The Spread of the Mountain Pine Beetle: Challenges and Solutions in Large-scale Spatial Ecological Modeling

by

Dean Cameron Koch

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Applied Mathematics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

The mountain pine beetle (MPB) is among the most destructive eruptive forest pests in North America. A recent increase in the frequency and severity of oubtreaks, combined with an eastward range expansion towards untouched boreal pine forests, has spurred a great interest by government, industry and academia into the population ecology of this tree-killing bark beetle. Modern approaches to studying the MPB often involve the analysis of large-scale, high resolution datasets on landscape level damage to pine forests. This creates a need for new modelling tools to handle the unique challenges associated with large sample sizes and spatial effects. In this thesis, I develop some of these tools and apply them to studying the spread and attack behaviour of MPB. Chapter 2 introduces a statistical framework for handling spatial autocorrelation based on the geostatistics paradigm of explicit covariance functions, known as covariograms. This extends previous work on computationally feasible models for covariances in lattice data, and introduces a powerful new estimator for the angle of anisotropy in stationary random field representations of autocorrelation. Chapter 3 unifies a number of previously unconnected results on redistribution kernels, by presenting a novel mechanistic derivation of a widely-applicable kernel for isotropic movement patterns. Phenomenological extensions are proposed to account for anisotropy as well as an approximation with superior computational properties, which is well suited to model-fitting on extremely large samples. In Chapter 4, these ideas are combined to construct a comprehensive model for the spatial spread of infested stems, coupled by MPB dispersal flights. This model facilitates the landscape-level inference of subtle properties of MPB attack behaviour based on aerial surveys of killed pine. Among these result is an accurate estimate of the size of the cryptic endemic MPB population, which formerly has been measurable only by means of costly and time-intensive ground surveys.

Preface

This thesis is an original work by Dean Koch. No part of this thesis has been previously published. However a version of Chapter 2 has been submitted to *Environmental and Ecological Statistics* and is currently in review. Mark Lewis and Subhash Lele were supervisory authors in all chapters, providing guidance and feedback in model development.

“To those who do not know mathematics it is difficult to get across a real feeling as to the beauty, the deepest beauty, of nature.”

-Richard P. Feynman

Acknowledgements

First I wish to thank my supervisor Dr. Mark Lewis for his unwavering support and confidence in me through the years. Mark is an unusually kind and patient role model, and has been a very positive force in my time as a graduate student. I also thank Dr. Subhash Lele, whose guidance, insight and feedback stoked my interest in statistics and probability. Both have gone out of their way to open doors for me in, and I am fortunate to have had the opportunity to work with them.

I am also indebted to all members (past and present) of the Lewis Research Group for their expert advice, feedback, and support through all stages of my research. I also thank Dr. Stan Butin, Dr. Murray Humphries, and Manuelle Landry-Cuerrier for their roles in leading the CREATE-EI training program, which has opened my eyes to the roles and responsibilities of ecologists outside of academia.

I wish to recognize my cohort of applied mathematics graduate students – especially Jody, Wafa, Carlos, and Nathan – whose friendship and support have been helped me overcome the challenges of a long graduate study program. Tessa, thank you for looking after me and putting up with me as I went through the most difficult part of my PhD. I couldn't have done this without you.

This PhD was supported by a grant from the Natural Science and Engineering Research Council of Canada (Grant No. NET GP 434810-12) to the TRIA Network; with contributions from Alberta Agriculture and Forestry; Foothills Research Institute; Manitoba Conservation and Water Stewardship; Natural Resources Canada-Canadian Forest Service; Northwest Territories Environment and Natural Resources; Ontario Ministry of Natural Re-

sources and Forestry; Saskatchewan Ministry of Environment; and industry partners West Fraser, and Weyerhaeuser.

Table of Contents

List of Tables

List of Figures

Glossary of Terms

AB Canadian province of Alberta.

Allee effect a phenomenon of reduced fitness induced by low population sizes, leading to a reduced per-capita population growth rate.

AOS aerial overview survey, in which forest damage is recorded from fixed-wing aircraft.

attack in the context of MPB ecology, the process by which female beetles bore through the bark of a tree, girdling and killing it.

BC Canadian province of British Columbia.

CAR conditional autoregressive, a type of Markov random field model.

condition number for a SPD matrix, the ratio of the largest to smallest eigenvalues.

covariogram (*i.e.* covariance function) defines covariances between pairs of points in a random field as a function of their positions.

endemic phase MPB populations which are too small to successfully attack healthy pine.

epidemic phase a MPB population which has grown to the point of widespread, sustained attacks on healthy pine.

eruptive in reference to a population, characterized by cycles of rapid growth and collapse.

fractal medium an irregular physical arrangement of space in which measurements are related to the scale of measurement by a power law.

fractal diffusion a diffusion model for movements within a fractal medium.

geometric anisotropy describing a SOS random field whose covariogram can be made isotropic through an affine transformation of coordinates.

GLS generalized least squares, a formula for unbiased and efficient estimates of linear regression parameters in the presence of correlated residuals.

ha hectare, a unit of area equal to 10,000 square metres.

incipient-epidemic phase the minimum population size at which MPB are capable of successfully attacking a healthy pine.

kriging a technique for spatial interpolation using a weighted sum of nearby known values.

MPB the mountain pine beetle *Dendoctronus ponderosae*.

multivoltinism having more than one brood of offspring per year.

MVN multivariate normal, a generalization of the univariate Gaussian distribution to *n*-dimensional random vectors.

PAC product anisotropic covariograms, a class of separable covariance functions depending on both distance and direction.

pseudoreplicate a replicate sample which cannot be viewed as a replicate treatment owing to statistical dependence.

pWMY product-WMY, a 2-dimensional redistribution kernel formulated as a product of two 1-dimensional WMY kernels.

redistribution kernel a probability density function for movement events.

refugium location where a small isolated population of a once-abundant species persists.

RMSPE root mean square prediction error, a measure of the (overall) agreement between model predictions and observations.

SAC spatial autocorrelation, or the property that pairs of measurements taken closer together are more (or less) similar than expected of a pair associated at random.

SAR simultaneous autoregressive, a type of Markov random field model.

scramble competition describing competition by multiple individuals (or species) for a common, shared, finite resource.

SOS second-order stationary, a class of random fields having constant mean and a covariance function that depends only on separation vectors (and not on absolute positions).

SPD symmetric positive definite, a class of square matrices having all positive eigenvalues.

stand susceptibility an index of the relative risk of MPB infestation among pine stands.

TSA timber supply area, a governmental division of forestland.

Type III functional response a sigmoid relationship between the production rate of a predator and the abundance of its prey.

WM the Whittle-Matérn covariogram, a covariance function in spatial statistics.

WMY the Whittle-Matérn-Yasuda redistribution kernel, a model for fractal diffusion with constant settling.

zonal anisotropy describing a SOS random field for which the covariogram has a sill that depends on direction.

Chapter 1

Introduction

1.1 Ecology of the mountain pine beetle

The mountain pine beetle (MPB), *Dendoctronus ponderosae* (Coleoptera: Curculionidae) is a tree-killing bark beetle native to pine forests of western North America. The MPB feeds on the phloem of pine trees, gaining access to this well protected inner layer of tissue through cooperative attacks involving dozens to hundreds of beetles that simultaneously bore through the bark *en masse* (Raffa and Berryman, 1983). The shock of this *mass attack* in combination with the introduction of mutualistic fungi (pathogenic to the tree), can overwhelm a pine's defense system, effectively girdling the tree and leading to its death (Safranyik and Carroll, 2006).

MPB ecologists have identified a number of distinct phases of MPB behaviour relating to their population density. The normative state is thought to be the cryptic *endemic phase*, in which low density populations cooperate with other bark beetle species to occupy a small niche of sick and dying pine (Boone et al., 2011; Lindgren and Raffa, 2013). This phase is so small as to have a virtually undetectable impact on pine populations. However, quasiperiodic eruptions in MPB populations (*outbreaks*) are a natural occurance, and, unlike the endemic phase, an outbreaking population is capable of spreading through a stand to kill the majority of mature healthy pine trees in a matter of years.

The frequency and severity of these outbreaks is on the rise – with forestry and fire suppression practices, along with climate change all likely playing important roles in driving an apparent regime shift for MPB over the past century (Raffa et al., 2008). In the past two decades, for example, an epidemic of MPB outbreaks has consumed more than half of the merchantable pine in western Canada, impacting over 18 million hectares of forestland (Dhar et al., 2015). This was the largest such event ever recorded, and it is expected to have severe consequences for the province of British Columbia (BC), both economic (Corbett et al., 2015) and environmental (Kurz et al., 2008).

Perhaps more worrisome, however, is that this latest epidemic was accompanied by a range expansion across the Rocky Mountains into Alberta (AB) during the mid-2000s (Giroday, Carroll, and Aukema, 2012), which threatens to continue eastward through the uninterrupted belt of jack-pine dominated boreal forest that spans much of the continent (Safranyik et al., 2010). In response there has been a large research effort towards better understanding the population dynamics of the MPB and the potential for anticipating and controlling future outbreaks using mathematical models (*eg*. Shore and Safranyik, 1992; Kunegel-Lion, McIntosh, and Lewis, 2018; Goodsman and Lewis, 2016).

However, MPB population modellers have been (and continue to be) challenged by the many nonlinear interactions and spatial effects that characterize this fascinating species (Nelson et al., 2008). The accurate forecasting and effective management of outbreaks remains extremely difficult (Six, Biber, and Long, 2014), in spite of an abundance of spatio-temporal outbreak data on which to calibrate parameters and test hypotheses (Wulder et al., 2010). This dissertation will contribute a suite of spatial modelling tools to assist MPB ecologists in connecting their models to data, in the hope that by better understanding the beetle today we might avert the continent-wide epidemic that looms tomorrow.

1.1.1 The cryptic endemic population phase

Endemic MPB subsist on an ephemeral group of pine trees whose defensive capability is compromised due to natural stressors – examples include aging, windthrow, disease, and suppression by more dominant vegetation (Berryman, 1979; Raffa and Berryman, 1983; Carroll et al., 2006). The colonization of these trees by bark beetles is met with little resistance. Mass attacks are not needed, so the Allee effect seen during outbreaks is largely absent (Bleiker et al., 2014). Because of this difference between the endemic and outbreak phases, most MPB population models either ignore the endemic population (viewing it to be negligeably small), or else represent MPB growth dynamics by joining two different recruitment curves into a single, often discontinuous, multi-equilibrium growth model (*eg*. Berryman, 1979; Cooke and Carroll, 2017). However, as time-series on endemic populations are scarce, such multi-equilibrium models are rarely confronted with data.

I describe a more parsimonious modelling approach in Chapter 4, where a spatially uniform endemic population (ϵ) is assumed to exist throughout the native range of the MPB, effectively elevating the intensity of all mass attack attempts. Mathematically, this amounts to shifting one of my state variables (the attacking beetle population B_t) by the fixed but unknown quantity ϵ. By fitting such outbreak models to large-scale aerial datasets on MPB damage and comparing with ground surveys, I show that this simple idea produces remarkably accurate estimates of the size of the endemic population – remarkable because aerial surveys detect only mass attacks, and therefore contain essentially no direct observations of the activity of MPB at the endemic-level.

1.2 Forest disturbance as a proxy for insect populations

MPB have a one-year life cycle and highly synchronized summer reproduction schedule, so it is mathematically convenient to model their generations as non-overlapping. More precisely, a generation begins in the summer of year $t - 1$ with the death of one or more host pine trees, within which the mated females of the parent generation have laid eggs. It ends in the summer of year t after the eggs have hatched and the new cohort has emerged and attempted attacks on different pines, with reproduction occurring only where an attack has killed the host. Exceptions such as strip (partial) attacks (Rasmussen, 1974) and multivoltinism (Safranyik et al., 2010) are biologically interesting, but rare enough that they are expected to have little bearing on stand level outbreak patterns under ordinary circumstances (Safranyik and Carroll, 2006).

Since each MPB reproduction event implies a host death, the task of monitoring outbreaks is usually made simpler and cheaper by tracking their effect on populations of pine. On-the-ground efforts to monitor MPB – typically by counts of bore holes and/or egg galleries in individual trees, or by pheromone trapping (Safranyik and Carroll, 2006) – are labour-intensive and location-specific. Host mortality on the other hand is relatively easy to track over large expanses of forestland via recordings of crown fade patterns (Westfall and Ebata, 2009).

Crown fade survey operators are trained to distinguish forest health issues affecting different types of trees. In particular, ϕ_t, the percent mortality in pines (in a given stand) due to MPB mass-attacks in summer t can be distinguished from other disturbance types. When data on pre-attack susceptible pine density H_t (in stems/ha), are also available (or estimable), the product $I_t = \phi_t H_t$ counts the density of stems becoming infested in summer t. A year later, each of the infested stems produces a cohort of adult beetles that emerge to join the attacking MPB population, which I will call B_t.

Thus MPB dynamics are sometimes better described in terms of infested tree units I_t, or *red-tops* (Heavilin and Powell, 2008), from which the beetle population can be estimated by multiplication with a production rate term (Nelson et al., 2008). For example, assuming dispersal to be negligeable, if the average number of female MPB to emerge from each infested stem is β then the attacking MPB population in summer $t + 1$ can be modelled as $B_{t+1} = \beta\phi_t H_t = \beta I_t$ (Berryman, 1974). Note that since attacks are initiated by females (Safranyik and Carroll, 2006), variable B_t refers (here and throughout this dissertation) to the density of female MPB (in beetles/ha) attacking in year t.

1.2.1 Large-scale surveys of eruptive forest insect populations

In the province of BC, the Aerial Overview Survey (AOS) measures $\phi_t(\boldsymbol{x})$ and other forest disturbance data over most of its landbase on a yearly basis (Westfall and Ebata, 2009). This provides a large time series of population data on the MPB, revealing their activity levels each summer across vast and remote areas. Although AOS data has precision issues arising from the manual delineation of damage patterns onto maps (Wulder et al., 2009; Robertson et al., 2009), its wide coverage and high level of detail can be leveraged to yield very large sample sizes in a data analysis. This approach has been instrumental in modelling the landscape-level properties of MPB outbreaks in BC that cannot be discerned from small localized surveys, such as synchrony (Aukema et al., 2006; Chen et al., 2015) and spread (Chen and Walton, 2011).

A growing number of models for eruptive forest pests are using aerial forest disturbance data in this way. Similar data have been used to fit contemporary MPB outbreak models all over the Rocky Mountains region; from AB and BC in the north (*eg.* Aukema et al., 2008; Goodsman et al., 2016) to Idaho, Wyoming, Utah and Colorado in the South (*eg.* Heavilin and Powell, 2008; Preisler et al., 2012; Chapman, Veblen, and Schoennagel, 2012; Powell and Bentz, 2014; Strohm, Reid, and Tyson, 2016). Aerial damage surveys are also routinely used to track populations of similar tree-killing bark beetle and defoliator species throughout North America (Hall et al., 2016), such as gypsy moth (*eg.* Hohn, Liebhold, and Gribko, 1993; Lele, Taper, and Gage, 1998) and spruce budworm (*eg.* Candau, Fleming, and Hopkin, 1998; Goodbody et al., 2018).

Improvements from recent decades in data-collection, analysis methodology, and computing technology have made large spatially referenced datasets on forest health such as the AOS an increasingly rich source of information for ecologists. Aerial surveys are just one example; others include the use of photography and remote sensing (Wulder et al., 2006); the interpolation of point data (Zhou and Liebhold, 1995); model extrapolations such as the forest inventory maps of Beaudoin et al. (2014); and simulation experiments such as the MPB population growth models of Raffa and Berryman (1986), Logan et al. (1998), and Carroll et al. (2003). As ecology evolves to view the spatial nature of data as more of an asset than a nuisance, these types of analyses should become more common (Legendre, 1993; Kareiva, 1994; Fortin and Dale, 2005).

With this new interest comes a need for new mathematical tools equipped to handle the unique challenges that come with large spatial ecological datasets. I elaborate on two of these challenges in Chapters 2 and 3 – autocorrelation and flight-based dispersal, respectively – and develop novel methodology for handling them. I then combine both methods in Chapter 4 to infer some difficult-to-measure characteristics of MPB populations, illustrating the depth of information that can be extracted from the AOS with the proper tools. First I will introduce the outbreak model used throughout this dissertation, and review its history in the MPB literature.

1.3 Aspatial models for stand-level MPB attack behaviour

The type of language found in the literature on MPB ecology (*eg.* host, susceptibility, outbreak, epidemic) speaks to its disease-like nature. A MPB closely resembles what disease ecologists call a *parasitoid* – an species that lives in close association with a host insect, only to ultimately kill it – except that in the case of the MPB, the host is a plant (Goodsman et al., 2016). MPB models are therefore often inspired by the classic parasitoid-host model of Nicholson and Bailey (1935). For example Goodsman, Cooke, and Lewis (2017) defined a general aspatial MPB model of this type:

$$B_{t+1} = \beta\phi(B_t, H_t)H_t, \tag{1.1}$$

$$H_{t+1} = \lambda\left(1 - \phi(B_t, H_t)\right)H_t, \tag{1.2}$$

relating pine (H_t) and MPB density (B_t) at time t (here yearly); with $\lambda \geq 1$ the rate parameter for geometric growth in the pine population; β the number of MPB brood to emerge per attacked tree, and ϕ a function summarizing a density dependent process determining the proportion of pine killed. Goodsman, Cooke, and Lewis (2017) defined ϕ to be an expected value for host mortality, and derived it mathematically from a stochastic description of the attack process. I define it more loosely here as a phenomenological description of average attack outcomes in stand level (aggregate) data on MPB.

1.3.1 Attack success curves

Rearranging equation (1.1) to emphasize the infested tree count ($I_t = B_t/\beta$) yields the equation $I_t/H_t = \phi(B_t, H_t)$. In the phenomenological approach, ϕ is usually chosen to have a simple mathematical form that reasonably matches with empirical relationships between pine mortality and attack density (*eg.* Berryman, 1979; Cooke and Carroll, 2017). In this context, ϕ is called an *attack-sucess* curve (Nelson et al., 2008). It relates attack density, as measured either by B_t, I_t or B_t/H_t, to the proportion of hosts that are successfully mass attacked in a given year.

Ground surveys suggest that during outbreaks this relationship should be sigmoid, *i.e.* it is S-shaped (Raffa and Berryman, 1983; Boone et al., 2011). Berryman et al. (1985), for example, fitted the probit function for ϕ in his study of cooperative attack dynamics. A mathematically simpler (but qualitatively similar) alternative is the following model with shape parameter $\kappa_t > 1$ and stand-level outbreak susceptibility parameter $a_t > 0$:

$$\phi(B_t, a_t, \kappa_t) = \frac{B_t^{\kappa_t}}{B_t^{\kappa_t} + a_t^{\kappa_t}} = \frac{I_t^{\kappa_t}}{I_t^{\kappa_t} + (a_t/\beta)^{\kappa_t}}. \tag{1.3}$$

In the special case $\kappa_t = 2$, equation (1.3) resembles the familiar type III functional response curve of Holling (1959), which describes a sigmoid relationship between predation levels and prey density when there is both prey-switching at low densities and saturation at high densities. Here the functional dependence in equation (1.3) is on predator (MPB) rather than prey (pine) density, better characterizing the types of functional responses found in parasitoid-prey systems (May, 1978; Hassell, 1978); In the case of MPB, cooperative attacks fail when the attackers are few, and the resulting Allee effect forces a switch to defensively compromised (*eg.* injured or dying) pine; At high densities, MPB tend to avoid sub-optimal hosts (*eg.* low vigour pine) until scramble competition forces their hand (as modelled in Lewis, Nelson, and Xu, 2010), leading to saturation.

Equation (1.3) has these essential features – a concavity near zero, and saturation at high attack densities – so it is often used to represent the host mortality rate in models for outbreaking insects. Examples include the spread-damage PDEs of Ludwig, Jones, and Holling (1978), Strohm, Tyson, and Powell (2013) and Strohm, Reid, and Tyson (2016). Upon dividing the top and bottom of equation (1.3) by β^{κ_t} (and replacing B_t with βI_t) we recover the red-top model of Heavilin and Powell (2008). This same model (with $\kappa_t = 2$) also appears in Goodsman and Lewis (2016) where it was used to study the Allee effect.

Sigmoid attack-sucess curves such as equation (1.3) lead to eruptive dynamics (Berryman, 1978; Raffa et al., 2008). This allows modellers to explain MPB outbreak initiation in terms of critical threshold points: if attack density initially lies below the threshold, beetle populations decline; if it lies above, they rapidly increase. Mathematical analyses of the properties of these thresholds (and the dynamical systems they drive) can lead to new insights in MPB ecology.

For example, the implicitly spatial analysis in Heavilin and Powell (2008) showed how, as pine stands mature, they can become more susceptible to attack, drawing the threshold down and making it increasingly likely that a random in-flight of MPB will spark an outbreak. The spatially explicit extension of this model in Goodsman and Lewis (2016) defined the threshold more precisely, using it to calculate a minimum founding population for MPB. I estimate a similar quantity in Chapter 4, the incipient-epidemic transition point, or the density at which MPB populations switch from endemic to outbreak-level attack behaviour (Carroll et al., 2006).

1.3.2 The generalized red-top model

The assumption of $\kappa_t = 2$ in Heavilin and Powell (2008) (and subsequent work based on the red-top model) appears to be motivated more by mathematical convenience than empirical evidence. The data analysis in Cooke and Carroll (2017) suggests that a range of κ_t values may be realized in practice, with year-to-year variations due to varying climatic conditions. My relaxation ($\kappa_t > 1$) therefore introduces into equation (1.3) some flexibility that was missing in the original red-top model: When the parameter κ_t becomes large, the sigmoid shape of the attack success curve becomes more pronounced; reflecting healthier, more defensively resilient pine. As $\kappa_t \to 1$ the Allee effect vanishes and equation (1.3) becomes more like a Beverton-Holt model (Kot, 2001), reflective of defensively compromised (*eg.* drought-stressed) pine. This generalized attack success curve, and the parameters β, a_t, and κ_t are discussed in more detail in Chapter 4.

Another weakness of the formulation of ϕ in the red-top model is that stand susceptibility a_t – *i.e.* the attack density B_t at which 50% of hosts in a stand are killed – is independent of the density of pine in the stand. Generally speaking, each successful attack requires the cooperation of many MPB, and each mass-attacked tree corresponds to a distinct group of MPB (Safranyik and Carroll, 2006). Therefore a_t should increase with pine density. For example a pine-leading stand with 1000 stems/ha would have a much higher a_t value than a stand with 100 stems/ha, because 10X more attackers would be required (at a minimum) to carry out the 500 mass-attacks.

One solution is to simply replace parameter a_t in equation (1.3) by $a_t H_t$; or equivalently, to scale B_t by $1/H_t$, and rewrite the functional response in equation (1.1) as $\phi(R_t; a_t, \kappa_t)$ where $R_t = B_t/H_t$ is the average number of attacking MPB per host. This is similar to the approach of Goodsman et al. (2016), where R_t is the mean of a random variable representing (per-tree) attack density, and ϕ is its cumulative distribution function. The system (1.1)-(1.2) with $\phi(B_t, H_t) = \phi(R_t; a_t, \kappa_t)$ defined as in equation (1.3), is more simplistic in that it uses a phenomonological attack success curve, but it has similar dynamics at the onset of outbreaks. I will refer to this system as the *generalized red-top model*:

$$B_{t+1} = \beta\phi\left(B_t/H_t; a_t, \kappa_t\right) H_t, \tag{1.4}$$

$$H_{t+1} = \lambda\phi\left(1 - (B_t/H_t; a_t, \kappa_t)\right) H_t, \tag{1.5}$$

When the parameters $a_t = a$ and $\kappa_t = \kappa$ are fixed in time, (1.4)-(1.5) can be solved analytically. I derive its solution and give a brief discussion of its properties in Appendix 1.6.1. Its description of the onset of mass attack behaviour is similar to that accompanying the fixed point analysis in Heavilin and Powell (2008). A critical threshold point is defined by the model parameters a, β, and κ; When initial attack density $R_0 = B_0/H_0$ exceeds this threshold, an outbreak takes hold and the host population is entirely consumed. When R_0 is initially below the threshold, mass attacks tend to fail and the MPB population rapidly declines to zero.

As time t progresses, the model (1.1)-(1.2) loses realism. For example in the absence of MPB the long-term behaviour of the red-top model in Heavilin and Powell (2008) is of geometric growth in H_t, with an outbreak threshold approaching zero. H_t also grows geometrically in my generalization (1.4)-(1.5), but the outbreak threshold increases without bound. Of course neither of these contradictory conclusions is biologically reasonable. The problem (in both models) is that there is no accounting of the complex process of succession in forest stand demographics. Susceptible pine density cannot increase without bound; it will saturate over time, and a_t will plateau and then decline as aging and crowding diminishes the defensive capacity of trees (Raffa et al., 2008). Moreover the use of a single variable to represent the host population (and a linear density dependence term for its growth), as assumed in the general template (1.1)-(1.2), is a gross oversimplification. A realistic approach to long-term dynamics would require, at a minimum, a structured representation of the host population (*eg.* as in Lewis, Nelson, and Xu, 2010).

However, from the analysis of long-term dynamics and steady state properties of systems based on (1.1)-(1.2) (*eg.* Heavilin and Powell, 2008; Goodsman et al., 2016; Goodsman, Cooke, and Lewis, 2017, and Appendix 1.6.1), we get a more complete picture of the model's behaviour, and a better understanding of the limitations of its descriptive ability. We see that (1.1)-(1.2) is plausible as a model for transient dynamics at the onset of outbreaks, but not long-term behaviour; and that it is most useful in situations where host population growth is known or assumed negligeable ($\lambda = 1$). Thus in Chapters 2-4, I use the generalized red-top model to focus on spatial (rather than temporal) aspects of MPB population dynamics in an area where H_t can be estimated at a high resolution.

1.4 Spatial extensions for dispersal and autocorrelation

In the spatially referenced version of the generalized red-top model, beetle and host variables are associated with a 2-dimensional position vector $\boldsymbol{x}$. This represents the central point of a pine stand over which population counts are aggregated to arrive at B_t and H_t. Thus in my 1 hectare (ha) resolution dataset, $B_t(\boldsymbol{x})$ (in females/ha) and $H_t(\boldsymbol{x})$ (in stems/ha) denote the average MPB and host density within the 100×100 m square block of land with centroid $\boldsymbol{x}$.

A model development for these data requires incorporating two mathematically challenging aspects of MPB damage patterns that are not addressed in (1.4)-(1.5): First, MPB populations are coupled by flight-based dispersal. Prior to attack, this dispersal acts among stands to redistribute the beetles counted by $B_t(\boldsymbol{x})$. Second, measurements of ecological data at this resolution will be spatially autocorrelated, so in spite of the mathematical appeal of an independence assumption it would be careless to ignore the dependence structure in model residuals (Legendre, 1993).

I studied both problems in some depth while developing the spatial generalized red-top model of Chapter 4. That research led to some novel modelling ideas and methodology with wider applications in spatial ecology. These results are introduced separately in Chapters 2 and 3, and then used together in Chapter 4 to show how movement and attack behaviour of MPB can be inferred at high resolution from historical AOS data.

1.4.1 Models for MPB dispersal

In the absence of immigration/emmigration flight events at block $\boldsymbol{x}$, the direct substitution of spatially referenced variables into equations (1.4)-(1.5) produces a reasonable model for the transient dynamics of an outbreak in a given stand. However, dispersal flights become very important to outbreak dynamics when working with data at this high resolution (Aukema et al., 2008). Large numbers of MPB may be expected to move *among* stands (particularly neighbouring ones) before attacking (Safranyik et al., 1992), and we know the model (1.3) to be highly sensitive to the attacking beetle density. Thus if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ lie close together, then their corresponding attack dynamics equations (1.4)-(1.5) must be coupled in a way that allows some of the population $B_t(\boldsymbol{x}_i)$ to move into block $\boldsymbol{x}_j$ and vice versa.

In a continuous time framework, the natural mathematical tool for modelling movements is the partial differential equation (PDE). This approach views $B_t(\boldsymbol{x})$ as a differentiable function of time, and models its time-evolution throughout the summer emergence and attack periods. For example, Powell and Bentz (2014) used PDEs based on diffusion and chemotaxis to predict attack damage; Strohm, Tyson, and Powell (2013) used them to explain the spacing of crown fade clusters; and Strohm, Reid, and Tyson (2016) studied the effectiveness of control efforts in simulation experiments based on PDEs.

In discrete time formulations such as (1.4)-(1.5), MPB dispersal is usually represented by a *redistribution kernel* – a probability density function for movement events, often derived as the solution to a PDE for movements. The model then becomes a system of *integrodifference* equations (Kot and Schaffer, 1986). For example in Heavilin and Powell (2008) and Goodsman and Lewis (2016), the emerging (pre-dispersal) MPB density is convolved with a Gaussian kernel to produce a post-dispersal attack density that is split across many neighbouring stands, coupling them together.

The Gaussian redistribution kernel – whose origins in ecology can be traced back to a PDE for animal movement solved by Skellam (1951) – has pleasant mathematical properties, making it extremely popular in spatial ecology. However it is just one example of many plausible kernels for movement. Research by Kot, Lewis, and Driessche (1996), for example, has shown that integrodifference systems for population spread can be quite sensitive to the subtle mathematical properties of the redistribution kernel. It is therefore important to consider the underlying movement mechanism that gives rise to the kernel – and as with the

Gaussian, this often done by constructing diffusion-based PDEs for movement.

There is a large body of mathematical theory associated with the PDE approach (*eg.* see Neubert, Kot, and Lewis, 1995; Okubo and Levin, 2001; Clobert et al., 2012). Building upon work by Yasuda (1975) and Hapca, Crawford, and Young (2008), I develop a PDE for biodiffusion in Chapter 3 that unifies some of this theory and produces a quite versatile family of functions to generalize and extend the Gaussian kernel. In particular, it includes as special (or limiting) cases three commonly used bark beetle dispersal kernels: The Gaussian, 2D Laplace, and Bessel kernels (Heavilin and Powell, 2008; Turchin and Thoeny, 1993; Goodsman et al., 2016), tying them together by a common mechanism.

Naturally it is important that the kernel function should also match reasonably well with patterns of redistribution seen in empirical data. Indeed kernels are often invoked by ecologists without mechanistic justification, as *phenomenological* models to this end. For example the 2D Laplace kernel appears often in the literature because, unlike the Gaussian kernel, it has a fat-tailed shape (Heavilin and Powell, 2008). Comparative reviews of other phenomenological models for various species can be found Taylor (1978) and Clark, Macklin, and Wood (1998) (for flying insects and seeds, respectively). In both of these reviews it is suggested that model fit is often substantially improved when using kernel families with more flexibility in tail-shape than the Gaussian. I show in Chapter 3 how my generalization of the Gaussian redistribution kernel achieves this flexibility by mechanistic rather than phenomenological arguments.

1.4.2 Models for covariance

Even with the dispersal mechanism properly accounted for, the model errors for a spatially referenced population variable such as $B_t(\boldsymbol{x})$ will often be correlated at nearby locations. There are several reasons: For one, errors in quantifying the source population (pre-dispersal) will be propagated in space by the dispersal model, leading to a predictable structure of spatial autcorrelation (SAC) in population levels post-dispersal. For example an error of omission at a single block in the AOS for attack year t can induce a cluster of errors in the model prediction for $t + 1$, as the (unaccounted for) MPB spread outward from an epicentre. This is discussed further in Section 5.1.1 of my concluding chapter.

More generally, the environment conditions encompassed by the blocks centered at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ tend to be more similar as the separation distance $|\boldsymbol{x}_i - \boldsymbol{x}_j|$ becomes smaller – this universal principle is often referred to as Tobler's First Law of Geography (Tobler, 1970). In the case of MPB, a shared environment drives the beetle to behave similarly in its interaction with pine at nearby locations. For example, drought stress or windthrow is unlikely to affect just a single hectare in isolation; Rather it would weaken pine defenses against MPB over a large set of contiguous blocks that are all exposed to similar weather patterns (Cooke and Carroll, 2017). For similar reasons, the drivers of measurement error in the AOS (*eg.* interpreter bias) may be more alike for measurements recorded at adjacent locations (Wulder et al., 2009).

SAC is thought to be ubiquitous in high resolution ecological data, so it is important that this phenomenon be scrutinized during model construction (Fortin and Dale, 2005; Beale et al., 2010). If SAC is dismissed by an assumption of spatial indendence, for example, the model will tend to underestimate standard errors, invalidating any planned hypothesis tests (Hawkins, 2012). Moreover, point estimators of model parameters lose precision in this situation (Cressie, 1993), in what essentially amounts to pseudoreplication (Legendre, 1993). This misspecification is often accepted as benign enough to ignore for the sake of simplicity in model-fitting (*eg.* as did Heavilin and Powell, 2008; Robertson et al., 2008; Powell and Bentz, 2014; Chen, 2014; Goodsman et al., 2016; Zheng and Aukema, 2010). However for my purposes – which involve inferring a very small endemic MPB population through indirect measurements – precision is paramount.

To account for SAC, a covariance structure can be specified either implicitly or explicitly. The implicit approach usually makes use of one of the autoregressive model families (commonly abbreviated as CAR and SAR), in which covariances are defined through partial correlations (Cressie, 1993; Ver Hoef et al., 2018). This means the precision matrix $\boldsymbol{V}^{-1}$ (the inverse covariance matrix) is specified directly, which allows speedy computations of the likelihood function, and leads to computationally simple model-fitting procedures (Fortin and Dale, 2005). Implicit covariance structures are quite popular in the bark beetle outbreak literature, appearing for example in the southern pine beetle outbreak model of Zhu, Huang, and Wu (2005), and the MPB outbreak models of: Zhu et al. (2008); Aukema et al. (2008); Robertson et al. (2009); Zhu, Huang, and Reyes (2010); Sambaraju et al. (2012); Reyes, Zhu, and Aukema (2012); and Preisler et al. (2012).

