daily lightning-caused wildfire ignitions. Due to the model selection method, all models had a non-linear log-intensity function which was very similar to the likelihood function for the logistic regression model in Chapter 5.

A lot of effort was spent interrogating the fit of the models. It was immediately apparent that the Poisson model had poor fit, struggling to stand up under both spatial and temporal interrogation. This was not surprising given the high degree of clustering that was suggested by the empircal K-function (Figure 8.1). Both cluster models showed improved fit. Neither the temporal residuals nor the K-functions of the cluster models suggested overall poor fit, although the cluster models frequently underpredicted the number of ignitions on days with large numbers of observed ignitions. Comparisons between simulated and observed data showed that all models overestimated the likelihood of a lightning-fire day and underestimated the the number of fires per lightning-fire day, although the cluster process models offered a clear improvement over the Poisson model.

In essence this misfit is being caused by the variety of behaviour shown by the observed process on similar fire weather days. That is, days with almost identical weather conditions can produce wildly varying numbers of ignitions, demanding that the models have large variance in the total numbers of points they generate. However, if the variance increases whilst the intensity remains the same then the Thomas and Cauchy cluster process can produce clusters with unrealistically large numbers of points. Indeed, this is already visible in the estimated maximum number of ignitions on a single day (Table 8.4), which had 95% confidence intervals of $[63, 247]$ for the Thomas model and $[58, 241]$ for the Cauchy model. None of the models were able to reliably reflect the annual variation in ignition behaviour, reflecting a poor fitting intensity function.

We revisited the case studies analysed using the logistic regression model,

focusing on the prediction intervals produced by the various models. The prediction intervals of the Poisson model generally behaved similarly to those from the logistic regression model in the sense that they were accurate on the same days and when they were inaccurate, they both overpredicted or both underpredicted. The cluster process models generally produced much wider prediction intervals than the Poisson model; the former more successfuly conveyed the uncertainty in the predictions.

Visual inspection of simulated point patterns from the cluster processes suggested that the shape of the spatial kernels may be inappropriate. The observed lightning ignition patterns often showed a degree of anisotropy which suggests that the spatial kernels should be asymmetric. Another notable behaviour of the observed lightning ignition patterns was the relatively frequent occurrence of pairs, or triples, of points which appeared anomalously close to each other. The high frequency of these pairs of points resulted in the estimated K-function being greater than πr^2 for very small values of r, something that the Thomas and Cauchy models struggled to replicate. Others have noted similar behaviour and model limitations. Turner (2008) wrote that "there are also many pairs which appeared to be at anomalously small (though non-zero) distances... I decided to eliminate all such pairs for the purpose of this exploratory analysis". Møller and Díaz-Avalos (2010) found that "the misfit suggests that the events in the data set show higher clustering at short distances than that estimated by our model". We are led to believe that these points are probably fairly accurate and should not be removed from the data set (Duff, T., 2018, personal communication).

Model validation exposed some issues with model fit and we are interested to see how the model compares with other models in the wildfire ignition literature. Turner (2008) considered annual, rather than daily, ignition point patterns and validated their Poisson models by inspecting residual fields as well as QQ plots generated from the residual field (Baddeley, Turner, et al.,

2005). Møller and Díaz-Avalos (2010) considered daily ignition point patterns and validated their spatio-temporal NSCP model by inspecting plots of raw spatial residuals as well as comparing estimated L-functions with model L-functions. Many papers compared plots of the empirical and model K- and L-functions (Aragó et al., 2016; Juan et al., 2012). Others seem to have no validation (Díaz-Avalos, Juan, et al., 2016; Serra et al., 2014). Our experience of model validation leads us to believe that a non-Poisson model can have elements of poor fit despite the model K- or L-functions showing good adherance to their empirical counterparts. We were only able to recognise flaws in model fit after interrogating diagnostics of the temporal process (Section 8.3.2), which other work in the literature does not do.

The poor fit of the Poisson model raises questions about the fit of the logistic regression model in Chapter 5. The link between the Poisson process and logistic regression suggests that we would have found evidence of poor fit for the logistic regression model had we investigated its spatial and temporal fit. However, there are a few important differences between the models that may influence model fit. Firstly, we are concerned that the point process models used in this study struggled with the relatively high frequency of observed pairs of points with anomalously small distance between them. This problem will disappear entirely when the point patterns are reduced to binary grids as they are in logistic regression and this may result in improved performance. More generally, logistic regression is focused on binary grids while point process models are focused on exact point locations. This is a classic example of the balance between precision and accuracy; we expect the logistic regression model to be more accurate than the Poisson model. Secondly, we used the same covariate data for both of our models so that we could more easily make comparisons. However, point process models are not limited to regular grids and can make use of the highest resolution data available for each covariate, which may improve fit. Thirdly, some ignition location data close to the bor-

der of the study region was lost in the rasterisation process. We expect that the point process models would perform better than indicated in the present study if we had not enforced parity in data between the regression and point process models.

In terms of providing practical outcomes for fire management, the cluster models are a clear improvement over the Poisson model. The cluster models produced more accurate prediction intervals than the Poisson model, which more accurately conveys uncertainty in the predicted output. All models struggled to produce accurate prediction intervals on severe fire days, with the Cauchy model performing best. In general the point process models underpredicted the number of ignitions on observed severe fire days; this is a problem that the point process models share with the logistic regression model.

Simulated point patterns from the cluster models were much more realistic than those from the Poisson model. Simulations from the Poisson model often stood out because they were overly regular. To our eyes, the Cauchy model tended to produce slightly more realistic simulated point patterns than the Thomas model. This was a result of the heavy tails of the spatial kernel, giving a more natural spread to the points in the clusters.

In this chapter we showed that cluster process models can offer improved fit over more standard models, such as regression models or Poisson process models. From the perspective of a fire management authority, the improved model fit will result in more accurate prediction intervals for operational forecasts and in more realistic simulated point patterns for long term risk analyses. It is important to note that cluster models do not improve the estimate of the intensity function. Close inspection of the point process models suggested many approaches to improving model fit, which we discuss in Section 9.1.

Chapter 9

Conclusion

9.1 Approaches to improving the models

During validation of the logistic regression (Section 5.6) and point process models (Section 8.7) some narratives emerged about when and how the models failed to fit. In this section we discuss possible improvements to the models based on our understanding of their weaknesses at modelling lightning ignition patterns. We have organised the ideas based on whether they apply to all models or just to the point process models. We do not list ideas that apply only to the logistic regression model since we believe that the point process models are the better approach.

9.1.1 All models

These suggestions for improving model fit apply to both the logistic regression and point process models.

Missing covariates

Perhaps the most impactful improvement to the models would be to acquire data for important covariates that are missing from this analysis. As we discussed in Section 5.6, the most notable omissions from our current analysis are a grassland curing index and a fuel load variable.

Both the preliminary data analysis (Section 2.2) and the annual fit of the point process models (Section 8.3.3) suggested that an index for the El Niño–Southern Oscillation (ENSO) could improve model fit. Ideally ENSO should not be a candidate covariate since it influences the wildfire regime only through its impact on the environmental covariates such as precipitation and temperature, however an ENSO index may act as a good substitute for missing or unknown environmental covariates.

Interaction terms

We could have explored using interaction terms in our models. In particular, there should be interaction terms between covariates linked with lightning, and the other weather covariates. These interactions may be important since lightning strikes are much more likely to result in detectable ignitions if the fire weather conditions are extreme. This suggests that the effect on ignition likelihood of the covariates linked with lightning strikes (CAPE index, 850 hPa dewpoint depression, etc.) should vary across the levels of the other weather covariates. Interaction terms may go some way towards improving model fit for periods such as the 2003 Alpine Fires (discussed in Section 8.3.2), where extreme drought magnified the effect of otherwise relatively benign conditions.

Random effects terms

The lightning ignition process proved to be highly variable. Models struggled to reproduce the between-year variation (Section 8.3.3) as well as the variance in the number of ignitions on a single day (Section 8.7). Mixed effect models (Wood, 2017) provide a way of introducing additional variance into a model. For example, a logistic regression model could be constructed such that

$$\pi(\mathbf{z}(y)) = \sum_{j=1}^{d} f_j(\mathbf{z}_j(y)) + r_y,$$

where $r_y \sim N(0, \sigma_y^2)$ is a *random effect* term and accounts for variation between the fire seasons and $\sigma_y^2 > 0$ is a constant. Random effect terms can also be added to the intensity function of a point process; this feature is im-

plemented in *spatstat* for log-linear intensity functions of Poisson and Gibbs process models. Random effect terms have already been implemented in point process models for wildfire ignitions; Juan et al. (2012) fitted a Gibbs process model with a random effect term for each year.

There are two notable consequences of a mixed effect model. Firstly, the estimate of the predictor term can be more accurate if the random effect term is appropriately specified; the random effect term can make a model resistant to the influence of 'outliers' in the data. Secondly, the random effect term will inflate the prediction intervals generated by the model. This is desirable, since we do not wish predictions to be unrealistically confident, but it is not as desirable as the addition of other significant covariates which could increase the accuracy without increasing the uncertainty of predictions.

9.1.2 Point process models

These suggestions for improving model fit apply only to the point process models.

High resolution data

Motivated by a desire to fairly compare the logistic regression model to the cluster process models, we used the same observation window and covariate data for all models. However, as we discussed in Section 8.2, point process models can make use of the highest resolution data available for each covariate. In particular, the point process models could have used the high resolution Victoria Fuel Layer (Section 2.1) to maximum effect, including vegetation type as a categorical covariate rather than the proportion covariate described in Section 5.2.

Spatio-temporal cluster process models

There was significant evidence of temporal interactions in the lightning ignition process. The case studies in Sections 5.5 and 8.4 showed periods of hightened lightning ignition activity that seemed to exceed the duration of

extreme fire weather, suggesting temporal dependence between the ignition patterns. As we discussed in Section 8.1.1, spatio-temporal cluster process models have been used to model wildfire ignitions (Møller and Díaz-Avalos, 2010), however this work assumes that the spatio-temporal kernel is seperable, $h(u,t) = h(u)h(t)$, where $h(u)$ and $h(t)$ is the spatial and temporal kernel respectively. In effect, this allows for spatial data and temporal data to be used, but prevents the use of spatial-temporal data such as daily gridded data. Clearly this is undesirable if the objective is operational forecasting. Non-separable spatio-temporal processes present significant barriers to inference and, although some headway is being made (A. Rodrigues and Diggle, 2010), they are far from mainstream.