In explicit specifications, the covariance structure is usually defined using a *covariogram*, which describes how correlations decay with distance, and therefore (explicitly) defines $\boldsymbol{V}$. The advantage in this approach is that SAC is modeled in a transparent and less constrained way (Wall, 2004), with the caveat that likelihood function evaluations typically involve a computationally expensive matrix inversion ($\boldsymbol{V}^{-1}$). For this reason covariograms seem to be used far less often in spatial ecology than autoregression (Simpson, Lindgren, and Rue, 2012). Exceptions include the gypsy moth outbreak models of Zhou and Liebhold (1995) and Lele, Taper, and Gage (1998); and the MPB outbreak model of Chapman, Veblen, and Schoennagel (2012).

Preferring the more intuitive covariogram, while recognizing its computational drawbacks, I introduce a family of novel covariance models in Chapter 2 for which the computation of $\boldsymbol{V}^{-1}$ becomes much simpler when analysing lattice data. These models are simpler to interpret than the CAR and SAR, while at the same time providing far more flexibility for accurately handling long-range correlations and anisotropy. My methodology provides a general statistical framework for handling SAC in ecological data, which I subsequently put to use in fitting models developed in Chapters 3 and 4.

1.5 Thesis overview

The expansive coverage of the AOS represents a huge sampling of pine stands and MPB populations with which to test ecological hypotheses using models. However when modelling data of such high resolution, spatial effects become important, and these can be difficult to capture elegantly without overcomplicating the model: Dispersal flights of the MPB are difficult to incorporate into population models, but are essential to explaining attack damage patterns in practice; SAC demands a more careful handling of spatial replicates, lest they become pseudoreplicates. Further, even with relatively simple representations of these effects, large sample sizes often lead to issues of computational complexity when fitting to data or running simulation experiments.

I discuss these challenges in more detail in Chapters 2 and 3, developing new modelling techniques to handle them, and demonstrating these techniques on AOS data using simplified spatial extensions of (1.4)-(1.5). Chapter 2 introduces some new methods for parametrizing and studying SAC structure, and Chapter 3 introduces a novel derivation for a flexible family of redistribution kernels. Chapter 4 develops the (spatial) generalized red-top model more fully, incorporating results from both of the earlier chapters to handle computational roadblocks, and to produce novel estimators of various stand-level properties of outbreak dynamics. Chapter 5 summarizes these results and discusses their implications in ecology.

1.6 Appendices to Introduction

1.6.1 Solving the generalized red-top equation for $\kappa \neq 1$

Given the attack-success curve $\phi(R; a, \kappa) = R^\kappa/(R^\kappa + a^\kappa)$, the generalized red-top model is

$$B_{t+1} = \beta\phi\left(B_t/H_t; a, \kappa\right) H_t,$$

$$H_{t+1} = \lambda\phi\left(1 - (B_t/H_t; a, \kappa)\right) H_t.$$

Assume both beetle (B_t) and host density (H_t) are initially nonzero, and note that this implies $B_t > 0$ and $H_t > 0$ for all t. I will solve (1.4)-(1.5) by first solving for the yearly ratio of attacking beetles to susceptible pine, or *attack density* $R_t = B_t/H_t$, given initial data $R_0 > 0$. This ratio satisfies the simpler first-order nonlinear difference equation,

$$R_{t+1} = \frac{B_{t+1}}{H_{t+1}} = \frac{\beta\phi(R_t; a, \kappa)H_t}{\lambda\left(1 - \phi(R_t; a, \kappa)\right)H_t} = \frac{\beta R_t^{\kappa}}{\lambda a^{\kappa}},$$

which is solved by straightforward substitution and induction to get

$$\begin{aligned} R_{t+1} &= \left(\frac{\beta}{\lambda a^{\kappa}}\right) R_t^{\kappa} = \left(\frac{\beta}{\lambda a^{\kappa}}\right)\left(\frac{\beta R_{t-1}^{\kappa}}{\lambda a^{\kappa}}\right)^{\kappa} = \left(\frac{\beta}{\lambda a^{\kappa}}\right)\left(\frac{\beta}{\lambda a^{\kappa}}\right)^{\kappa}\left(\frac{\beta}{\lambda a^{\kappa}} R_{t-2}^{\kappa}\right)^{\kappa^2} \\ \ldots &= \left(\frac{\beta}{\lambda a^{\kappa}}\right)^{\sum_{i=0}^{t}\kappa^i} R_0^{\kappa^{t+1}}. \end{aligned}$$

Recalling the formula for sums of geometric progressions,

$$\sum_{i=0}^{t}\kappa^i = \kappa^0 + \kappa^1 + \ldots + \kappa^t = \begin{cases} t+1 & \text{for } \kappa = 1 \\ \frac{\kappa^{t+1}-1}{\kappa-1} & \text{for } \kappa \neq 1 \end{cases},$$

the solution for the case $\kappa \neq 1$ simplifies to become

$$R_t = \left(\frac{\beta}{\lambda a^{\kappa}}\right)^{\frac{\kappa^t-1}{\kappa-1}} R_0^{\kappa^t} = R_0\left(\frac{R_0}{\gamma}\right)^{\kappa^t-1} = \gamma\left(\frac{R_0}{\gamma}\right)^{\kappa^t}, \quad \text{where } \gamma := \left(\frac{\lambda a^{\kappa}}{\beta}\right)^{\frac{1}{\kappa-1}}. \tag{1.6}$$

It is clear from (1.6) that the constant γ is an unstable fixed point of the dynamical system in R_t. For $\gamma \neq R_0$ the ratio grows or decays rapidly (faster than exponential), depending on the signs of $\kappa - 1$ and $R_0 - \gamma$. For example when $\kappa > 1$, low initial attack densities ($R_0 < \gamma$) imply that R_t decreases monotonically to zero (the stable fixed point); when $R_0 > \gamma$ it increases monotonically. Plugging (1.6) into the attack-success curve yields the trajectory of pine mortality levels,

$$\phi\left(R_t; a, \kappa\right) = \frac{\left(\gamma\left(R_0/\gamma\right)^{\kappa^t}\right)^{\kappa}}{\left(\gamma\left(R_0/\gamma\right)^{\kappa^t}\right)^{\kappa} + a^{\kappa}} = \frac{\left(\left(R_0/\gamma\right)^{\kappa^t}\right)^{\kappa}}{\left(\left(R_0/\gamma\right)^{\kappa^t}\right)^{\kappa} + \left(a/\gamma\right)^{\kappa}} = \phi\left(\left(R_0/\gamma\right)^{\kappa^t}; a/\gamma, \kappa\right). \tag{1.7}$$

The trajectory of H_t for $t > 0$ follows by induction on (1.5) and the direct substitution of the formula in (1.7),

$$H_t = H_0\lambda^t\prod_{i=0}^{t-1}\left(1 - \phi\left(R_t; a, \kappa\right)\right) = \frac{\left(\beta/\gamma\right)^t H_0}{\prod_{i=1}^{t}\left(\left(a/\gamma\right)^{\kappa} + \left(R_0/\gamma\right)^{\kappa^i}\right)}. \tag{1.8}$$

The beetle density B_t is the product of (1.6) and (1.8), which for $t > 0$ simplifies to:

$$B_t = R_t H_t = \frac{(\beta/\gamma)^t \, (R_0/\gamma)^{\kappa^t - 1} \, B_0}{\prod_{i=1}^{t} \left((a/\gamma)^\kappa + (R_0/\gamma)^{\kappa^i} \right)}. \tag{1.9}$$

1.6.2 Asymptotic behaviour of MPB and pine populations in outbreaks

These solutions of (1.4)-(1.5) describe two possible outcomes for the host pine population when $\kappa > 1$, depending on whether initial attack density R_0 is above the threshold $\gamma = (\lambda a^\kappa/\beta)^{\frac{1}{\kappa-1}}$,

1. *host depletion* ($R_0 > \gamma$, or $R_0 = \gamma$ and $\lambda < 1 + (\gamma/a)^\kappa$)

   If attack density R_t initially exceeds γ it will increase monotonically, and the entire host population will be killed in the limit $t \to \infty$. This is straightforward from (1.6)-(1.8), since $R_0 > \gamma \implies R_t \to \infty \implies \phi(R_t; a, \kappa) \to 1 \implies H_t \to 0$. The MPB population vanishes with it, but an initial phase of increase ($B_{t+1} > B_t$) may precede the collapse (Figure 1.1)

   If $R_0 = \gamma$, attack density R_t is initialized at its nonzero fixed point, so we have constant host mortality $\phi(\gamma; a, \kappa)$ in each year. This decay in H_t is tempered by the geometric growth rate (λ) of the pine: Thus (1.8) shows that hosts H_t decline when λ is below a threshold:

   $$\lambda < 1 + (\gamma/a)^\kappa \implies \lambda\phi(a; \gamma, \kappa) < 1 \implies \lambda \left(1 - \phi(\gamma; a, \kappa)\right) < 1 \implies H_t \to 0 \tag{1.10}$$

   Thus both $H_t = H_0 \left(\lambda\phi(a; \gamma, \kappa)\right)^t$ and $B_t = \beta\phi(\gamma; a, \kappa) H_t$ decay (exponentially) to zero.

2. *host persistence* ($R_0 < \gamma$, or $R_0 = \gamma$ and $\lambda \geq 1 + (\gamma/a)^\kappa$)

   When $R_0 = \gamma$ and the forest growth rate λ lies at or above the threshold $1 + (\gamma/a)^\kappa$, the host population will either stay constant or increase. By the same chain of arguments as in (1.10) above, one verifies that if $\lambda = 1 + (\gamma/a)^\kappa$, then H_t remains fixed at

H_0 (and B_t at B_0), with regeneration exactly matching MPB damage each year; If $\lambda > 1 + (\gamma/a)^\kappa$ then regeneration exceeds mortality ($\lambda\phi(a;\gamma,\kappa) > 1$), and so both H_t and B_t grow exponentially.

In the case $R_0 < \gamma$, equations (1.6)-(1.7) show that both attack density and pine mortality decrease with time. MPB populations must eventually approach zero, since in (1.9) the denominator is positive and the numerator is dominated in the limit $t \longrightarrow \infty$ by the doubly exponential term $(R_0/\gamma)^{\kappa^t - 1}$. This however does not preclude an initial period of increase in the MPB population as in the example of Figure 1.1.

The host population always persists in this $R_0 < \gamma$ case (Appendix 1.6.3), though it may decline substantially in early years before $\phi(R_t; a, \kappa)$ becomes negligeably small.

1.6.3 Proof of host persistence

I have shown that host mortality declines monotonically when $R_0 < \gamma$. However it is perhaps still unclear whether the product in equation (1.8) in fact approaches zero (in which case $H_t \longrightarrow 0$), or is bounded below by a positive limit (implying host persistence). This question can be answered by taking logarithms and considering the convergence of the resulting infinite series.

Without loss of generality, assume $\lambda = 1$ and rewrite (1.8) on the log scale as

$$\log(H_t) = \log(H_0) - \sum_{n=0}^{t-1} A(n) \text{ where } A(n) := \log\left(1 + (\gamma/a)^\kappa \, (R_0/\gamma)^{\kappa^n}\right). \tag{1.11}$$

Note that $H_t \longrightarrow 0$ if and only if the infinite series $\sum A(n)$ diverges to $+\infty$. Note also that the $A(n)$ are strictly positive. Therefore by the ratio test, $\sum A(n)$ converges if $\lim_{n\to\infty} |A(n)/A(n-1)| < 1$. This limit has the indeterminate form $0/0$. However by viewing $A(n)$ as a differentiable function of $n \in \mathbb{R}$, and finding its derivative $A'(n) = (\gamma/a)^\kappa \log(R_0/\gamma) \log(\kappa) \kappa^n (R_0/\gamma)^{\kappa^n}$ to be nonzero, I can apply l'Hôpital's rule:

$$\lim_{n\to\infty} \left|\frac{A(n)}{A(n-1)}\right| = \lim_{n\to\infty} \left(\frac{A(n)}{A(n-1)}\right) = \lim_{n\to\infty} \left(\frac{A'(n)}{A'(n-1)}\right) = \lim_{n\to\infty} \kappa \, (R_0/\gamma)^{\kappa^n - \kappa^{n-1}} = 0.$$

Thus by the ratio test $\sum A(n)$ converges, and I conclude that H_t does not converge to 0.

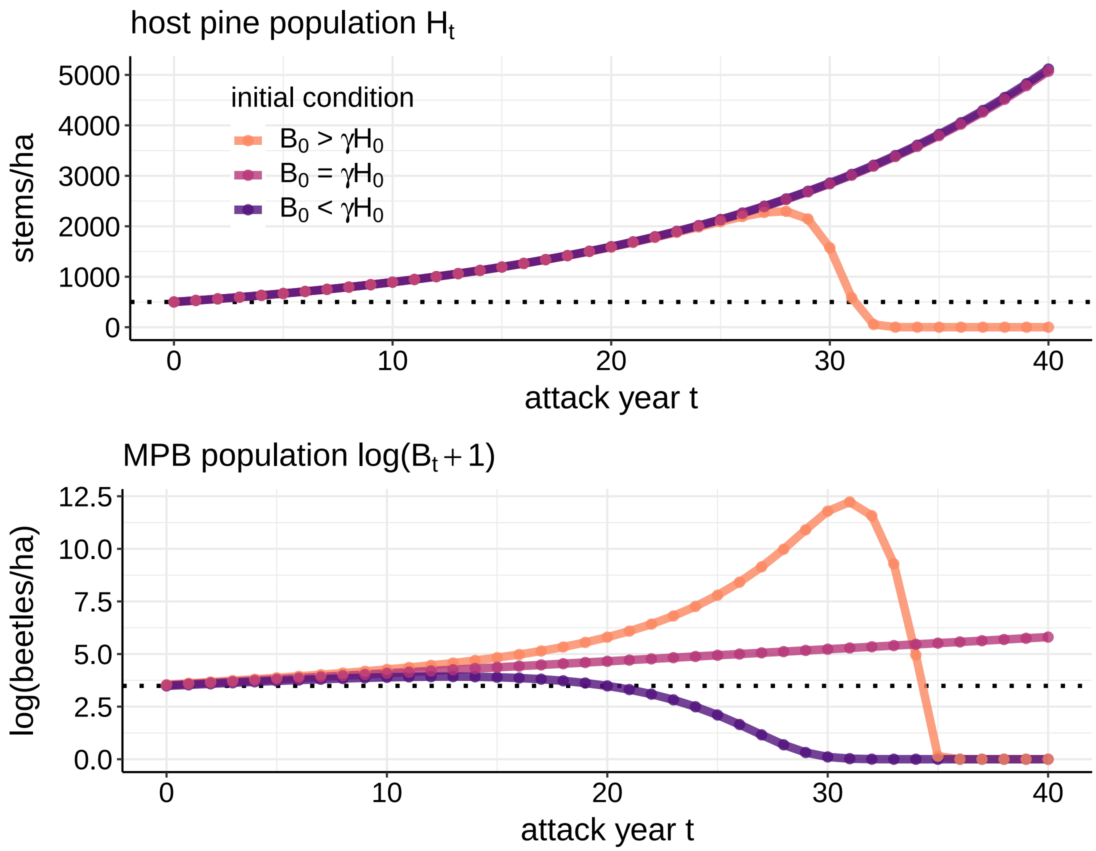

Figure 1.1: Solutions of the generalized red-top system (1.4)-(1.5) with parameters: $b = 200$; $a = 50$; $\kappa = 1.2$; $\lambda = 1.05$; and initial host density $H_0 = 500$ stems. The beetle population B_0 is initialized to three slightly different levels at $t = 0$ to illustrate the distinct qualitative behaviours discussed above: MPB eruption and host depletion ($B_0 = 33.7 > \gamma H_0$); coexistence with exponential growth ($R_0 = 32.7 = \gamma H_0$); and MPB collapse with host persistence ($R_0 = 31.7 < \gamma H_0$)

Chapter 2

Computationally Simple Anisotropic Lattice Covariograms

2.1 Introduction

In the analysis of large scale spatial ecological data, researchers frequently encounter special statistical challenges that preclude the use of more traditional models. Examples include incomplete data (Nakagawa and Freckleton, 2008); extremely large datasets (Simpson, Lindgren, and Rue, 2012); complex dependencies among model residuals (Legendre, 1993); an inability to replicate measurements; and complex underlying ecological systems that obscure the relationship between covariate and response (Hilborn and Mangel, 1997).

A large number of spatial regression methods have been developed to address these challenges, including generalized least squares (GLS), the autoregressive model family, generalized additive mixed models, as well as fully Bayesian approaches. In fact some authors (eg. Beale et al., 2010; Hawkins, 2012) suggest the sheer diversity of methods in the literature and resulting paralysis-by-analysis has been an impediment to the widespread adoption of spatial techniques by ecologists. A framework that is both immediately intuitive and easy to implement will therefore be useful to researchers outside the field of statistics.

We believe the geostatistical approach is appropriate for this role. Geostatistics, whose name reflects early origins in the mining industry, has over time grown into a quite general and mature spatial statistical framework (see *eg*. Cressie, 1993; Banerjee, Carlin, and

Gelfand, 2014). Its theory is built from regionalized random variables $Z(\boldsymbol{s})$ whose properties depend on a location index $\boldsymbol{s}$ that varies continuously through some spatial domain $\mathcal{D}$. Dependencies among the n sample points $Z(\boldsymbol{s}_i)$ are explicitly specified by the *covariogram* (also known as a covariance function or kernel), which maps coordinate pairs $\boldsymbol{s}_i$, $\boldsymbol{s}_j$ to the entries of the $n \times n$ covariance matrix $\boldsymbol{V}$. In applications, covariograms that decrease with separation distance can serve as convenient models for data exhibiting spatial autocorrelation (SAC).

While this transparent representation of covariance has intuitive appeal, a large n will quickly lead to implementation difficulties owing to the $O(n^3)$ complexity of important matrix computations involving $\boldsymbol{V}$ (Simpson, Lindgren, and Rue, 2012). If, however, the sample sites form a rectangular lattice, the structure of $\boldsymbol{V}$ can often be simplified, and these computational difficulties largely avoided. Such sampling designs are increasingly common in ecology today with the widespread adoption of remote sensing methods (Wulder et al., 2006).

In this paper we revisit the longstanding idea of simplifying $\boldsymbol{V}$ using Kronecker products and show how this leads to a novel family of anisotropic covariance models; as well as an estimator of the direction of range anisotropy in geometrically anisotropic data; and a graphical tool for studying nonstationary covariance structures. Though we emphasize ecological data and SAC, these models are quite general, and their computational simplicity makes them attractive in broader applications.

2.1.1 Why model spatial autocorrelation?

SAC describes when the random components of measurements that are near in space tend to be more (or less) similar than expected for a spatially distant pair. This phenomenon is extremely common in ecological studies, where the data-generating process is often driven by environmental factors shared among nearby sites, and/or demographic processes intrinsically tied to distance, such as aggregation or dispersal (Beale et al., 2010).

It is widely accepted that neglecting SAC (in favour of an independence assumption) amounts to pseudoreplication, and leads to precision issues for inference (Legendre, 1993). Though some authors dispute the importance of this misspecification, it is clear that SAC should be examined in the course of model development (Keitt et al., 2002). Examples from our research area include Klutsch *et al.* (2009), who used SAC as a model diagnostic in a study of environmental predictors for outbreaks of the mountain pine beetle, and Robertson *et al.* (2009), who improved their outbreak model with a simple SAC model.

The autocorrelation patterns themselves are sometimes of scientific interest. For example Aukema *et al.* (2008) used nonparametric covariance functions to study how spatial synchrony drives different stages of beetle outbreaks. SAC may illuminate features of the underlying ecological system that generates a dataset. We will touch on this idea in a case study of beetle outbreak data in Sect. 2.4.

2.1.2 Covariograms in practice

In explicit covariance models, all n^2 entries of $\boldsymbol{V}$ must be specified, of which up to $n(n+1)/2$ can be distinct (by symmetry). Since these values are typically unknown, $\boldsymbol{V}$ is constructed using a parametric covariogram $c : (\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}) \mapsto \text{Cov}\left\{Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)\right\}$, and the parameters $\boldsymbol{\theta}$ inferred from the data. Many functional forms have been proposed for c, but some care is required to ensure a well defined distribution (Guillot et al., 2014). In particular c (and $\boldsymbol{V}$) must be symmetric and positive definite (SPD). The reader is directed to Roberts et al. (2013) for a discussion of design principles and admissible forms for c.

In spatial ecology, modellers often use covariograms that are a functions of separation distance $d_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ only, ignoring directionality and position. This reflects the assumption that the $Z(\boldsymbol{s}_i)$ are drawn from a second-order stationary (SOS) and isotropic random field (Myers, 1989); that is, the covariance structure is invariant to translations and/or rotations of $\mathcal{D}$. A popular example is the Whittle-Mátern (WM) covariogram which, in 2-dimensional (2D) space, can be written:

$$c\left(d_{ij};\sigma,\lambda,\nu\right)=\sigma^{2}\left(2^{1-\nu}/\Gamma(\nu)\right)\left(2\sqrt{\nu}d_{ij}/\lambda\right)^{\nu}H_{\nu}\left(2\sqrt{\nu}d_{ij}/\lambda\right), \tag{2.1}$$

where H_ν is the modified Bessel function of the second kind. Figure 2.1 (top-left) illustrates the type of spatial patterns generated by the WM. This covariogram has many names, and many desireable properties, as chronicled in Guttorp and Gneiting (2006). Given its ubiquity and importance in statistics, we will make use of the WM later on as a reference model for generating data in our simulation studies.

Isotropy is however seldom justified, except as a means to a parsimonious model. A more robust model should allow directionality, and one of the simplest ways of building this into a covariogram is to assume *geometric anisotropy*. This extends the isotropic SOS covariogram by applying an affine transformation $\boldsymbol{A}$ to the coordinate system, and measuring distances by $d_{ij} := \|\boldsymbol{A}(\boldsymbol{s}_i - \boldsymbol{s}_j)\|$. In 2D, $\boldsymbol{A}$ can be understood as the product of a diagonal scaling matrix $\boldsymbol{S}(s^x, s^y)$, and a rotation matrix $\boldsymbol{R}_{-\alpha}$: Circular contours of constant covariance get mapped to ellipses whose axes are stretched by factors s^x, s^y, and which are oriented along the counterclockwise rotation of the x, y axes by angle α. Figure 2.1 (bottom-left) illustrates the resulting pattern of SAC.

2.1.3 Computations with covariance matrices

Among the simplest implementations of the covariogram in regression is GLS. This extends ordinary least squares (OLS) on the residuals vector $\boldsymbol{Z}$, when $\boldsymbol{V}$ is given. The model is:

$$\boldsymbol{Y}=\boldsymbol{\beta}\boldsymbol{X}+\boldsymbol{Z} \quad \text{where} \quad \mathbb{E}\left(\boldsymbol{Z}\mid\boldsymbol{X}\right)=\boldsymbol{0} \quad \operatorname{Cov}\left(\boldsymbol{Z}\mid\boldsymbol{X},\boldsymbol{\theta}\right)=\boldsymbol{V}(\boldsymbol{\theta}) \tag{2.2}$$

where $\boldsymbol{X}$ is the data matrix of covariates.

GLS uses the Cholesky factor $\boldsymbol{L}$ of $\boldsymbol{V}$ to define a transformed response $\tilde{\boldsymbol{Y}} = \boldsymbol{L}^{-1}\boldsymbol{Y}$ for which the problem of estimating $\boldsymbol{\beta}$ reduces to OLS. However since $\boldsymbol{V}$ is seldom known, a parametric form $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta})$ is often proposed, and $\boldsymbol{\theta}$ estimated by numerical likelihood

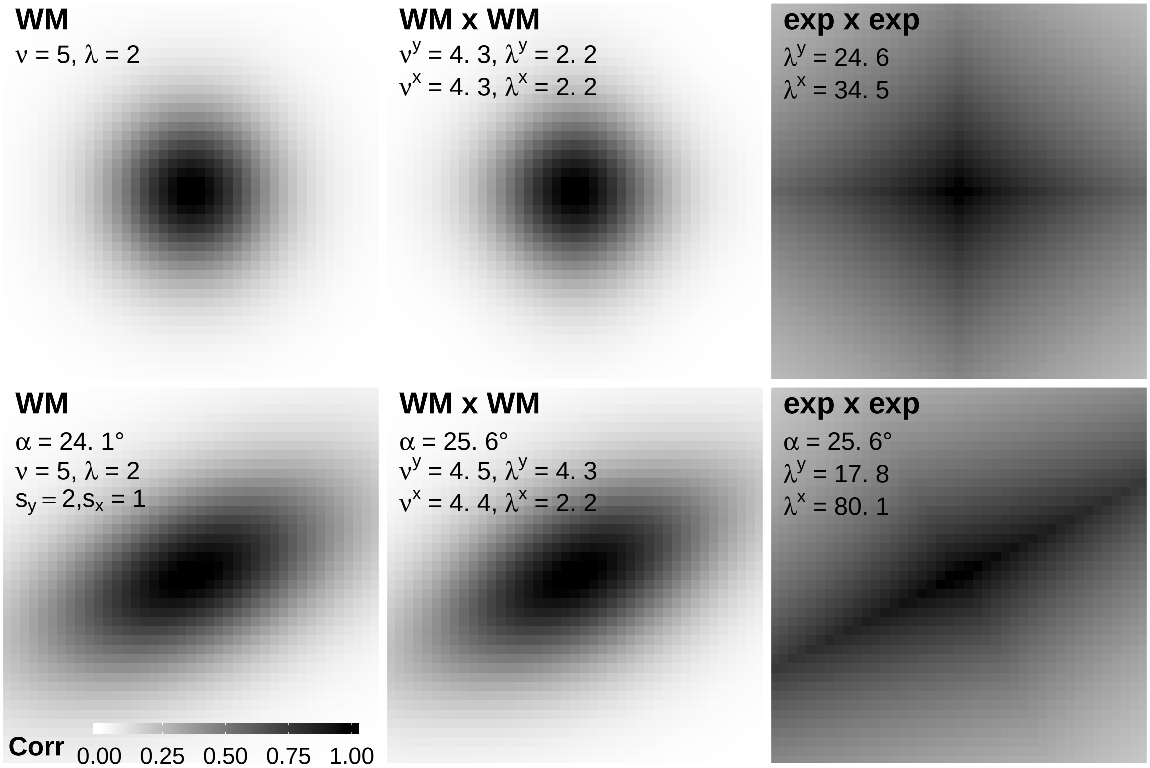

Figure 2.1: Examples of covariograms in 2D illustrated by heatmaps of the correlation with the central point. The two leftmost panels illustrate isotropy (top) and geometric anisotropy (bottom). The others illustrate separable (top) and rotated product (bottom) covariograms, as introduced in Sect. 2.2.3-2.3. The two rightmost panels are examples of separable kernels fitted in simulations in Sect. 2.2.4, 2.3.3

maximization. For example when Z is multivariate normal (MVN), we minimize:

$$-\log \mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\beta} \mid \boldsymbol{X}, \boldsymbol{Y} = \boldsymbol{y}\right) \propto \log \left|\boldsymbol{V}(\boldsymbol{\theta})\right| + \left(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{X}\right)^{*} \boldsymbol{V}(\boldsymbol{\theta})^{-1} \left(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{X}\right). \qquad (2.3)$$

Note that here (and throughout the paper) we use an asterix to denote transposes.

When n is large, evaluations of (2.3) can be computationally demanding because $\boldsymbol{V}(\boldsymbol{\theta})$ is usually dense, and analytic forms for $\boldsymbol{V}(\boldsymbol{\theta})^{-1}$ and $|\boldsymbol{V}(\boldsymbol{\theta})|$ are rarely available. Thus each time the optimizer adjusts $\boldsymbol{\theta}$, it must solve a factorization problem with arithmetic complexity $O(n^3)$. This problem arises in universal kriging interpolation methods (Simpson, Lindgren, and Rue, 2012), as well as in more sophisticated extensions of GLS such as spatial generalized estimating equations (Dormann et al., 2007); spatial generalized linear mixed models (Heagerty and Lele, 1998); and Bayesian MCMC based techniques (Banerjee, Carlin, and Gelfand, 2014).

The autoregressive model family (CAR, SAR) avoids this problem by defining $\boldsymbol{V}$ implicitly. A weights matrix is used to specify partial correlations rather than covariances, thus constructing the precision matrix $\boldsymbol{V}^{-1}$ directly. Ver Hoef et al. (2018) and Beale et al. (2010) make compelling cases for the autoregressive approach. There are drawbacks, however. The implied correlation structure in $\boldsymbol{V}$ is often unclear (Wall, 2004); and some unintuitive restrictions on the weights are required to ensure valid joint distribution for $\boldsymbol{Z}$.

Authors preferring the more intuitive covariogram approach have developed various structured forms for $\boldsymbol{V}$ that avoid the large-n difficulties (Banerjee, Carlin, and Gelfand, 2014, chap. 12). These include kernel convolution and stochastic partial differential equation based methods (Simpson, Lindgren, and Rue, 2012), as well as low-rank or sparse approximations of $\boldsymbol{V}$ (Ambikasaran et al., 2016). Following Genton (2007) we will be interested in the highly patterned forms of $\boldsymbol{V}$ generated by a *separable* covariance model.

2.1.4 Chapter outline

We will begin by reviewing a well-known result for SOS spatial processes on rectangular grids in Sect. 2.2, showing this implies the useful property of bisymmetry in $\boldsymbol{V}$. We then discuss separable covariance structures, demonstrating their flexibility in a simulation study. In Sect. 2.3 we propose a broader class of covariograms incorporating range anisotropy, and use them to develop a novel estimator of the angle of geometric anisotropy. Sect. 2.4 demonstrates the method on a spatial ecological dataset.

2.2 Covariograms on the rectangular lattice

A random field is called SOS (or weakly stationary) when its covariogram has the form $c(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\theta})$, and the data are detrended ($\mathbb{E}\{Z(\boldsymbol{s}_i)\} = \mu$), for all $\boldsymbol{s}_i, \boldsymbol{s}_j$ (Cressie, 1993, sec. 2.3). We focus on the application of this model to 2D lattice data, where the random field has been sampled at a fixed set of locations $\mathcal{G} \subset \mathbb{R}^2$ that together form a spatially regular $n^y \times n^x$ rectangular grid (with sample size $n = n^x n^y$).

For convenience assume that $\boldsymbol{Z} := \{Z(\boldsymbol{s}_1), \dots Z(\boldsymbol{s}_n)\}^*$ is indexed in column-vectorized order, so that $\text{Vec}_{n^y}^{-1}(\boldsymbol{Z})$ recovers the natural $n^y \times n^x$ matrix representation for the data (*ie.* as a raster image). This ordering introduces computationally useful patterns in $\boldsymbol{V}$.

2.2.1 Symmetry structures in covariance matrices for lattice data

In an important early paper, Zimmerman (1989) showed that for SOS models on $\mathcal{G}$, the covariance matrix $\boldsymbol{V}$ is block-Toeplitz with Toeplitz blocks (BTTB). He concluded with an algorithm that reduces the arithmetic complexity of solving $\boldsymbol{V}^{-1}$ by a factor of n^x. Dietrich (1993) subsequently showed that the Cholesky factor of $\boldsymbol{V}$ can be computed for the same complexity cost, providing a shortcut for the determinant and quadratic form in (2.3).

Unfortunately these algorithms can become inaccurate when $\boldsymbol{V}$ is numerically singular (Golub and Van Loan, 2012, sec. 2.7). This problem is not uncommon with large-n covariance models. For example the popular Gaussian covariogram (the $\nu \longrightarrow \infty$ limit of the WM) is known for generating matrices with extremely large condition numbers. It is prudent in this situation to avoid explicit inversions and Cholesky factorizations in favour of more robust and numerically stable methods. A common workaround involves the singular value decomposition (SVD) (Neumaier, 1998), however this has $O(n^3)$ complexity.

We can nevertheless speed computations by appealing to *bisymmetry* in $\boldsymbol{V}$. This is symmetry about both the diagonal and counterdiagonal. More precisely, if $\boldsymbol{J}_n$ is the $n \times n$ *exchange matrix* (with ones on the counterdiagonal and zeros otherwise), then $\boldsymbol{V}$ is bisymmetric if $\boldsymbol{V} = \boldsymbol{J}_n \boldsymbol{V} \boldsymbol{J}_n$ and $\boldsymbol{V} = \boldsymbol{V}^*$. The SOS assumption implies bisymmetry in $\boldsymbol{V}$, since the BTTB property implies *persymmetry* (symmetry about the counterdiagonal).

For bisymmetric covariance matrices $\boldsymbol{V}$, a unitary similarity transformation cuts the dimensionality of factorizations in half, speeding up computations by a factor of four. We present the even-n case here, and refer readers to Abu-Jeib (2002) for the odd case. Suppose that $n = 2m$, and $\boldsymbol{V}$ is $n \times n$ and bisymmetric. Writing $\boldsymbol{V}_a$ for the $(m \times m)$ top-left block and $\boldsymbol{V}_b$ for the bottom-left block of $\boldsymbol{V}$, we have the block-diagonalization:

$$\boldsymbol{H}_n \boldsymbol{V} \boldsymbol{H}_n^* = \begin{pmatrix} \boldsymbol{V}_a + \boldsymbol{J}_m \boldsymbol{V}_b & \\ & \boldsymbol{V}_a - \boldsymbol{J}_m \boldsymbol{V}_b^* \end{pmatrix}, \quad \text{where } \boldsymbol{H}_n = \frac{1}{\sqrt{2}} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{J}_m \\ -\boldsymbol{J}_m & \boldsymbol{I} \end{pmatrix}. \tag{2.4}$$

The diagonal blocks of $\boldsymbol{H}_n \boldsymbol{V} \boldsymbol{H}_n^*$ have two important properties: They are symmetric, as a consequence of persymmetry in $\boldsymbol{V}_b$ and symmetry in $\boldsymbol{V}_a$; and the union of their spectra is the spectrum of $\boldsymbol{V}$, since $\boldsymbol{H}_n$ is orthogonal and $\boldsymbol{H}_n \boldsymbol{V} \boldsymbol{H}_n^*$ block-diagonal. This implies that both blocks inherit positive definiteness from $\boldsymbol{V}$, and hence their Cholesky factors exist. It also implies that their condition numbers are as good or better than that of $\boldsymbol{V}$.

Equation (2.4) therefore stably converts an n-dimensional factorization problem in $\boldsymbol{V}$ into a pair of $n/2$-dimensional ones in $\boldsymbol{H}_n \boldsymbol{V} \boldsymbol{H}_n^*$. In statistical applications, the $O(n^3)$ complexity of the relevant factorization (*eg.* SVD, spectral, Cholesky) is therefore reduced by a factor of $2^3/2 = 4$. Moreover since $\boldsymbol{H}_n$ has only two nonzero elements per row/column, the overhead of performing this transformation is $O(n^2)$, a negligible cost compared to the factorization step when n is large.

2.2.2 A geometrical perspective on bisymmetry

The action of $\boldsymbol{H}_n$ on vectorized lattice data from Gaussian random fields leads to an interesting theoretical aside. Since $\boldsymbol{V}_a + \boldsymbol{J}_m \boldsymbol{V}_b$ and $\boldsymbol{V}_a - \boldsymbol{J}_m \boldsymbol{V}_b^*$ are SPD, we can view them as covariance matrices for a pair of m-dimensional random vectors, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$. Indeed by standard MVN theory, if $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{V})$ then $\boldsymbol{H}_n\boldsymbol{Z} := (\boldsymbol{Z}_1^*, \boldsymbol{Z}_2^*)^* \sim \mathcal{N}\left(\boldsymbol{H}_n\boldsymbol{\mu}, \boldsymbol{H}_n\boldsymbol{V}\boldsymbol{H}_n^*\right)$.

When vector $\boldsymbol{Z}$ represents a raster image, the transformed components $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ have a simple geometrical interpretation. Let $\boldsymbol{Z}^F$ denote the transformation of the raster $\boldsymbol{Z}$ by a left-right, and an up-down reflection. Then $\boldsymbol{Z}_1$ is simply the left half of the superimposition $\boldsymbol{Z}^F + \boldsymbol{Z}$. Similarly $\boldsymbol{Z}_2$ is the right half of the superimposition $\boldsymbol{Z}^F - \boldsymbol{Z}$.

Despite a striking symmetry between the left and right halves of $\boldsymbol{H}_n\boldsymbol{Z}$, they are statistically independent under the Gaussian model, since by (2.4) their cross-covariance matrices are zero. While Gaussian data can always be linearly transformed into independent subsets using eigenvectors of $\boldsymbol{V}$, (2.4) shows that for SOS lattice data, half of these eigenvectors are symmetric and the other half are skew-symmetric (Abu-Jeib, 2002).

2.2.3 Separable SOS covariograms

We have seen that a simple transformation $\boldsymbol{H}_n$ speeds computations for the general SOS model by a factor of four. If the covariance is separable, the speedup can be made closer to 16x by using transformation, $\boldsymbol{H}_{n^x \times n^y} := \boldsymbol{H}_{n^x} \otimes \boldsymbol{H}_{n^y}$ (where $\otimes$ denotes the Kronecker product). This partitions $\mathcal{G}$ into four subsets, each constructed through superimpositions of $\pm\boldsymbol{Z}$ and $\pm\boldsymbol{Z}^F$. Using $\boldsymbol{H}_{n^x \times n^y}$ one finds that $\boldsymbol{V}$ is similar to a Kronecker product of block-diagonal matrices, and this leads easily to a block-diagonal form having *four* SPD blocks whose condition numbers are as good or better than that of $\boldsymbol{V}$.

Nevertheless, with large enough n, constant factor improvements like these are of little consequence (numerical stability notwithstanding) and even the $O(n^x)$ improvements of Dietrich (1993) may be inadequate to make computations feasible. In that case we suggest that modellers consider an *a priori* assumption of separability in order to exploit computational efficiency in the well-known algebra of Kronecker products (Van Loan, 2000). In Sect. 2.2.4 we provide some justification for the robustness of these models.

The idea of separability is to disentangle the x and y component distances by applying

1D covariograms (say c^x and c^y) to each component separately, before taking their product. Thus if $\boldsymbol{s}_i = (x_i, y_i)$, a SOS separable spatial covariogram can be written:

$$c(\boldsymbol{s}_i - \boldsymbol{s}_j) = \sigma^2 c^x \left(x_i - x_j\right) c^y \left(y_i - y_j\right), \tag{2.5}$$

where the marginal variance parameters from c^x and c^y have been combined into σ^2 (as they are not separately identifiable). The resulting covariance matrix decomposes into a Kronecker product $\boldsymbol{V} = \boldsymbol{V^x} \otimes \boldsymbol{V^y}$. Whereas the $(n \times n)$ matrix $\boldsymbol{V}$ has one row per sample site, matrices $\boldsymbol{V}^x$ and $\boldsymbol{V}^y$ ($n^x \times n^x$ and $n^y \times n^y$, respectively) have only one row per grid line. $\boldsymbol{V}^x$ and $\boldsymbol{V}^y$ are themselves covariance matrices, for a pair of 1D processes with covariograms σc^x and σc^y. Indeed this is how separable covariance was originally formalized by Martin (1979), though he presented it in the framework of autoregression.

Martin (1979) recognized a number of desireable computational properties in (2.5), and we will mention some of them before moving to less familiar results. In brief, most of the matrix algebraic computations on $\boldsymbol{V}$ that arise in spatial inference and prediction can be applied instead to the lower-dimensional components $\boldsymbol{V}^x$ and $\boldsymbol{V}^y$. This includes the inverse and determinant; as well as matrix-vector multiplications; and the SVD, Cholesky, and spectral decompositions. For example the negative log-likelihood for an observation of $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{V})$ is proportional to:

$$n^y \log|\boldsymbol{V}^x| + n^x \log|\boldsymbol{V}^y| + (\boldsymbol{Z} - \boldsymbol{\mu})^* \text{vec}\left\{(\boldsymbol{V}^y)^{-1} \text{vec}_{n^y}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})(\boldsymbol{V}^x)^{-1}\right\}. \tag{2.6}$$

Comparing with (2.3), this reduces the arithmetic complexity from $O\left\{(n^x n^y)^3\right\}$ to $O\left\{(n^x)^3 + (n^y)^3\right\}$. Computer memory requirements are also reduced; Only the components $\boldsymbol{V}^x$ and $\boldsymbol{V}^y$ must be stored in memory, and never the full covariance matrix $\boldsymbol{V}$. Moreover since $\boldsymbol{V}^x$ and $\boldsymbol{V}^y$ are bisymmetric, we can use (2.4) to further speed computations by 4X.

Formula (2.6) is often exploited in analyses of spatio-temporal datasets, where a Kronecker product of spatial and temporal covariance matrices is commonly viewed as the simplest baseline model (Genton, 2007). It is also well-established in pattern recognition applications of machine learning (Wilson et al., 2014). We are however aware of very few examples in the applied statistics literature of *spatially* separable covariograms.

Statisticians may prefer isotropic models like the WM (or its geometric anisotropy extension) for reasons of parsimony. However the more computationally attractive separable covariogram seems to mimic these standard models quite well, as we demonstrate next in a simulation study. In the appendix (Online Resource 1), we address computational aspects of marginal distributions, since these are characterized by submatrices of $\boldsymbol{V}$ that lack separability.

2.2.4 Simulation study comparing separable and isotropic covariograms

Given the scarcity of empirical results on the performance of separable spatial covariograms, we sought to evaluate their flexibility using simulations. We compared the root mean-squared prediction error (RMSPE) from separable covariograms against two standard isotropic ones in simulations of a (40×40) lattice of SAC data from a linear model (2.2).

For each of 300 replicates, we generated spatial error terms using two covariance models: the (isotropic) WM ($\nu = 5$, $\lambda = 2$; Figure 2.1, top-left), and the separable (and highly anisotropic) product of two 1D WMs ($\nu^x = 5$, $\lambda^x = 1$, $\nu^y = 5$, $\lambda^y = 3$). We then fit the model (2.2) by maximum likelihood for five different covariogram types, including the correct one, and compared precision and predictive ability.

For each replicate we used a design matrix $\boldsymbol{X}$ with four covariates: two independent standard normal variates (uncorrelated); and two Gaussians drawn from the isotropic WM with strong SAC ($\nu = 6$, $\lambda = 4$). We drew regression coefficients $\boldsymbol{\beta}$ uniformly at random from $(-1, 1)$ in each replicate, computed the linear predictor $\boldsymbol{X\beta}$, then added spatial error Z to form the dataset $\{\boldsymbol{y}, \boldsymbol{X}\}$. We then fitted $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}$ by maximum likelihood under the following covariance models: OLS (independence); exponential; WM; a separable product of 1D exponentials; and a separable product of 1D WMs.

To estimate RMSPE in each replicate we randomly generated a new dataset from the

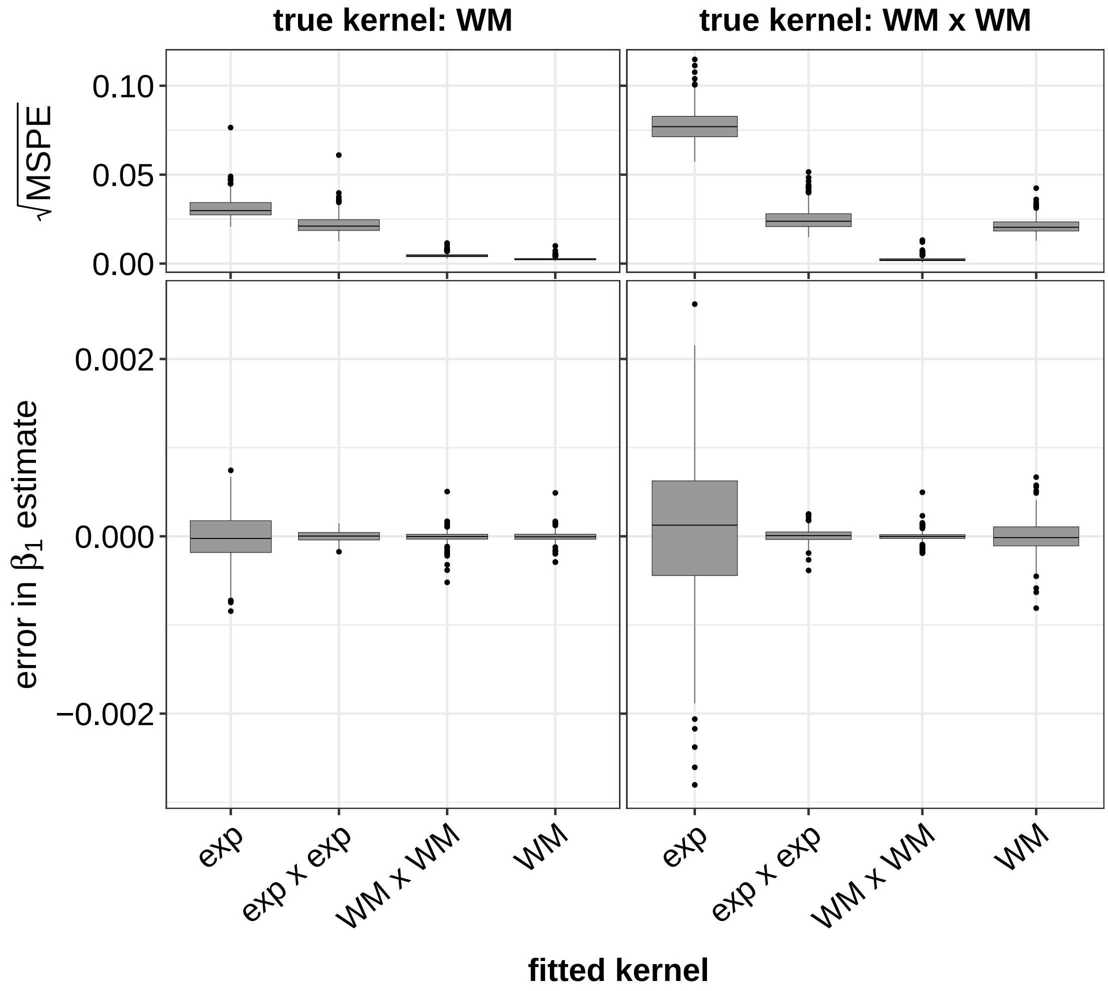

Figure 2.2: Inference and prediction on simulated data ($n = 40 \times 40$) from a linear model with covariance given by: the WM (left); and a separable product of 1D WMs (right). In each of 300 replicates, we fit MLEs for 5 types of covariogram. We plot errors in one of the estimated regression parameters (bottom); and RMSPEs in predictions on a second simulated dataset (top)

same distribution, selected half of the points to condition over (uniformly at random), and predicted on the other half. These results, and the errors in parameter inference, are summarized in Figure 2.2. For brevity we plot only the regression parameter estimate of the first covariate, X_1, and omit the OLS errors. Unsurprisingly these OLS errors were extremely large by comparison, making it difficult to discern differences among the boxplots for the other four models. Results for the autocorrelated covariates showed little difference among the five test models.

Although the exponenential is nested in the WM family ($\nu = 0.5$), it did relatively poorly in nearly all of our simulations, outperforming only the naive OLS estimates. Nevertheless, the separable product of exponentials produced surprisingly robust MLEs for both types of spatial error. Indeed, when the data were drawn from the isotropic WM covariogram, both of the separable models were about as precise in their estimates of β_1 as the correct one (Figure 2.2, bottom-left). Meanwhile, in datasets with separable spatial error (Figure 2.2, bottom-right), both isotropic models performed noticeably worse than the separable ones.

The RMSPE results were also favourable to separable covariograms. In datasets from the isotropic WM (Figure 2.2, top-left), the separable product of 1D WMs did nearly as well as the correct model. The separable product of 1D exponentials performed worse than the correct model, but better than the isotropic exponential. This was surprising given the markedly anisotropic nature of the fitted covariograms from this family (Figure 2.1, top-right).

These results suggests a remarkable flexibility in separable products of the 1D WM. In our simulations they did well to approximate their isotropic counterparts, but the converse was not true. Moreover, the likelihood maximization problem was far simpler with separability. The reduction in computational complexity is borne out even on this relatively small dataset, where separable models fit around 400X faster than nonseparable ones.

2.3 Product anisotropic covariograms

In Sect. 2.1.2 we saw that isotropic covariograms can be generalized to incorporate directionality by a modification of the coordinate system, called geometric anisotropy. What happens if we do the same with separable covariograms?

Separable covariograms are already equipped to handle some degree of range anisotropy, since c^x and c^y can be assigned different range parameters. However, the directionality is constrained to align with the coordinate system, making this approach relatively inflexible. This motivates an extension that we will call the *product anisotropic covariogram* (PAC), in which we compose a separable covariogram with an affine transformation ($\boldsymbol{A}$) of coordinates. Using the notation of Sect. 2.1.2, we write the general 2D spatial PAC as a function of the transformed coordinate differences:

$$c(\boldsymbol{s}_i - \boldsymbol{s}_j) = \sigma^2 c^x \left(\tilde{x}_i - \tilde{x}_j\right) c^y \left(\tilde{y}_i - \tilde{y}_j\right) \text{ where } (\tilde{x}_i - \tilde{x}_j, \tilde{y}_i - \tilde{y}_j)^* = \boldsymbol{A}(\boldsymbol{s}_i - \boldsymbol{s}_j). \quad (2.7)$$

Some examples are illustrated in Figure 2.1 (middle and righthand panes). Just as geometric anisotropy assumes that $\boldsymbol{A}^{-1}$ leads to an isotropic process, the product form covariogram supposes it leads to a separable process. Taking the simulation results from Sect. 2.2.4 as an indication that the (unrotated) exponential PAC adequately approximates the WM, it follows that its rotated analogue should adequately approximate geometric anisotropy. We propose that the former can serve as a computationally efficient surrogate in situations where the latter is a reasonable model.

This approach allows more flexibility in covariance structure than does the classical approach of geometric and/or zonal anisotropy. For example, c^x and c^y need not be from the same covariogram family; the contours of constant covariance are not restricted to ellipses; and sill anisotropy is naturally accommodated in addition to range anisotropy. These are highly desirable features since, as discussed in Zimmerman (1993), subtly different types of anisotropy can have important consequences for predictions.

Recalling Sect. 2.2.1, the covariance matrix defined by (2.7) is BTTB (and bisymmetric). However, it will usually not be separable. Thus in generalizing the separable covariogram we seem to have lost its main selling point of computational efficiency. However, for Gaussian data it turns out that in some cases we can partition the data into subsets whose marginal covariance *is* separable, as we show next.

2.3.1 Special transformations of the coordinate system

The key insight here is to consider transformations $\boldsymbol{A}$ for which α, s^x, and s^y satisfy:

$$\alpha = \arctan\left(\alpha^y/\alpha^x\right), \qquad \text{where } \alpha^x, \alpha^y \in \mathbb{Z}^+, \text{ and } s^x = s^y = \|(\alpha^x, \alpha^y)\|. \qquad (2.8)$$

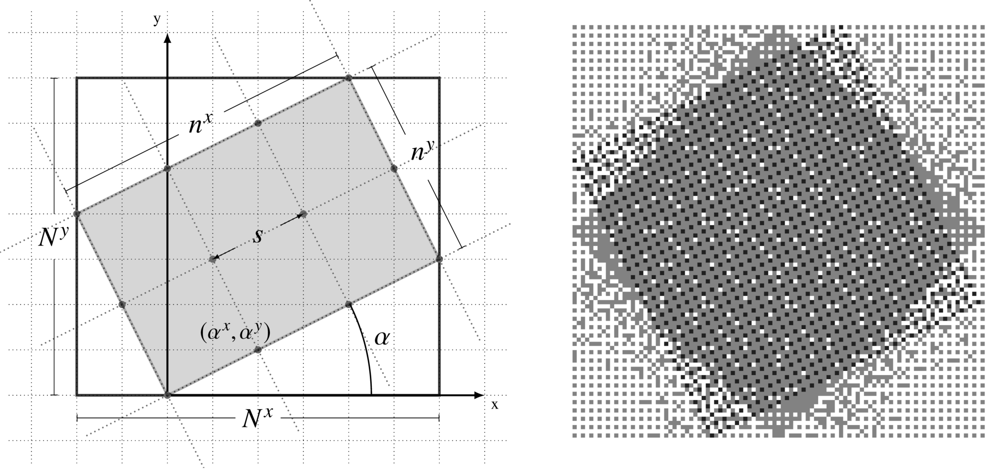

Figure 2.3: Special transformations of lattice coordinates. The left diagram illustrates the geometry of a rotation by $\alpha = \arctan(1/2)$ and a scaling of $s = \sqrt{5}$. The shaded rectangle is the rotated $n^y \times n^x$ subgrid $\mathcal{G}_{\boldsymbol{\kappa}}$, which lies on the (unrotated) $N^y \times N^x$ grid $\mathcal{G}$. The right panel shows $\mathcal{G}_{\boldsymbol{\kappa}}$ (black cells) on a much larger domain. The gray and black cells together make up the full set of subgrids corresponding to the 12 special angles in (2.13). Our proposed range anisotropy estimator uses these shaded cells to fit covariograms to each of the 12 angles, and their complement (white cells) to estimate error.

The scaling $s = \|(\alpha^x, \alpha^y)\|$ ensures that the entries of $\boldsymbol{A}$ are integers, so that the transformed coordinates lie in the square lattice $\mathbb{Z}^2$. Thus if the original coordinate system $\mathcal{G}$ is an $N^y \times N^x$ regular rectangular grid then we can always choose a subset of locations $\mathcal{G}_{\boldsymbol{\kappa}} \subset \mathcal{G}$ that forms a $n^y \times n^x$ regular rectangular grid with respect to the *transformed* coordinates, provided N^x and N^y are large enough. Specifically, it is necessary that

$$N^x - 1 \geq \alpha^x(n^x - 1) + \alpha^y(n^y - 1) \quad \text{and} \quad N^y - 1 \geq \alpha^y(n^x - 1) + \alpha^x(n^y - 1). \tag{2.9}$$

For example Figure 2.3 (left) shows how $\alpha = \arctan(1/2)$ produces a 3×4 rotated inner subset within an 8×9 outer grid. It is fairly straightforward (though tedious) to find the indexing vector $\boldsymbol{\kappa}$ that pulls this inner subset from $\mathcal{G}$ in column-vectorized order, so we state the general case here without proof and refer the reader to the appendix (Online Resource 1) for a derivation. Let $\boldsymbol{V}$ be the covariance matrix for MVN vector $\boldsymbol{Z}$, as in (2.7), with $\boldsymbol{A}$ defined as in (2.8). If the subgrid dimensions $n^y \times n^x$ satisfy (2.9), we define the indexing vector $\boldsymbol{\kappa}$ by:

$$\left[\mathrm{vec}_{n^y}^{-1}(\boldsymbol{\kappa})\right]_{ij} = 1 + \alpha^y(n^x - 1) + (N^y\alpha^y + \alpha^x)(i-1) + (N^y\alpha^x - \alpha^y)(j-1). \qquad (2.10)$$

Referring to Figure 2.3 (left), if the unshaded rectangle is $\mathcal{D}$, then the indexing $\boldsymbol{\kappa}$ selects the inner rectangular lattice $\mathcal{G}_{\boldsymbol{\kappa}}$ of grey points. By construction, $\boldsymbol{\kappa}$ selects these points in column-vectorized order. Their covariance matrix is therefore a Kronecker product (similar to that of $\mathcal{G}$ except with the spacing of grid lines increased by a factor of s).

Notice that through translations $\boldsymbol{\tau}_k$ following the rotation and scaling, the entire integer lattice $\mathbb{Z}^2$ can be partitioned into s^2 disjoint subsets of the form $\boldsymbol{A}^T\boldsymbol{s}_i + \boldsymbol{\tau}_k$, a consequence of Pick's theorem (Pick, 1899). From each subset, one may select a rectangular grid $\mathcal{G}_{\boldsymbol{\kappa}^{(k)}}$ of sample sites that lies in $\mathcal{G}$, by a suitable choice of the indexing vector $\boldsymbol{\kappa}^{(k)}$. Thus up to s^2 disjoint subsets $\mathcal{G}_{\boldsymbol{\kappa}^{(k)}} \subset \mathcal{G}$ are available, each with the same separable covariance matrix $\boldsymbol{V}_{\boldsymbol{\kappa}}$.

2.3.2 Applications of product anisotropic covariograms

An immediate application for the ideas of Sect. 2.3.1 is in analyses where a MVN with geometric anisotropy is a suitable model, and the angle α is known or has been estimated (*eg.* from directional semivariograms). If N is large enough that computational complexity becomes an issue, we suggest using a PAC with a nearby special angle of the form (2.8). Separability can then be exploited over the subsets $\mathcal{G}_{\boldsymbol{\kappa}^{(k)}}$.

Of course when using one of the $\mathcal{G}_{\boldsymbol{\kappa}^{(k)}}$ and discarding points from $\mathcal{G} \setminus \mathcal{G}_{\boldsymbol{\kappa}^{(k)}}$, we lose efficiency. However much of this efficiency can be recovered by using a composite marginal likelihood function $\mathcal{L}_C$ (Lindsay, 1988) that combines information from all s^2 subsets:

$$\mathcal{L}_C\left(\boldsymbol{\mu}, \boldsymbol{\theta} \mid \left\{Z_i; \boldsymbol{s}_i \in \cup_{k=1}^{s^2}\mathcal{G}_{\boldsymbol{\kappa}}^{(k)}\right\}\right) = \prod_{k=1}^{s^2} \mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\theta} \mid \boldsymbol{Z}[\boldsymbol{\kappa}^{(k)}]\right). \qquad (2.11)$$

This can be evaluated using (2.6), with the factorization of $\boldsymbol{V}_{\boldsymbol{\kappa}}$ reused for all s^2 terms.

Typically the direction of range anisotropy α will be unknown and so must be estimated from the data. For a second application we propose a simple cross-validation-like method for constructing such an estimator. The idea is to assemble a set of test angles α_j ($j = 1, \ldots m$), each of the form (2.8), that define a suite of candidate PAC models. We then fit each of these candidates to its corresponding rotated subgrid $\mathcal{G}_{\kappa}(\alpha_j)$ (defined precisely in Appendix 2.6.5, equation 2.28). The model-fitting points $\cup_{j=1}^{m} \mathcal{G}_{\kappa}(\alpha_j)$ do not cover all of $\mathcal{G}$, so from the unused portion we can select at random a test set, $\mathcal{G}_{\text{pred}}$ to predict over, conditional on the remaining data. Figure 2.3 (right) illustrates this partition of $\mathcal{G}$ into test and training data.

Having estimated the prediction errors for each α_j, a preferred angle can be chosen by lowest RMSPE. Alternatively one can compute the circular mean of the α_j's, inversely weighted by their RMSPE; Letting ω_j denote the j^{th} weight, the circular mean is:

$$\hat{\alpha_{\omega}} = (1/2) \arg \left\{ \sum_{j=1}^{m} \omega_j \left(\cos(2\alpha_j) + i \sin(2\alpha_j) \right) \right\}. \tag{2.12}$$

This simply maps each angle to a vector on the unit circle, scaling lengths according to RMSPE, before finding the angle of the resulting vector sum. Note that because covariance functions are symmetric, we cannot distinguish between α and $\alpha \pm \pi$. Hence we double each $\alpha_j \in [0, \pi)$ before mapping it to the unit circle in equation (2.12), dividing the final result by 2 to return to $[0, \pi)$.

The more conventional method of investigating the angle of range anisotropy involves studying empirical directional (co)variograms for ad-hoc sets of angles and spatial lags (Sherman, 2011). This approach is both computationally fast and intuitive, and remains an important part of model selection. However, as an informal graphical diagnostic it suffers from issues of subjective interpretation (Guan, Sherman, and Calvin, 2004). By comparison our method requires very little calibration on the part of the user. It also appears quite robust to model misspecification, as we demonstrate in the next section.

2.3.3 Simulation study for range anisotropy detection

We examined the performance of the angle estimators of the preceding section using simulated data exhibiting range anisotropy. As we are mainly interested in situations where the covariance structure is unknown, we considered a situation of model misspecification: simulating data from a WM with geometric anisotropy ($\nu = 3$, $\lambda = 1$, $s^y/s^x = 2$), but using exponential PACs to fit each candidate angle.

We used the experimental setup of Sect. 2.2.4: a linear model with two autocorrelated covariates and two independent ones, on a 40×40 spatial domain. We tested 50 values of $\alpha \in [0, \pi)$ for the true spatial error covariogram (chosen uniformly at random). For each of these angles we replicated the simulation with new data 50 times, following the procedure outlined in the previous section to construct two angle estimates per replicate. Our candidate models comprised the 12 angles of the form (2.8) for which the interpoint distance $s \leq 5$, or:

$$\alpha_j = \{\arctan(\alpha^y/\alpha^x) \mid \alpha^y, \alpha^x \in \{1, \ldots 4\}, \|(\alpha^x, \alpha^y)\| \leq 5\} \bigcup \{0\}\,. \tag{2.13}$$

From this list we constructed the subsets $\mathcal{G}(\alpha_j)$ using equation (2.28). In the case of $\alpha_j = 0$, we defined $\mathcal{G}(0)$ by simply omitting all even-numbered gridlines from $\mathcal{G}$. For the nonzero angles, note that the dimensions $N^y \times N^x$ of the full dataset can be decremented by discarding outer rows/columns, and the dimensions $n^y \times n^x$ of the subgrids $\mathcal{G}(\alpha_j)$ adjusted as needed, until the dimensional constraints (2.9) are met. A more cautious implementation could seek to ensure equal sample sizes in each $\mathcal{G}(\alpha_j)$. However, we found this had little impact on our simulation results.

Because we used the same covariogram family for c^x and c^y (the 1D exponential), the model for $\mathcal{G}(\alpha_j)$ simultaneously tests both α_j and $\alpha_j + \pi/2$. Our test set (2.13) therefore encompasses 24 angles, whose positions on the (mod π) compass rose are indicated by the gray bars in Figure 2.4 (left). For a given fitted covariogram, we distinguished α_j and $\alpha_j + \pi/2$ by taking the larger of the two fitted range parameters ($\hat{\lambda}^x$ or $\hat{\lambda}^y$) to indicate the major axis direction.

In the final step, we selected from the unused data (white cells in Figure 2.3, right) a subset of size $\sqrt{n^x n^y} = 40$ to set aside as a conditioning set, and predicted the remaining points $\mathcal{G}_{\text{pred}}$ under each of the 12 candidate models. We then determined the angle with least

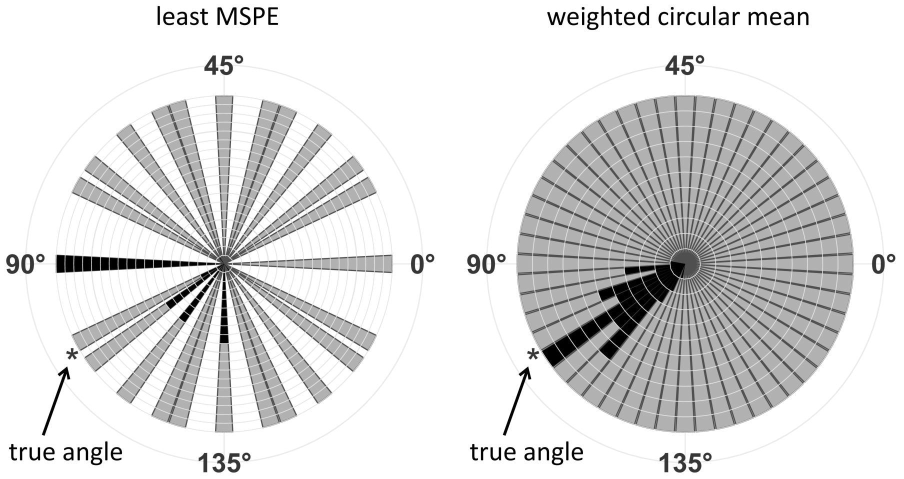

Figure 2.4: Rose diagrams summarizing estimates of the angle of range anisotropy over 50 simulation replicates. Datasets were generated from a model with a geometrically anisotropic WM covariogram, oriented at angle ≈ 105°. Estimates by least RMSPE select the best performing angle from 24 specially chosen candidates (grey bins, left). Weighted circular mean (right) combines information from all 24 candidates to form a continuous estimate

RMSPE, and the weighted circular mean $\hat{\alpha_\omega}$, and recorded the error (mod π) in each case. Figure 2.4 plots the results for one of the 50 tested angles (the other 49 can be generated using the R code files in Online Resource 2). In Figure 2.5 we show the pooled errors over all 2500 simulations.

While Figure 2.5 suggests that both estimators are reasonably unbiased, least-RMSPE tended to favour angles with higher sampling density (specifically $\alpha_j = 0, \pi/4, \pi/2$), leading to a multimodal error distribution. Moreover, the individual histograms often exhibited an interesting (but unwanted) dip near the true angle, as in Figure 2.4, left.

Weighted circular mean performed far better. Its pooled error distribution appears unimodal, as do the individual histograms, and there was noticeably less variance. Even for the worst performing simulation angle (in terms of error variance), $\hat{\alpha_\omega}$ fell within 15° of the true angle in 70% of the repetitions. Over all simulations it was within 45° of the true angle 99.1% of the time, and 77% of the time it was within 15° (Figure 2.5). This is remarkable given that our estimator derives from only 12 candidate angles. Recall that an

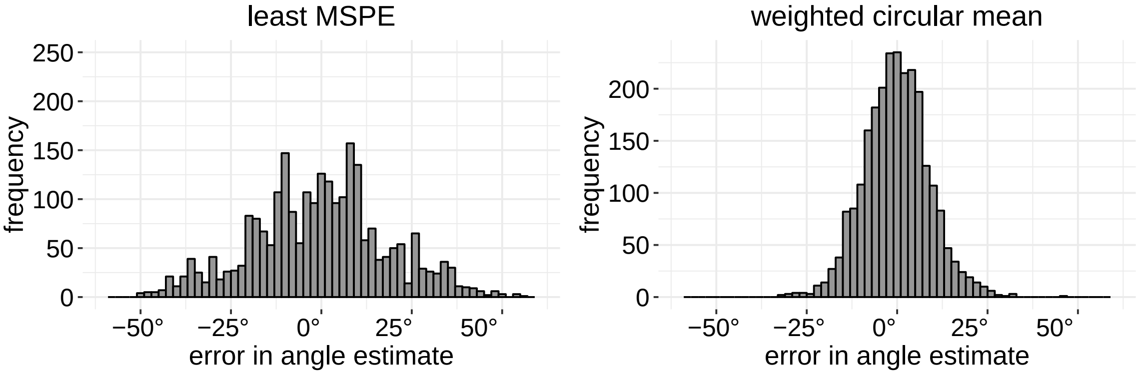

Figure 2.5: Errors in two estimators of the angle of range anisotropy for data generated from a model of geometric anisotropy (WM covariogram). Left and right histograms are the pooled results over the same set of simulations, with 50 repetitions for each of 50 randomly chosen true angles

empirical rose plot over 12 evenly spaced angles in $[0, \pi)$ would have a detection tolerance of 15°.

By design, the computing resources needed to estimate α are quite reasonable for large sample sizes. A desktop PC required only around 1-3 seconds with the 40×40 sample size. Moreover the algorithm is completely automated. Unlike the empirical rose plot there is no need to pick an ad-hoc collection of spatial lags or contour levels, nor does the output rely on any kind of subjective visual inspection.