Advanced intensity estimation

The standard approach to inference for the inhomogeneous NSCP (Section 6.7.4) is to approximate the intensity function of the NSCP using the intensity function of an inhomogeneous Poisson process. This method is statistically inefficient, making no use of the second moment information in the data. Methods exist that make use of the second moment information (Baddeley, Rubak, et al., 2015, Section 12.4.4) and they are implemented in *spatstat* for single point patterns, but are not yet available for replicated point patterns. Using these methods may improve the accuracy of the estimated intensity function.

Mechanisms for 'anomalously' close pairs of points

As noted in Section 8.7, lightning ignition point patterns often have very close pairs of points. The point process models explored in this thesis struggled to reproduce this kind of behaviour. This suggests that the Gaussian and Cauchy spatial kernels, used in the Thomas and Cauchy models respectively (Section 6.7.2), may have been inappropriate.

An alternative class of spatial kernels that could be interesting to investi-

gate are of the form

$$h(u) = p\phi(u) + (1-p)\delta_R(u),$$

where $\phi(u)$ is a symmetric bivariate density such as the Gaussian or Cauchy density and $\delta_R(u)$ is the uniform density on the circle of radius R centred on the origin (which is the spatial kernel for the Matérn process, Section 6.7.2). If $R \approx 0$, then this kernel is essentially the mixture of the 'bulky' spatial kernel $\phi(u)$ and the 'spike' spatial kernel $\delta_R(u)$. This spike at the origin will encourage more points within very small distances of one another, although it may result in clusters with far more than 2 anomalously close points.

Another approach to dealing with these close points is to construct *marked* point patterns (Baddeley, Rubak, et al., 2015, Chapter 15), which are point patterns where each point has an associated number, known as a *mark*. In this case, each lighting ignition point pattern would be converted into a marked point pattern where the 'anomalously' close clusters of points would be replaced by a single point with mark equal to the number of points in the 'anomalously' tight cluster. A cluster process model could be fitted to the marked point process and another distribution could be fitted to the marks. The process of converting the lightning ignition point patterns into marked point patterns would require some subjective decisions about what constitutes 'anomalously' close.

9.2 Final words

This thesis looked at two approaches to modeling lightning-caused wildfire ignition locations. The logistic regression model presented in Chapter 5 showed an ability to model lightning ignition location that would benifit operational planning as well as long term risk analysis. We demonstrated how to deconstruct model forecasts into contributions from the individual covariates (Section 5.3.1), making it easier to relate model output back to the environmental

conditions driving it. We used a similar idea to construct an approximating model, which we implemented in Section 5.4.3 and showed good performance, although we adopted the full model out of concern that the approximating model may struggle on extreme days. Nonetheless, we believe this is an interesting and simple method of generating approximating models.

In Chapter 8 we presented point process models for lightning ignition. There was evidence of significant clustering in the lightning ignition point patterns and this led to both cluster process models having improved fit over the Poisson process model. We found significant misfit between the observed and simulated number of lightning-fire days for all models (Section 8.3.3), although the cluster models were the least affected. We used a new weighted pooled empirical K-function (Section 7.4), giving an improved estimate for the K-function (Figure 8.1).

Ultimately, point process theory offers a more natural framework to model wildfire ignitions. A Poisson process model is roughly equivalent to a logistic or Poisson regression model, but allows the use of data with varying resolutions. Although we didn't discuss it, concepts of covariate contributions (Sec 3.3.1, 3.5 and 3.6) are easily extended to point process models, where the terms in the log-intensity function are the contributions from the covariates.

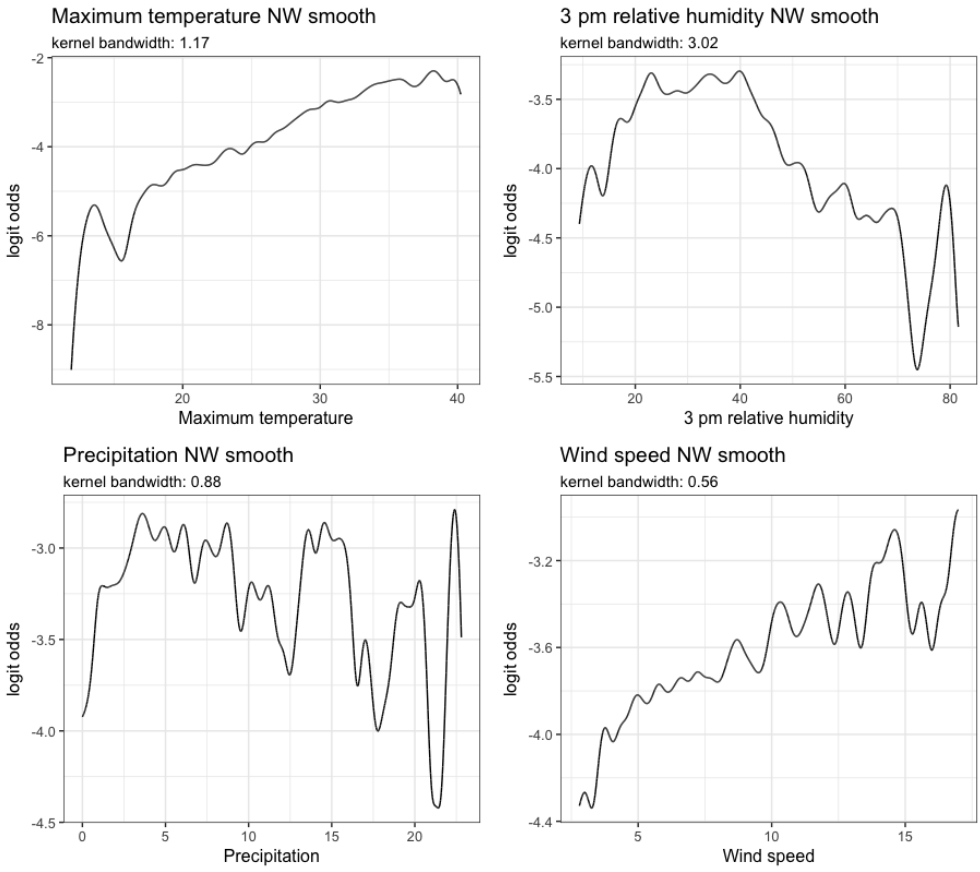
Like most large modeling projects, the list of possible improvements to the model is lengthy (Section 9.1). We believe that the most promising routes to improvement are also the simplest: acquiring data for other important covariates, including interaction terms and using high resolution data.

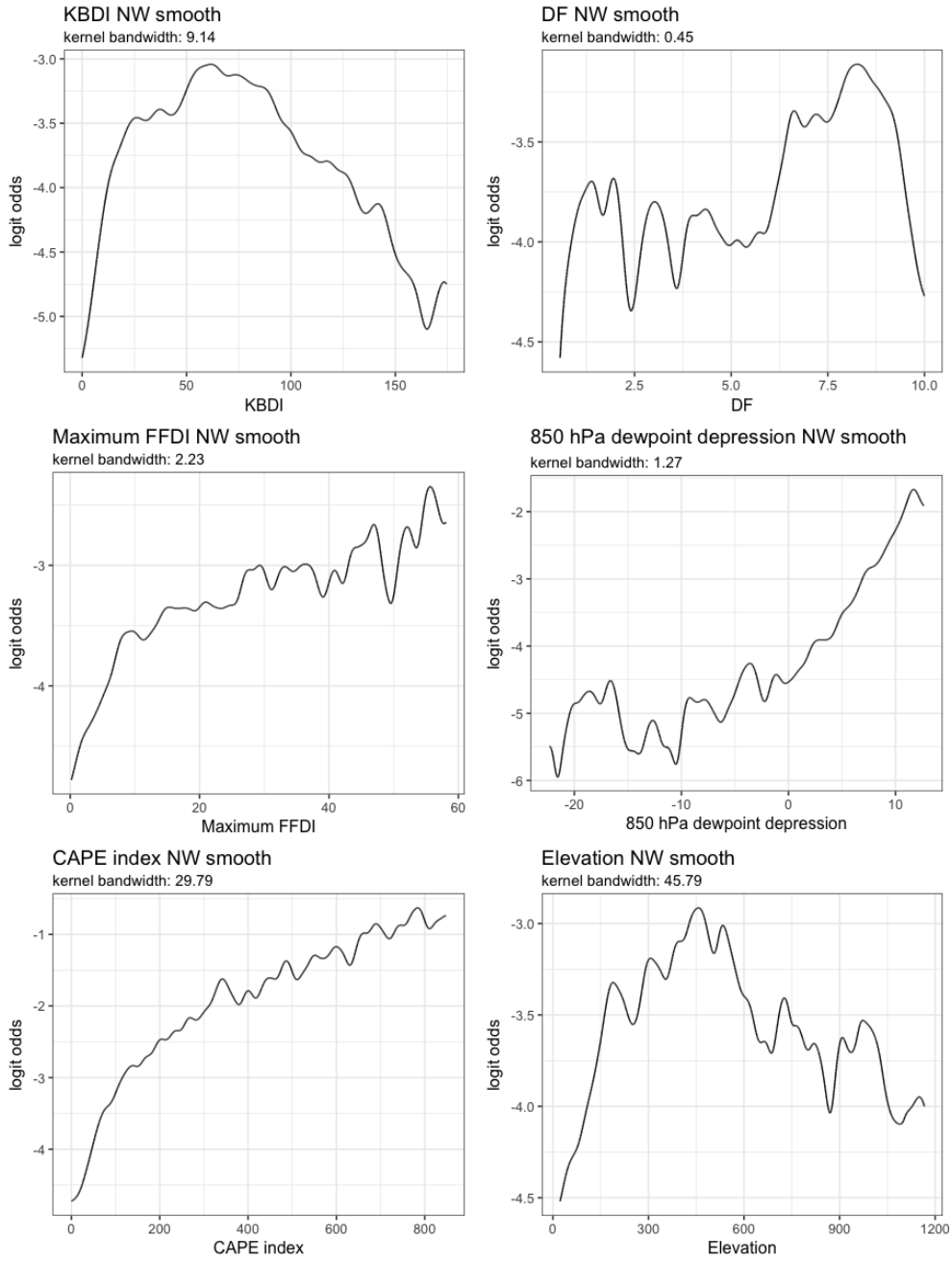
Lightning-caused wildfire ignition is a significant threat in Australia and abroad. The process governing lightning-caused wildfire ignition is complex and highly variable, posing a significant modelling challenges. Models will always struggle to accurately capture the behaviour of wildfire ignitions and users should be very critical of models. However, with a thorough understanding of any misfit, models such as those presented in this thesis are valuable