Needless to say, these estimators are only meaningful if it is reasonable to assume a SOS process with range anisotropy. A number of nonparametric tests can detect departures from isotropy (Weller, Hoeting, et al., 2016), but we are aware of very few such tests for stationarity. Modellers will more often seek to detrend the data using a careful constructed mean function. This is not always possible however, and sometimes it is informative to study the nonstationary covariance structure itself, particularly with ecological data.

For example Sampson and Guttorp (1992) described how nonstationarity can be explored visually using smooth nonlinear deformations of the spatial domain. We will do something similar with $\hat{\alpha_\omega}$. Suppose that the spatial process over the $N^y \times N^x$ domain $\mathcal{D}$ is nonstationary, but exhibits local stationarity on the scale of a much smaller $n^y \times n^x$ subgrid.

We propose using our range anisotropy detection method repeatedly on a sliding window of size $n^y \times n^x$ that moves across $\mathcal{D}$, estimating at each position the angle $\hat{\alpha_\omega}$ and the range parameters $\hat{\lambda}^x$, $\hat{\lambda}^y$. These values define a pair of orthogonal vectors for each location. Plotted together, these depict graphically how the covariance changes through space, much like a biorthogonal grid. This idea is demonstrated in the following case study.

2.4 Case study: mountain pine beetle damage

We applied our angle detection method to analyse damage to pine forests caused by the mountain pine beetle (MPB). Populations of this tree-killing bark beetle have in recent decades grown to unprecedented levels, leading to an epidemic of mortality in pines throughout its vast native habitat in Western North America. The economical and ecological consequences of the epidemic will be severe and long-lasting (Dhar, Parrott, and Heckbert, 2016).

In response, a large body of research has sought to reveal the factors that give rise to MPB outbreaks and allow them to spread, including microclimate, altitude, pine density and proximity to infested stands (Safranyik and Carroll, 2006). Nevertheless, the large-scale dispersal habits of the MPB are difficult to assay, and remain poorly understood. We applied our methods here to better understand the movements of this forest pest, in the hope that spatially explicit predictions of future outbreaks can be improved.

Monitoring efforts by the Canadian province of British Columbia (BC) are a source of unusually detailed and comprehensive spatial data on MPB activity (Westfall and Ebata, 2009). These data comprise yearly sketch maps of the severity of damage by the beetle (% of pines killed per hectare). We rasterized these maps to produce regular gridded data covering almost the entire treed area of the province, at a one hectare resolution.

In the preliminary analysis we looked at a 300×300 subset, using the linear model (2.2) with an exponential PAC. To avoid the complications of temporal dependence, we only fitted the spatial process from a single year, $t = 2007$ (around the peak of the pine beetle epidemic in Southern BC). Our response variable $Y_t(\boldsymbol{s}_i)$ is the logit-transformed beetle

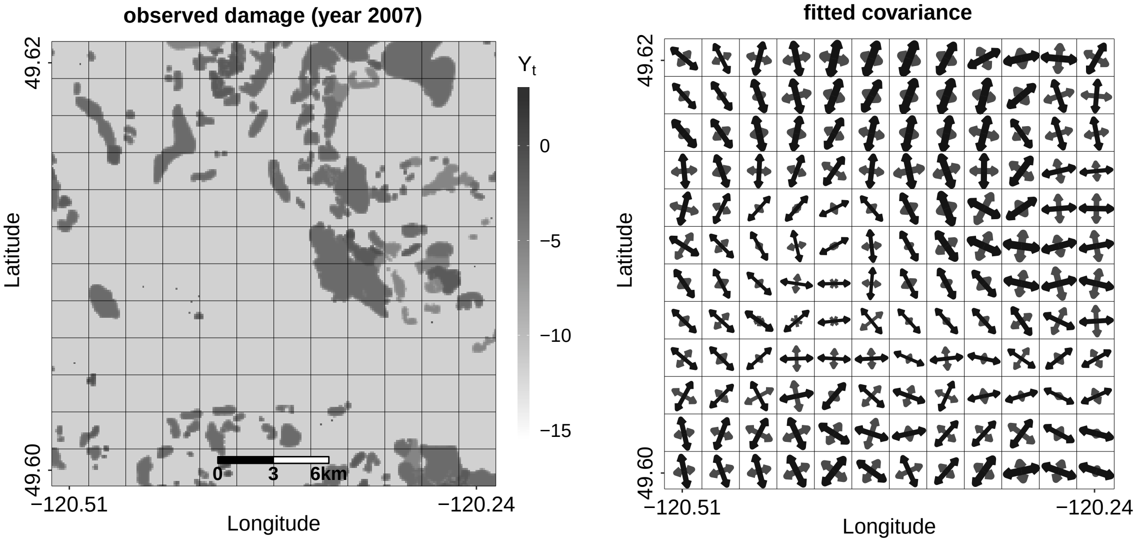

Figure 2.6: Beetle damage data from from 2007 (left) used to fit a nonstationary covariance structure (right). Arrows indicate the directions $\hat{\alpha}_\omega$ identified by a blockwise application of the angle-detection technique of Sect. 2.3.1 to model residuals. Black arrows indicate the major axis direction (largest $\hat{\lambda}$) and arrow thickness indicates the magnitude of the sill ($\hat{\sigma}$)

damage measurement for site $\boldsymbol{s}_i$ in year t, after adding a small offset ϵ_t to adjust for zeroes (Warton and Hui, 2011) (Figure 2.6, left). The design matrix $\boldsymbol{X}$ comprised 29 covariates – mostly climate and weather related – known to influence MPB attack dynamics. The full list can be found in Appendix 2.6.6.

To begin we fit the covariogram to the full 300 x 300 domain by maximum likelihood, given an initial set of OLS estimates for $\boldsymbol{\beta}$. We then used GLS, as described below equation (2.2), to obtain $\hat{\boldsymbol{\beta}}_{GLS}$, and refitted the covariogram using the updated regression parameters.

Next, to examine nonstationarity we constructed a 12 x 12 layout of spatial blocks, each of size 80 x 80, with an overlap of 60 cells in each direction. Within each block (and with $\boldsymbol{\beta}$ fixed to $\hat{\boldsymbol{\beta}}_{GLS}$) we fitted the linear model (2.2) using exponential PACs corresponding to each of the angles in (2.13). We estimated the within-block angle by $\widehat{\alpha_\omega}$ using the method of Sect. 2.3.3, picking the nearest special angle from the set (2.13) and using its corresponding fitted covariogram to predict over the unseen data in that block. We then compared RMSPE values of these blockwise predictions with those of the separable model fitted to the full

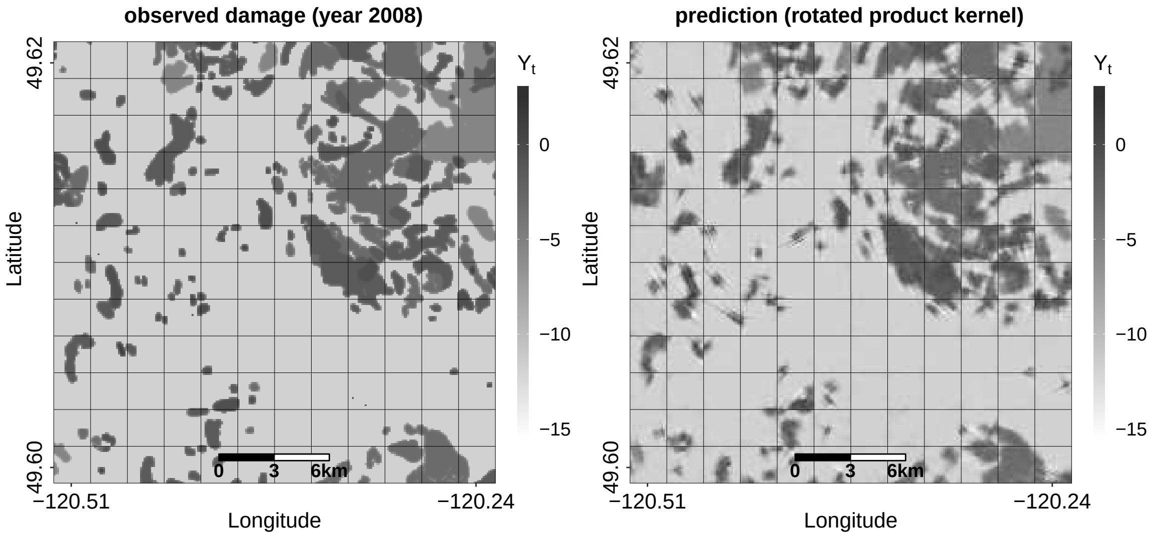

Figure 2.7: Observed pine beetle damage in a 30 x 30 km area of Southern BC in year 2008 (left). A random sample of 20% of these datapoints was used to predict the remainder (right), using a linear model with rotated product anisotropic covariograms fitted blockwise to previous-year data

domain.

Lastly, to gauge future predictive ability we used the 2007 (blockwise) models to estimate damage in the year 2008 (Figure 2.7, left), conditional on a subset of the response data from that year. The conditioning set comprised a random subsample of 20% of the points from the non-overlapping 20 x 20 subsets at the center of each block (black grid in Figures 2.6 and 2.7). We then predicted on the unobserved points to compute RMSPE. This arrangement ensured that the spatial locations of the predictions lay entirely within the subset used to fit the covariogram of the previous year (Figure 2.3, right). Again, we formulated blockwise predictions from both the PAC (within block) and the ordinary separable covariogram (full domain) models, and compared prediction error.

The fitted values of $\widehat{\alpha_\omega}$ (Figure 2.6, right) revealed an interesting pattern of range anisotropy varying through space, indicating nonstationarity. However, large parts of the study area exhibited a consistent directionality and effective range, suggesting that an assumption of local stationarity is reasonable over these areas. This is reflected in the within-block predictions for the model fitting year, for which the product covariogam (RMSPE = 0.026) outperformed the (full domain) separable covariogram (RMSPE = 0.027). Unsurprisingly, both spatial models improved on the nonspatial OLS predictions (RMSPE = 0.032), highlighting the strong SAC in these data.

If the same spatial covariance structure persists to some degree between years, we can expect to improve 2008 predictions by incorporating information from the blockwise estimated covariograms from the previous year. This was indeed the case, with OLS estimates producing much larger errors in next-year predictions (RMSPE = 0.069) when compared with kernel-based predictions from the separable (RMSPE = 0.044) and PAC (RMSPE = 0.046) models.

Note that overall, the separable covariogram outperformed the blockwise PAC on these next-year test data. This could be a consequence of the decreased sampling resolution induced by (2.28) (see discussion), or of year-to-year changes in the covariance structure. The PAC, however, performed far better on blocks heavily damaged by the pine beetle, producing the lowest MSPE in 40 of the 64 blocks (63%) for which damage extended over 25% or more of the area. Nevertheless, in spite of the variability in $\hat{\alpha_{\omega}}$, both spatial models performed adequately. The predicted sizes and shapes of the beetle damaged areas in 2008 appear quite reasonable given the sparsity of the conditioning set.

Computationally, this analysis was very simple. The entire process - from GLS, angle-detection and fitting, to prediction, for both models on all 144 blocks - was completed by an ordinary desktop PC in about 7 minutes. By contrast it took closer to an hour to fit a standard model of geometric anisotropy to a *single* block, even with α known. Moreover, the blockwise approach is easily parallelized, and thus if needed it could be sped up even further by using a cluster of computers.

2.5 Discussion

The computational complexity involved with explicit representations of covariance can be a formidable obstacle. Building upon work by Zimmerman (1989), we have argued for the unconventional solution of using covariograms that are separable in space. These models are not without their drawbacks. Stein (2005) noted issues related to ridges along the coordinate axes that lead to undesirable correlations in linear predictors. These ridges are visible in the plot of the exponential product covariogam in Figure 2.1 (top-right), for example, as well as the predicted values in Figure 2.7. Moreover, separable covariograms are never mathematically isotropic (except in the special, but problematic case of the Gaussian).

In many applications, however, we believe these issues have minimal impact on prediction and inference, and that the drawbacks are outweighed by the substantial reduction in computational complexity. In ecology, this complexity often prevents a SAC-corrected analysis in the first place – here it bears repeating the observation of Keitt et al. (2002), that "making any correction is more important than quibbling about which correction to make". Our simulations suggests that separable WM product covariogams are adequate surrogates for more conventional isotropic models. Thus, we suggest them as a simpler alternative to the more sophisticated approximations developed in Genton (2007) and Hirano (2014).

On simulating data exhibiting range anisotropy with a known direction, we found that separable covariograms substantially improved MSPE compared with isotropic ones, in spite of a model misspecification. This kind of flexibility will be desirable to modellers with reason to doubt the assumption of isotropy in their data, a common situation in ecology. Note that the separable product of WM covariograms limits to a Gaussian (as $\nu \to \infty$), so not only is this extremely common model well approximated, it is generalized to include a range of heavier-tailed alternatives. Readers interested in separable approximations are directed to Wilson et al. (2014), who showed that *any* SOS covariogram can be approximated to arbitrary precision using sums of separable covariograms.

While the direction of range anisotropy will rarely be known *a priori*, we showed in

Sect. 2.3 how separability leads to a fast estimator of this angle, $\hat{\alpha_\omega}$. The simulation results in Sect. 2.3.3 demonstrated a satisfactory level of accuracy and precision, in spite of model misspecification. We believe this method will nicely complement more standard data exploration techniques such as windrose plots, and formal tests of anisotropy, such as in Guan, Sherman, and Calvin (2004). It could be used to automate the analysis of a large number of datasets; to identify specific angles and lags to study in more detail using directional (co)variograms; and to provide an objective verification of the conclusions of the graphical analyst.

Note that because the dimensional constraints in (2.9) preclude large values of α^x, α^y, not all angles of the form (2.8) can be feasibly tested in a given domain $\mathcal{D}$. Moreover the interpoint distance s of points in the rotated subgrid increases with both α^x and α^y, making its sampling layout increasingly grainy, and hindering the detection of small-scale covariances over the data in $\mathcal{D}_\kappa$. For example, in our case study the largest interpoint distance was $s = 5$. At a one-hectare resolution this was acceptable, given that the clusters of beetle damage of greatest concern were much larger than 500 metres. However, depending on the application, modellers may need to upsample their raster data, or shrink the set of candidate angles (2.13), until the scale of interest is smaller than the largest s. Alternatively one could modify the composite likelihood function in (2.11) to incorporate information on both the large and small scale, much like the hybrid method proposed by Varin, Reid, and Firth (2011).

By building $\hat{\alpha_\omega}$ into a sliding-window estimator, we revealed a remarkably smooth pattern of directional dependence resembling a vector field (Figure 2.6) in the pine beetle damage dataset. An ecological explanation for these patterns would be an interesting topic for further research, as it could shine a light on the dispersal habits of the beetle and assist in future predictions. In future work, we hope to explore the connection between local estimates $\hat{\alpha_\omega}$ and local covariates such as wind direction or connectivity of forest corridors - both likely drivers of directionality in pine beetle damage patterns (Aukema et al., 2006).

Though $\hat{\alpha_\omega}$ served as an exploratory tool in our demonstration, it could be useful in other roles. For example, covariance plots like Figure 2.6 might lead to a graphical means of model selection similar to that described in Das et al. (2002); candidate covariate sets can be compared in terms of fitted covariograms, with the aim of finding a minimal set of explanatory variables that yields a stationary process. Another interesting avenue of research would be to construct a predictor that incorporates information from nonlocal covariance estimates. For example, one could take an average of the blockwise predictions, weighted by distance to the block centroid, to obtain a smoothed prediction surface.

We have throughout this paper used the WM covariogram in demonstrations because it is extremely common in spatial statistics. However, in future research, it will be important to compare against other covariance families to get a more complete picture of the robustness of the product-form alternatives proposed here. Although our empirical results are promising, a healthy skepticism of $\hat{\alpha_\omega}$ is wise until its statistical properties are investigated more formally in a theoretical setting.

Until then we would simply argue that separable PACs are a viable means of accounting for SAC, while speeding up analyses by many orders of magnitude on large-n problems. Though our focus here is spatial, we remind the reader that Kronecker product decompositions also apply to separable spatio-temporal covariance matrices. It is our hope that by adopting this trick in the *spatial* domain, practitioners can continue to use the easily interpreted geostatistical model in the rapidly developing world of big data. However, those preferring to use more conventional nonseparable models such as geometric anisotropy can nevertheless enjoy a stable 4X speedup by exploiting bisymmetry.

2.6 Appendices to Chapter 2

2.6.1 Matrix symmetries

All covariance matrices are symmetric about their diagonal, but a regular arrangement of sampling locations can lead to additional useful structures in $\boldsymbol{V}$:

Definition 1 (Toeplitz structure) *$\boldsymbol{A}$ is Toeplitz when its entries satisfy $\boldsymbol{A}[i, j] = \boldsymbol{A}[i + k, j + k]$ for all feasible k. Equivalently, the entries along its k-diagonals are all equal. $\boldsymbol{A}$ is called block-Toeplitz when it has identical blocks along all of its block-diagonals.*

Definition 2 (The exchange matrix $\boldsymbol{J}_m$) *Also called the counteridentity, $\boldsymbol{J}_m$ is the $m \times m$ matrix with 1's on the counterdiagonal and zeroes everywhere else.*

$$\underset{m\times m}{\boldsymbol{J}_m} = \begin{pmatrix} 0 & \dots & 0 & 1 \\ \vdots & & 1 & 0 \\ 0 & \iddots & & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix}$$

Note $\boldsymbol{J}_m$ is symmetric and orthogonal. When left-multiplied by matrix $\boldsymbol{A}$, it reverses the order of the columns in $\boldsymbol{A}$, and when right-multiplied it reverses the order of the rows. This leads to the following symmetry definitions:

Definition 3 (Persymmetry and bisymmetry) *$\boldsymbol{A}$ is persymmetric if $\boldsymbol{J}_m\boldsymbol{A}\boldsymbol{J}_m = \boldsymbol{A}^T$. This is symmetry about the counterdiagonal. If $\boldsymbol{A}$ is symmetric then $\boldsymbol{A}\boldsymbol{J}_m$ and $\boldsymbol{J}_m\boldsymbol{A}$ are both persymmetric, and vice versa. $\boldsymbol{A}$ is bisymmetric if it is both symmetric and persymmetric.*

Definition 4 (Centrosymmetry) *$\boldsymbol{A}$ is centrosymmetric if $\boldsymbol{J}_m\boldsymbol{A}\boldsymbol{J}_m = \boldsymbol{A}$. This is symmetry about the centre of the matrix. Centrosymmetric matrices need not be symmetric, but when they are, bisymmetry is implied. If $\boldsymbol{J}_m\boldsymbol{A}\boldsymbol{J}_m = -\boldsymbol{A}$ then $\boldsymbol{A}$ is called skew-centrosymmetric.*

2.6.2 Kronecker products

the Kronecker product of $m \times n$ matrix $\boldsymbol{A}$ (with entries a_{ij}) and $p \times q$ matrix $\boldsymbol{B}$ is the matrix:

$$\underbrace{\boldsymbol{A} \otimes \boldsymbol{B}}_{mp\times nq} := \begin{pmatrix} a_{11}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & \dots & a_{mn}\boldsymbol{B} \end{pmatrix}. \tag{2.14}$$

Kronecker products admit the following useful identities (Golub and Van Loan, 2012):

Proposition 5 (Determinant) *If $\boldsymbol{A}$ is order m and $\boldsymbol{B}$ is order n, then $|\boldsymbol{A} \otimes \boldsymbol{B}| = |\boldsymbol{A}|^n |\boldsymbol{B}|^m$.*

Proposition 6 (Inverse) *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonsingular, then $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.*

Proposition 7 (Mixed product) *If the products $\boldsymbol{AC}$ and $\boldsymbol{BD}$ are conformable, then*

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC} \otimes \boldsymbol{BD}).$$

Proposition 8 (Decompositions of Kronecker products) *If $\boldsymbol{A} = \boldsymbol{U}_A \boldsymbol{C}_A \boldsymbol{V}_A$ and $\boldsymbol{B} = \boldsymbol{U}_B \boldsymbol{C}_B \boldsymbol{V}_B$,*

$$\boldsymbol{A} \otimes \boldsymbol{B} = (\boldsymbol{U}_A \otimes \boldsymbol{U}_B)(\boldsymbol{C}_A \boldsymbol{V}_A \otimes \boldsymbol{C}_B \boldsymbol{V}_B) = (\boldsymbol{U}_A \otimes \boldsymbol{U}_B)(\boldsymbol{C}_A \otimes \boldsymbol{C}_B)(\boldsymbol{V}_A \otimes \boldsymbol{V}_B).$$

Proposition 9 (Matrix equation representation) *If the product $\boldsymbol{ABC}$ is conformable, then*

$$vec(\boldsymbol{ABC}) = \left(\boldsymbol{C}^T \otimes \boldsymbol{A}\right) vec(\boldsymbol{B}).$$

2.6.3 Indexing via selection matrices

We use the square-bracket notation of Horn and Johnson (2013) to refer to submatrices and subvectors: If $\boldsymbol{A}$ is $M \times N$, and $\boldsymbol{\kappa}$ indexes a subset of rows and $\boldsymbol{\gamma}$ a subset of columns, then we write $\boldsymbol{A}[\boldsymbol{\kappa}, \boldsymbol{\gamma}]$ for the corresponding submatrix. However, we also allow duplications and re-orderings of rows and/or columns in $\boldsymbol{A}[\boldsymbol{\kappa}, \boldsymbol{\gamma}]$ when this is indicated in $\boldsymbol{\kappa}$ and/or $\boldsymbol{\gamma}$. This concept is likely familiar already to users of the R programming language, in which arrays can be manipulated by this syntax. Let us make this precise:

Definition 10 (Selection matrices) *If $\boldsymbol{\gamma}$ is a length-n vector with entries γ_i drawn from $\{1, \ldots, N\}$, and $\mathbf{e}_i^N$ is the i^{th} standard basis vector of $\mathbb{R}^N$, the $N \times n$ selection matrix $\boldsymbol{S}_{\boldsymbol{\gamma}}$ is:*

$$\boldsymbol{S}_{\boldsymbol{\gamma}} := \begin{pmatrix} \boldsymbol{e}_{\gamma_1}^N & \boldsymbol{e}_{\gamma_2}^N & \ldots & \boldsymbol{e}_{\gamma_n}^N \end{pmatrix} \tag{2.15}$$

Thus if $\boldsymbol{\kappa}$ (length-m) has entries drawn from $\{1, \ldots, M\}$, and $\boldsymbol{\gamma}$ (length-n) has entries drawn from $\{1, \ldots, N\}$, then we write $\boldsymbol{A}[\boldsymbol{\kappa}, \boldsymbol{\gamma}]$ to mean the $m \times n$ matrix product $\boldsymbol{S}_{\boldsymbol{\kappa}}^T \boldsymbol{A} \boldsymbol{S}_{\boldsymbol{\gamma}}$.

Using square-bracket notation, we can also express (2.14) as a Hadamard (entrywise) product. Writing $\mathbb{1}_m$ for the length-m vector of ones, and letting $\boldsymbol{m} = (1, \dots m)^T$ denote the column vector of the first m positive integers (and similar for $\boldsymbol{n}$, $\boldsymbol{p}$, and $\boldsymbol{q}$), we write

$$\boldsymbol{A} \otimes \boldsymbol{B} = \boldsymbol{A}[\boldsymbol{m} \otimes \mathbb{1}_p, \boldsymbol{n} \otimes \mathbb{1}_q] \odot \boldsymbol{B}[\mathbb{1}_m \otimes \boldsymbol{p}, \mathbb{1}_n \otimes \boldsymbol{q}]. \tag{2.16}$$

This will be convenient when dealing with submatrices of Kronecker products, such as in the marginal covariance formula of Section 2.6.5. We present this idea in a lemma:

Lemma 11 (Selections) *Suppose $\boldsymbol{A}$ is $m \times n$; and $\boldsymbol{B}$ is $p \times q$; where $M = mp$, $N = nq$. If γ_r (length-r) has entries drawn from $\{1, \dots M\}$, and γ_c (length-c) from $\{1, \dots N\}$, then*

$$(\boldsymbol{A} \otimes \boldsymbol{B})[\gamma_r, \gamma_c] = \boldsymbol{A}[(\boldsymbol{m} \otimes \mathbb{1}_p)[\gamma_r], (\boldsymbol{n} \otimes \mathbb{1}_q)[\gamma_c]] \odot \boldsymbol{B}[(\mathbb{1}_m \otimes \boldsymbol{p})[\gamma_r], (\mathbb{1}_n \otimes \boldsymbol{q})[\gamma_c]]. \tag{2.17}$$

Proof. *This is immediate from (2.16) on recognizing that the square-bracket indexing admits a composition rule $\boldsymbol{X}[\boldsymbol{\kappa}_r, \boldsymbol{\kappa}_c][\boldsymbol{\eta}_r, \boldsymbol{\eta}_c] = \boldsymbol{X}[\boldsymbol{\kappa}_r[\boldsymbol{\eta}_r], \boldsymbol{\kappa}_c[\boldsymbol{\eta}_c]]$, as well as a distributive rule over Hadamard products $(\boldsymbol{X} \odot \boldsymbol{Y})[\boldsymbol{\kappa}, \boldsymbol{\eta}] = \boldsymbol{X}[\boldsymbol{\kappa}, \boldsymbol{\eta}] \odot \boldsymbol{Y}[\boldsymbol{\kappa}, \boldsymbol{\eta}]$. These properties become obvious upon switching to the selection matrix representation in (2.15).* ■

2.6.4 Distance matrices for rectangular grids

We will focus on covariance models for lattice data where the random field variable $Z(\boldsymbol{s})$ is observed at a fixed set of locations $\mathcal{G} = \{\boldsymbol{s}_k\}_{k=1,\dots n} \subset \mathbb{R}^2$ that together form a spatially regular $n^y \times n^x$ rectangular grid in the Cartesian plane (where $n = n^x n^y$). Assuming our random vector $\boldsymbol{Z} := (Z(\boldsymbol{s}_1), \dots Z(\boldsymbol{s}_n))^*$ is in column-vectorized order, the position (i, j) in the grid is mapped to the index $k = \text{vec}_{n^y}(i, j) := i + n^y(j-1)$ in $\boldsymbol{Z}$. For convenience we will assume that $\mathcal{G}$ has been scaled/translated to lie on the integer lattice $\mathbb{Z}^2$ with its bottom-left corner on the origin, so that its point coordinates are given by $\boldsymbol{s}_k = (x_k, y_k) = (j-1, n^y-1)$.

In this configuration, pairwise distances between points in $\mathcal{G}$ can be succinctly represented using Kronecker products. To see this, number the grid lines of $\mathcal{G}$ using vectors $\boldsymbol{g^x} = (1, \dots n^x)^*$ and $\boldsymbol{g^y} = (n^y, \dots 1)^*$, so that if $\boldsymbol{s}_k$ is at the $(i, j)^{th}$ position in the grid, its x and y coordinates are $x_k = [(\mathbb{1}_{n^y})(\boldsymbol{g^x})^*]_{i,j}$ and $y_k = [(\boldsymbol{g^y})(\mathbb{1}_{n^x})^*]_{i,j}$ (where $\mathbb{1}_m$ is the length-m column vector of ones). Vectorizing these outer products, one obtains the full set of coordinates for $\mathcal{G}$ (in column-vectorized order) as the vectors $\boldsymbol{x} = (x_1, ...x_n)^* = \boldsymbol{g^x} \otimes \mathbb{1}_{n^y}$ and $\boldsymbol{y} = (y_1, ...y_n)^* = \mathbb{1}_{n^x} \otimes \boldsymbol{g^y}$. Denote the matrix of differences among the x coordinates as $[\boldsymbol{\delta}^x]_{i,j} := x_i - x_j$, and similar for y. Then a straightforward application of Property 7 yields:

$$\boldsymbol{\delta}^x = \boldsymbol{x}\mathbb{1}_n^* - \mathbb{1}_n\boldsymbol{x}^* = \left(\boldsymbol{g^x}\mathbb{1}_{n^x}^* - \mathbb{1}_{n^x}\boldsymbol{g^{x*}}\right) \otimes \left(\mathbb{1}_{n^y}(\mathbb{1}_{n^y})^*\right), \tag{2.18}$$

$$\boldsymbol{\delta}^y = \boldsymbol{y}\mathbb{1}_n^* - \mathbb{1}_n\boldsymbol{y}^* = \left(\mathbb{1}_{n^x}(\mathbb{1}_{n^x})^*\right) \otimes \left(\boldsymbol{g^y}\mathbb{1}_{n^y}^* - \mathbb{1}_{n^y}\boldsymbol{g^{y*}}\right). \tag{2.19}$$

This reveals a $(n^x \times n^x)$ pattern of $(n^y \times n^y)$ blocks in both matrices: in $\boldsymbol{\delta}^x$ the i, j^{th} block is $(i - j)\boldsymbol{I}$; and in $\boldsymbol{\delta}^y$ it is $\boldsymbol{g^y}\mathbb{1}_{n^y}^* - \mathbb{1}_{n^y}\boldsymbol{g^{y*}}$. These matrices provide an algebraically convenient decomposition of $\boldsymbol{s}_i - \boldsymbol{s}_j$ into its projections along the x and y axes, leading naturally to the idea of separability. They also allow us to write the matrix of pairwise Euclidean distances $[\boldsymbol{D}]_{ij} = d_{ij}$ as the (entrywise) square root of:

$$\boldsymbol{D} \odot \boldsymbol{D} = \boldsymbol{\delta}^x \odot \boldsymbol{\delta}^x + \boldsymbol{\delta}^y \odot \boldsymbol{\delta}^y. \tag{2.20}$$

Structure in covariance matrices for nonseparable SOS models

Note that both of the coordinate difference matrices in (2.18)-(2.19) are block-Toeplitz with Toeplitz blocks (BTTB). This is because $\boldsymbol{g}\mathbb{1}_m^* - \mathbb{1}_m\boldsymbol{g}^*$ expresses the differences between adjacent entries of the length-m vector $\boldsymbol{g}$, and therefore becomes Toeplitz when the entries of $\boldsymbol{g}$ are regularly spaced (eg. as when $\mathcal{G}$ is a regular grid).

In an SOS isotropic model, where covariance is defined by applying the covariogram function c to the vector $\boldsymbol{s}_i - \boldsymbol{s}_j = ([\boldsymbol{\delta}^x]_{i,j}, [\boldsymbol{\delta}^y]_{i,j})^*$, the covariance matrix $\boldsymbol{V}$ inherits the BTTB structure. This is because the entrywise application of any function to a matrix (for instance to $\boldsymbol{\delta}^x$ and $\boldsymbol{\delta}^y$) preserves all of its unsigned symmetry structures. Thus the entrywise squares, sums, and square-roots implied in (2.20) are all structure-preserving, which means that $\boldsymbol{D}$ inherits the BTTB property from $\boldsymbol{\delta}^x$ and $\boldsymbol{\delta}^y$, and $\boldsymbol{V}$ inherits it from $\boldsymbol{D}$ in turn. These arguments also apply with geometric anisotropy, where the $\boldsymbol{\delta}^x$ and $\boldsymbol{\delta}^y$ terms in (2.20) are replaced by linear combinations $s^x(\cos(\alpha)\boldsymbol{\delta}^x + \sin(\alpha)\boldsymbol{\delta}^y)$ and $s^y(\cos(\alpha)\boldsymbol{\delta}^y - \sin(\alpha)\boldsymbol{\delta}^x)$. Linear combinations preserve BTTB structure, and thus $\boldsymbol{V}$ is also BTTB.

The BTTB result was proven in more generality by Zimmerman (1989). However by emphasizing more directly how matrix symmetry patterns emerge from the regular spacing of grid lines, we believe our description has pedogogical value to readers unfamiliar with lattice designs. Moreover (2.18)-(2.20) holds more generally for irregular grid lines (with spacings defined in $\boldsymbol{g}^x$, $\boldsymbol{g}^y$) and so could prove useful in expressing the block structure in $\boldsymbol{V}$ common to any SOS model for gridded data.

2.6.5 Covariance matrices for marginal distributions

Applications often call for marginal distributions over subsets of the data. Examples include cross-validation; conditional expectation, such as in Kriging; and inference with missing data. We will consider the Gaussian case, for which subsets of $\boldsymbol{Z}$ are themselves Gaussian with covariances given as submatrices of $\boldsymbol{V}$. Let $\boldsymbol{\kappa}$ denote the indexing vector for the subset of interest, so that $\boldsymbol{Z}_{\boldsymbol{\kappa}} = \boldsymbol{Z}[\boldsymbol{\kappa}]$ is the subset, and $\boldsymbol{V}_{\kappa} = \boldsymbol{V}[\boldsymbol{\kappa}, \boldsymbol{\kappa}]$ is its covariance matrix.

Generally speaking $\boldsymbol{V}_{\boldsymbol{\kappa}}$ inherits neither separability, BTTB structure, nor bisymmetry from $\boldsymbol{V}$, unless the subset in question forms a regular rectangular grid (note that the angle estimator from Section 3 of the main text exploits this exception.) Thus $\boldsymbol{V}_{\boldsymbol{\kappa}}$ may be more difficult to work with than $\boldsymbol{V}$, in spite of its smaller dimensionality. However if the subset is not too small, one can use a shortcut suggested in Ver Hoef et al. (2018): Leting $\boldsymbol{\kappa}^c$ index the observations not in $\boldsymbol{\kappa}$, the determinant and inverse of $\boldsymbol{V}_{\boldsymbol{\kappa}}$ can be computed using:

$$\boldsymbol{V}_{\boldsymbol{\kappa}}^{-1} = \boldsymbol{V}^{-1}[\boldsymbol{\kappa}, \boldsymbol{\kappa}] - \left(\boldsymbol{V}^{-1}[\boldsymbol{\kappa}, \boldsymbol{\kappa}^c]\right)\left(\boldsymbol{V}^{-1}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}^c]\right)^{-1}\left(\boldsymbol{V}^{-1}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}]\right) \tag{2.21}$$

$$|\boldsymbol{V}_{\boldsymbol{\kappa}}| = |\boldsymbol{V}||\boldsymbol{V}^{-1}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}^c]|. \tag{2.22}$$

When it is computationally cheap to evaluate the inverse and determinant of $\boldsymbol{V}$, but not $\boldsymbol{V}_{\boldsymbol{\kappa}}$, these formulae can lead to substantial speedups; If m is the length $\boldsymbol{\kappa}$ and n the length of $\boldsymbol{Z}$, we trade a problem of dimension m for one of dimension $n - m$ (the length of $\boldsymbol{\kappa}^c$).

These formulae are proved as follows. Suppose $\{1, \ldots n\}$ is partitioned into sets of size m and $n - m$, identified by the index vectors $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^c$, respectively. Then if $\boldsymbol{V}$ is any $n \times n$ matrix, the permutation $\boldsymbol{P} = (\ \boldsymbol{S}_{\boldsymbol{\kappa}} \quad \boldsymbol{S}_{\boldsymbol{\kappa}^c}\ )$ takes $\boldsymbol{V}$ and its inverse to the block forms:

$$\boldsymbol{P}^T \boldsymbol{V} \boldsymbol{P} = \begin{pmatrix} \boldsymbol{V}_{\boldsymbol{\kappa}} & \boldsymbol{V}[\boldsymbol{\kappa}, \boldsymbol{\kappa}^c] \\ \boldsymbol{V}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}] & \boldsymbol{V}_{\boldsymbol{\kappa}^c} \end{pmatrix}$$

$$\boldsymbol{P}^T \boldsymbol{V}^{-1} \boldsymbol{P} = \begin{pmatrix} \boldsymbol{V}^{-1}[\boldsymbol{\kappa}, \boldsymbol{\kappa}] & \boldsymbol{V}^{-1}[\boldsymbol{\kappa}, \boldsymbol{\kappa}^c] \\ \boldsymbol{V}^{-1}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}] & \boldsymbol{V}^{-1}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}^c] \end{pmatrix} \tag{2.23}$$

Writing $[\boldsymbol{A}]_{ij}$ for the four blocks of a square partitioned matrix $\boldsymbol{A}$, the matrix $\boldsymbol{A}/[\boldsymbol{A}]_{22} := [\boldsymbol{A}]_{11} - [\boldsymbol{A}]_{12}[\boldsymbol{A}]_{22}^{-1}[\boldsymbol{A}]_{21}$ is called the *Schur complement of* $[\boldsymbol{A}]_{22}$ *in* $\boldsymbol{A}$. Ver Hoef et al. (2018) note that $[\boldsymbol{A}]_{11}^{-1} = \boldsymbol{A}^{-1}/[\boldsymbol{A}^{-1}]_{22}$, and suggest this identity as a shortcut to invert a block of a covariance matrix, after having computed its full inverse. Thus equation (2.21) follows from letting $\boldsymbol{A} = \boldsymbol{P}^T \boldsymbol{V} \boldsymbol{P}$ and writing its inverse in the form (2.23). Equation (2.22) follows from the identity $|\boldsymbol{A}/\boldsymbol{B}| = |\boldsymbol{A}|/|\boldsymbol{B}|$ for Schur complements.

A Bayesian perspective

A more general version of (2.21) due to Pukelsheim is presented in Searle, Casella, and McCulloch (2009, pp. 447-452), in the context of likelihood functions for hierarchical models. Before discussing applications, let us revisit this Bayesian perspective. Suppose we put an improper prior on the mean of $\boldsymbol{Z}$. In particular consider the normal hierarchical model,

$$\boldsymbol{Z} \mid \boldsymbol{\mu} \sim N(\boldsymbol{X}_{\boldsymbol{\kappa}^c} \boldsymbol{\mu}, \boldsymbol{V}) \text{ with } \mu_i \overset{iid}{\sim} \text{unif}(-\infty, \infty), \tag{2.24}$$

where the $n \times n$ design matrix $\boldsymbol{X}_{\boldsymbol{\kappa}^c}$ has a one at each diagonal entry indexed in $\boldsymbol{\kappa}^c$, and is zero otherwise. Using (2.15) we can write $\boldsymbol{X}_{\boldsymbol{\kappa}^c} = \boldsymbol{S}_{\boldsymbol{\kappa}^c}(\boldsymbol{S}_{\boldsymbol{\kappa}^c})^T$ where the $N \times (N - n)$ selection matrix $\boldsymbol{S}_{\boldsymbol{\kappa}^c}$ is generated from the length-$(N - n)$ indexing vector for unobserved datapoints.

The likelihood function $\mathcal{L}(\boldsymbol{V} \mid \boldsymbol{Z} = \boldsymbol{z})$ is then equal to $\mathcal{L}(\boldsymbol{V} \mid \boldsymbol{Y} = \boldsymbol{y})$ for the model $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\kappa}})$, over all $\boldsymbol{z}$ for which $\boldsymbol{z}[\boldsymbol{\kappa}] = \boldsymbol{y}$. This is because the effect of $\boldsymbol{X}_{\boldsymbol{\kappa}^c} \boldsymbol{\mu}$ is to add noise to the mean-zero version of $\boldsymbol{Z}$, but only to those components indexed in $\boldsymbol{\kappa}^c$. As the variance of this noise increases, the corresponding entries in the precision matrix for (2.24) shrink, and in the limit of unbounded variance they vanish.

The improper Bayesian posterior likelihood completely ignores these components of $\boldsymbol{Z}$, as their values are washed out by the noise of the flat prior. From an information-theoretical perspective, it is equivalent to simply omit the noisy, unreliable components. Indeed the likelihood function for (2.24) contains exactly those formulae (2.21)-(2.22) that specify $\boldsymbol{Z}_{\boldsymbol{\kappa}}$.

Applications of the marginal covariance formula

Equations (2.21)-(2.22) will speed evaluations of the MVN likelihood function (for a marginal distribution) whenever it is computationally simpler to find the inverse and determinant of $\boldsymbol{V}$ and $\boldsymbol{V}^{-1}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}^c]$ instead of $\boldsymbol{V}_{\boldsymbol{\kappa}}$. This happens when n is large and the number $(n - m)$ of points omitted from $\boldsymbol{\kappa}$ is relatively small. For example, if we consider only the highest order (cubic) terms in the arithmetic complexity, then computation time is reduced whenever $\boldsymbol{V}$ (but not $\boldsymbol{V}_{\boldsymbol{\kappa}}$) is bisymmetric and $(n - m)/n < \sqrt[3]{(m/n)^3 - 1/4}$, or (approximately) $(n - m)/n < 0.37$. If $\boldsymbol{V}$ (but not $\boldsymbol{V}_{\boldsymbol{\kappa}}$) is separable and we assume for simplicity that $\mathcal{G}$ is a *square* grid, the bound becomes $(n - m)/n < \sqrt[3]{(m/n)^3 - 2/n^{(3/2)}}$. Asymptotically (for large n), the bound is $(n - m)/n < 0.5$.

An obvious application is in analyses where the locations $\mathcal{D}$ lie on a lattice, but do not form a complete rectangular subset. In this case one simply formulates $\boldsymbol{V}$ over a grid $\boldsymbol{G}$ covering the bounding box for the data, by introducing empty cells and indexing them in $\boldsymbol{\kappa}^c$. Other applications include cross-validation and leave-one-out estimators, where subsets of the data are withheld and the marginal distribution over the remainder is of interest. For example Ver Hoef *et al.* (2018) implemented (2.21)-(2.22) in a predictions over unobserved data. More generally these formulae can simplify computations of the conditional expectation when a partial observation $\boldsymbol{Z}_{\boldsymbol{\kappa}}$ is used to inform estimates of the unobserved components $\boldsymbol{Z}_{\boldsymbol{\kappa}^c}$, since the conditional distribution of $\boldsymbol{Z}_{\boldsymbol{\kappa}^c}$ given $\boldsymbol{Z}_{\boldsymbol{\kappa}} = \boldsymbol{z}[\boldsymbol{\kappa}]$

is:

$$\mathcal{N}\left\{\boldsymbol{\mu}[\boldsymbol{\kappa}^c] + \boldsymbol{V}^*_{\boldsymbol{\kappa}\boldsymbol{\kappa}^c}\boldsymbol{V}_{\boldsymbol{\kappa}}^{-1}\left(\boldsymbol{z}[\boldsymbol{\kappa}] - \boldsymbol{\mu}[\boldsymbol{\kappa}]\right), \boldsymbol{V}_{\boldsymbol{\kappa}^c} - \boldsymbol{V}^*_{\boldsymbol{\kappa}\boldsymbol{\kappa}^c}\boldsymbol{V}_{\boldsymbol{\kappa}}^{-1}\boldsymbol{V}_{\boldsymbol{\kappa}\boldsymbol{\kappa}^c}\right\}, \tag{2.25}$$

where $\boldsymbol{V}_{\boldsymbol{\kappa}^c}$ is the marginal covariance matrix for $\boldsymbol{Z}_{\boldsymbol{\kappa}^c}$, and $\boldsymbol{V}_{\boldsymbol{\kappa}\boldsymbol{\kappa}^c}$ is the cross covariance of $\boldsymbol{Z}_{\boldsymbol{\kappa}}$ and $\boldsymbol{Z}_{\boldsymbol{\kappa}^c}$. Note that in the case of separable covariance, the computationally prohibitive step of forming the $N \times N$ matrix $\boldsymbol{V} = \boldsymbol{V}^x \otimes \boldsymbol{V}^y$ can be avoided. The required matrices can easily be computed directly from $\boldsymbol{V}^x$ and $\boldsymbol{V}^y$ using the identity (2.17), which leads to

$$\boldsymbol{V}_{\boldsymbol{\kappa}\boldsymbol{\kappa}^c} = \boldsymbol{V}[\boldsymbol{\kappa}, \boldsymbol{\kappa}^c] = \boldsymbol{V}^x[\boldsymbol{\kappa}^x, \boldsymbol{\kappa}^x] \odot \boldsymbol{V}^y[\boldsymbol{\kappa}^y, \boldsymbol{\kappa}^y], \tag{2.26}$$

$$\boldsymbol{V}_{\boldsymbol{\kappa}^c} = \boldsymbol{V}[\boldsymbol{\kappa}^c, \boldsymbol{\kappa}^c] = \boldsymbol{V}^x[\boldsymbol{\kappa}^x, (\boldsymbol{\kappa}^x)^c] \odot \boldsymbol{V}^y[\boldsymbol{\kappa}^y, (\boldsymbol{\kappa}^y)^c], \tag{2.27}$$

where the sites indexed in $\boldsymbol{\kappa}$ have x-coordinates $\boldsymbol{g^x}[\boldsymbol{\kappa}^x]$ and y-coordinates $\boldsymbol{g^y}[\boldsymbol{\kappa}^y]$, and the $\boldsymbol{\kappa}^x, \boldsymbol{\kappa}^y$ can be found using the inverse vectorization operator: $(\boldsymbol{\kappa}^x[i], \boldsymbol{\kappa}^y[i]) = \mathrm{vec}_{n^y}^{-1}(\boldsymbol{\kappa}[i])$,

$$\mathrm{vec}_{n^y}^{-1}(k) = (i_k, j_k) = \left(k - n^y\left(\left\lceil \frac{k}{n^y}\right\rceil - 1\right), \left\lceil \frac{k}{n^y}\right\rceil\right).$$

Indexing of the rotated sublattice

Here, we derive the indexing vector $\boldsymbol{\kappa}$ for the rotated subgrid in $\mathcal{G}$ described in Section 3.1. Without loss of generality we will assume that N^x, N^y, n^x, and n^y have been suitably chosen to satisfy equation (18).

Begin by applying the transformation $\boldsymbol{A}^T$ to the coordinates $\boldsymbol{s}_k$ of points in the $n^y \times n^x$ sublattice $\mathcal{G}_0$ with grid line vectors $\boldsymbol{g}_0^x = (0, \ldots n^x - 1)^T$ and $\boldsymbol{g}_0^y = (n^y - 1, \ldots 0)^T$. By construction, these new coordinates $\boldsymbol{A}^T\mathcal{G}_0 = \left\{\tilde{\boldsymbol{s}}_k = \boldsymbol{A}^T\boldsymbol{s}_k; k = \mathrm{vec}_{n^y}(i, j), i = 1, \ldots n^y, j = 1, \ldots n^x\right\}$ are a subset of $\mathbb{Z}^2$ that forms a regular rectangular grid in the rotated coordinate system $\mathcal{D}_\alpha = \left\{\boldsymbol{R}_\alpha\boldsymbol{v}_k; \boldsymbol{v}_k \in \mathbb{R}^2\right\}$, with inter-point spacing $s = \|(\alpha^x, \alpha^y)\|$. Thus if we apply equation (15) to the $\tilde{\boldsymbol{s}}_k$, we recover the separable kernel in equation (8), since $\boldsymbol{A}\boldsymbol{A}^T = I$. It follows from (8) that after rescaling the coordinate difference matrices $\boldsymbol{\delta}^x$ and $\boldsymbol{\delta}^y$ by the new inter-point distance s, we obtain $c(\tilde{\boldsymbol{s}}_i - \tilde{\boldsymbol{s}}_j) = \sigma^2 c^x(s\,[\boldsymbol{\delta}^x]_{ij})c^y(s\,[\boldsymbol{\delta}^y]_{ij})$, which implies the Kronecker product representation in equation (20).

Now let us track one of the grid points along the transformation. The position $\boldsymbol{s}_k = (j-1, n^y - i)^T$ of the $(i,j)^{th}$ element in $\mathcal{G}_0$ gets mapped to the new position $\tilde{\boldsymbol{s}}_k = \boldsymbol{A}^T \boldsymbol{s}_k = (\alpha^x(j-1) - \alpha^y(n^y - i), \alpha^x(n^y - i) + \alpha^y(j-1))^T$. In order to find the row/column index in $\boldsymbol{V}$ corresponding to $\tilde{\boldsymbol{s}}_k \in \boldsymbol{A}^T\mathcal{G}_0$, we must identify the element of $\mathcal{G}$ that coincides with $\tilde{\boldsymbol{s}}_k$. However, following the transformation by $\boldsymbol{A}^T$, some of the leftmost points of $\boldsymbol{A}^T\mathcal{G}_0$ will have fallen out of $\mathcal{G}$ (they have negative x coordinates). So, we simply translate all of the points $\boldsymbol{A}^T\mathcal{G}_0$ to the right by distance $\alpha^y(n^y - 1)$ to ensure that the least x coordinate becomes zero. Note that such translations of coordinates have no bearing on the covariance matrix for points in $\boldsymbol{A}^T\mathcal{G}_0$ under the SOS assumption.

Following this translation, the new positions are $(\tilde{x}_{ij}, \tilde{y}_{ij}) = (\alpha^x(j-1) + \alpha^y(i-1), \alpha^x(n^y - i) + \alpha^y(j-1))^T$, where $i = 1, \dots n^y$, $j = 1, \dots n^x$. With reference to $\mathcal{G}$, we identify their row and column indices as $\tilde{i} = N^y - \tilde{y}_{ij}$ and $\tilde{j} = 1 + \tilde{x}_{ij}$. This simplifies to $\tilde{i} = 1 + \alpha^x(i-1) - \alpha^y(j-1) + \alpha^y(n^x - 1)$ and $\tilde{j} = 1 + \alpha^x(j-1) + \alpha^y(i-1)$. The vectorized index for the element of $\mathcal{G}$ in row $\tilde{i}$ and column $\tilde{j}$ is $\mathrm{vec}^{-1}_{N^y}(\boldsymbol{\kappa})[i,j] = \tilde{i} + N^y(\tilde{j} - 1)$. Substituting the above expressions for $\tilde{i}$ and $\tilde{j}$, and collecting terms, we obtain:

$$\left[\mathrm{vec}^{-1}_{n^y}(\boldsymbol{\kappa})\right]_{ij} = 1 + \alpha^y(n^x - 1) + (N^y\alpha^y + \alpha^x)(i-1) + (N^y\alpha^x - \alpha^y)(j-1) \tag{2.28}$$

$$\implies \boldsymbol{V}_{\boldsymbol{\kappa}} = \boldsymbol{V}[\boldsymbol{\kappa}, \boldsymbol{\kappa}] = \sigma^2 c^x \left(s(\boldsymbol{n^x}\mathbb{1}^*_{n^x} - \mathbb{1}_{n^x}\boldsymbol{n^{x*}})\right) \otimes c^y \left(s(\boldsymbol{n^y}\mathbb{1}^*_{n^y} - \mathbb{1}_{n^y}\boldsymbol{n^{y*}})\right) \tag{2.29}$$

where $\boldsymbol{n^x}$ is the length-n^x vector $(1, 2, \dots n^x)^*$ (representing the grid lines of $\mathcal{G}_0$), and similar for y.

2.6.6 Mountain pine beetle covariates

Category	Covariates	Notes
topography	altitude, slope, aspect, lakes indicator	
host quality	tree density, pine density, stand age	baseline 2001 estimates from Beaudoin et al. (2014), with pine density adjusted for cumulative mortality
beetle activity	beetle pressure (200, 400, 800, 1600, 3200 metre radius)	a (WM) kernel convolution of the product of the previous year beetle damage and the host pine density. Radius indicates the effective range of the kernel
	beetles/host (200, 400, 800, 1600, 3200 metre radius)	beetle pressure divided by pine density (attack density)
microclimate	seasonal precipitation (winter and spring); precipitation as snow; degree days (below 0, below 18); seasonal temperatures: mean in coldest winter month; min, max, & averages (spring, winter)	weather station data interpolated using ClimateBC software (Wang et al., 2007)

Table 2.1: Covariates included in the linear regression model for mountain pine beetle damage patterns.

Chapter 3

A unifying theory for 2D spatial redistribution kernels with applications to model-fitting in ecology

3.1 Introduction

Given that environmental heterogeneity and movement is present virtually everywhere in the natural world, ecologists are inherently concerned with questions of a spatial nature. They are therefore often rewarded by new insights when the mechanism underlying a spatial effect can be worked into models (Kareiva, 1994). *Redistribution kernels* are a popular means to this end, with applications as diverse as predator-prey interactions (Neubert, Kot, and Lewis, 1995); range expansion and invasion biology (Kot, Lewis, and Driessche, 1996); grouping/swarming behaviour (Okubo and Levin, 2001); chemical communication (Powell and Bentz, 2014); and cellular movements (Painter and Hillen, 2018).

Statistical ecologists are concerned with the many of the same questions. However the focus in statistics is how to characterize the randomness in measurements of an ecological system. These measurements are often spatially autocorrelated, a phenomenon that (if ignored) can dramatically reduce the precision of estimators. A common solution is to use phenomenological models, known as *covariance kernels* (Chiles and Delfiner, 2012), that tie correlation to separation distance.

There is a remarkable degree of overlap between the families of functions that are typically chosen in statistics to serve as kernels for covariance and those chosen in ecology for redistribution. An example that we find particularly interesting is the Whittle-Matérn covariance kernel (Matérn, 1986). Its versatility and mathematical elegance makes it one of the most important kernels in spatial statistics (Lindgren, Rue, and Lindström, 2011). However, as a redistribution kernel, it is rarely seen in ecology, in spite of many advantages over more familiar alternatives.

Up to a normalization constant and a restriction on the shape parameter, the Whittle-Matérn is identical to a redistribution kernel first described by Yasuda (1975), and later by Yamamura (2002) and Hapca, Crawford, and Young (2008). We focus in this paper on its use in redistribution, but in recognition of the theory contributed from spatial statisticians we will refer to it as the Whittle-Matérn-Yasuda (WMY) kernel.

In the context of redistribution, a kernel $D(\boldsymbol{x}, \boldsymbol{x}')$ maps a pair of coordinates (source $\boldsymbol{x}'$ and destination $\boldsymbol{x}$) to a probability density for the redistribution event from $\boldsymbol{x}'$ to $\boldsymbol{x}$. When modeling movement events having a random character, kernels provide a simple means of parametrizing their probability density functions (PDFs). The simplest such models are stationary, meaning their kernels are functions only of the separation vector $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{x}'$, independent of location.

When isotropy (radial symmetry) is also assumed, kernels can be defined more simply as a function of $r = |\boldsymbol{r}|$. Such kernels assign probability densities to movement distances. In this case one must take care to differentiate between the density $D(r)$ at point (r, θ) and the marginal density $D_r(r)$ at radius r. For example in 2-dimensional (2D) space the redundant angular coordinate is sometimes conditioned out from D, and the kernel for marginal density is written:

$$D_r(r) := \int_0^{2\pi} r D(r) \, \mathrm{d}\theta = 2\pi r D(r). \tag{3.1}$$

The distinction is important but sometimes unclear in the literature. Our notation will distinguish the two by always using an r subscript to indicate marginal density functions, as in (3.1).

The WMY is an example of a stationary and isotropic kernel. In the 2D case it is written:

$$\mathcal{D}(r;\kappa,\rho) = A\left(\kappa,\rho\right)\left(r/\rho\right)^{\kappa} K_{\kappa}\left(r/\rho\right), \quad \text{with } 1/A\left(\kappa,\rho\right) = 2^{\kappa+1}\pi\rho^{2}\Gamma(\kappa+1), \qquad (3.2)$$

where K_κ denotes the κ^{th} order modified Bessel function of the second kind (Appendix 3.7.1), $\rho > 0$ is a distance-scaling (range) parameter, and κ is a shape parameter. The domain of κ depends on the application: For covariance, $\kappa > 0$; and for redistribution, $\kappa > -1$. By studying κ, we will see that the WMY is closely related to a number of other kernels in common use among ecologists. Indeed, it generalizes some the most prominent ones (Table 3.1), in addition to providing a spectrum of others that combine fat tails with Gaussian-like behaviour near the origin (Appendix 3.7.2).

shape (κ)	kernel name	density* $D(r)$	mechanistic derivation
$(-1,0)$	-	$(r/\rho)^{\kappa}K_{\kappa}(r/\rho)$	2D fractal diffusion with constant settling hazard (Section 3.2.1)
0	Bessel	$K_0(r/\rho)$	2D Fickian diffusion with constant settling hazard (Broadbent and Kendall, 1953)
1/2	2D Laplace	$\exp(-r/\rho)$	2D turbulent diffusion with instantaneous settling (Joseph and Sendner, 1958)
∞	Gaussian	$\exp(-(r/\rho)^2)$	2D Fickian diffusion with instantaneous settling (Skellam, 1951)

Table 3.1: Notable examples from the WMY family $\mathcal{D}(r;\kappa,\rho)$. All arise from 2D Fickian diffusion with gamma-distributed settling times (Yasuda, 1975). The special cases listed here have been derived independently under various movement models. Alternatively, the full WMY family can be derived by repeated iterations of the kernel in the top row (Section 3.2.2). *for brevity the normalization $1/2\pi\int_{-\infty}^{\infty} rD\,\mathrm{d}r$ is omitted.

This makes the WMY an unusually flexible model for the spread of populations. An example that we will use throughout the paper is the mountain pine beetle (MPB). Dispersal flights of this forest pest allow population outbreaks to spill outward into neighbouring areas, and are therefore a key part of spatially explicit modeling efforts. At the conclusion of the paper we will show how the WMY can be used in modeling MPB outbreak patterns, using example data from the height of a recent string of infestations in British Columbia (BC), Canada.

First, to motivate the use of equation (3.2) more generally, we begin Section 3.2 by deriving the WMY kernel in a novel way, constructing it as the solution to a partial differential equation (PDE) for diffusion through inhomogeneous habitat.

We then discuss some of the appealing mathematical properties of WMY as an isotropic model, before showing in Section 3.3 how simple extensions can be used to accomodate departures from isotropy. In Section 3.4 we discuss how kernels like the WMY can be used in applications, emphasizing practical considerations of computational complexity. This motivates a data analysis in Section 3.5 demonstrating how approximations can be used in data-fitting on large-samples, where computing time is otherwise problematic.

3.2 The WMY as a model for diffusion with settling

Yasuda (1975) derived the WMY as the settled density in a Fickian diffusion process with gamma distributed settling times. Equation (3.2) can therefore be understood as a macro-scale description of random walks taking place in unrestricted 2D space. However, in the spatial statistics community, the WMY is known as a covariance model. It describes covariances in the stationary random field solution of a fractional stochastic PDE resembling a generalized Helmholtz equation (Whittle, 1954). This suggests that as a redistribution kernel, the WMY might also solve a similar deterministic PDE involving fractional derivatives.

These exotic dynamical systems often appear in connection with the statistical mechanics of random walks through a complex medium that hinders movement. They are studied in Metzler, Glöckle, and Nonnenmacher (1994) for example, to explain the physical and chemical properties of porous substrates. Here we adapt these results in the ecological context to find a versatile description of disperser movements.

3.2.1 Diffusion over fractal media

In our application, we imagine the flight of a forest-dwelling insect as a random walk in 2D. Unlike most random walk models, however, we suppose that obstacles in the environment inhibit movements to some degree. This better reflects flights constrained to lie within the complex network of gaps in the forest vegetation, or ones that track patchy distributions of habitat.

Our constrained domain of movement can be viewed abstractly as a porous medium. Its porosity, like many other aspects of forest structure, is more conveniently described using the mathematics of fractals (Zeide, 1991). The idea is not unfamiliar in forest ecology; Goodchild and Mark (1987) discussed fractal aspects of tree crown cover; hydrological networks; and topography. These structures exhibit power laws under scaling that allow their pertinent features to be summarized by simple parameters, such as the fractal (Hausdorff) dimension d_f (Seuront, 2010). For example, Jonckheere et al. (2006) improved predictions of light penetration in pine forests by estimating the d_f value in hemispherical images of scots pine canopy gaps, finding them to be highly fractal. We will use d_f to summarize the network of habitat navigated by an insect in flight.

d_f summarizes space-filling properties. In full 2D space the area enclosed in a disk of radius r scales as πr^2, whereas in an embedded fractal space it would scale as πr^{d_f} (Méndez, Campos, and Bartumeus, 2014). Unlike the topological dimension (in our case, $d = 2$), fractal dimension can assume non-integer values. One can define spaces with $0 < d_f < 2$ that nearly fill the plane, yet leave a complex arrangement of patches inaccessible, with the availability of habitat decreasing with d_f. For example, Hargis, Bissonette, and Turner (1999) used fractal dimension to characterize the quality of marten habitat, estimating a d_f in the range of 1.7-1.9 for pine and spruce forests in their Utah study area.

In the context of random walks, a fractal medium offers less space for movement. This prompts some adjustments of the balance law behind Fickian diffusion: Suppose $u(r,t)$ is a PDF for occupancy within the available space at radius r, at time t. If all of $\mathbb{R}^2$ were available, the density within the annulus Ω would be measured by $2\pi \int_\Omega u(r,t) r \, \mathrm{d}r$. On a d_f-dimensional fractal it is:

$$2\pi \int_\Omega u(r,t) p(r) r \, \mathrm{d}r \quad \text{where } p(r) = r^{d_f - 2}, \text{ and } 0 < d_f \leq 2. \tag{3.3}$$

The scaling function $p(r) = \pi r^{d_f} / \pi r^2$ is the proportion of the area inside radius r that is available for movement. For notational convenience we will suppress this dependence on r (and d_f) and simply write p. The function $u(r,t)$ is therefore an occupancy PDF with the (unusual) distance-scaled probability measure $pr \, \mathrm{d}r$. Under the more familiar Lebesgue measure, the PDF is $D(r,t) = u(r,t)p$.

O'Shaughnessy and Procaccia (1985) explained how the usual equation for Fickian diffusion may be modified in order to remain consistent with the scaling property (3.3). Their generalized 2D heat equation describes the time-evolution of $D(r,t)/p = u(r,t)$:

$$\frac{\partial u}{\partial t} = \frac{1}{pr} \frac{\partial}{\partial r} \left(\alpha p r \frac{\partial u}{\partial r} \right) \quad \text{with } \alpha > 0, \; u(r,0) = \frac{\delta(r)}{2\pi p r}, \tag{3.4}$$

where $\delta(r)$ denotes the 1D Dirac delta function (and $\delta(r)/2\pi pr$ its 2D analogue), representing the initial departure of the disperser from the origin. Notice the diffusivity αp scales with distance, approximating correlations in movements to due geometrical constraints (Metzler, Glöckle, and Nonnenmacher, 1994).

Under the condition $u \to 0$ as $r \to \infty$ (required for a valid PDF), the general solution to equation (3.4) is given in O'Shaughnessy and Procaccia (1985). Switching to Lebesgue measure, it can be written:

$$D(r,t) = u(r,t)p = \left(\pi \Gamma(d_f/2) (4\alpha t)^{d_f/2} \right)^{-1} r^{d_f - 2} \exp\left(-r^2/4\alpha t \right), \tag{3.5}$$

where $D(r,t)$ is the PDF for the position (r,θ) at time t of a random walker that departs the origin at $t = 0$ and diffuses through a medium with fractal dimension d_f. The effect of decreasing d_f is to make this density function more peaked and fatter-tailed relative to a 2D Gaussian kernel.

Note that $u(r,t)$ is simply a Gaussian kernel renormalized for measure $pr\,\mathrm{d}r$. Indeed when $d_f = 2$ we have $D(r,t) = u(r,t)$ (since $p = 1$), and equation (3.4) is the usual 2D heat equation in radial coordinates, with the 2D Gaussian kernel its well-known solution (Okubo and Levin, 2001). When $d_f = 1$ equation (3.4) simplifies to become a 1D diffusion equation (since $pr = 1$), so in the resulting *marginal* density function $D_r = 2\pi u(r,t)pr = 2\pi u(r,t)$ we find the 1D Gaussian kernel.

Settling *via* constant hazard

When building kernels directly from equation (3.4)-(3.5), one must assume that movement proceeds until a particular fixed time t. However, in reality the duration of dispersal is often stochastic. Unpredictable environmental factors such as temperature can compel dispersers to wait out unfavourable conditions (Jackson et al., 2008). Moreover, settling events can be prompted by chance encounters, such as the detection of a prey item (Turchin and Thoeny, 1993) or mate (Yasuda, 1975).

Preferring a model that accounts for randomly cued settling events, we suggest a simple extension of equation(3.4) that introduces a constant settling hazard $\lambda > 0$. We then define our kernel as the total settled density over all time. Thus we have $D = \int_0^\infty \lambda u p\,\mathrm{d}t$, with p defined as in equation (3.3), and:

$$\frac{\partial u}{\partial t} = \frac{\alpha}{pr}\frac{\partial}{\partial r}\left(pr\frac{\partial u}{\partial r}\right) - \lambda u \quad \text{with} \quad u(r,0) = \frac{\delta(r)}{2\pi pr} \quad \text{and} \quad \lim_{r\to\infty} u(r,t) = 0. \tag{3.6}$$

Here, the disperser moves about the domain as in (3.4) but settles at a randomly determined time, drawn from an exponential distribution with mean $1/\lambda$. $D(r)$ now expresses the PDF for position at the time of settling. To find D, one can integrate the PDE (3.6) over all time and consider weak solutions u (Appendix 3.7.3). The resulting kernels are the singular members of the WMY family:

$$D(r) = \int_0^\infty \lambda u p\,\mathrm{d}t = \mathcal{D}(r;\kappa,\rho) \quad \text{where } \kappa = d_f/2 - 1 \quad \text{and} \quad \rho^2 = \alpha/\lambda. \tag{3.7}$$

These WMY kernels ($-1 < \kappa \leq 0$) also emerge as the long-time limit of $u(r,t)p$ when the point source is stationary in time rather than instantaneous. To see this, we modify

(3.6) by viewing λ as a mortality hazard and adding a source term $F = \lambda\delta(r)/2\pi pr$ to the right-hand-side of the PDE. Dispersers are therefore continuously introduced from the origin, and continously removed throughout $\mathbb{R}^2$ in a density-dependent manner. The steady state in this smokestack-like system is $\lim_{t\to\infty} u(r,t)p = \mathcal{D}(r;\kappa,\rho)$, with κ, ρ defined as in (3.7) (Appendix 3.7.3).

The (non-fractal) case of $d_f = 2$ produces a Bessel kernel, or $\mathcal{D}(r;0,\rho)$. This model has a long history in ecology, with Broadbent and Kendall (1953) and Williams (1961) first using it to explain movements of worms and moths, respectively; and Awerbuch, Samson, and Sinskey (1979) later proposing it as a model for the biological activity of a chemical diffusing outward from the center of a petri dish.

In the data analysis of the latter study – and more recently in the bark beetle flight model of Turchin and Thoeny (1993) – a (long distance) asymptotic form $\mathcal{D}(r;0,\rho) \sim r^{-1/2}\exp(-r/\rho)$ was used in lieu of the Bessel, for computational convenience. Interestingly, this approximation is yet another special case of the WMY, with $\kappa = -1/2$; It arises from diffusion with constant settling when $d_f = 1$. Thus the WMY family generalizes both the Bessel and a commonly used approximation, with the PDE (3.6) revealing a common mechanistic explanation for both.

3.2.2 Multi-stage extensions

Redistribution events may naturally split into multiple stages. For example diurnal periods of flight activity occur in many insect orders (Hu et al., 2016). Moreover forest-dwelling insects like the MPB may initially fly in the unrestricted space above the canopy before switching to subcanopy dispersal (Jackson et al., 2008). In this section we look at two simple ways of extending (3.6) to model multi-stage processes.

Switching *via* constant hazard

The first idea is to connect each stage by a switching hazard with rate constant $\lambda > 0$. Writing $u_m(r,t)p$ for the density in the m^{th} stage, we assume an instantaneous point release of unit density initializes dispersers in the first stage, so that the dynamics of u_1 are the same as u in (3.6). As time progresses, this initial impulse trickles through subsequent stages, eventually exiting the n^{th} stage at rate λ (as settled density). We will be interested in the long-term settled density:

$$D(r) = U_n(r)p = \lambda \int_0^\infty u_n(r,t)p \, \mathrm{d}t, \text{ where} \tag{3.8a}$$

$$\frac{\partial u_1}{\partial t} = \frac{\alpha}{pr}\frac{\partial}{\partial r}\left(pr\frac{\partial u_1}{\partial r}\right) - \lambda u_1, \text{ with } u_1(r,0) = \frac{\delta(r)}{2\pi pr}, \text{ and} \tag{3.8b}$$

$$\frac{\partial u_m}{\partial t} = \frac{\alpha}{pr}\frac{\partial}{\partial r}\left(pr\frac{\partial u_m}{\partial r}\right) + \lambda u_{m-1} - \lambda u_m \tag{3.8c}$$

with $u_m(r,0) = 0$, for $1 < m \leq n$.