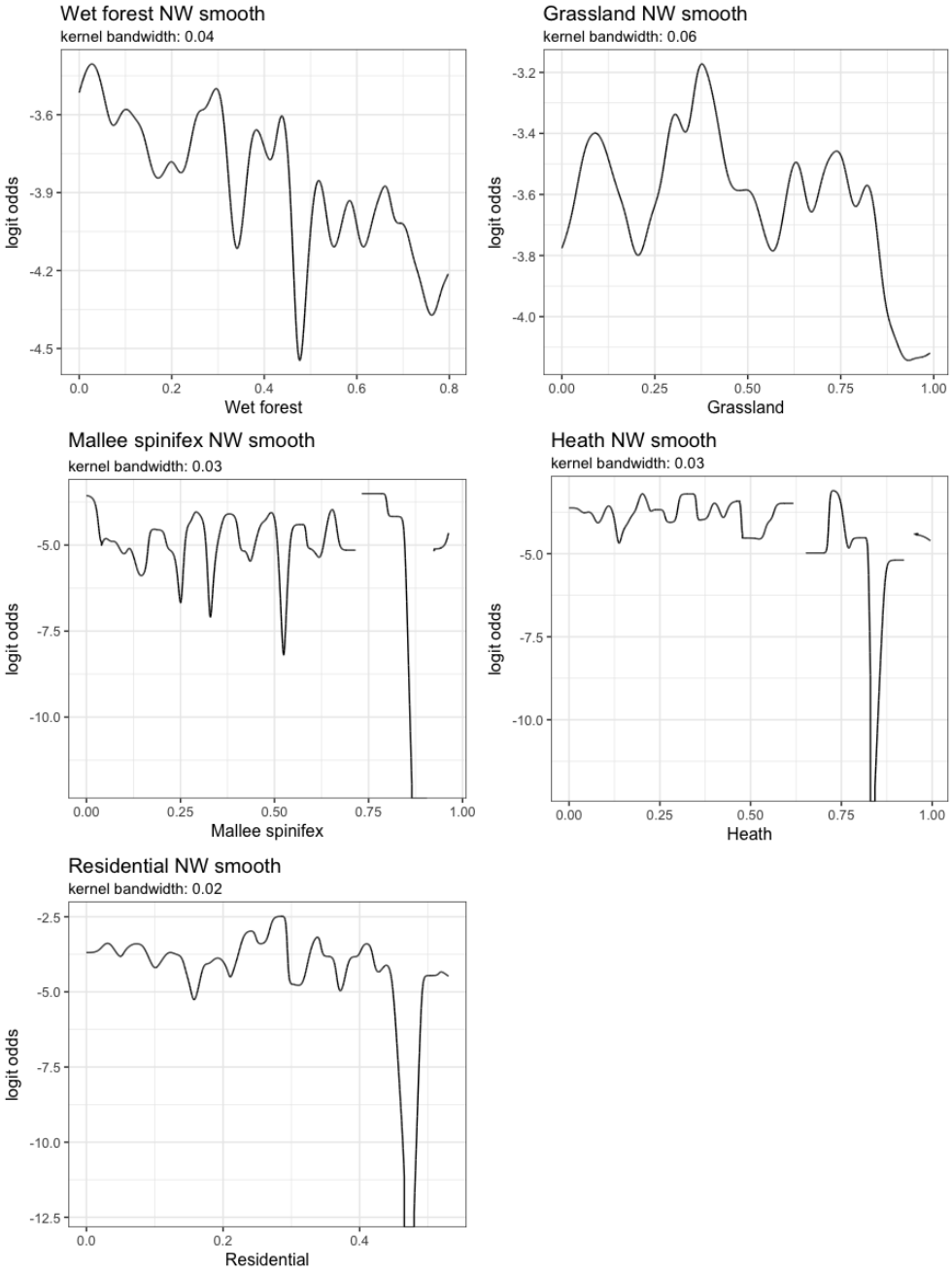
tools and could be used to to inform operational decision making and to improve long-term risk forecasts.

Appendix A

Nadaraya Watson smooths of covariate/response pairs

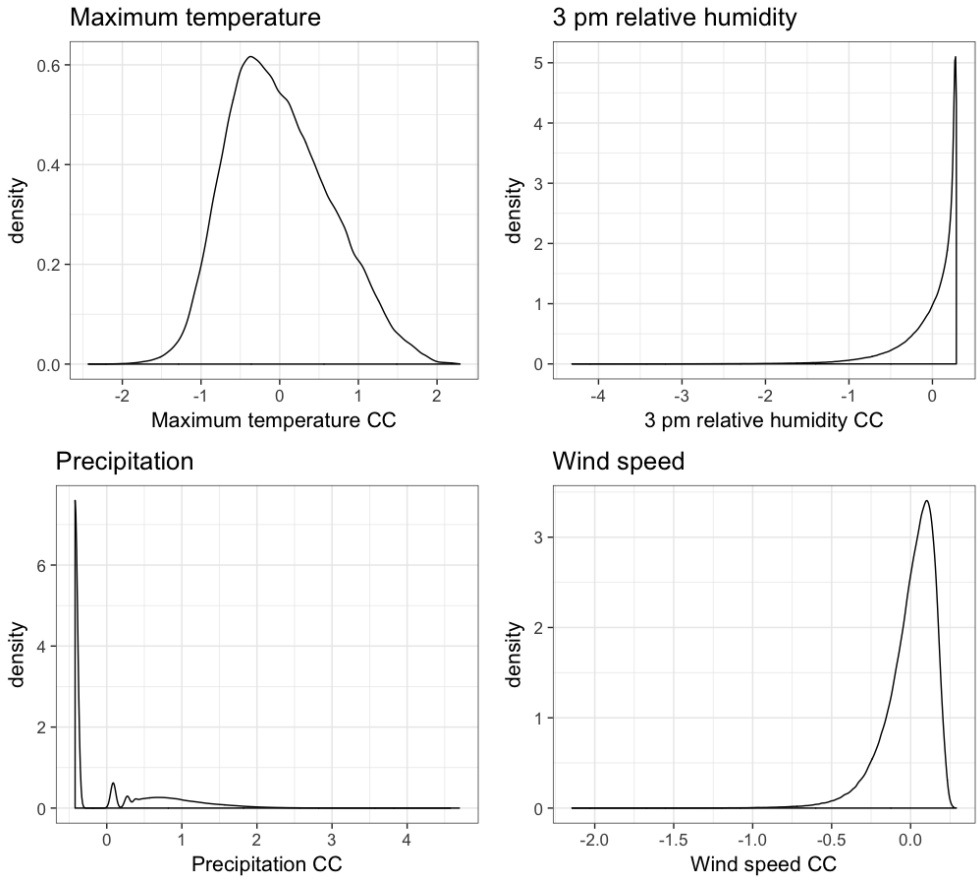


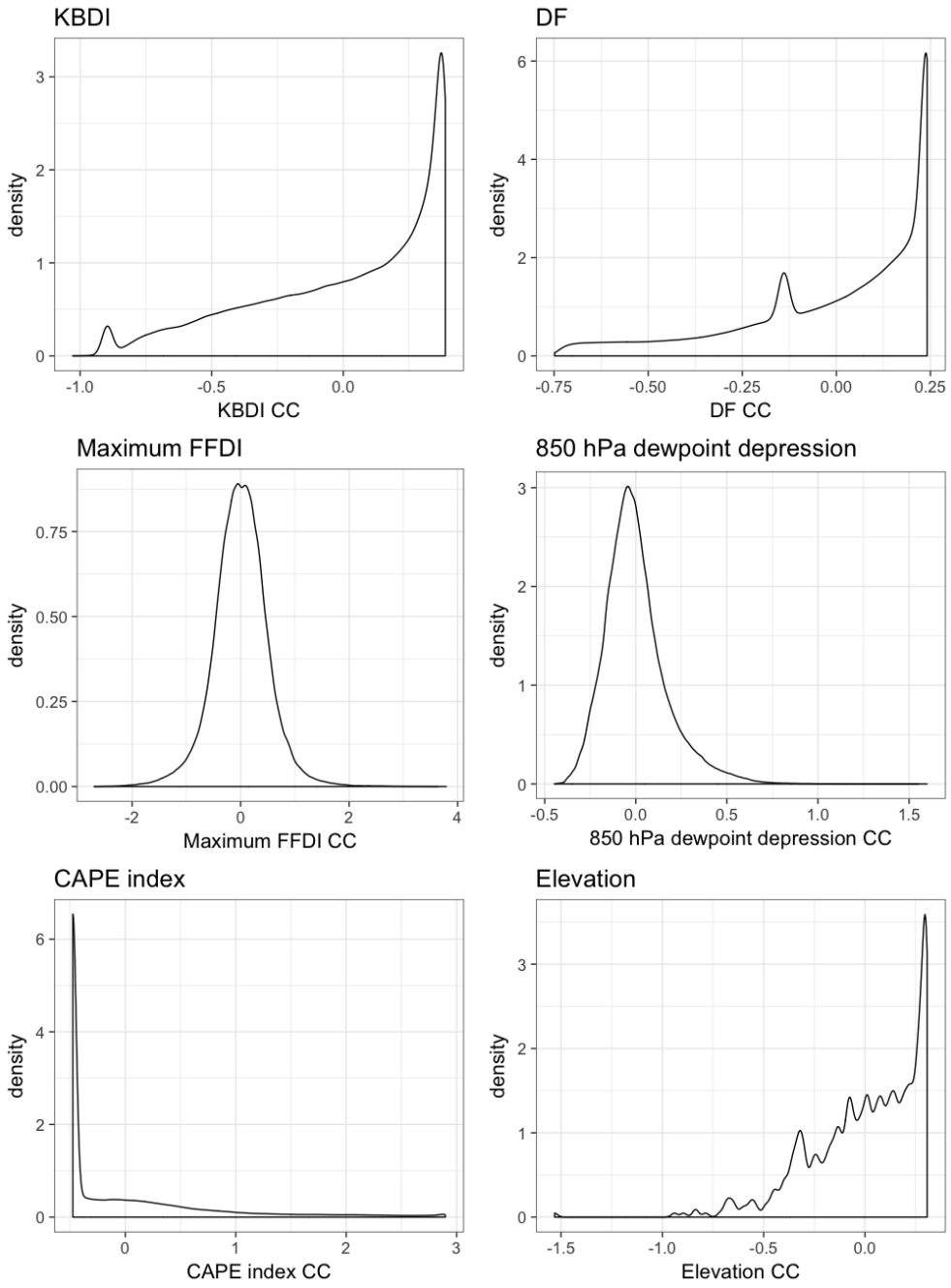

KBDI NW smooth
kernel bandwidth: 9.14
logit odds
KBDI
DF NW smooth
kernel bandwidth: 0.45
logit odds
DF
Maximum FFDI NW smooth
kernel bandwidth: 2.23
logit odds
Maximum FFDI
850 hPa dewpoint depression NW smooth
kernel bandwidth: 1.27
logit odds
850 hPa dewpoint depression
CAPE index NW smooth
kernel bandwidth: 29.79
logit odds
CAPE index
Elevation NW smooth
kernel bandwidth: 45.79
logit odds
Elevation