Assuming $\lim_{r\to\infty} u_m(r,t) = 0$, an analytic solution is available (Appendix 3.7.3):

$$D(r) = U(r)p = \left(2^{n-1+d_f/2}\pi\rho^2\Gamma(n)\Gamma(d_f/2)\right)^{-1}(r/\rho)^{n-2+d_f/2}K_{n-d_f/2}(r/\rho). \tag{3.9}$$

where $\rho^2 = \alpha/\lambda$. The PDF in (3.9) is simply the (renormalized) product of the WMY kernel $\mathcal{D}(r; n - d_f/2, \rho)$ with the scaling function p. Thus in the non-fractal case of $d_f = 2$ it produces the kernel family $D = U = \mathcal{D}(r; n-1, \rho)$. This extends the Bessel kernel ($n = 1$) to yield a sequence of distributions that are concave and bounded in their approach to the origin.

For $d_f < 2$, however, the distribution (3.9) remains singular at the origin for all n. In general, by increasing n we shift density away from the tails, effectively stalling dispersers near the origin. Decreasing d_f has the opposite effect, producing fatter tails and a highly peaked shape.

This approach of linking PDEs for n-stage processes was suggested by Neubert, Kot, and Lewis (1995) to describe (non-fractal) diffusion with settling on the real line. In their 1D system, the solution is a product of a Laplace kernel and a polynomial of order $n-1$. These same kernels were proposed by Schlägel and Lewis (2016) as extensions of the Laplace that are robust to changes in sampling frequency. We show in Appendix 3.7.3 that they emerge also from (3.8)-(3.9), as the *marginal* density functions D_r for the case $d_f = 1$.

Convolutions of WMY kernels

A simpler approach to the n-stage modeling problem is to suppose a long time delay separates stages. Thus within each stage, we assume the fractal diffusion process (3.6) operates until (nearly) all density has settled. The $t \to \infty$ limit in (3.7) then becomes initial data for the next stage. Under this assumption, the n^{th} stage settled density can be written as a convolution of n WMY kernels.

This is because when a population independently undergoes the redistribution process represented by D twice in succession, the resulting composite kernel is the autoconvolution $D * D$. More generally if a population undergoes a sequence of n independent redistribution stages described by the kernels $D^{(1)}, D^{(2)}, \ldots D^{(n)}$ then their combined effect is $D = D^{(1)} * D^{(2)} * \cdots * D^{(n)}$.

Certain kernel families have the property of closure under n-part convolutions. Chesson and Lee (2005) explained how this property aids interpretability, using it to develop redistribution kernels for lattice data. The Gaussian kernel is an example. The WMY is another, provided the range parameter $\rho > 0$ is fixed in all stages (Appendix 3.7.2). In particular if $\mathcal{D}^{(m)} = \mathcal{D}(r; \kappa_m, \rho)$ is a 2D WMY kernel,

$$\mathcal{D}^{(1)} * \mathcal{D}^{(2)} * \cdots * \mathcal{D}^{(n)} = \mathcal{D}\left(r; \kappa, \rho\right), \quad \text{where } \kappa = n - 1 + \sum_{m=1}^{n} \kappa_m. \tag{3.10}$$

Note that we may assume the shape parameters κ_m belong to $(-1, 0]$, since $0 < d_f \leq 2$. So by allowing arbitrary $n > 0$, the model (3.10) generates the complete set of thinner-tailed and bounded WMY kernels ($0 < \kappa < \infty$) not captured by the single stage model (3.6) in the previous section.

By adding stages we increase κ, shifting density away from the tails and origin, and towards the shoulders of the distribution. Thus as κ increases through zero, $\mathcal{D}$ becomes bounded at the origin; and as κ increases through $1/2$, its approach to the origin switches from convex to concave (Appendix 3.7.2). With further increases in κ, the WMY increasingly resembles a 2D Gaussian kernel; In fact if we parametrize $\rho^2 \propto 1/\kappa$, then $\mathcal{D}$ limits to a Gaussian kernel as $\kappa \to \infty$ (Stein, 1999).

When $\kappa = 1/2$ the WMY simplifies to $\mathcal{D}(r; 1/2, \rho) = (2\pi\rho^2)^{-1} \exp(-r/\rho)$, which has the same functional form as a (1D) Laplace kernel. The Laplace is the 1D analogue of the Bessel; it arises from diffusion with constant-hazard settling in $\mathbb{R}^1$ (Okubo and Levin, 2001). In the 2D setting, exponential decay of this form arises from turbulent limnological diffusion (Joseph and Sendner, 1958). However, in most ecological applications it is invoked simply as a phenomenological model (*eg.* as in Heavilin and Powell, 2008; Gilbert et al., 2017). Equations (3.6)-(3.7) and (3.10) provide a new mechanistic origin for this *2D Laplace* kernel. $\kappa = 1/2$ can arise, for example, in a two-stage process where the first stage of diffusion takes place in unrestricted 2D space (with $d_f = 2$) and the second in a fractal medium with $d_f = 1$ (so that $\kappa_1 = 0$ and $\kappa_2 = -1/2$).

Note that if all of the κ_m are identical, (3.10) expresses that the WMY family is robust to changes in sampling frequency. In fact, Schlägel and Lewis (2016) described the WMY family implicitly (in Fourier space) as their first example of a 2D kernel with this property. Equation (3.2) gives the explicit (back-transformed) density function for this kernel.

The kernel convolution approach – and resulting WMY family – appears well suited to modeling MPB flight patterns. In radar-based MPB flight surveys, Ainslie and Jackson (2011) observed a diurnal cycle of above-canopy dispersal events lasting multiple days. The $\mathcal{D}^{(m)}$ in that case would represent movements during the m^{th} day, and $\mathcal{D}$ in (3.10) their net outcome after n days of flying.

3.2.3 Flexibility in kurtosis

The foregoing derivations are meant to illustrate how WMY patterns of redistribution might arise in a very wide range of ecological systems. Inhomogeneous environments are the rule rather than the exception in ecology; It makes sense to relax the assumption of unrestricted movement while retaining it as a special case (the Bessel), as we did in the model (3.6). The natural extension in (3.10) generates a wide spectrum of kernel shapes, ranging from highly singular and heavy-tailed examples like the Bessel, to the highly smooth and thin-tailed Gaussian kernel. Let us now consider how this flexibility in shape is also attractive from a phenomenological standpoint.

One of the more important factors to consider when modeling redistribution is the balance of peakedness and tailedness in a kernel, or its *kurtosis*. In $\mathbb{R}^1$, kurtosis is the (standardized) fourth moment. In $\mathbb{R}^2$, we use the bivariate kurtosis measure $k(D)$ suggested in Clark et al. (1999). This is based on the zero-centered fourth *marginal* moment of distance r, or D_r, as defined in (3.1):

$$k(D) = \frac{\int_0^\infty r^4 D_r \,\mathrm{d}r}{\left(\int_0^\infty r^2 D_r \,\mathrm{d}r\right)^2}. \tag{3.11}$$

This definition is similar to the univariate case, but differs in that: we use D_r rather than D; and the moments in (3.11) are centered on zero, rather than the (nonzero) marginal mean of r. However, it summarizes the relevant behaviour of radially symmetric kernels in a concise way: Larger values of $k(D)$ indicate fatter tails (a higher density assigned to long-distance events) and a sharper peak at the origin (Clark et al., 1999). For the WMY, kurtosis is determined entirely by κ (Appendix 3.7.2):

$$k(\mathcal{D}) = 2\left(\frac{\kappa+2}{\kappa+1}\right), \text{ where } \kappa \in (-1, \infty). \tag{3.12}$$

The 2D Gaussian kernel sits at the low end of this range with a kurtosis of 2 (it is the $\kappa \to \infty$ limit of $\mathcal{D}$). The WMY family extends it to capture a wide range of leptokurtic (higher than Gaussian) alternatives. The 2D Laplace, for example, has kurtosis $10/3$; For the Bessel it is 4. The 1-stage diffusion model (3.7) spans $(4, \infty)$, with extremely fat-tailed examples emerging as d_f becomes small. This kind of flexibility in kurtosis is of particular importance in dynamical systems for population spread, where the tail behaviour can determine both the success and speed of invasions during a range expansion (Kot, Lewis, and Driessche, 1996). For example, a model that uses a (thin-tailed) Gaussian kernel out of mathematical convenience might fail to account for tail-like long-distance range expansion events.

Range expansions are highly relevant in models of MPB populations, which carry out yearly dispersal flights. While this flight is often adequately characterized by Gaussian kernels (Heavilin and Powell, 2008), radar evidence suggests that wind-assisted movements, carrying MPB very long distances, are not uncommon (Ainslie and Jackson, 2011). One such long-distance event recently carried populations across a mountain range from BC into central Alberta, Canada – a range expansion with severe ecological and economic consequences (Giroday, Carroll, and Aukema, 2012). Equation (3.12) shows how by using WMY kernels, the probability of tail-like events in a model can be finely tuned using κ.

3.3 Anisotropic extensions

While mathematically pleasant, the assumption of isotropic redistribution is often unsatisfactory in ecological applications. Wind, for example, is of course directional, and wind-assisted migratory flights occur regularly in insects (Jackson et al., 2008). Indeed such events are central to explaining the recent range expansion of the MPB (Giroday, Carroll, and Aukema, 2012).

Isotropic kernels are nevertheless useful as a jumping-off point in model development. We show in this section how they can be used as building blocks towards phenomenological kernels with a more realistic degree of flexibility.

3.3.1 Advection

Our first extension approximates of the effect of drift. This is the passive transport of individuals along a certain direction for a certain distance – as a result of wind, for example – moving them along a displacement vector which we will call $\boldsymbol{\tau} := \boldsymbol{x}' - \boldsymbol{x}$. We will assume spatial homogeneity in this effect, so that all individuals drift in the same way, regardless of their position.

In the absence of diffusion, drift amounts to a translation of coordinates. We denote this *advection* operator by $D_{\boldsymbol{\tau}}$, and define it by its action on a population density $N(\boldsymbol{x})$:

$$D_{\boldsymbol{\tau}} * N(\boldsymbol{x}) = N\left(\boldsymbol{x} + \boldsymbol{\tau}\right). \tag{3.13}$$

Drift is typically incorporated into diffusion models by introducing an advection term into the PDE for movement density (Okubo and Levin, 2001). Doing so allows the two drivers of movement – advection and diffusion – to act simultaneously. By contrast, we assume the two drivers act in sequence. Composing advection with a kernel representing diffusion, we obtain the model $D * D_{\boldsymbol{\tau}}$.

This is of course a simplification. However, in the case of bark beetles (like the MPB) it may be closer to the reality. Experimental observations by Shegelski, Evenden, and Sperling (2019) and survey data in Jackson et al. (2008) suggested behavioural adaptations in this genus for "drifting": individuals ride thermals by extending – but not flapping – their wings, exerting substantial flight effort only to maintain an optimal altitude. During these drifting periods we may view the diffusion process to have slowed, as the beetle ceases its horizontal flight effort and is passively transported by the wind.

The kernel $D * D_{\boldsymbol{\tau}}$ can be interpreted as a two-part MPB flight model: individuals first fly upwards to catch the wind and drift along vector $\boldsymbol{\tau}$. This allows them to escape heavily infested stands where host depletion would limit their ability to reproduce. After some time they descend back below the canopy and begin a diffusive flight (D), in search of new hosts and mates.

3.3.2 Geometric anisotropy

Another type of anisotropy in redistribution is introduced by directed search behaviours like phototaxis and chemotaxis. Isotropic diffusion might reasonably characterize an undirected search, but a MPB search flight is coordinated by a range of cues, including temperature, stem silhouettes, and chemical signatures of hosts and conspecifics (Safranyik and Carroll, 2006).

A full accounting of such a complex system is a formidable modeling problem (Powell and Bentz, 2014). We propose a far simpler phenomenological model: Suppose the net effect of these directional cues is to deform the circular contours of an isotropic kernel D into ellipses with by semi-axis lengths Δ_x and Δ_y, and oriented along angle θ. This is accomplished by a linear transformation of coordinates, defining a directional version of D:

$$D^{\boldsymbol{a}}(\boldsymbol{r}) := D(r_{\boldsymbol{a}})/(\Delta_x \Delta_y) \tag{3.14}$$

$$\text{where} \quad r_{\boldsymbol{a}} = |\boldsymbol{A}\boldsymbol{r}| \quad \text{and} \quad \boldsymbol{A} = \begin{pmatrix} \cos\theta/\Delta_x & -\sin\theta/\Delta_x \\ \sin\theta/\Delta_y & \cos\theta/\Delta_y \end{pmatrix}.$$

The same technique is often suggested in statistics for extending isotropic covariograms to incorporate geometric anisotropy (Stein, 1999). We see it far less often with redistribution kernels. However, the observation of ellipsoid clusters of MPB damage in practice (*eg.* Figure 3.4 of Section 3.5) suggests that the contours of a MPB flight kernel should admit these shapes. Equation (3.14) introduces the necessary flexibility. It is easy to implement, and its effect on D is easy to visualize and understand. Figure 3.1 (left and right panels) illustrates the two-part kernel $\mathcal{D}^{\boldsymbol{a}} * D_{\boldsymbol{\tau}}$, where the WMY is used for the diffusive stage.

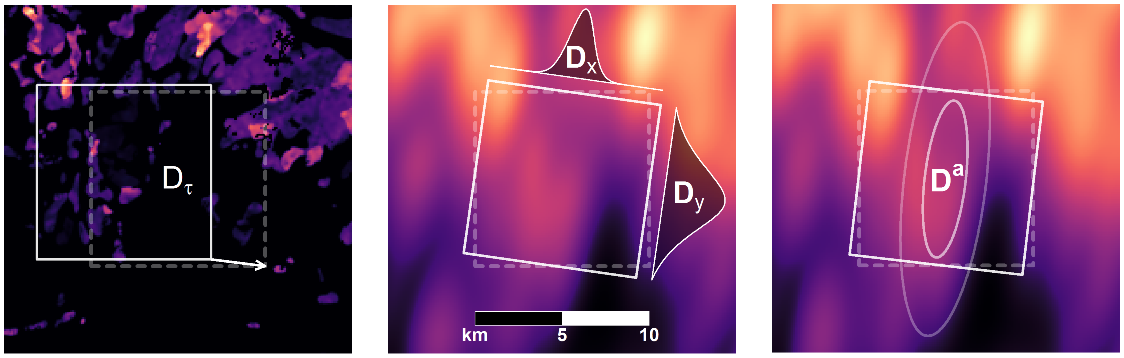

Figure 3.1: Geometry of advection ($D_{\boldsymbol{\tau}}$) and geometric anisotropy ($D^{\boldsymbol{a}}$) demonstrated for two redistribution kernels: the WMY and the separable product-WMY. Source densities are first translated by the vector $\boldsymbol{\tau}$ to represent initial drift (left). Geometric anisotropy then rotates/scales the coordinate system (right) prior to the kernel convolution, producing diffusion ellipses. With the product-WMY (middle), this convolution is formulated as a composition of two 1D WMY kernels.

.

3.3.3 Product-WMY kernels

Our last extension employs a simple trick for constructing anisotropic 2D kernels from 1D isotropic ones; We take the product of a 1D kernel applied along the x-axis followed by another (possibly different) one applied along the y-axis. This builds 2D kernels with the computationally desirable property of *separability*. We will see in the next section how this property can resolve issues of computational complexity that would otherwise make the kernel-based analyses of large datasets infeasible.

Separability is the property that $D(\boldsymbol{r})$ factors into the product of two 1D kernels, $D_x(r_x)$ and $D_y(r_y)$, where $\boldsymbol{r} = (r_x, r_y)^T$, each depending on only one of the spatial dimension. The 2D Gaussian kernel is an example: it has the form $D_x(r_x)D_y(r_y)$, where $D_x(r) = D_y(r) = (\sqrt{\pi}\rho)^{-1}\exp(-(r/\rho)^2)$ are 1D Gaussian kernels. The 2D WMY kernel, on the other hand, is not separable. In seeking a computationally efficient alternative we propose the *product-WMY*:

$$\mathcal{D}_{\otimes}(\boldsymbol{r}; \kappa_x, \kappa_y, \rho_x, \rho_y) = \frac{\mathcal{D}(|r_x|; \kappa_x, \rho_x)\mathcal{D}(|r_y|; \kappa_y, \rho_y)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{D}(|r_x|; \kappa_x, \rho_x)\mathcal{D}(|r_y|; \kappa_y, \rho_y)\, \mathrm{d}r_x\, \mathrm{d}r_y}. \tag{3.15}$$

where $\boldsymbol{r} = (r_x, r_y)^T$. Here, the component kernels D_x and D_y are 1D WMY kernels. These have the same functional form as the 2D WMY (3.2), differing only in normalization constants. Thus both component kernels limit to a 1D Gaussian kernel in the limit of large shape and small scale parameters (Stein, 1999). This implies that $\mathcal{D}_{\otimes}$ limits to a 2D Gaussian kernel (Figure 3.2), much like the WMY. Moreoever it is easily shown via Fourier transforms that, with fixed range parameters, the family $\mathcal{D}_{\otimes}$ is closed under n-part convolutions (Appendix 3.7.2), leading to an identity analogous to (3.10).

Rewriting $\mathcal{D}_{\otimes}$ in polar coordinates reveals an angular dependence for any choice of parameters. The product-WMY is always anisotropic to some degree. For example, setting $\kappa_x = \kappa_y = 1/2$ and $\rho_x = \rho_y = \rho$ yields $\mathcal{D}_{\otimes} = \exp(-(|r_x| + |r_y|)/\rho)$, a 2D Laplace kernel imbued with a Manhattan distance metric, for which the anisotropy is quite pronounced (Figure 3.2, left). Such a model might for example be appropriate for animals following

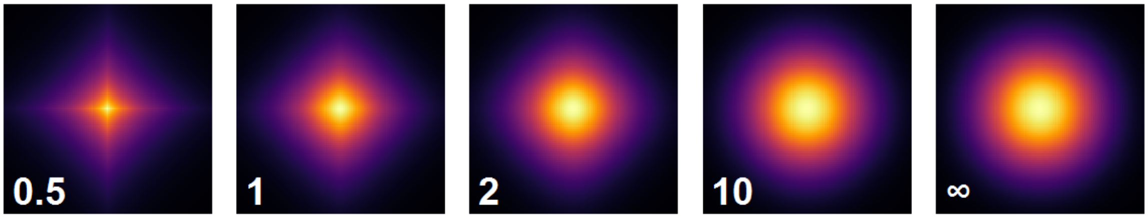

Figure 3.2: Heatmaps of the product-WMY kernel redistributing a point mass about the center of a 100 × 100 lattice of sites. Shape parameter values, $\kappa = \kappa_y = \kappa_y$ are indicated in the plots. With range parametrized as $\rho_x = \rho_y \propto 1/\sqrt{\kappa}$, the kernel approaches isotropy as κ increases while the effective range stays fixed, with fast convergence to a Gaussian kernel for large κ (right).

anthropogenic features in the landscape, such as seismic lines (Painter and Hillen, 2018), provided they form a dense grid.

In spite of the inherent anisotropy in $\mathcal{D}_\otimes$, visual inspection suggests it closely resembles $\mathcal{D}$ when κ is not too small (*eg.* Figure 3.1, middle). A numerical investigation of pointwise relative error (RE) between the two kernels supports this observation, indicating that for WMY kernels with moderate shape values of around $\kappa \approx 2$ and higher, $\mathcal{D}_\otimes$ is a good approximation, provided the sampling density is high enough (Appendix 3.7.4). However it also suggests poor approximations to the Bessel ($\kappa = 0$) and 2D Laplace ($\kappa = 1/2$), two popular choices for modeling MPB movements. As there is very little empirical data on MPB dispersal habits to inform the choice of κ, we sought to estimate it in the next section, by fitting $\mathcal{D}$ to historical records of MPB activity.

3.4 Fitting redistribution kernels to data

Redistribution kernels are the central feature of the integrodifference equation (IDE), a model that combines growth (G) with spatially explicit dispersal (D) for a population density variable $N_t(\boldsymbol{x})$ indexed continuously in space and discretely in time (Kot and Schaffer, 1986). D is stationary in the most well-studied case, and the IDE reads $N_t(\boldsymbol{x}) = (D * G(N_{t-1}))(\boldsymbol{x})$, where $*$ denotes convolution. A rich theory has emerged in connecting the mathematical properties of the function D with the spread dynamics of N_t as t becomes large (see *eg*. Kot, Lewis, and Driessche, 1996; Lewis, 2000).

However, redistribution kernels are also often applied in analyses of data from individual and aggregated years (*eg*. Heavilin and Powell, 2008; Clark et al., 1999). In this more general setting, $\tilde{N}(\boldsymbol{x}) = (D * N)(\boldsymbol{x})$ simply connects the pre- and post-dispersal levels in a population, and the modeler proposes a function g to connect $\tilde{N}(\boldsymbol{x})$ with the mean of a measurable response $Y(\boldsymbol{x})$,

$$\mathbb{E}\left[Y(\boldsymbol{x}) \mid N(\boldsymbol{x}), \boldsymbol{\beta}\right] = g\left(\boldsymbol{x}, \tilde{N}(\boldsymbol{x})\right) = g\left(\boldsymbol{x}, (D * N)(\boldsymbol{x})\right). \tag{3.16}$$

where $\boldsymbol{\beta}$ is the vector of all model parameters, including those of D and g.

3.4.1 Discretization and error structure

In practice, the response $Y(\boldsymbol{x})$ is observed over a finite set of sites on the plane, with centroids $\{\boldsymbol{x}_k\}$ $(k \in 1, \ldots n)$. The (continuous) kernel convolution $D * N$ must therefore be replaced with a discretized version: In place of D one specifies an $n \times n$ *redistribution matrix* $\boldsymbol{D}$, where entry $[\boldsymbol{D}]_{ij}$ assigns density to the redistribution event from location $\boldsymbol{x}_j$ to $\boldsymbol{x}_i$. The simplest way to construct this matrix is by means of the pointwise kernel density values in D (Chipperfield et al., 2011):

$$[\boldsymbol{D}]_{ij} = \frac{D(\boldsymbol{r}_{ij})}{\sum_{i,j} D(\boldsymbol{r}_{ij})} \quad \text{where} \quad \boldsymbol{r}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j \tag{3.17}$$

Let vector $\boldsymbol{y}$, with entries $\boldsymbol{y}_k = Y(\boldsymbol{x}_k)$, be the observed response data. Similarly, define the pre and post-dispersal vectors by $\boldsymbol{n}_k = N(\boldsymbol{x}_k)$ and $\tilde{\boldsymbol{n}}_k = \tilde{N}(\boldsymbol{x}_k)$. The discretized analogue of model (3.16) then becomes $\mathbb{E}\left[\boldsymbol{y}_k \mid \boldsymbol{n}, \boldsymbol{\beta}\right] = g\left(\boldsymbol{x}_k, \tilde{\boldsymbol{n}}_k\right)$, where through (3.17), the parameters of D define $\tilde{\boldsymbol{n}} = \boldsymbol{D}\boldsymbol{n}$.

Real measurements of $\boldsymbol{y}$ will of course deviate from this expected value. Model-fitting techniques are built around notions of minimizing these *model residuals*, $\epsilon_k = \boldsymbol{y}_k - g\left(\boldsymbol{x}_k, \tilde{\boldsymbol{n}}_k\right)$, given their collective sampling distribution. This error distribution is context-dependent, and reflects a careful consideration of how process and measurement error are manifested in the $Y(\boldsymbol{x})$.

Modelers will seek a balance of mathematical convenience and realism that works for their choice of $Y(\boldsymbol{x})$, and their data. For example in the IDE of Goodsman et al. (2016), the ϵ_k are independent Poisson random variables, reflecting uncertainty in the number of infested pines; In the "red-top" model of Heavilin and Powell (2008) they are independent Gaussians, as justified by (large-sample) asymptotic results when $Y(\boldsymbol{x})$ is an *estimating function* (Lele, Taper, and Gage, 1998); and in Appendix 3.7.5 we describe a modified red-top model where $\boldsymbol{\epsilon}$ is multivariate Gaussian, reflecting both the multiplicative process error in beetle populations and their spatial-autocorrelation.

3.4.2 Computational aspects of maximum likelihood

Having characterized the error distribution, the modeler obtains a likelihood function $\mathcal{L}(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{n})$, that is maximized to estimate $\boldsymbol{\beta}$. Regardless of how the distribution of ϵ_k has been defined, $\mathcal{L}$ will be functionally dependent on $\boldsymbol{\epsilon}$, which in turn depends on $\tilde{\boldsymbol{n}}$. Thus if $\mathcal{L}$ is to be optimized numerically, we encounter a fundamental issue of computability with large n: The matrix product $\tilde{\boldsymbol{n}} = \boldsymbol{D}\boldsymbol{n}$ (in $\mathcal{L}$) must be repeatedly evaluated, and this operation has $O(n^2)$ complexity.

However there are remedies: When the sampling locations $\{\boldsymbol{x}_k\}$ are arranged in a regular lattice of dimensions $n_y \times n_x = n$, multiplications with $\boldsymbol{D}$ can be done using fast Fourier transforms (FFTs) (Andersen, 1991), reducing the complexity from $O(n^2)$ to $O(n \log n)$. Better still, if D is spatially separable, the complexity can be further reduced to $O(n_x^2 + n_y^2)$, without the use of FFTs. On square domains, that is $O(n)$, a considerable improvement. The implementation of this trick is straightforward, but notationally awkward, so we relegate the details to Appendix 3.7.5.

Two practical advantages of the product-WMY kernel defined in (3.15) are now clear. Like the Gaussian kernel, it can be evaluated far more quickly than a nonseparable kernel such as the WMY (Figure 3.3), allowing a more comprehensive and expedient search of the parameter space for the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$. Unlike the Gaussian kernel, it also has the flexibility to closely mimic the more realistic WMY family over a large part of the its parameter space. In the next section we show how this flexibility can also lead to improvements in model performance.

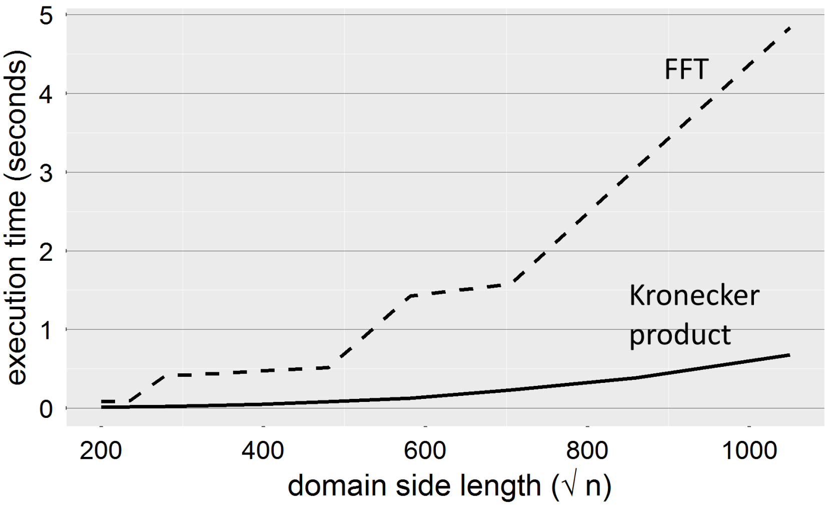

Figure 3.3: Time to compute the length-n post-dispersal vector $\boldsymbol{Dn}$, for redistribution matrices from the WMY (dashed line) and product-WMY (solid line) kernels. The product-WMY is computed using Kronecker products (Appendix 3.7.5), while the nonseparable WMY uses FFTs.

3.5 Application: damage patterns of the mountain pine beetle

To demonstrate the ideas of Section 3.2-3.4 on a real ecological problem, we use the example of MPB outbreaks. These insects kill pine trees on an annual basis in forests of Western North America, exhibiting eruptive population dynamics with outbreak periods of elevated activity that can persist for decades (Safranyik and Carroll, 2006). Tree mortality becomes visible on the landscape during these outbreaks. Clusters of faded crowns (dead pines, or "red tops") spread through space from year to year, reflecting macro-scale movement patterns of the beetle (Chen and Walton, 2011).

Aerial overview surveys (AOS) of MPB damage are carried out by the provincial government annually. We rasterized these AOS data (*sensu* Chen et al., 2015) to a 100×100

m (1 ha) resolution for the years 2007-2008, over a 100 × 100 km area ($n = 10^6$) centered on the pine-rich Merritt Timber Supply Area (Figure 3.4). This provides a window into redistribution at the height of the latest MPB outbreak in Southern BC, in one of its most heavily damaged areas (Boone et al., 2011).

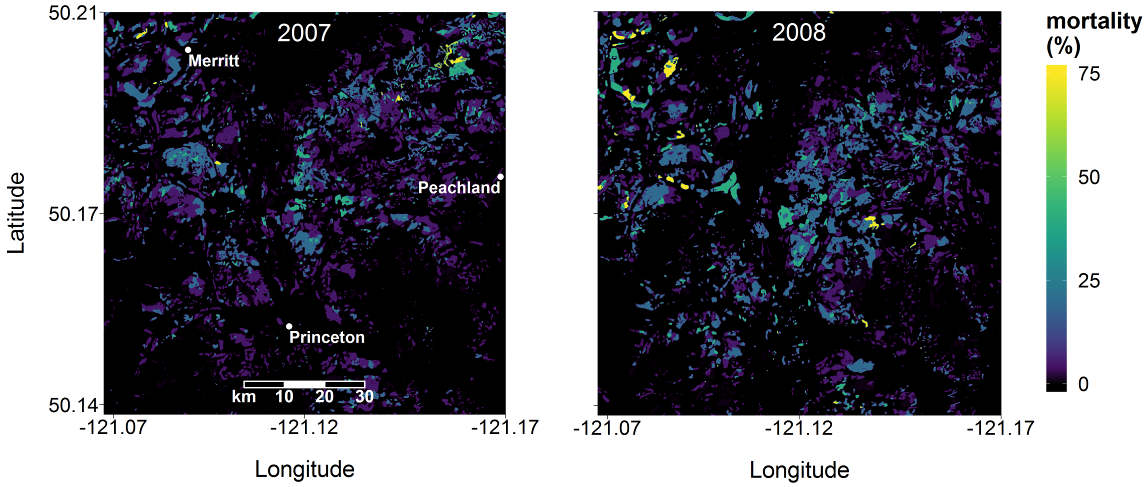

Figure 3.4: AOS survey of pine mortality due to MPB attacks in the summers of 2007-2008.

3.5.1 Growth and dispersal model

We used a slight modification of the red-top model of Heavilin and Powell (2008) to model the AOS data. Their IDE relates the pine mortality in a given year to its value in the following year. Pine deaths are used as a proxy for the beetle population in this model, and dispersal across the landscape is represented by the kernel D. A single growth dynamics parameter $\alpha > 0$ parametrizes a nonlinear growth function representing the dynamics of beetle-pine interactions from year to year. A more detailed description of the red-top model is given in Appendix 3.7.5.

To better suit our BC dataset, we make three changes: First, based on observations by Boone et al. (2011), we introduce a small local endemic MPB population that is constant and uniform in space; Second, we relax the Gaussian kernel redistribution assumption, testing it along with the alternative hypotheses of WMY and product-WMY kernels; Third, finding the model errors to be highly correlated in space, we assume a multivariate Gaussian structure for the residuals. This allows us to specify a covariance matrix $\boldsymbol{V}$, parametrized by variance σ^2 and range $c > 0$, to account for spatial auto-correlation. This model and its likelihood function are detailed in Appendix 3.50.

In fitting the modified red-top model to these data, we considered three questions: Which WMY kernel (*ie.* which κ) best characterizes MPB spread? Is the product-WMY a useful surrogate for the WMY? And do the extensions in Section 3.3 increase explanatory power?

3.5.2 Kernel comparisons on blocks

Although we have focused on stationary redistribution patterns in this paper, we must accept that local conditions influence MPB movements to some degree. To account for these variations we divided the data into a set of $N = 81$ nonoverlapping blocks, each of size 10×10 km. We assumed that beetle flight was reasonably homogeneous within each of these smaller regions. This allowed us to exploit the results on stationary kernels developed in Sections 3.3-3.4, by fitting the modified red-top model separately to each block. Each model fit is viewed as a trial in which the within-block errors are evaluated to measure predictive power. By pooling the results over all blocks, we can compare the overall performance of different kernels (Figure 3.5).

Within each block, we randomly selected 25% of the response data to withhold as a test set. Likelihoods were then numerically optimized over the remaining within-block training data, and the resulting MLE of $\boldsymbol{\beta}$ used to compute $\mathbb{E}\left[Y(\boldsymbol{x})\right]$ over the test set, conditional on the training data.

Comparing these predictions with the observed data $Y(\boldsymbol{x})$, we calculated two model performance statistics: root mean squared prediction error (RMSPE) which views the errors as independent points; and log-likelihood (LL), which corrects RMSPE for spatial auto-correlation using $\boldsymbol{V}$.

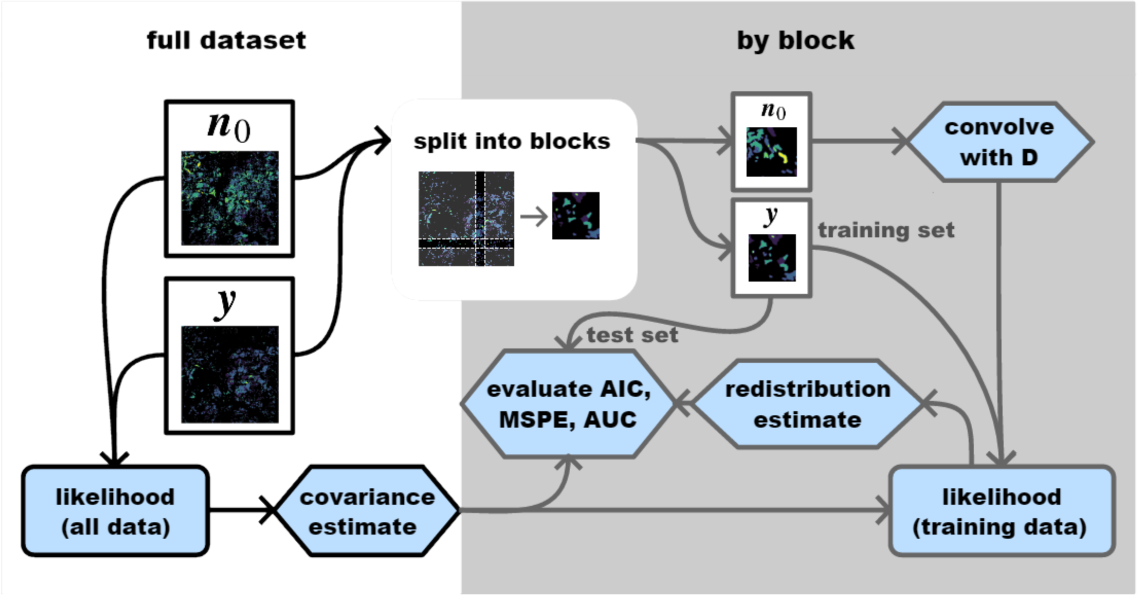

Figure 3.5: A process for comparing redistribution kernels. At left: the covariance matrix is estimated by maximizing likelihood assuming no dispersal. The dataset is then split into nonoverlapping blocks, each randomly partitioned into test and training sets. At right: within-block likelihood is optimized over the training data, and subsequent test set predictions are used to estimate error. Pooled errors from all blocks are compared for different redistribution kernels (D).

The parameters of interest are those of D, whereas the variance σ^2 and range c which determine the covariance matrix $\boldsymbol{V}$ are essentially nuisance parameters. We therefore estimated $\boldsymbol{V}$ prior to the blockwise analysis (using the combined training data from all blocks), by numerically maximizing the likelihood function (Appendix 3.50, equation 3.52) with D fixed to an initial guess. For this initial D we used the Gaussian kernel with a biologically plausible range of $\rho \approx 250$m (Heavilin and Powell, 2008).

Thus, we assumed that each datapoint shares a common covariance structure, and that the difference between this true $\boldsymbol{V}$ and the one estimated under an initial (simpler) redistribution model is negligeable. This is reasonable given that when using the red-top model in practice, we have found the MLEs for $\boldsymbol{V}$ (estimated jointly with D) to be very similar under quite different redistribution kernels. Moreover this approach avoids the complication of accounting for differences – among blocks and kernels – in the sampling distribution of our

MLE for $\boldsymbol{V}$. It also reduced the number of unknowns to be estimated later over the relatively small sample sizes ($n = 10^4$) within blocks, allowing the numerical solver to converge more quickly in the blockwise analysis.

In each block, we separately fitted kernels from three families: the Gaussian; the WMY (with $\kappa > 0$, to exclude singularities at $r = 0$); and the product-WMY (with $\kappa_x, \kappa_y > 0$). For each of these families we fitted the kernel with and without the geometric anisotropy and advection extensions of Sections 3.3.2 and 3.3.1 (*eg.* for the product-WMY we fit both $\mathcal{D}_\otimes$ and $\mathcal{D}_\mathcal{T} * \mathcal{D}_\otimes^{\boldsymbol{a}}$).

For both the covariance and within-block MLEs, the likelihood function was optimized numerically using the *hjkb* function (Hooke and Jeeves, 1961) in the R package *dfoptim*, to a maximum of 10^4 iterations. Translations and rotations of coordinates were computed using the image processing libraries in package *imager*. Separable kernel convolutions were computed using Kronecker products (Appendix 3.50) and nonseparable ones using FFTs as implemented in the package *smoothie*. The Bessel function in (3.2) was calculated using R's built-in function *besselK*.

3.5.3 Kernel comparison results

Figure 3.6 summarizes the results from the 81 trials by centering each test statistic on that of the best performing model for that block. Like Akaike's information criterion (AIC), the boxplots provide a ranking of model performance. Indeed negative LL on test data produces the ranking that is intended by AIC, as the two statistics approximate the same thing – (relative) expected Kullback-Liebler information loss (Burnham and Anderson, 2004). Averaging LL over all blocks, the extended product-WMY ($\mathcal{D}_\mathcal{T} * \mathcal{D}_\otimes^{\boldsymbol{a}}$) ranked highest (best) among the six kernels.

For all three kernel familes, the advection and geometric anisotropy extensions improved test set performance substantially, increasing mean LL and decreasing RMSPE. Though evaluations of $\mathcal{L}$ were only marginally slower, convergence of the optimizer required more

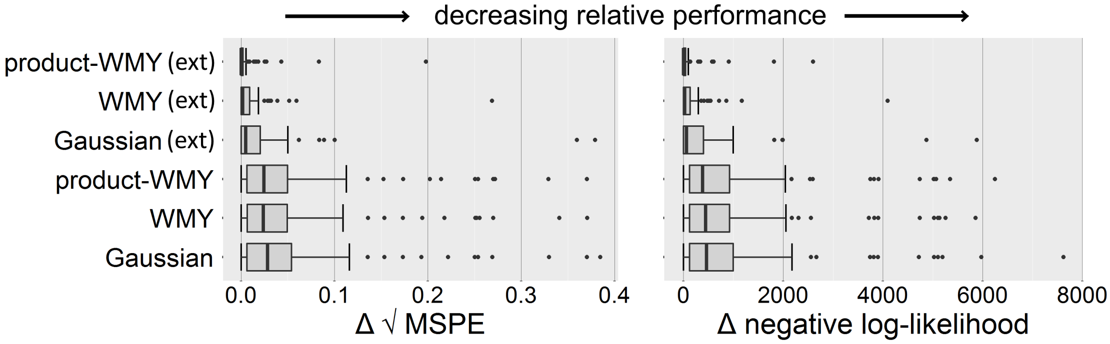

Figure 3.6: Model performance on hold-out sets for six redistribution kernels over 81 model-fitting trials: The Gaussian, WMY, and product-WMY; along with extended versions (indicated by "ext") incorporating geometric anisotropy and advection. RMSPE and negative log-likelihood were computed in each trial, and their differences (Δ) with the best model within that trial are summarized as boxplots. Extreme outliers (points to the right of the boxplot whiskers) were associated with blocks having very low levels of MPB activity, where it appears there was insufficient information to reliably parametrize movement patterns.

iterations, so computing time increased by around 4X. This was in the neighbourhood of 8 minutes for the (separable) Gaussian kernel and product-WMY to 16 minutes for the (nonseparable) WMY.

Comparing among the three kernels without the advection and geometric anisotropy extensions, the WMY was slightly favoured by both LL and RMSPE (in averages over blocks). The performance of the product-WMY was very similar to that of the isotropic WMY, with both kernels reducing errors dramatically compared to the Gaussian kernel on certain blocks having low estimated κ.

We observed a wide spread of κ estimates among the isotropic WMY kernels. On only 23% of blocks did the fitted shape parameter indicate a Laplace/Bessel-like spread pattern with convex behaviour at the origin ($\kappa \leq 1/2$). On 34% of blocks, κ stalled at its upper bound ($\kappa = 25$), indicating a tendency towards Gaussian-like spread patterns; and on the remaining 43% it was intermediate, with a combination of fat tails and a concave behaviour at the origin.

Interestingly, only 5% of the fitted isotropic WMY kernels lay in the region of good approximability (RE $<$ 5%) indicated by numerical simulations (Appendix 3.7.4). Nevertheless the product-WMY appeared to match or exceed the performance of the WMY overall.

3.6 Discussion

Model selections by Heavilin and Powell (2008) and Goodsman et al. (2016) on MPB damage datasets similar to ours have led to somewhat contradictory conclusions on the character of MPB movement patterns: In the latter, the Bessel is favoured over the Gaussian kernel, and the authors argued for a movement mechanism with constant settling hazard; In the former, the Gaussian kernel is favoured over a 2D Laplace, suggesting abrupt settling events that lead to thin-tailed movement patterns.

This inconsistency can be resolved by simply dropping the assumption of single-stage redistribution. We saw in Section 3.2.2 that the n-stage model $\mathcal{D}(r; n-1, \rho)$ for Fickian diffusion with constant settling hazard captures the behaviour of both the Bessel ($n = 1$) and thinner-tailed kernels like the Gaussian (large n). These two classes of redistribution patterns, though qualitatively very different, can emerge from the same mechanism, differing only in the number of times the flight event is interrrupted and restarted (*eg.* by fluctuations in temperature). Since it is reasonable to assume that this number may differ when populations are sampled in different geographical areas, we can retain the parsimonious assumption of constant settling hazard – a hypothesis supported by the laboratory flight-mill experiments of Evenden, Whitehouse, and Sykes (2014) and Shegelski, Evenden, and Sperling (2019).

Notice that because a sum of exponential settling times has a gamma distribution, our n-stage model is a particular case of the original formulation of the WMY by Yasuda (1975). However, unless we abandon the constant settling hazard assumption, neither interpretation accounts for WMY kernels with non-integer shape parameters. We showed in Sections 3.2.1 and 3.2.2 how these gaps in the parameter space of the WMY family can be filled by relaxing the assumption of unobstructed diffusion. By modeling the heterogeneous medium of a forest using fractal scaling properties (summarized by dimension d_f), we recovered the full complement $-1 < \kappa < \infty$ in the multi-stage formulation (3.10), along with a novel family of singular kernels in (3.9).

This explains the wide range of WMY shape parameter estimates in the analysis of Section 3.5 – most of which were non-integer. At different geographic locations, we expect both n and d_f to vary, and thus so does κ. The standard models listed in Table 3.1 (Bessel, 2D Laplace, Gaussian kernels), on their own, lack the flexibility to capture the full range of redistribution patterns exhibited by the AOS data. The WMY, which extends all three, can therefore be expected to improve model performance. Though our results suggested a Gaussian kernel was appropriate over large parts of the study area, the more flexible WMY led to improvements overall, in both RMSPE and LL.

In Section 3.2.3, we explained how this notion of flexibility can be made more precise, using 2D kurtosis. The WMY admits an exceptionally large range of kurtosis values. To our knowledge, the only other redistribution kernel with this flexibility is the 2Dt, a phenomenological model of seed shadow distributions (Clark et al., 1999). In fact these two kernels are closely related: the 2Dt and WMY are Fourier duals; *ie.* the spectral density of the WMY has the same functional form as the 2Dt kernel (and vice versa). This is because the WMY can be derived by placing a gamma prior on the range parameter of the Gaussian kernel, whereas the 2Dt is derived using an *inverse* gamma.

Kurtosis is useful in quantifying the degree to which a kernel assigns importance to long-distance movements, or its tail behaviour. Fat-tailed kernels (with high kurtosis) would assign a higher redistribution density to the types of long-distance flight events that led to the recent expansion of the historical range of the MPB. On the other hand, it appears the vast majority of MPB dispersal flights are short-range, for which a thin-tailed kernel is more appropriate. Given that models for MPB outbreak risk are unavoidably sensitive to local beetle density estimates (Nelson et al., 2008), we therefore suggest that MPB modelers in search of an isotropic kernel family should prefer those having a high degree of flexibility in tail behaviour, such as the WMY or 2Dt.

For the MPB, long distance flight events are most likely driven by wind (Giroday, Carroll, and Aukema, 2012). We conceptualize this as a drifting behaviour followed by a diffusive flight, motivating the development of an advection extension in Section 3.3.1 that differs from the more typical advection-diffusion PDE based models. We think this could prove useful in spatial models for the flights of a wide range of other insects. Hu et al. (2016) described how a surprisingly diverse assemblage of flying insects appear to orient themselves towards a downwind direction, suggestive of the same type of drifting behaviour. Moreoever, our formulation may also help to account for measurement errors, since in visual surveys (like the AOS), precision issues may arise simply as a result of operators marking the wrong location on the map. In addition to representing the length and direction of a drifting event, the advection vector $\boldsymbol{\tau}$ can serve to re-center imprecise datapoints.

In Section 3.3.2, we proposed geometric anisotropy as an extension of general isotropic kernels to better match the diffusion ellipses often seen in real-world survey data. Though we present it as a phenomenological model, a mechanistic justification may be forthcoming: Note that if D is the Gaussian kernel then (3.14) solves the standard 2-dimensional anisotropic diffusion equation with diffusion tensor $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$, as derived in Painter and Hillen (2018). In future work we hope to pursue a similar derivation incorporating the scaling properties discussed in Section 3.2.1.

The results in Figure 3.6 showed that these two anisotropy extensions improved model predictive performance across the board. This is important because it demonstrates that overfitting was not an issue on the MPB data, in spite of the introduction of an additional 4 parameters. In Chapter 4, we show how a complex nonstationary pattern of MPB movements can be revealed by fitting anisotropic kernels blockwise in the manner of Section 3.5.2. Since these anisotropy extensions are simple to understand and implement with any stationary kernel, we think they could be used to study directionality in a wide range of ecological redistribution problems.

Moreover, in combining these extensions with the novel product-WMY kernel, we achieved a higher level of predictive performance than even the WMY. This separable kernel family takes the best of both worlds: by design it is computationally fast and simple, like the Gaussian kernel, yet it closely approximates the more flexible WMY over a large part of its parameter space, and appears to consistently match it in predictive performance on test data of MPB damage patterns.

It is often pointed out that difficulties in tracking the flight habits of an insect as small as the MPB hinders the study of its life cycle and its impacts on the forest ecosystem (Nelson et al., 2008). Our hope is that by providing new modeling tools for the redistribution phase, we might open new avenues for research on cryptic dispersal processes in ecology.

3.7 Appendices to Chapter 3

3.7.1 The modified Bessel functions

We define the WMY kernel by means of $K_\kappa(z)$, a modified Bessel function of the second kind. These functions emerge from the following ordinary differential equation (ODE) for $y(z) : \mathbb{C} \to \mathbb{C}$,

$$z^2 y'' + z y' - \left(z^2 + \kappa^2\right) y = 0.$$

The general solution of this ODE can be written in the form $c_1 I_\kappa(z) + c_2 K_\kappa(z)$ (for arbitrary $c_1, c_2 \in \mathbb{C}$), where I_κ and K_κ are the modified Bessel functions of order $\kappa \in \mathbb{C}$, defined by:

$$I_\kappa(z) = z^\kappa \sum_{m=0}^{\infty} \frac{z^{2m}}{2^{2m+\kappa} m! \Gamma\left(\kappa + 1 + m\right)} \quad \text{and, for } \kappa \notin \mathcal{Z},\ K_\kappa(z) = \frac{I_{-\kappa}(z) - I_\kappa(z)}{2\sin(\pi\kappa)/\pi}. \tag{3.18}$$

In the case $\kappa = n$ an integer, K_κ is defined as the limit $\kappa \to n$. The series representation for I_κ shows that K_κ is continuous with respect to its order for all κ. In the special case of half-integer orders, K_κ becomes the product of an exponential and a polynomial, with

(3.18) reducing to

$$K_{n+1/2}(z) = \exp(-z)\left(\frac{\pi}{2z}\right)^{1/2}\sum_{m=0}^{n}\frac{(n+m)!}{m!(n-m)!(2z)^m}. \tag{3.19}$$

In physical applications, K_κ is the more useful solution because it decays exponentially through positive values as $z \longrightarrow \infty$ (Watson, 1995). An asymptotic series expansion shows that for large z,

$$K_\nu(z) \sim (2z/\pi)^{-1/2}\exp(-z)\,(1 + O(1/z)) \tag{3.20}$$

At the origin, K_κ diverges, behaving either like $K_\kappa(z) \propto z^{-|\kappa|}$ for $\kappa \neq 0$ or $K_0(z) \propto -\log(z)$ when $z \longrightarrow 0$. For $|\kappa| < 1$ however, K_κ is absolutely integrable on $\mathbb{R}_+ = (0, \infty)$. It is also regular enough near the origin to admit the following useful integral (Gradshteyn and Ryzhik, 1965, p. 6.561.16):

$$\int_0^\infty z^\mu K_\kappa(z/\rho)\,\mathrm{d}z = 2^{\mu-1}\rho^{\mu+1}\Gamma\left(\frac{1+\mu+\kappa}{2}\right)\Gamma\left(\frac{1+\mu-\kappa}{2}\right) \quad \text{for } 1+\mu\pm\kappa > 0 \tag{3.21}$$

K_κ also satisfies the following for all κ and $z \neq 0$ (Gradshteyn and Ryzhik, 1965, p. 8.486):

$$z\,(K_{\kappa-1}(z) - K_{\kappa+1}(z)) = -2\kappa K_\kappa(z) \tag{3.22}$$

$$\frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}\left(z^\kappa K_\kappa(z)\right) = -z^{\kappa-1}K_{\kappa-1}(z) \tag{3.23}$$

3.7.2 Properties of the WMY kernel family

For simplicity we define an unnormalized WMY family for $\rho > 0$, $\kappa > -1$,

$$w_{\kappa,\rho}(r) = (r/\rho)^\kappa K_\kappa(r/\rho) = \mathcal{D}(r;\kappa,\rho)/A(\kappa,\rho)\,. \tag{3.24}$$

These are strictly positive functions belonging to the Lebesgue space $L^1(\mathbb{R}^2)$. The norm of $w_{\kappa,\rho}$ can be calculated using (3.21) after integrating out the redundant angular coordinate:

$$\|w_{\kappa,\rho}\|_1 = \int_{\mathbb{R}^2} w_{\kappa,\rho}(|\boldsymbol{x}|)\,\mathrm{d}\boldsymbol{x} = 2\pi\rho^{-\kappa}\int_0^\infty r^{\kappa+1}K_\kappa(r/\rho)\,\mathrm{d}r = 2^{\kappa+1}\pi\rho^2\Gamma(\kappa+1). \tag{3.25}$$

From this one obtains the WMY kernel normalization constant, $1/A(\kappa,\rho) = \|w_{\kappa,\rho}\|_1$. A similar calculation yields $\mu_n = \int_0^\infty r^n\mathcal{D}(r;\kappa,\rho)\,\mathrm{d}r$, the n^{th} marginal moment of the WMY kernel:

$$\begin{aligned}\mu_n &= 2\pi\rho^{-\kappa} A\left(\kappa, \rho\right) \int_0^\infty r^{\kappa+n+1} K_\kappa(r/\rho)\,\mathrm{d}r \\ &= (2\rho)^n \Gamma(\kappa+1+n/2)\Gamma(1+n/2)/\Gamma(\kappa+1). \end{aligned} \tag{3.26}$$

For nonzero even n, this reduces to $(2\rho)^n (n/2)!\prod_{m=1}^{n/2}(\kappa+m)$, from which the kurtosis expression in Section 2.3 of the main text can be derived.

The spectral density (d-dimensional Fourier transform) of the d-dimensional WMY kernel is

$$\hat{\mathcal{D}}\left(\boldsymbol{z};\kappa,\rho\right) = \int_{\mathbb{R}^d} \mathcal{D}\left(|\boldsymbol{x}|;\kappa,\rho\right)\exp\left(-2\pi i\boldsymbol{z}\cdot\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x} \propto \left(|\boldsymbol{z}|^2+\rho^{-2}\right)^{-(2\kappa+d)/2} \quad \text{where } \boldsymbol{z}\in\mathbb{R}^d. \tag{3.27}$$

This relationship is widely known in spatial statistics community (see *eg*. Guttorp and Gneiting, 2006; Stein, 1999) but less so in mathematical ecology.

Note that in the 2D case ($d = 2$), the Hankel transform of $\mathcal{D}$ is identical to the spectral density (3.27). Thus, in the sense of Schlägel and Lewis (2016), $\mathcal{D}$ is a *robust kernel*: That is, the n^{th} power of (3.27) is again the spectral density of a WMY kernel, with the same range parameter, and shape $n(\kappa+1)-1$. More generally, since convolutions become products in Fourier space, the spectral density of the n-part convolution $\mathcal{D}\left(r;\kappa_1,\rho\right) * \mathcal{D}\left(r;\kappa_2,\rho\right) * \ldots \mathcal{D}\left(r;\kappa_n,\rho\right)$ is proportional to:

$$\prod_{m=1}^{n}\left(|\boldsymbol{z}|^2+\rho^{-2}\right)^{-(2\kappa_m+d)/2} = \left(|\boldsymbol{z}|^2+\rho^{-2}\right)^{-\sum_{m=1}^{n}(2\kappa_m+d)/2} = \left(|\boldsymbol{z}|^2+\rho^{-2}\right)^{-(2\kappa+d)/2} \tag{3.28}$$

where $\kappa = \sum_{m=1}^{n}\kappa_m + (n-1)d/2$. Therefore the WMY family is closed under convolutions.

The WMY family is bounded everywhere in $r > 0$. It becomes unbounded at the origin when $\kappa \leq 0$, and otherwise we have $w_\kappa(0) = 2^{\kappa-1}\Gamma(\kappa) < \infty$. This is easily shown by setting $a_{m,\kappa} = 2^{2m+\kappa} m!\Gamma\left(\kappa+1+m\right)$ and $z = r/\rho$ and using the representation (3.18) to write for $\kappa \notin \mathbb{Z}$:

$$w_{\kappa,\rho} = \frac{\pi}{2\sin(\pi\kappa)}\left(\left(\frac{1}{a_{0,-\kappa}}+\frac{z^2}{a_{1,-\kappa}}+\frac{z^4}{a_{2,-\kappa}}+\ldots\right)-\left(\frac{z^{2\kappa}}{a_{0,\kappa}}+\frac{z^{2(\kappa+1)}}{a_{1,\kappa}}+\frac{z^{2(\kappa+2)}}{a_{2,\kappa}}\ldots\right)\right). \tag{3.29}$$

The WMY family is strictly decreasing in r, since by (3.23) its first derivative can be written:

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(w_{\kappa,\rho}\right) = -\frac{r}{\rho^2} w_{\kappa-1,\rho}. \tag{3.30}$$

From the expansion in (3.29), one sees that this derivative becomes unbounded at the origin whenever $\kappa < 1/2$ and otherwise is finite. Similarly, after differentiating a second time we obtain:

$$\frac{\mathrm{d}^2}{\mathrm{d}r^2}\left(w_{\kappa,\rho}\right) = \left(1/\rho^2\right)\left((r/\rho)^2 w_{\kappa-2,\rho} - w_{\kappa-1,\rho}\right) = \left(1/\rho^2\right)\left(w_{\kappa,\rho} + (1-2\kappa) w_{\kappa-1,\rho}\right) \tag{3.31}$$

where (3.22) is used for the second equality. From (3.31) the second derivative is finite and negative at $r = 0$ when $\kappa > 1$. When $1/2 < \kappa \leq 1$, the $(1-2\kappa)w_{\kappa-1}$ term in (3.31) diverges to $-\infty$ as $r \longrightarrow 0$; whereas for $\kappa < 1/2$ it diverges to $+\infty$. Thus the special case of the exponential kernel ($\kappa = 1/2$) marks a qualitative change in behaviour; For $\kappa > 1/2$ the approach to the origin is concave (and bounded), whereas for $\kappa < 1/2$ it is convex with an unbounded first derivative.

3.7.3 The modified Laplace operator

We now introduce a PDE to serve as our general isotropic dispersal model for a population density $u(r,t)p(r)$. It describes the time evolution of u by the sum of three terms: an fractal diffusion operator $L_\kappa u$; a decay term of the form $-\lambda u$ with rate constant $\lambda > 0$, and a source term $F(r,t) \geq 0$:

$$\frac{\partial u}{\partial t} = \alpha L_\kappa u - \lambda u + F \quad \text{where} \quad L_\kappa u = \frac{1}{r^{d_f-1}}\frac{\partial}{\partial r}\left(r^{d_f-1}\frac{\partial u}{\partial r}\right). \tag{3.32}$$

The linear differential operator L_κ is an extension of the Laplacian to diffusion occuring in fractal media. The Hausdorf dimension $0 < d_f \leq 2$ reflects geometric constraints on movement and $\alpha > 0$ the diffusivity, so that in the edge case $d_f = 2$ we recover Fickian diffusion, or $L_\kappa = \nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)$. For notational convenience, define the shape parameter $\kappa = (2-d_f)/2$, and distance scaling factor $p(r) = r^{d_f-2} = r^{-2\kappa}$. Note that the operator

pL_κ can expressed in the form:

$$pL_\kappa = p\left((1-2\kappa)\frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial}{\partial r^2}\right) = \left(\frac{\partial p}{\partial r}\right)\left(\frac{\partial}{\partial r}\right) + \frac{p}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right). \tag{3.33}$$

For radially symmetric u, this means $pL_\kappa u = \nabla p \cdot \nabla u + p\nabla^2 u = \nabla \cdot (p\nabla u)$. Thus (3.32) can also be written in a normal form for parabolic problems in Euclidean coordinates, as in Zauderer (2006):

$$\left(\frac{p}{\alpha}\right)\frac{\partial u}{\partial t} = \nabla \cdot (p\nabla u) - \left(\frac{\lambda p}{\alpha}\right)u + \frac{pF}{\alpha} \tag{3.34}$$

When u and/or its derivatives have singularities at the origin (*eg.* $u \propto w_{\kappa,\rho}$ for $\kappa < 1/2$) we use a weaker formulation. Integrating (3.34) over a closed and bounded region $\Omega \in \mathbb{R}^2$ then applying the divergence theorem, we obtain a balance law for density in Ω in terms of the measure $\int_\Omega upr\,\mathrm{d}r$,

$$\int_\Omega (1/\alpha)\frac{\partial u}{\partial t}p\,\mathrm{d}\boldsymbol{x} = \int_{\partial\Omega}\frac{\partial u}{\partial n_{\boldsymbol{x}}}p\,\mathrm{d}s_{\boldsymbol{x}} + \int_\Omega (1/\alpha)\,(F - \lambda u)\,p\,\mathrm{d}\boldsymbol{x}, \tag{3.35}$$

where $ds_{\boldsymbol{x}}$ is the surface differential over $\partial\Omega$ and $\partial u/\partial n_{\boldsymbol{x}}$ the exterior normal derivative of u on this boundary. The law states that the rate of change of disperser density in Ω is equal to the flux through the boundary ($\partial\Omega$) plus the sum of sources (F) and density-dependent losses (at rate λ). Weak solutions of (3.32) satisfy (3.35) along with the appropriate initial and boundary conditions.

A stationary point emission of dispersers

Begin by considering the density pattern that arises under a continuous release of dispersers from a point source. We will assume that new dispersers are released from the origin at a constant rate of $\gamma > 0$ (in density per unit time). Recalling the density measure $\int_\Omega up\,\mathrm{d}\boldsymbol{x}$, the source term F in (3.32) will be a renormalized 2D Dirac delta function of the form $\gamma\delta(r)/2\pi rp$.

Upon their release, dispersers diffuse through $\mathbb{R}^2$ according to L_κ, and are subject to a constant mortality hazard with rate λ. To obtain biologically reasonable solutions we further assume disperser density $u(r,t)p(r)$ vanishes as $r \to \infty$; and for consistency with the Bessel kernel ($d_f = 2$, or $\kappa = 0$) we also assume this density becomes unbounded in the limit $r \to 0$.

Supposing the system has reached its steady state, we can set $\partial u/\partial t = 0$ in (3.34) and solve the resulting elliptic problem in $u(r)$. Away from the origin there are no sources, so the PDE becomes:

$$0 = \alpha L_\kappa u - \lambda u \quad \text{for } r > 0, \tag{3.36}$$

$$\text{where } \lim_{r\to\infty} u(r)p(r) = 0, \text{ and } \lim_{r\to 0} u(r)p(r) = \infty.$$

At the origin, the source term $\int_\Omega pF/\alpha$ in (3.35) equals γ/α whenever $\mathbf{0} \in \Omega$ and vanishes otherwise. Following Chapter 6.7 of Zauderer (2006) a sufficient condition for this behaviour can be found by letting Ω be a small disk centered at the origin, and taking the limit of (3.35) as $\Omega \to \mathbf{0}$,

$$\lim_{r\to 0}\left(2\pi p r \frac{\partial u}{\partial r}\right) = -\gamma/\alpha. \tag{3.37}$$

WMY functions with $0 < \kappa < 1$ satisfy both conditions, as we show next. First note that $w_{\kappa,\rho}(r)p(r) \to 0$ as $r \to \infty$ for all κ, by (3.20). Further, if $\rho^2 = \alpha/\lambda$, then $u = c w_{\kappa,\rho}$ satisfies (3.36) for arbitrary normalization constants c, since by (3.30)-(3.31) $w_{\kappa,\rho}$ is an eigenfunction of $\rho^2 L_\kappa$ for all κ:

$$\begin{aligned}\rho^2 L_\kappa w_{\kappa,\rho} &= \rho^2(1-2\kappa)\frac{1}{r}\frac{\partial w_{\kappa,\rho}}{\partial r} + \frac{\partial w_{\kappa,\rho}}{\partial r^2} \\ &= -(1-2\kappa)w_{\kappa-1,\rho} + w_{\kappa,\rho} + (1-2\kappa)w_{\kappa-1,\rho} = w_{\kappa,\rho}.\end{aligned} \tag{3.38}$$

Next we use (3.30) and the series representation (3.29) to show that if $u = c w_{\kappa,\rho}$ (and $\kappa \notin \mathbb{Z}^+$),

$$\begin{aligned}\lim_{r\to 0}\left(2\pi p r\frac{\partial u}{\partial r}\right) &= -2\pi c \rho^{-2\kappa}\lim_{r\to 0}\left((r/\rho)^{-2(\kappa-1)} w_{\kappa-1,\rho}\right) = -c\left[2^{1-\kappa}\pi\rho^{-2\kappa}\Gamma(1-\kappa)\right] \\ &= -c\left[(1/\rho^2)\int_{\mathbb{R}^2} w_{\kappa,\rho}(r)p(r)\,\mathrm{d}\boldsymbol{x}\right]\end{aligned} \tag{3.39}$$

The square-bracketed expression in (3.39) is finite for $0 < \kappa < 1$, so by setting the normalization constant $c = (\gamma/\lambda)/\int_{\mathbb{R}^2} w_{\kappa,\rho} p \,\mathrm{d}\boldsymbol{x}$, the required behaviour (3.37) at the origin is established. By (3.29), the density $w_{\kappa,\rho}p$ is unbounded in the limit $r \to 0$ for all $0 \leq \kappa < 1$. Moreover, because $w_{\kappa,\rho}$ is continuous in κ, equation (3.39) also holds for $\kappa = 0$. Thus for all reasonable values ($0 < d_f \leq 2$) of the fractal dimension, the free-space Green's functions of (3.36) belong to the WMY family.

Switching back to Lebesgue measure, solutions $u(r)p(r)$ remain in the WMY family. This is because the modified Bessel functions satisfy $K_\kappa = K_{-\kappa}$ for all κ, and so when $0 < \kappa \leq 1$, we may rewrite $w_{\kappa,\rho}(r)p(r) = \rho^{-2\kappa}(r/\rho)^{-\kappa}K_\kappa(r/\rho) = \rho^{-2\kappa}w_{-\kappa,\rho}$. Using (3.25), this simplifies to:

$$w_{\kappa,\rho}(r)p(r) = cr^{-2\kappa}w_{-\kappa,\rho}(r) = (\gamma/\lambda)A_{-\kappa,\rho}w_{-\kappa,\rho}(r) \tag{3.40}$$

Thus when $\gamma = \lambda$ we recover the WMY kernel $\mathcal{D}(r; -\kappa, \rho)$ with shape parameter $d_f/2-1$.

Settling after an instantaneous point release

Next suppose that mortality is negligeable, and consider a one-time point release of dispersers that diffuse in space and settle with constant hazard. We therefore interpret λ in (3.32) as a settling (rather than mortality) rate, and $u(r,t)p(r)$ as the remaining density of active dispersers. All individuals eventually settle, so as $t \to \infty$ the accumulated density of settled individuals approaches:

$$U(r)p(r) \text{ where } U(r) = \lambda \int_0^\infty u(r,t)\,\mathrm{d}t\,. \tag{3.41}$$

As before we assume $U(r)p(r)$ diverges at zero and vanishes in the limit $r \to \infty$. Assume further that all active dispersers are initially concentrated at the origin, with total density N_0. The initial condition for u is therefore the renormalized 2D Dirac delta function $u(0,r) = N_0\delta(r)/2\pi rp$. The PDE for U is then found by setting $F = 0$ in (3.32) and integrating both sides over all time:

$$\begin{aligned} \lambda \int_0^\infty \frac{\mathrm{d}u}{\mathrm{d}t}\,\mathrm{d}t &= -\lambda u(0,r) = -\lambda N_0\delta(r)/\pi rp \\ &= \lambda \int_0^\infty (\alpha L_\kappa u - \lambda u)\,\mathrm{d}t = \alpha L_\kappa U - \lambda U \qquad (3.42) \\ &\text{where } \lim_{r\to\infty} U(r)p(r) = 0, \text{ and } \lim_{r\to 0} U(r)p(r) = \infty. \end{aligned}$$

Up to a relabeling of variables, this is equivalent to the stationary point emission problem of the previous section: Away from the origin, the left hand side of (3.42) vanishes and we have the elliptic problem of (3.36), with U taking the place of u; At the origin we can view the left hand side of (3.42) as a stationary source term F, and by setting $\gamma = \lambda N_0$ we recover the condition (3.37). Thus:

$$U(r) = \left(\frac{N_0}{\int_{\mathbb{R}^2} w_{\kappa,\rho} p \, \mathrm{d}\boldsymbol{x}} \right) w_{\kappa,\rho}(r), \text{ where } \rho^2 = \alpha/\lambda. \tag{3.43}$$

Notice that by the same arguments as in (3.40), when $N_0 = 1$ the settled density becomes $U(r)p(r) = \mathcal{D}(r; -\kappa, \rho)$; *ie.* it is identical to the steady state of the stationary emissions problem.