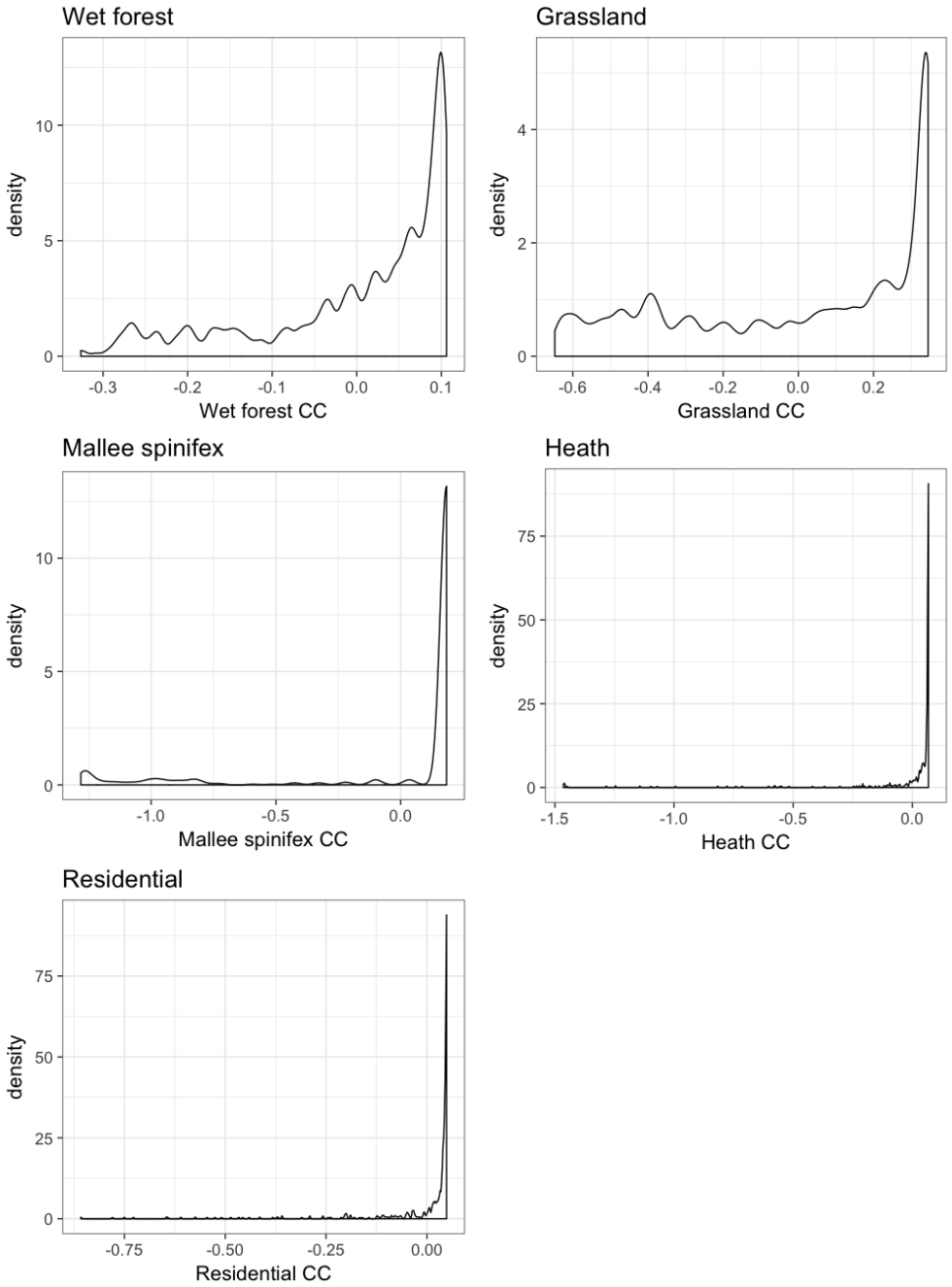

Wet forest NW smooth
kernel bandwidth: 0.04
logit odds
-3.6
-3.9
-4.2
-4.5
0.0
0.2
0.4
0.6
0.8
Wet forest
Grassland NW smooth
kernel bandwidth: 0.06
logit odds
-3.2
-3.4
-3.6
-3.8
-4.0
0.00
0.25
0.50
0.75
1.00
Grassland
Mallee spinifex NW smooth
kernel bandwidth: 0.03
logit odds
-5.0
-7.5
-10.0
0.00
0.25
0.50
0.75
1.00
Mallee spinifex
Heath NW smooth
kernel bandwidth: 0.03
logit odds
-5.0
-7.5
-10.0
0.00
0.25
0.50
0.75
1.00
Heath
Residential NW smooth
kernel bandwidth: 0.02
logit odds
-2.5
-5.0
-7.5
-10.0
-12.5
0.0
0.2
0.4
Residential

Appendix B

Covariate contribution density estimates



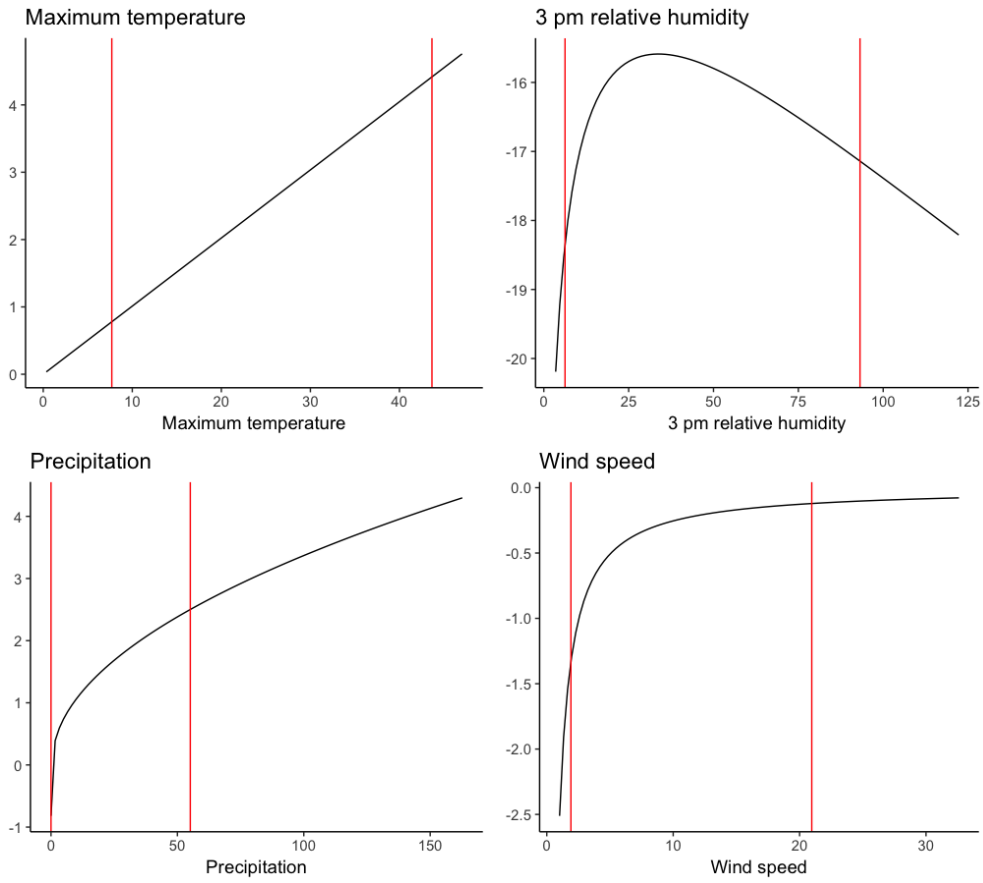

KBDI
density
KBDI CC
DF
density
DF CC
Maximum FFDI
density
Maximum FFDI CC
850 hPa dewpoint depression
density
850 hPa dewpoint depression CC
CAPE index
density
CAPE index CC
Elevation
density
Elevation CC

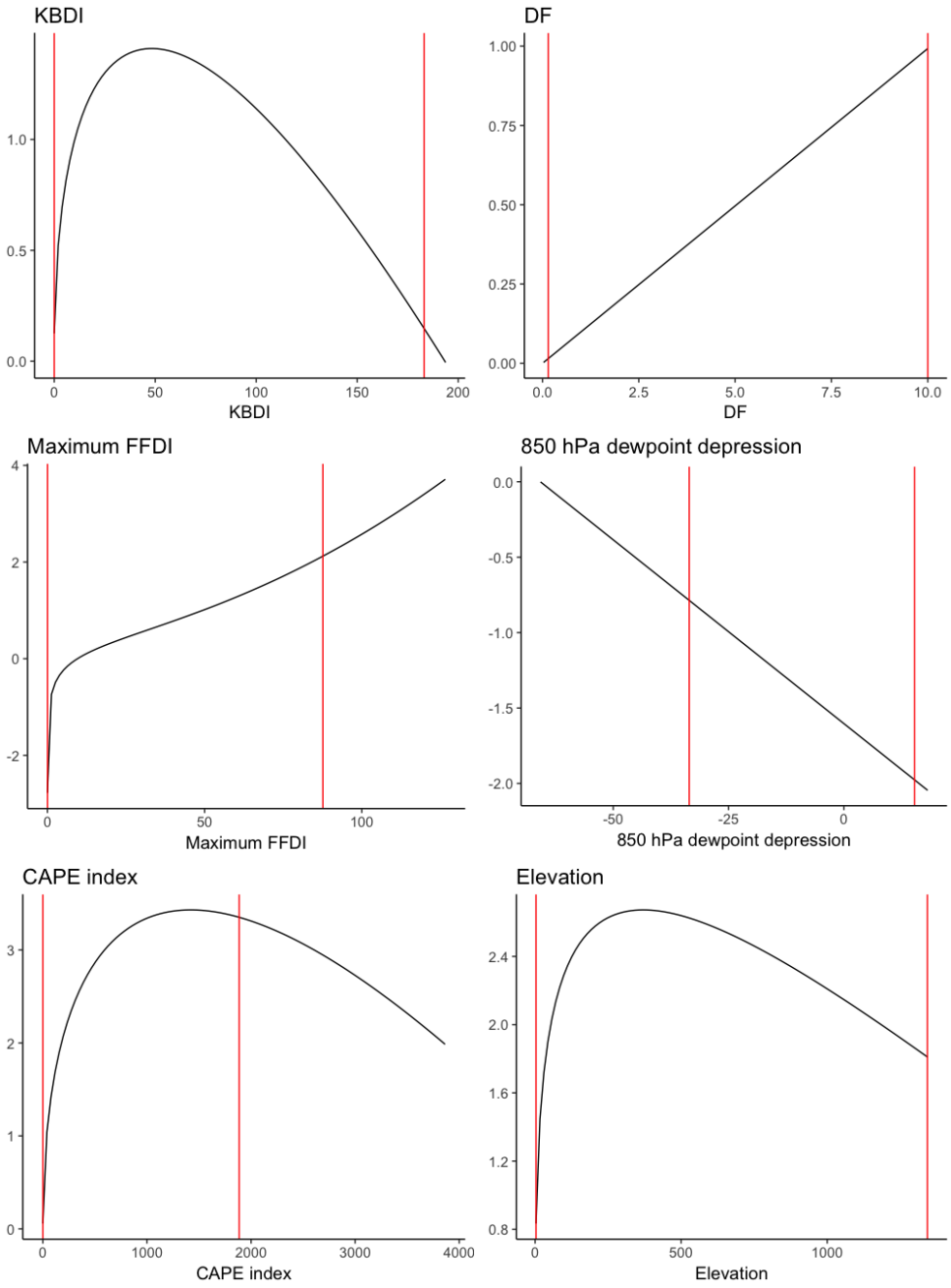

Wet forest
density
Wet forest CC
Grassland
density
Grassland CC
Mallee spinifex
density
Mallee spinifex CC
Heath
density
Heath CC
Residential
density
Residential CC

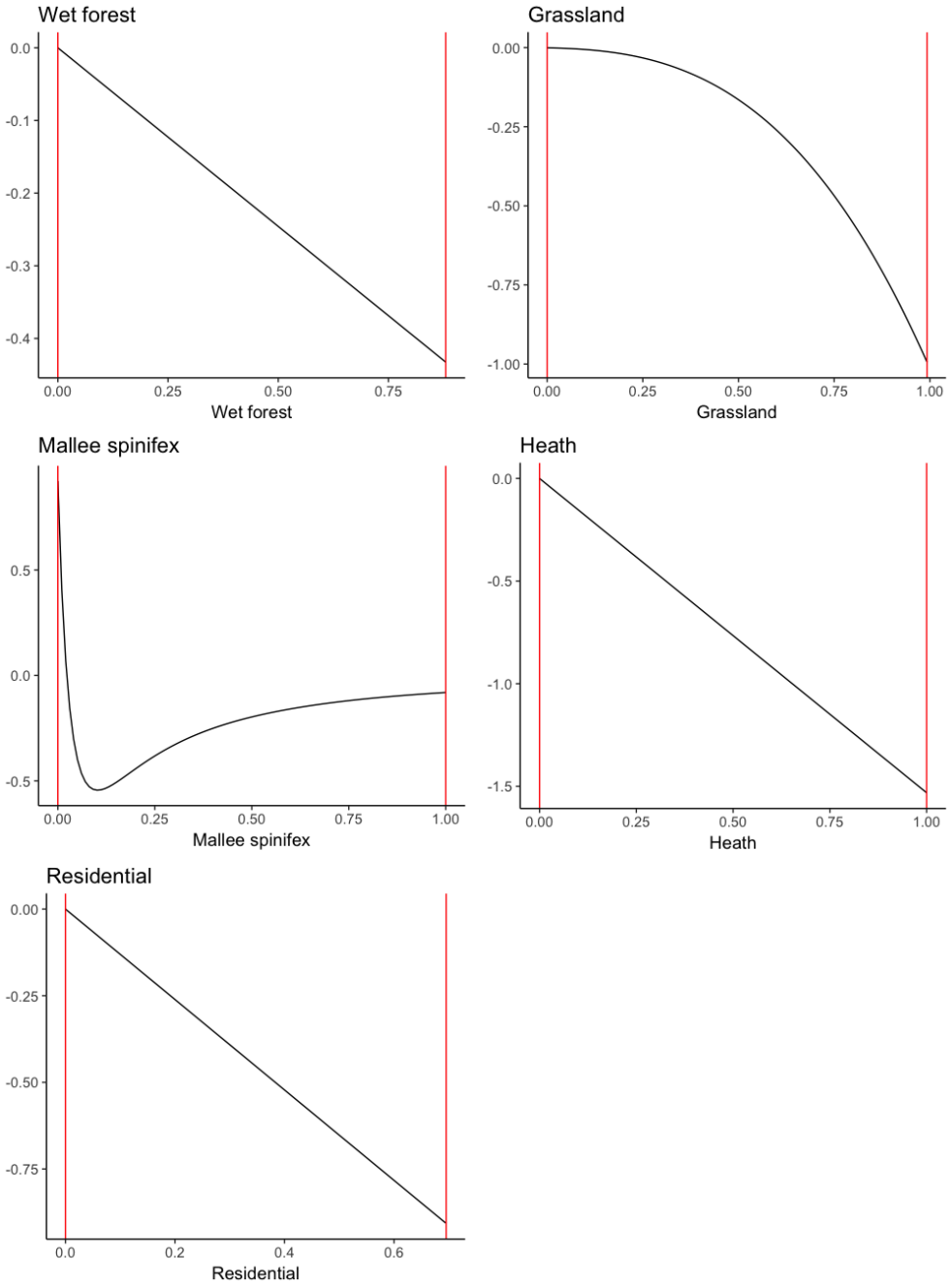
Appendix C

Scaling functions

The vertical red lines show the 0.001 and 0.999 quantiles of the observed covariates.




KBDI
1.0
0.5
0.0
0
50
100
150
200
KBDI
DF
1.00
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0
DF
Maximum FFDI
4
2
0
-2
0
50
100
Maximum FFDI
850 hPa dewpoint depression
0.0
-0.5
-1.0
-1.5
-2.0
-50
-25
0
850 hPa dewpoint depression
CAPE index
3
2
1
0
0
1000
2000
3000
4000
CAPE index
Elevation
2.4
2.0
1.6
1.2
0.8
0
500
1000
Elevation


Wet forest
0.0
-0.1
-0.2
-0.3
-0.4
0.00
0.25
0.50
0.75
Wet forest
Grassland
0.00
-0.25
-0.50
-0.75
-1.00
0.00
0.25
0.50
0.75
1.00
Grassland
Mallee spinifex
0.5
0.0
-0.5
0.00
0.25
0.50
0.75
1.00
Mallee spinifex
Heath
0.0
-0.5
-1.0
-1.5
0.00
0.25
0.50
0.75
1.00
Heath
Residential
0.00
-0.25
-0.50
-0.75
0.0
0.2
0.4
0.6
Residential

Index

Bibliography

Abtew, W. and A. Melesse (2013). "Vapor Pressure Calculation Methods". en. In: *Evaporation and Evapotranspiration*. Dordrecht: Springer, pp. 53–62. DOI: 10.1007/978-94-007-4737-1_5. (Visited on 09/26/2016).

Albert-Green, A., C. Dean, D. L. Martell, and D. G. Woolford (2012). "A methodology for investigating trends in changes in the timing of the fire season with applications to lightning-caused forest fires in Alberta and Ontario, Canada". In: *Canadian Journal of Forest Research* 43.1, pp. 39–45. DOI: 10.1139/cjfr-2011-0432. (Visited on 08/22/2016).

Albertson, K., J. Aylen, G. Cavan, and J. McMorrow (2009). "Forecasting the outbreak of moorland wildfires in the English Peak District". In: *Journal of Environmental Management* 90.8, pp. 2642–2651. DOI: 10.1016/j.jenvman.2009.02.011. (Visited on 08/22/2016).

Ambler, G. and A. Benner (2015). *mfp: Multivariable Fractional Polynomials*.

Anagnostopoulos, C. and D. J. Hand (2012). *hmeasure: The H-measure and other scalar classification performance metrics*.

Anderson, G. B., M. L. Bell, and R. D. Peng (2013). "Methods to calculate the heat index as an exposure metric in environmental health research". In: *Environemtental Health Perspectives* 121.10, pp. 1111–1119. DOI: 10.1289/ehp.1206273.

Anderson, K. (2002). "A model to predict lightning-caused fire occurrences". In: *International Journal of Wildland Fire* 11.4, pp. 163–172. DOI: https://doi.org/10.1071/WF02001. (Visited on 08/22/2016).

Aragó, P., P. Juan, C. Díaz-Avalos, and P. Salvador (2016). "Spatial point process modeling applied to the assessment of risk factors associated with forest wildfires

incidence in Castellón, Spain". en. In: *European Journal of Forest Research* 135.3, pp. 451–464. DOI: 10.1007/s10342-016-0945-z. (Visited on 03/21/2018).

Baddeley, A., I. Bárány, and R. Schneider (2007). *Stochastic Geometry: Lectures given at the C.I.M.E. Summer School held in Martina Franca, Italy, September 13-18, 2004*. en. Berlin: Springer-Verlag.

Baddeley, A., J. Møller, and R. P. Waagepetersen (2000). "Non- and semi-parametric estimation of interaction in inhomogeneous point patterns". en. In: *Statistica Neerlandica* 54.3, pp. 329–350. DOI: 10.1111/1467-9574.00144.

Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial Point Patterns: Methodology and Applications with R*. en. 1st ed. Boca Raton: CRC Press.

Baddeley, A., R. Turner, J. Møller, and M. Hazelton (2005). "Residual analysis for spatial point processes (with discussion)". en. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.5, pp. 617–666. DOI: 10.1111/j.1467-9868.2005.00519.x.

Berman, M. and R. Turner (1992). "Approximating Point Process Likelihoods with GLIM". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 41.1, pp. 31–38. DOI: 10.2307/2347614.

Binder, H., W. Sauerbrei, and P. Royston (2013). "Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response". en. In: *Statistics in Medicine* 32.13, pp. 2262–2277. DOI: 10.1002/sim.5639.

Bradstock, R. A., J. S. Cohn, A. M. Gill, M. Bedward, and C. Lucas (2009). "Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather". In: *International Journal of Wildland Fire* 18.8, pp. 932–943. DOI: 10.1071/WF08133. (Visited on 08/22/2016).

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and Regression Trees*. en. Belmont: Wadsworth.

Brillinger, D. R., H. K. Preisler, and J. W. Benoit (2003). "Risk Assessment: A Forest Fire Example". In: *Lecture Notes-Monograph Series* 40. www.jstor.org/stable/4356185., pp. 177–196. (Visited on 10/05/2016).

Brown, T. J., G. A. Mills, S. Harris, D. Podnar, H. Reinbold, and M. Fearon (2016). "A bias corrected WRF mesoscale fire weather dataset for Victoria, Australia 1972-

2012". In: *Journal of Southern Hemisphere Earth Systems Science* 66, pp. 281–313. (Visited on 02/28/2017).

Brunsdon, C., S. Fotheringham, and M. Charlton (1998). "Geographically Weighted Regression". en. In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.3, pp. 431–443. DOI: 10.1111/1467-9884.00145. (Visited on 02/13/2018).

Buja, A., D. Cook, H. Hofmann, M. Lawrence, E.-K. Lee, D. F. Swayne, and H. Wickham (2009). "Statistical inference for exploratory data analysis and model diagnostics". en. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367.1906, pp. 4361–4383. DOI: 10.1098/rsta.2009.0120. (Visited on 05/02/2018).

Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference*. 2nd ed. New York: Springer. (Visited on 11/14/2017).

Casella, G. and R. L. Berger (2002). *Statistical inference*. 2nd ed. Delhi: Wads-Worth.

Chow, J. Y. J. and A. C. Regan (2011). "Resource Location and Relocation Models with Rolling Horizon Forecasting for Wildland Fire Planning". In: *INFOR: Information Systems and Operational Research* 49.1, pp. 31–43. DOI: 10.3138/infor.49.1.031.

Collins, K. M., O. F. Price, and T. D. Penman (2015). "Spatial patterns of wildfire ignitions in south-eastern Australia". en. In: *International Journal of Wildland Fire* 24.8, p. 1098. DOI: 10.1071/WF15054. (Visited on 10/20/2016).

Cook, R. D. (1977). "Detection of Influential Observation in Linear Regression". In: *Technometrics* 19.1, pp. 15–18. DOI: 10.2307/1268249. (Visited on 12/18/2017).

Costermans, L. (2009). *Native trees and shrubs of south-eastern Australia*. Chatswood: New Holland Publishers.

Daley, D. J. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. en. 2nd ed. Vol. 1. Probability and its Applications. New York: Springer-Verlag. (Visited on 11/02/2016).

— (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. 2nd ed. Vol. 2. Probability and its Applications. New York: Springer-Verlag. (Visited on 11/02/2016).

Dee, D. P. et al. (2011). "The ERA-Interim reanalysis: configuration and performance of the data assimilation system". en. In: *Quarterly Journal of the Royal*

Meteorological Society 137.656, pp. 553–597. DOI: 10.1002/qj.828. (Visited on 02/08/2018).

Department of Environment, Land, Water and Planning (2015). *Measuring bushfire risk in Victoria*. Tech. rep. Melbourne: Department of Environment, Land, Water and Planning. (Visited on 11/28/2017).

Department of Environment, Land, Water & Planning (2008). *Vicmap Elevation DTM 20m*.

Díaz-Avalos, C., P. Juan, and L. Serra-Saurina (2016). "Modeling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes". In: *Forest Ecology and Management* 381, pp. 360–369. DOI: 10.1016/j.foreco.2016.09.013. (Visited on 03/21/2018).

Díaz-Avalos, C., D. L. Peterson, E. Alvarado, S. A. Ferguson, and J. E. Besag (2001). "Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon". In: *Canadian Journal of Forest Research* 31.9, pp. 1579–1593. DOI: 10.1139/x01-089. (Visited on 04/05/2017).

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. en. Boca Raton: CRC Press.

Diggle, P. J., V. Gómez-Rubio, P. E. Brown, A. G. Chetwynd, and S. Gooding (2007). "Second-Order Analysis of Inhomogeneous Spatial Point Processes Using Case–Control Data". en. In: *Biometrics* 63.2, pp. 550–557. DOI: 10.1111/j.1541-0420.2006.00683.x.

Diggle, P. J., N. Lange, and F. M. Beneš (1991). "Analysis of Variance for Replicated Spatial Point Patterns in Clinical Neuroanatomy". In: *Journal of the American Statistical Association* 86.415, pp. 618–625. DOI: 10.1080/01621459.1991.10475087.

Dissing, D. and D. L. Verbyla (2003). "Spatial patterns of lightning strikes in interior Alaska and their relations to elevation and vegetation". In: *Canadian Journal of Forest Research* 33.5, pp. 770–782. DOI: 10.1139/x02-214.

Dodd, L. E. and M. S. Pepe (2003). "Partial AUC Estimation and Regression". en. In: *Biometrics* 59.3, pp. 614–623. DOI: 10.1111/1541-0420.00071. (Visited on 01/10/2018).

Dowdy, A. J. (2015). "Large-scale modelling of environments favourable for dry lightning occurrence". In: *21st International Congress on Modelling and Simulation*. Vol. 21. Gold Coast, QLD. (Visited on 10/19/2016).

Dowdy, A. J. and G. A. Mills (2009). *Atmospheric states associated with the ignition of lightning-attributed fires*. Technical Report 19. The Centre for Australian Weather and Climate Research. (Visited on 11/28/2016).

Duff, T. J., D. M. Chong, B. A. Cirulis, S. F. Walsh, T. D. Penman, and K. G. Tolhurst (2014). "Understanding risk: representing fire danger using spatially explicit fire simulation ensembles". In: *Advances in Forest Fire Research*. Ed. by D. X. Viegas. Coimbra: Imprensa da Universidade de Coimbra, pp. 1286–1294. (Visited on 05/20/2017).

Duff, T. J., R. E. Keane, T. D. Penman, and K. G. Tolhurst (2017). "Revisiting Wildland Fire Fuel Quantification Methods: The Challenge of Understanding a Dynamic, Biotic Entity". en. In: *Forests* 8.9, p. 351. DOI: 10.3390/f8090351. (Visited on 11/28/2017).

Elith, J. and J. R. Leathwick (2009). "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time". In: *Annual Review of Ecology, Evolution, and Systematics* 40.1, pp. 677–697. DOI: 10.1146/annurev.ecolsys.110308.120159. (Visited on 12/04/2017).

Feikema, P. M., C. B. Sherwin, and P. N. J. Lane (2013). "Influence of climate, fire severity and forest mortality on predictions of long term streamflow: Potential effect of the 2009 wildfire on Melbourne's water supply catchments". In: *Journal of Hydrology* 488, pp. 1–16. DOI: 10.1016/j.jhydrol.2013.02.001. (Visited on 11/25/2016).

Flach, P., J. Hernandez-Orallo, and C. Ferro (2011). "A coherent interpretation of AUC as a measure of aggregated classification performance". In: *Proceedings of the 28th International Conference on Machine Learning*. Bellevue, WA. (Visited on 11/03/2017).

Flannigan, M. D. and B. M. Wotton (1991). "Lightning-ignited forest fires in northwestern Ontario". In: *Canadian Journal of Forest Research* 21.3, pp. 277–287. DOI: 10.1139/x91-035. (Visited on 05/19/2017).

Gabriel, E. and P. J. Diggle (2009). "Second-order analysis of inhomogeneous spatio-temporal point process data". en. In: *Statistica Neerlandica* 63.1, pp. 43–51. DOI: 10.1111/j.1467-9574.2008.00407.x. (Visited on 02/20/2018).

Gaetan, C. and X. Guyon (2009). *Spatial Statistics and Modeling*. en. 1st ed. New York: Springer-Verlag.

Gill, A. M., S. L. Stephens, and G. J. Cary (2013). "The worldwide "wildfire" problem". en. In: *Ecological Applications* 23.2, pp. 438–454. DOI: 10.1890/10-2213.1. (Visited on 11/30/2017).

Griffiths, D. (1999). "Improved Formula for the Drought Factor in McArthur's Forest Fire Danger Meter". In: *Australian Forestry* 62.2, pp. 202–206. DOI: 10.1080/00049158.1999.10674783. (Visited on 11/28/2016).

Guo, F., Z. Su, G. Wang, L. Sun, F. Lin, and A. Liu (2016). "Wildfire ignition in the forests of southeast China: Identifying drivers and spatial distribution to predict wildfire likelihood". In: *Applied Geography* 66, pp. 12–21. DOI: 10.1016/j.apgeog.2015.11.014. (Visited on 10/20/2016).

Haines, D. A., W. A. Main, J. S. Frost, and A. J. Simard (1983). "Fire-Danger Rating and Wildfire Occurrence in the Northeastern United States". In: *Forest Science* 29.4, pp. 679–696.

Hall, B. L. and T. J. Brown (2006). "Climatology of positive polarity flashes and multiplicity and their relation to natural wildfire ignitions". In: Tucson, AZ. (Visited on 05/19/2017).

Hand, D. J. (1997). *Construction and assessment of classification rules*. Chichester: John Wiley & Sons.

— (2005). "Good practice in retail credit scorecard assessment". en. In: *Journal of the Operational Research Society* 56.9, pp. 1109–1117. DOI: 10.1057/palgrave.jors.2601932. (Visited on 02/07/2018).

— (2009). "Measuring classifier performance: a coherent alternative to the area under the ROC curve". en. In: *Machine Learning* 77.1, pp. 103–123. DOI: 10.1007/s10994-009-5119-5. (Visited on 08/02/2017).

— (2012). "Assessing the Performance of Classification Methods". en. In: *International Statistical Review* 80.3, pp. 400–414. DOI: 10.1111/j.1751-5823.2012.00183.x. (Visited on 01/17/2018).

Hand, D. J. and C. Anagnostopoulos (2012). "A better Beta for the H measure of classification performance". In: *arXiv:1202.2564 [cs, stat]*. arXiv: 1202.2564. (Visited on 01/17/2018).

— (2013). "When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance?" In: *Pattern Recognition Letters* 34.5, pp. 492–495. DOI: 10.1016/j.patrec.2012.12.004.

Harrell, F. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. en. 2nd ed. Cham: Springer.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. 2nd ed. New York: Springer. (Visited on 12/07/2017).

Hilbe, J. M. (2011). *Negative binomial regression*. 2nd ed. Cambridge: Cambridge University Press.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant (2013). *Applied Logistic Regression*. en. 3rd ed. Hoboken, NJ: John Wiley & Sons.

Jalilian, A., Y. Guan, and R. P. Waagepetersen (2013). "Decomposition of Variance for Spatial Cox Processes". en. In: *Scandinavian Journal of Statistics* 40.1, pp. 119–137. DOI: 10.1111/j.1467-9469.2012.00795.x.

Johnson, R. A. and D. W. Wichern (2007). *Applied Multivariate Statistical Analysis*. 6th ed. USA: Pearson Prentice Hall.

Juan, P., J. Mateu, and M. Saez (2012). "Pinpointing spatio-temporal interactions in wildfire patterns". en. In: *Stochastic Environmental Research and Risk Assessment* 26.8, pp. 1131–1150. DOI: 10.1007/s00477-012-0568-y. (Visited on 10/12/2016).

Kahle, D. and H. Wickham (2013). "ggmap: Spatial Visualization with ggplot2". In: *The R Journal* 5.1, pp. 144–161.

Kallenberg, O. (1983). *Random Measures*. New York: Academic Press.

Keane, R. E. (2015). *Wildland fuel fundamentals and applications*. Switzerland: Springer.

Keetch, J. J. and G. M. Byram (1968). "A Drought Index for Forest Fire Control". In: *U.S.D.A. Forest Service Research Paper* SE 38. (Visited on 11/28/2016).

Kilinc, M. (2017). *Bushfire spatial data models and ignition data — Bushfire & Natural Hazards CRC*. (Visited on 06/14/2017).

Kilinc, M. and J. Beringer (2007). "The Spatial and Temporal Distribution of Lightning Strikes and Their Relationship with Vegetation Type, Elevation, and Fire Scars in the Northern Territory". In: *Journal of Climate* 20.7, pp. 1161–1173. DOI: 10.1175/JCLI4039.1.

Lele, S. R., J. L. Keim, and P. Solymos (2017). *ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data.*

Liu, Z., J. Yang, Y. Chang, P. J. Weisberg, and H. S. He (2012). "Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of Northeast China". en. In: *Global Change Biology* 18.6, pp. 2041–2056. DOI: 10.1111/j.1365-2486.2012.02649.x. (Visited on 10/12/2016).

Lobo, J. M., A. Jiménez-Valverde, and R. Real (2008). "AUC: a misleading measure of the performance of predictive distribution models". en. In: *Global Ecology and Biogeography* 17.2, pp. 145–151. DOI: 10.1111/j.1466-8238.2007.00358.x.

Magnussen, S. and S. W. Taylor (2012). "Prediction of daily lightning- and human-caused fires in British Columbia". In: *International Journal of Wildland Fire* 21.4, pp. 342–356. DOI: 10.1071/WF11088. (Visited on 08/22/2016).

Mandallaz, D. and R. Ye (1997). "Prediction of forest fires with Poisson models". In: *Canadian Journal of Forest Research* 27.10, pp. 1685–1694. DOI: 10.1139/x97-103. (Visited on 10/05/2016).

Martell, D. L., E. Bevilacqua, and B. J. Stocks (1989). "Modelling seasonal variation in daily people-caused forest fire occurrence". In: *Canadian Journal of Forest Research* 19.12, pp. 1555–1563. DOI: 10.1139/x89-237. (Visited on 09/05/2016).

Martell, D. L., S. Otukol, and B. J. Stocks (1987). "A logistic model for predicting daily people-caused forest fire occurrence in Ontario". In: *Canadian Journal of Forest Research* 17.5, pp. 394–401. DOI: 10.1139/x87-068. (Visited on 08/22/2016).

Martínez-Fernández, J., E. Chuvieco, and N. Koutsias (2013). "Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression". In: *Nat. Hazards Earth Syst. Sci.* 13.2, pp. 311–327. DOI: 10.5194/nhess-13-311-2013. (Visited on 08/22/2016).

Matthews, S. (2014). "Dead fuel moisture research: 1991–2012". en. In: *International Journal of Wildland Fire* 23.1, pp. 78–92. DOI: 10.1071/WF13005. (Visited on 11/28/2017).

McArthur, A. G. (1967). *Fire Behaviour in Eucalypt Forest*. 107. Canberra: Forestry and Timber Bureau.

McCaw, W. L., J. E. Neal, and R. H. Smith (2002). "Stand characteristics and fuel accumulation in a sequence of even-aged Karri (Eucalyptus diversicolor) stands in south-west Western Australia". In: *Forest Ecology and Management* 158.1, pp. 263–271. DOI: 10.1016/S0378-1127(00)00719-2.

Meteorology, B. of (2012). *Record-breaking La Nina events*. en. Tech. rep. Melbourne: Bureau of Meteorology, p. 24. (Visited on 10/03/2017).

Møller, J. and C. Díaz-Avalos (2010). "Structured Spatio-Temporal Shot-Noise Cox Point Process Models, with a View to Modelling Forest Fires". en. In: *Scandinavian Journal of Statistics* 37.1, pp. 2–25. DOI: 10.1111/j.1467-9469.2009.00670.x. (Visited on 08/15/2016).

Møller, J. and R. P. Waagepetersen (2003). *Statistical Inference and Simulation for Spatial Point Processes*. en. Boca Raton: CRC Press.

Moradi, M., N. Read, J. Mateu, and A. Baddeley (n.d.). *Normalisation of the inhomogeneous K-function and pair correlation function*.

Mundo, I. A., T. Wiegand, R. Kanagaraj, and T. Kitzberger (2013). "Environmental drivers and spatial dependency in wildfire ignition patterns of northwestern Patagonia". In: *Journal of Environmental Management* 123, pp. 77–87. DOI: 10.1016/j.jenvman.2013.03.011. (Visited on 10/12/2016).

Nichols, K., F. P. Schoenberg, J. E. Keeley, A. Bray, and D. Diez (2011). "The application of prototype point processes for the summary and description of California wildfires". en. In: *Journal of Time Series Analysis* 32.4, pp. 420–429. DOI: 10.1111/j.1467-9892.2011.00734.x. (Visited on 10/05/2016).

Oliveira, S., F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira (2012). "Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest". In: *Forest Ecology and Management* 275.Supplement C, pp. 117–129. DOI: 10.1016/j.foreco.2012.03.003.

Otway, S. G., E. W. Bork, K. R. Anderson, and M. E. Alexander (2008). "Predicting sustained smouldering combustion in trembling aspen duff in Elk Island National Park, Canada". en. In: *International Journal of Wildland Fire* 16.6, pp. 690–701. DOI: 10.1071/WF06033. (Visited on 11/28/2017).

Ouyang, W., C. Y. Han, and S. T. Y. Tong (2016). "Modeling Wildfire Ignition Distribution and Making Prediction of Human-caused Wildfire". English. In: *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. Athens, USA: CSREA Press, pp. 335–341. (Visited on 10/20/2016).

Paterson, G. and D. M. Chong (2011). "Implementing the Phoenix fire spread model for operational use". In: *Proceedings of the Surveying and Spatial Sciences Biennial Conference 2011*. Wellington, p. 111. (Visited on 04/03/2017).

Peng, R. D., F. P. Schoenberg, and J. A. Woods (2005). "A Space–Time Conditional Intensity Model for Evaluating a Wildfire Hazard Index". In: *Journal of the American Statistical Association* 100.469, pp. 26–35. DOI: 10.1198/016214504000001763. (Visited on 10/05/2016).

Penman, T. D., R. A. Bradstock, and O. F. Price (2013). "Modelling the determinants of ignition in the Sydney Basin, Australia: implications for future management". en. In: *International Journal of Wildland Fire* 22.4, p. 469. DOI: 10.1071/WF12027. (Visited on 10/05/2016).

Plucinski, M. P. (2013). "Modelling the probability of Australian grassfires escaping initial attack to aid deployment decisions". en. In: *International Journal of Wildland Fire* 22.4, pp. 459–468. DOI: 10.1071/WF12019. (Visited on 11/28/2017).

Plucinski, M. P., W. L. McCaw, J. S. Gould, and B. M. Wotton (2014). "Predicting the number of daily human-caused bushfires to assist suppression planning in south-west Western Australia". In: *International Journal of Wildland Fire* 23.4, pp. 520–531. DOI: 10.1071/WF13090. (Visited on 08/22/2016).

Podur, J., D. L. Martell, and F. Csillag (2003). "Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998". In: *Ecological Modelling* 164.1, pp. 1–20. DOI: 10.1016/S0304-3800(02)00386-1. (Visited on 10/05/2016).

Pregibon, D. (1981). "Logistic Regression Diagnostics". In: *The Annals of Statistics* 9.4, pp. 705–724. (Visited on 12/14/2017).

Preisler, H. K., D. R. Brillinger, R. E. Burgan, and J. W. Benoit (2004). "Probability based models for estimation of wildfire risk". In: *International Journal of Wildland Fire* 13.2, pp. 133–142. DOI: 10.1071/WF02061. (Visited on 08/22/2016).

Preisler, H. K., S. Chen, F. Fujioka, J. W. Benoit, and A. L. Westerling (2008). "Wildland fire probabilities estimated from weather model-deduced monthly mean fire

danger indices". In: *International Journal of Wildland Fire* 17.3, pp. 305–316. DOI: 10.1071/WF06162. (Visited on 08/26/2016).

Preisler, H. K. and A. L. Westerling (2007). "Statistical Model for Forecasting Monthly Large Wildfire Events in Western United States". In: *Journal of Applied Meteorology and Climatology* 46.7, pp. 1020–1030. DOI: 10.1175/JAM2513.1. (Visited on 08/26/2016).

Preisler, H. K., A. L. Westerling, K. M. Gebert, F. Munoz-Arriola, and T. P. Holmes (2011). "Spatially explicit forecasts of large wildland fire probability and suppression costs for California". In: *International Journal of Wildland Fire* 20.4, pp. 508–517. DOI: 10.1071/WF09087. (Visited on 05/12/2016).

Raupach, M., P. Briggs, V. Haverd, E. King, M. Paget, and C. Trudinger (2009). *Australian water availability project (AWAP): CSIRO Marine and Atmospheric Research component : final report for phase 3*. en. CAWCR Technical Report No. 013. Melbourne: The Centre for Australian Weather and Climate Research, 67 pp.

Raupach, M., V. Haverd, E. King, M. Paget, and C. Trudinger (2015). *Australian Water Availability Project*. Tech. rep. Canberra, Australia: CSIRO Marine and Atmospheric Research. (Visited on 06/01/2015).

Read, N., T. J. Duff, and P. G. Taylor (2018). "A lightning-caused wildfire ignition forecasting model for operational use". In: *Agricultural and Forest Meteorology* 253-254, pp. 233–246. DOI: 10.1016/j.agrformet.2018.01.037. (Visited on 03/16/2018).

Renner, I. W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic, and D. I. Warton (2015). "Point process models for presence-only analysis". en. In: *Methods in Ecology and Evolution* 6.4, pp. 366–379. DOI: 10.1111/2041-210X.12352. (Visited on 02/13/2018).

Rodrigues, A. and P. J. Diggle (2010). "A Class of Convolution-Based Models for Spatio-Temporal Processes with Non-Separable Covariance Structure". In: *Scandinavian Journal of Statistics* 37.4, pp. 553–567. DOI: 10.1111/j.1467-9469.2009.00675.x. (Visited on 04/11/2018).

Rodrigues, M., J. de la Riva, and S. Fotheringham (2014). "Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using ge-

ographically weighted logistic regression". In: *Applied Geography* 48.Supplement C, pp. 52–63. DOI: 10.1016/j.apgeog.2014.01.011.

Rorig, M. L. and S. A. Ferguson (1999). "Characteristics of Lightning and Wildland Fire Ignition in the Pacific Northwest". In: *Journal of Applied Meteorology* 38.11, pp. 1565–1575. DOI: 10.1175/1520-0450(1999)038<1565:COLAWF>2.0.CO;2. (Visited on 10/19/2016).

Royston, P. and D. G. Altman (1994). "Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 43.3, pp. 429–467. DOI: 10.2307/2986270. (Visited on 12/05/2017).

— (2010). "Visualizing and assessing discrimination in the logistic regression model". en. In: *Statistics in Medicine* 29.24, pp. 2508–2520. DOI: 10.1002/sim.3994. (Visited on 01/12/2018).

Royston, P. and W. Sauerbrei (2008). *Multivariable Model - Building: A Pragmatic Approach to Regression Anaylsis based on Fractional Polynomials for Modelling Continuous Variables*. en. Chichester: John Wiley & Sons.

Rubin, D. B. and S. Weisberg (1974). "The Variance of a Linear Combination of Independent Estimators Using Estimated Weights". en. In: *ETS Research Bulletin Series* 1974.2, pp. i–5. DOI: 10.1002/j.2333-8504.1974.tb00860.x.

Schoenberg, F. P. (2005). "Consistent parametric estimation of the intensity of a spatial–temporal point process". In: *Journal of Statistical Planning and Inference* 128.1, pp. 79–93. DOI: 10.1016/j.jspi.2003.09.027. (Visited on 04/25/2018).

Schoenberg, F. P., J. Pompa, and C.-H. Chang (2008). "A note on non-parametric and semi-parametric modeling of wildfire hazard in Los Angeles County, California". en. In: *Environmental and Ecological Statistics* 16.2, pp. 251–269. DOI: 10.1007/s10651-007-0087-z. (Visited on 10/05/2016).

Serra, L., M. Saez, J. Mateu, D. Varga, P. Juan, C. Díaz-Ávalos, and H. Rue (2014). "Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994–2008". en. In: *Environmental and Ecological Statistics* 21.3, pp. 531–563. DOI: 10.1007/s10651-013-0267-y. (Visited on 04/05/2017).

Stephenson, C. (2010). *The Impacts, Losses and Benefits Sustained from Five Severe Bushfires in South-Eastern Australia*. Tech. rep. 88. Melbourne: Bushfire Cooperative Research Centre. (Visited on 11/21/2016).

Tjur, T. (2009). "Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination". In: *The American Statistician* 63.4, pp. 366–372. DOI: 10.1198/tast.2009.08210. (Visited on 12/19/2017).

Todd, B. and P. H. Kourtz (1992). "Predicting the daily occurrence of people-caused forest fires". In: *Proceedings of the Eighth Central Region Fire Weather Committee Scientific and Technical Seminar*. Winnipeg.

Turner, R. (2008). "Point patterns of forest fire locations". en. In: *Environmental and Ecological Statistics* 16.2, pp. 197–223. DOI: 10.1007/s10651-007-0085-1. (Visited on 10/05/2016).

Vasconcelos, M. J. P. d., S. Silva, M. Tome, M. Alvim, and J. M. C. Pereira (2001). "Spatial prediction of fire ignition probabilities: Comparing logistic regression and neural networks". In: *Photogrammetric Engineering and Remote Sensing* 67.1, pp. 73–81. (Visited on 10/21/2016).

Vasilakos, C., K. Kalabokidis, J. Hatzopoulos, G. Kallos, and Y. Matsinos (2007). "Integrating new methods and tools in fire danger rating". en. In: *International Journal of Wildland Fire* 16.3, p. 306. DOI: 10.1071/WF05091. (Visited on 10/21/2016).

Vasilakos, C., K. Kalabokidis, J. Hatzopoulos, and Y. Matsinos (2008). "Identifying wildland fire ignition factors through sensitivity analysis of a neural network". en. In: *Natural Hazards* 50.1, pp. 125–143. DOI: 10.1007/s11069-008-9326-3. (Visited on 10/21/2016).

Vega Garcia, C., P. M. Woodard, S. J. Titus, W. L. Adamowicz, and B. S. Lee (1995). "A Logit Model for Predicting the Daily Occurrence of Human Caused Forest-Fires". en. In: *International Journal of Wildland Fire* 5.2, p. 101. DOI: 10.1071/WF9950101. (Visited on 10/05/2016).

Wang, Y. and K. R. Anderson (2010). "An evaluation of spatial and temporal patterns of lightning- and human-caused forest fires in Alberta, Canada, 1980–2007". en. In: *International Journal of Wildland Fire* 19.8, p. 1059. DOI: 10.1071/WF09085. (Visited on 10/05/2016).

Williams, D. (1991). *Probability with Martingales*. en. Cambridge: Cambridge University Press.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. 2nd. Boca Raton: CRC Press.

Woolford, D. G., D. R. Bellhouse, W. J. Braun, C. B. Dean, D. L. Martell, and J. Sun (2011). "A Spatio-temporal Model for People-Caused Forest Fire Occurrence in the Romeo Malette Forest". In: *Journal of Environmental Statistics* 2.1, pp. 1–26.

Woolford, D. G., J. Cao, C. B. Dean, and D. L. Martell (2010). "Characterizing temporal changes in forest fire ignitions: looking for climate change signals in a region of the Canadian boreal forest". en. In: *Environmetrics* 21.7-8, pp. 789–800. DOI: 10.1002/env.1067. (Visited on 08/22/2016).

Woolford, D. G., C. B. Dean, D. L. Martell, J. Cao, and B. M. Wotton (2014). "Lightning-caused forest fire risk in Northwestern Ontario, Canada, is increasing and associated with anomalies in fire weather". en. In: *Environmetrics* 25.6, pp. 406–416. DOI: 10.1002/env.2278. (Visited on 08/15/2016).

Worboys, G. (2003). "A Brief Report on the 2003 Australian Alps Bushfires". In: *Mountain Research and Development* 23.3, pp. 294–295. DOI: 10.1659/0276-4741(2003)023[0294:ABROTA]2.0.CO;2. (Visited on 04/11/2018).

Wotton, B. M. and D. L. Martell (2005). "A lightning fire occurrence model for Ontario". In: *Canadian Journal of Forest Research* 35.6, pp. 1389–1401. DOI: 10.1139/x05-071. (Visited on 08/22/2016).

Wotton, B. M., D. L. Martell, and K. A. Logan (2003). "Climate Change and People-Caused Forest Fire Occurrence in Ontario". en. In: *Climatic Change* 60.3, pp. 275–295. DOI: 10.1023/A:1026075919710. (Visited on 09/02/2016).

Xu, H. and F. P. Schoenberg (2011). "Point process modeling of wildfire hazard in Los Angeles County, California". EN. In: *The Annals of Applied Statistics* 5.2A, pp. 684–704. DOI: 10.1214/10-AOAS401. (Visited on 10/05/2016).

Yang, J., H. S. He, S. R. Shifley, and E. J. Gustafson (2007). "Spatial Patterns of Modern Period Human-Caused Fire Occurrence in the Missouri Ozark Highlands". In: *Forest Science* 53.1, pp. 1–15.

Yang, J., P. J. Weisberg, T. E. Dilts, E. L. Loudermilk, R. M. Scheller, A. Stanton, and C. Skinner (2015). "Predicting wildfire occurrence distribution with spatial point process models and its uncertainty assessment: a case study in the Lake Tahoe Basin, USA". en. In: *International Journal of Wildland Fire* 24.3, p. 380. DOI: 10.1071/WF14001. (Visited on 10/05/2016).

Zhang, H., X. Han, and S. Dai (2013). "Fire Occurrence Probability Mapping of Northeast China With Binary Logistic Regression Model". In: *IEEE Journal of*

Selected Topics in Applied Earth Observations and Remote Sensing 6.1, pp. 121–127. DOI: 10.1109/JSTARS.2012.2236680.

Zhang, Y., S. Lim, and J. J. Sharples (2016). "Modelling spatial patterns of wildfire occurrence in South-Eastern Australia". In: *Geomatics, Natural Hazards and Risk* 7.6, pp. 1800–1815. DOI: 10.1080/19475705.2016.1155501.