Extensions to multiple stages

Lastly, we consider augmenting equation (3.42) with additional stages of dispersal, each connected by a stage-switching hazard term with rate constant $\lambda > 0$. Writing $u_m(r,t)p(r)$ for the density in the m^{th} stage, we will assume that an instantaneous point release of unit density initializes dispersers in the first stage, so that the dynamics of u_1 are the same as u in the previous section (with $N_0 = 1$). As time progresses, the impulse moves through the stages and exits the n^{th} stage at rate λ (as settled density). We will be interested in the long-term accumulation of settled density:

$$U_n(r)p(r) = \lambda \int_0^\infty u_n(r,t)p(r) \, \mathrm{d}t, \text{ where} \tag{3.44a}$$

$$\frac{\partial u_1}{\partial t} = \alpha L_\kappa u_1 - \lambda u_1, \text{ with } u_1(r,0) = \delta(r)/2\pi r p, \text{ and} \tag{3.44b}$$

$$\frac{\partial u_m}{\partial t} = \alpha L_\kappa u_m + \lambda u_{m-1} - \lambda u_m, \text{ with } u_m(r,0) = 0, \text{ for } 1 < m \leq n. \tag{3.44c}$$

In the previous section we solved the one-stage analogue of this system. We now show how this more general $n \geq 1$ stage case can be solved more directly using survival analysis. Similar to Yasuda (1975) and Lutscher, Pachepsky, and Lewis (2005), we will write the settled density function as $U_n = \int_{t=0}^\infty u(r,t) f(t) \, \mathrm{d}t$, where $f(t)$ is the PDF for time to settling out the n^{th} stage, and $u(r,t)$ is the PDF for disperser position (in terms of measure $up \, \mathrm{d}r$).

As before, we assume the stage-wise density functions $u_m(r)p(r)$ vanish in the limit of large distances.

Recall that in the absence of stage-switching/settling ($\lambda = 0$), the PDF for disperser position is:

$$u(r,t)p(r) = \left((1/4\alpha t)^{1-\kappa}\exp(-r^2/4\alpha t)/\pi\Gamma(1-\kappa)\right)\left(r^{-2\kappa}\right)$$

When $\lambda > 0$, this becomes the conditional PDF for position at time t, given that the disperser has not yet settled. Meanwhile, the stage switching process operates independently of the movement process; At time t, the probability of belonging to the m^{th} stage is the integrated density $u_m(r,t)p(r)$:

$$v_m(t) = 2\pi \int_0^\infty u_m p r \,\mathrm{d}r = 2\pi \int_0^\infty r^{1-2\kappa} u_m \,\mathrm{d}r\,. \tag{3.45}$$

Let T denote the (random) duration of movement before settling, and let $f(t)$ be its PDF. The survival function for T is the probability that the disperser remains in any of the n active stages at time t, or $S(t) = \Pr(T > t) = \sum v_m$. This is related to the settling time PDF by $f(t) = -\,\mathrm{d}S/\mathrm{d}t$. Thus we can find f via the v_m. Beginning with v_1, we rearrange (3.44b) and integrate over space:

$$\int_{\mathbb{R}^2} \frac{\partial}{\partial t}(pu_1)\,\mathrm{d}\boldsymbol{x} = \alpha \int_{\mathbb{R}^2} \nabla\cdot(p\nabla u_1)\,\mathrm{d}\boldsymbol{x} - \lambda \int_{\mathbb{R}^2} pu_1\,\mathrm{d}\boldsymbol{x} \implies \frac{\mathrm{d}v_1}{\mathrm{d}t} = -\lambda v_1, \tag{3.46}$$

This follows by letting B_r be a disk of radius r centred at the origin, with normal vector $\boldsymbol{n_x}$ and length element $\mathrm{d}s_{\boldsymbol{x}}$, and using the divergence theorem to simplify the second integral in (3.46):

$$\int_{\mathbb{R}^2} \nabla\cdot(p\nabla u_1)\,\mathrm{d}\boldsymbol{x} = \lim_{r\to\infty}\int_{B_r} \nabla\cdot(p\nabla u_1)\,\mathrm{d}\boldsymbol{x} = \lim_{r\to\infty}\int_{\partial B_r} p\nabla u_1\cdot\boldsymbol{n}_x\,\mathrm{d}s_{\boldsymbol{x}} = \lim_{r\to\infty}\left(2\pi pr\frac{\partial u}{\partial r}\right) = 0.$$

The ODE for v_1 in (3.46) has initial data $v_1(0) = 1$ and so is solved by $v_1(t) = \exp(-\lambda t)$. Integrating equation (3.44b) in the same way we obtain for $m > 1$ a set of ODEs, $\frac{\mathrm{d}v_m}{\mathrm{d}t} = \lambda(v_{m-1} - v_m)$ with $v_m(0) = 0$. These are solved in sequence using integrating factors to obtain the recurrence relation $v_m = (\lambda t)^{m-1}\exp(-\lambda t)/(m-1)!$ Thus $f(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\sum v_m = \lambda v_n$, *ie* $f \sim \mathrm{Gamma}(n, \lambda)$.

With f and u so defined, and setting $\rho^2 = \alpha/\lambda$ and $\tilde{\kappa} = \kappa + n - 1$, we obtain:

$$\begin{aligned}
U(r) &= \int_0^\infty f(t)u(r,t)\,\mathrm{d}t \\
&= \left((4\alpha)^{1-\kappa}\pi\lambda^{-n}\Gamma(n)\Gamma(1-\kappa)\right)^{-1}\int_0^\infty t^{\tilde{\kappa}-1}\exp\left(-\lambda t - \left(r^2/4\alpha\right)/t\right)\mathrm{d}t \\
&= \left((4\alpha)^{1-\kappa}\pi\lambda^{-n}\Gamma(n)\Gamma(1-\kappa)\right)^{-1}\left(2\left(\left(r^2/4\alpha\right)/t\right)^{\tilde{\kappa}/2}K_{\tilde{\kappa}}\left(2\sqrt{\left(r^2/4\alpha\right)/\lambda}\right)\right) \\
&= \left(2^{n-\kappa}\pi\rho^{2-2\kappa}\Gamma(n)\Gamma(1-\kappa)\right)^{-1}(r/\rho)^{\tilde{\kappa}}K_{\tilde{\kappa}}\left(r/\rho\right) \propto w_{\tilde{\kappa},\rho}.
\end{aligned}$$

where the identity in the second line is found in Gradshteyn and Ryzhik (1965) (3.471.9). Using (3.21) one can verify that the normalization factor in the third line ensures $1 = 2\pi\int_0^\infty Upr\,\mathrm{d}r$. Thus

$$D(r) = U(r)p(r) = \left(2^{n-\kappa}\pi\rho^2\Gamma(n)\Gamma(1-\kappa)\right)^{-1}(r/\rho)^{n-1-\kappa}K_{n-1+\kappa}\left(r/\rho\right) \tag{3.47}$$

is a properly normalized PDF under Lebesgue measure. Notice that in the case $d_f = 2$ (or $\kappa = 0$), equation (3.47) defines the subset of WMY kernels with integer-valued shape parameters. In the case $d_f = 1$ (or $\kappa = 1/2$), we use (3.19) to simplify (3.47) and express the marginal distribution as:

$$\begin{aligned}
D_r(r) &= 2\pi Upr = 2\pi U \\
&= \left(2^{n-1}\rho\Gamma(n)\right)^{-1}\left(\sum_{m=0}^{n-1}\frac{(n-1+m)!}{2^m m!(n-1-m)!}(r/\rho)^{n-1-m}\right)\exp\left(-r/\rho\right).
\end{aligned} \tag{3.48}$$

Notice that this kernel family is identical to the one described in Neubert, Kot, and Lewis (1995) for multi-stage 1D diffusion with settling (in the case of equal stage switching rates).

3.7.4 Error in approximations of $\mathcal{D}$ by $\mathcal{D}_\otimes$

Lacking an analytic result on approximations of the WMY by the product-WMY, we instead investigated the relative errors (RE) in the entries of the redistribution matrices, $\left([\boldsymbol{D}]_{ij} - [\boldsymbol{D}_\otimes]_{ij}\right)/[\boldsymbol{D}]_{ij}$, over a large parameter grid. For each of 10^4 WMY kernels

$\mathcal{D}(r;\rho,\kappa)$, we numerically optimized the parameters of $\mathcal{D}_\otimes$ with the constraint $\rho_x = \rho_y$ and $\kappa_x = \kappa_y$, to minimize the mean absolute RE in the redistribution matrices for a 100×100 lattice of sample sites. Figure 3.7 plots the results.

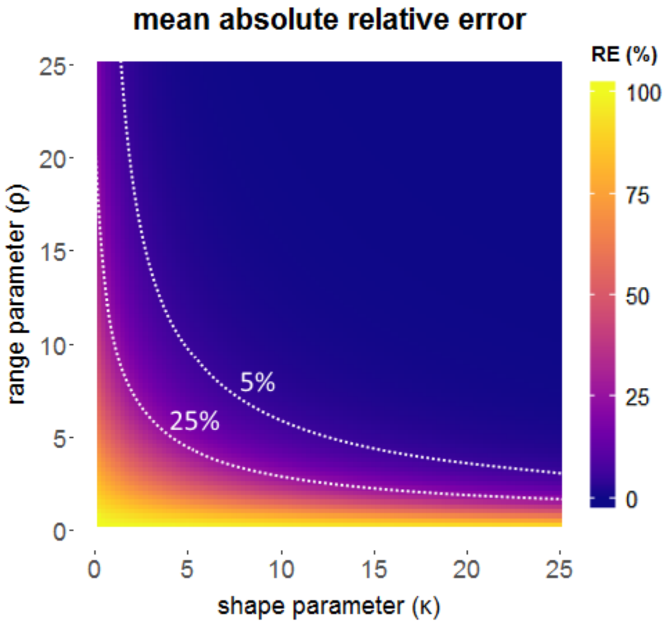

Figure 3.7: Heatmap of the error in approximation by product-WMY kernels parametrized to resemble the (isotropic) WMY. Parameters were set by minimizing mean absolute RE for each of $100 \times 100 = 10,000$ parameter pairs. Approximations appear to be adequate provided ρ and κ are not too small.

Point clouds of the signed relative errors (not shown) indicate that the product-WMY tends to assign less density to short distances than the WMY. At long distances, the errors are more evenly spread out, with separation vectors orthogonal to the coordinate axes overestimating the WMY, and those at a 45 degree angles underestimating. For low κ_x and κ_y values, this introduces the characteristic diamond-shaped contours that are obvious in Figure 2 of the main text (left panel).

Unsurprisingly, these approximations are poor when the shape value of $\mathcal{D}$ is low. However, as long as the range parameter ρ is not too small, approximations rapidly improve as κ increases through moderate values. Note that ρ is directly tied to sampling resolution.

For example if $\mathcal{D}(r; \kappa, \rho = 5)$ is an appropriate model for points sampled at 1 km intervals, then $\mathcal{D}(r; \kappa, \rho = 10)$ should be used for data sampled at 500 meter intervals.

3.7.5 Redistribution kernels in practice: the red-top model

Our modeling approach closely follows that of Heavilin and Powell (2008), in which beetle population sizes are represented implicitly through measurements of their contribution to pine mortality.

Let the (pre-attack) number of susceptible pines in summer t at site $\boldsymbol{x}$ be $S_t(\boldsymbol{x})$. MPB must attack and kill a fraction of them, $M_t(\boldsymbol{x})$, so their offspring can overwinter under the bark. By the next summer these infested pines become obvious due to their faded (red) crowns, and their number $I_{t+1}(\boldsymbol{x})$ is more easily counted. Pine mortality in year t can thus be written $M_t(\boldsymbol{x}) = I_{t+1}(\boldsymbol{x})/S_t(\boldsymbol{x})$.

In summer $t + 1$, the (now mature) offspring emerge in large numbers from each of the $I_{t+1}(\boldsymbol{x})$ infested host trees to disperse and attack new pines. For now we assume the number of surviving offspring per tree, λ, is uniform across the landscape. Thus $B_{t+1} = \lambda I_{t+1}(\boldsymbol{x})$ emerge from site $\boldsymbol{x}$.

After dispersal, the attacking cohort may be joined by *endemic* MPB populations. The endemic population phase is cryptic and low-density, so its impact is too slight to observe directly by aerial survey. However it appears to be ubiquitous in our study area, at a level of around one endemic host tree per hectare (Boone et al., 2011). Thus we write the post-dispersal MPB density in year $t + 1$ as the sum $\tilde{B}_{t+1}(\boldsymbol{x}) = \lambda(1 + (D * B_{t+1})(\boldsymbol{x}))$. These MPB will infest susceptible pines with varying success; Pines can repel small numbers but are overwhelmed in mass attacks. This leads to a nonlinear *attack curve* expressing the pine mortality rate as a sigmoid function of $\tilde{B}_{t+1}(\boldsymbol{x})$:

$$M_{t+1}(\boldsymbol{x}) = \underbrace{\frac{\tilde{B}_{t+1}(\boldsymbol{x})^2}{\tilde{B}_{t+1}(\boldsymbol{x})^2 + a^2}}_{\text{attack curve}} \quad \text{with} \quad \underbrace{\tilde{B}_{t+1}(\boldsymbol{x}) = \lambda \left(1 + (D * M_t S_t)(\boldsymbol{x})\right),}_{\text{endemic + post-dispersal MPB}} \tag{3.49}$$

where parameter $a > 0$ is the attacking beetle density at which 50% pine mortality occurs.

Note that (3.49) is identical to the red-top model of Heavilin and Powell (2008), except that we have added the endemic component and changed the notation for B_t. Inverting the attack curve in (3.49) and taking logarithms of both sides, we find the equivalent model for the mortality *log-odds*:

$$\operatorname{logit}\left(M_{t+1}(\boldsymbol{x})\right) = \log\left(\frac{M_{t+1}(\boldsymbol{x})}{1 - M_{t+1}(\boldsymbol{x})}\right) = 2\log\left(1 + \left(D * (M_t S_t)\right)(\boldsymbol{x})\right) - 2\log\left(\alpha\right). \quad (3.50)$$

where $\alpha = a/\lambda$ (in units of trees/ha) expresses the 50% mortality level of attacking beetle density in terms of infested tree equivalents. Thus (3.49) and (3.50) provide two plausible expressions for the expected damage to pine stands.

Error structure

We should expect that variations in beetle productivity and mortality introduce process errors into the post-dispersal MPB counts. These errors likely have a multiplicative effect so it is reasonable to assume that $\tilde{B}_{t+1}(\boldsymbol{x})$ will be log-normally distributed (Limpert, Stahel, and Abbt, 2001). Equation (3.50) becomes:

$$\operatorname{logit}\left(M_{t+1}(\boldsymbol{x})\right) = 2\log\left(1 + \left(D * (M_t S_t)\right)(\boldsymbol{x})\right) - 2\log\left(\alpha\right) + Z(\boldsymbol{x}), \quad \text{where } Z(\boldsymbol{x}) \sim N(0, \sigma^2). \quad (3.51)$$

The measurement process also contributes error. We model this by adding a further mean-zero Gaussian term to equation (3.51). However, as measurements are presumably independent of the growth and dispersal process, it is (mathematically) equivalent to simply increase the variance of $Z(\boldsymbol{x})$. Thus we merge the two error terms and hereafter view σ^2 in (3.51) as their combined variance.

The $Z(\boldsymbol{x})$ are likely to be spatially autocorrelated; pairs of errors located near to each other in space will be more highly correlated than ones spaced far apart. Thus we specify these correlations as a decreasing function of distance, by means of a parametric covariance kernel C. Recall that these functions are very similar to redistribution kernels. For simplicity, we use the Gaussian, so that $\operatorname{Cov}\left[Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)\right] = C(r_{ij}; \sigma^2, c) = \sigma^2 \exp(-r_{ij}^2/c)$, where $r_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|$.

Finally, we assume that $Z(\boldsymbol{x})$ is a *Gaussian random field*. This property conveniently implies that every finite sample $\boldsymbol{\epsilon} := (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n))^T$ is distributed as a multivariate Gaussian $\boldsymbol{Z} \sim \mathcal{N}(0, \boldsymbol{V})$, where $[\boldsymbol{V}]_{ij} = C(r_{ij})$ (Chiles and Delfiner, 2012). Having specified the distributional properties of the model residuals from (3.51) in full, we now turn to the problem of inference.

Discretization and likelihood

If the observed data are $\boldsymbol{m}_t = (M_t(\boldsymbol{x}_1), \ldots M_t(\boldsymbol{x}_n))^T$ and $\boldsymbol{s}_t = (S_t(\boldsymbol{x}_1), \ldots S_t(\boldsymbol{x}_n))^T$, the (expected) post-dispersal MPB density then becomes $\lambda\left(\mathbb{1} + \boldsymbol{D}\left(\boldsymbol{m}_t \odot \boldsymbol{s}_t\right)\right)$, where $\mathbb{1}$ is the length-n vector of 1's and $\odot$ is the pointwise product. Letting boldface **log** and **logit** denote entrywise transforms, the likelihood function L for (3.51) can now be written compactly in terms of the data vectors:

$$L \propto \det \boldsymbol{V} + \exp\left(\boldsymbol{\epsilon}^T \boldsymbol{V}^{-1} \boldsymbol{\epsilon}\right) \quad \text{where } \boldsymbol{\epsilon} = \mathbf{logit}\left(\boldsymbol{m}_{t+1}\right) - 2\mathbf{log}\left(\mathbb{1} + \boldsymbol{D}\left(\boldsymbol{m}_t \odot \boldsymbol{s}_t\right)\right) + 2\log\left(\alpha\right)\mathbb{1}. \tag{3.52}$$

L is a function of the attack parameter α; the covariance parameters, c and σ^2 (defining $\boldsymbol{V}$); and the redistribution parameter(s) (defining $\boldsymbol{D}$). Thus, given a dataset $\{\boldsymbol{m}_{t+1}, \boldsymbol{m}_t, \boldsymbol{s}_t\}$, it appears fairly straightforward to implement (3.52) in code and maximize it (numerically) to fit the model (3.51).

However optimization algorithms must evaluate (3.52) many times, so it becomes important in large-n problems that the matrix operations of determinant, inverse, and product are done efficiently. Our application, like many in spatial ecology, has data in the form of a raster, where the sampling locations $\{\boldsymbol{x}_k\}$ are arranged in a regular lattice of dimensions $n_y \times n_x = n$. This introduces symmetries in $\boldsymbol{V}$ and $\boldsymbol{D}$ that permit certain computational shortcuts. For a given $\boldsymbol{\epsilon}$, the efficient minimization of $\det \boldsymbol{V} + \boldsymbol{\epsilon}^T \boldsymbol{V}^{-1} \boldsymbol{\epsilon}$ in this situation is a well-studied problem in spatial statistics (*eg*. Lindgren, Rue, and Lindström, 2011). We focus instead on the evaluations of $\boldsymbol{D}\left(\boldsymbol{m}_t \odot \boldsymbol{s}_t\right)$ needed to construct $\boldsymbol{\epsilon}$.

Whenever the $\{\boldsymbol{x}_k\}$ form a lattice, separability implies that the redistribution matrix $\boldsymbol{D}_\otimes$

has a Kronecker product factorization:

$$\underbrace{D(\boldsymbol{r}) = D_x(r_x)D_y(r_y)}_{\text{spatial separability}} \implies \underbrace{\boldsymbol{D}_{\otimes} = \boldsymbol{D}_x \otimes \boldsymbol{D}_y}_{\text{Kronecker product}} \implies \boldsymbol{D}_{\otimes}\boldsymbol{u} = \underbrace{\operatorname{vec}\left(\boldsymbol{D}_y^T \operatorname{vec}_{n_y \times n_x}^{-1}(\boldsymbol{u})\, \boldsymbol{D}_x\right)}_{O\left(n_y^2 + n_x^2\right) \text{ complexity}} \tag{3.53}$$

The x-component redistribution matrix $\boldsymbol{D}_x$ is computed as the pointwise discretization of D_x, with $r_x = x_i - x_j$ replacing $\boldsymbol{r}$, and similar for $\boldsymbol{D}_y$. vec() denotes column-vectorization (stacking the columns of a matrix to form a vector) and $\operatorname{vec}_{m \times n}^{-1}()$ denotes its inverse, with target dimensions indicated in the subscript. This formula is discussed in more detail in Chapter 2, along with block-diagonalizations of $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ (not discussed here) that facilitate additional constant-factor reductions in complexity.

Chapter 4

The signature of endemic populations in the spread of mountain pine beetle outbreaks

4.1 Introduction

The mountain pine beetle (MPB) *Dendoctonus ponderosae* Hopkins (Coleoptera Curculionidae), is a tree-killing species of bark beetle native to pine forests of Western North America. Each year for a short period in summer, adult MPB seek to complete their life cycle by attacking a suitable living host pine. During attacks, MPB bore into the bark, introducing fungal pathogens in the process, and ultimately girdle the tree (Safranyik and Carroll, 2006). Death follows swiftly for a pine whose defense systems fail to eject these attackers. When an attack succeeds, MPB use the host to feed and reproduce, laying eggs in galleries excavated underneath its bark. Outbreaks of these insects can be devastating to pine forests, prompting modeling efforts to explain their origin and how they spread across the landscape.

With few exceptions, the adult MPB die after reproduction, and their progeny emerge as teneral adults the following summer to begin the cycle again. This semelparity, and the approximately linear relationship between reproductive success and host death, are mathematically convenient properties when constructing models to track large-scale year-to-year changes in MPB populations. For example if a total of B beetles have attacked a stand containing H susceptible pines, killing a fraction ϕ of them, then a rough estimate of the MPB population emerging in the next year is $\lambda\phi H$, where $\lambda > 0$ is a suitably chosen constant representing average per-stem productivity (Heavilin and Powell, 2008).

Two major complications in MPB dynamics must be addressed by modelers attempting to link ϕ with the underlying beetle population (Nelson et al., 2008). First, any plausible recruitment curve relating ϕ to the attack density B must be highly nonlinear to accomodate the eruptive nature of MPB populations (Berryman, 1978) and the distinctive behaviours that occur in different population phases: During the incipient-epidemic phase, attacks occur at densities low enough to be defended by hosts, so cooperative efforts in overcoming these defenses leads to positive density dependence (or Allee effect) in ϕ (Boone et al., 2011); However as the number of attacking individuals rises, and the MPB enters epidemic and post-epidemic phases, the density dependence turns negative as a result of scramble competition (Peters and Peters, 1991).

Thus empirical data on $\phi(B)$ (Raffa and Berryman, 1983) reveals an S-shaped, or sigmoid relationship. This form is reminiscent of the familiar type-III functional responses for parasitism behaviour (Holling, 1959), and indeed many aspects of MPB population dynamics are well described by parasitoid-prey systems theory (Goodsman et al., 2016).

Second, models describing the evolution of an outbreak over multiple years (t) cannot easily relate B_{t+1} and $\lambda\phi(B_t)H$ without incorporating the spatial effects of dispersal. Given the sensitive and nonlinear dependence of recruitment on attack density B_t, it is clear that an oversimplification of the redistribution process can have strong (and undesireable) effects on predictions of pine mortality ϕ. Dispersal flights of the MPB allow it to escape depleted stands, spark outbreaks in neighbouring areas, and expand its range (Giroday, Carroll, and Aukema, 2012). By modeling B_t as the outcome of a spatially explicit dispersal event, we are better equipped to capture these interesting and important ecological phenomena, and achieve a higher precision in fitting $\phi(B)$ to data. With these advantages however, comes mathematical complexity.

A variety of MPB dispersal models can be found in the literature (*eg.* Goodsman et al., 2016; Preisler et al., 2012; Aukema et al., 2008; Heavilin and Powell, 2008), but most make two simplifying assumptions for mathematical convenience: that movements occur in all directions with equal probability (isotropy); and that patterns of redistribution do not vary with spatial location (stationarity). The main novelty in our model is that our dispersal model has the flexibility to capture directed and location-dependent (anisotropic and nonstationary) events – it is meant as a phenomenological alternative to dynamical systems based approaches to the same problem (*eg*. Garlick et al., 2011; Powell and Bentz, 2014; Powell et al., 2018), but with a simpler mathematical representation that borrows computationally efficient methods from spatial statistics.

One of our goals in presenting this modeling framework is to demonstrate the remarkable amount of information that can be extracted from visual aerial surveys, which are both cost-effective and cover large spatial and temporal extents. Our observations of pine mortality (ϕ) are derived from the Aerial Overview Survey (AOS), which covers most of the forestland in the province of British Columbia (BC), and is published annually. Similar datasets are available for the neighbouring province of Alberta (AB), in which a MPB range expansion of great economic and scientific interest is currently underway. Since our methods do not rely on expensive and time-consuming ground surveys, they offer a cost-effective way of conducting large-scale analyses of MPB attack dynamics in contemporary areas of concern in Western Canada.

Modelers using AOS data must contend with errors – both positional and temporal – introduced by a reliance on imprecise visual evidence (*eg*. Kautz, 2014; Wulder et al., 2006). However, in spite of its shortcomings, the AOS covers an impressively large extent and timeline of forest damage patterns in BC. For this reason, a considerable body of landscape-level MPB research draws from the AOS and its predecessor, the Forest Insect and Disease Survey (*eg*. Aukema et al., 2006; Robertson et al., 2009; Chen and Walton, 2011; Reyes, Zhu, and Aukema, 2012; Sambaraju et al., 2012; Chen, 2014; Chen et al., 2015). We will show that by employing a statistical framework well-equipped to handle spatial error, we regain precision and accuracy by exploiting the large sample sizes available in the AOS.

In particular our model accurately estimates the size of the cryptic, low-density endemic MPB population using only spatial data on outbreaks. This is remarkable given that pine mortality caused by the endemic phase happens at levels far below the operational detection threshold of the AOS (Cooke and Carroll, 2017). Studies of endemic MPB more typically rely on intensive ground surveys of attacked pine (*eg*. Boone et al., 2011; Bleiker et al., 2014). Rather than observing these attacked hosts directly, our model infers them using data on outbreak-level pine mortality.

Section 4.2.1 begins by reviewing a popular mathematical representation for $\phi(B)$, with several refinements introduced in Sections 4.2.2-4.2.4. The flight model is outlined in Section 4.2.5, and an error model suitable for the AOS dataset is proposed in Section 4.2.7. We demonstrate the model in Section 4.3 by fitting to data on outbreaks of the MPB in BC.

4.2 Methods

Our case study covers a pine-rich area of roughly 10,000 km^2, centered over the Merritt Timber Supply Area (TSA) of Southern BC (Figure 4.1). We divided this into a 1 hectare (ha) resolution grid (*sensu* Aukema et al., 2006) to form a 1000 x 1000 lattice of cells, with matching layers provided by the province (http://www.hectaresbc.org) on wildfire, cutblocks, and topography.

Since we are interested in how dispersal patterns are related to outbreak development, we analysed the attack years 2006-2008, in which a large number of pine-leading stands would see transitions from endemic to epidemic behaviour (the incipient-epidemic population phase). This period captures the peak of the larger epidemic in the Merritt TSA (in terms of basal area damaged) at a time when around one out of four cells in the area exhibited crown-fade due to MPB activity.

Our analysis tracks four state variables, indexed by year (t) and location (i): Only two of

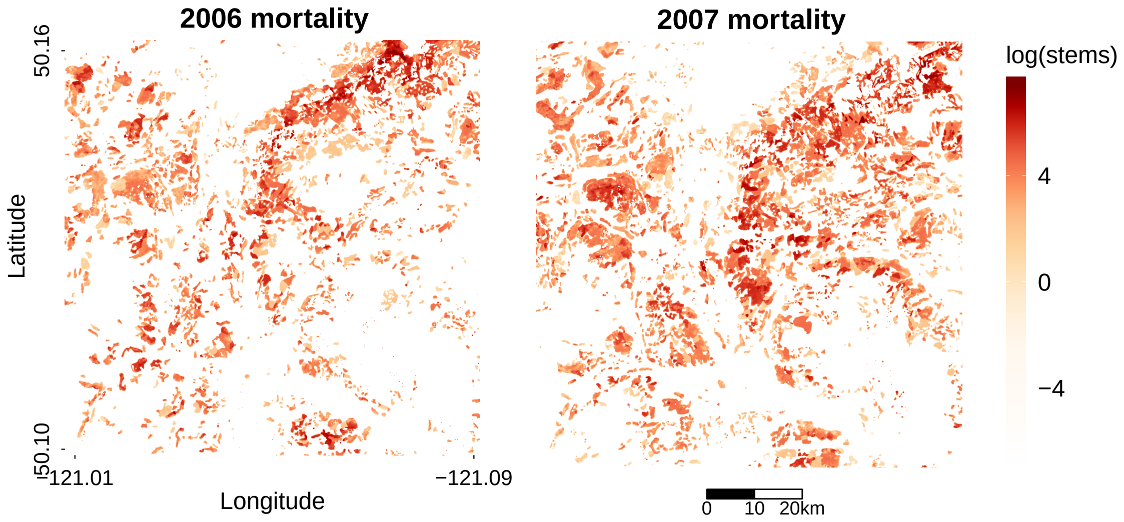

Figure 4.1: Hosts killed by MPB ($\phi_{i,t}H_{i,t}$, in stems/ha) in the summers of 2006-2007. AOS data on damage severity were rasterized to approximate susceptible host mortality ($\phi_{i,t}$). Host density $H_{i,t}$ was derived from pine volume estimates in Beaudoin et al. (2014), as described in Appendix 4.5.1

them can observed on large scales in practice: pine mortality ($\phi_{i,t}$) and host density ($H_{i,t}$, in stems/ha) (Appendix 4.5.1); The others, MPB density *pre*-dispersal $\tilde{B}_{i,t}$ and *post*-dispersal $B_{i,t}$ (in females/ha), are latent variables, inferred by the model but never observed (Table 4.1). In Sections 4.2.1-4.2.5, we develop a model relating these four variables. Sections 4.2.7-4.3 connect the model to data.

location i	vectorized	definition	units	type
$H_{i,t}$	$\boldsymbol{H}_t$	pre-attack susceptible pine density	stems/ha	observed
$\phi_{i,t}$	$\boldsymbol{\phi}_t$	proportion of $H_{i,t}$ killed by MPB		
$\tilde{B}_{i,t}$	$\tilde{\boldsymbol{B}}_t$	emerging MPB density (pre-dispersal)	females/ha	latent
$B_{i,t}$	$\boldsymbol{B}_t$	MPB attack density (post-dispersal)		

Table 4.1: Notation for state variables in the MPB attack dynamics model. Indexing is by year t and location i, and boldface denotes the vector of all n locations, *eg.* $\boldsymbol{\phi}_t = (\phi_{1,t}, \phi_{2,t}, \ldots, \phi_{n,t})'$.

4.2.1 Attack dynamics (ϕ)

Our description of pine mortality $\phi_{i,t}(B_{i,t})$ will generalize the red-top model of Heavilin and Powell (2008) to better match the types of recruitment curves fitted in Cooke and Carroll (2017). The red-top model is more easily introduced by focusing at first on a particular location and year; so for notational convenience I omit the subscripts i and t until they are needed again in Section 4.2.4. Thus, for i and t fixed, we relate the attack density B (females/ha) to pine mortality ϕ by:

$$\text{proportion of } H \text{ killed} = \phi(B) = \frac{B^\kappa}{a^\kappa + B^\kappa} \quad \text{where } a > 0,\ \kappa > 0. \tag{4.1}$$

Parameter a is the half-saturation value, or attack density (in females/ha) at which 50% mortality occurs; and κ is a shape parameter controlling the density dependence. The special case $\kappa = 2$ recovers the red top model of Heavilin and Powell (2008) (after multiplying both sides by H). Other κ values reflect alternative regimes of density dependence. For example larger κ (and/or a) values would coincide with a stronger defensive response by pines, whereas when $\kappa \leq 1$ the Allee effect vanishes, reflecting compromised defenses as might occur for example during a drought.

Parameter estimation becomes simpler if (4.1) can be made linear in its parameters. Observing that the odds-ratio of pine mortality $\phi/(1-\phi)$ is $(B/a)^\kappa$, we can take logarithms to get:

$$\text{logit}(\phi) = -\kappa \log(a) + \kappa \log(B), \tag{4.2}$$

a linear equation on the logit-log scale. This also happens to be the mathematical form of the recruitment curve fitted in Cooke and Carroll (2017) to the data reported in Boone et al. (2011) on attacked pines in our study area. Their analysis estimated $\hat{\kappa} = 1.66$ for the 2 years leading up to 2006. In years prior, a much lower value (0.56) was estimated, suggesting that environmental stressors on pine may have relaxed the Allee effect and bolstered endemic populations to spark the large-scale outbreaks we observe in 2006-2008 (Figure 4.1).

Once started, outbreaks are not easily stopped. Though pine vigour recovers from periods of stress, irruptions from the endemic to epidemic behaviour may continue as in-flights from neighbouring infested stands push local populations higher. Figure 4.2 illustrates how $\phi(B)$ changes in relation to MPB attack density for healthy pines ($\kappa > 1$). Above a certain density threshold, the MPB have sufficient numbers (B_T) to cooperatively attack a healthy pine (a mass attack), releasing them from the ordinary pressures of the Allee effect and marking the beginning of the incipient-epidemic phase. Empirical data from our study area suggest that a density of $\hat{B}_T \approx$ 300-600 females/ha is sufficient to instigate this change in an area of $1/T = 15.3$ ha (Carroll et al., 2006; Cooke and Carroll, 2017).

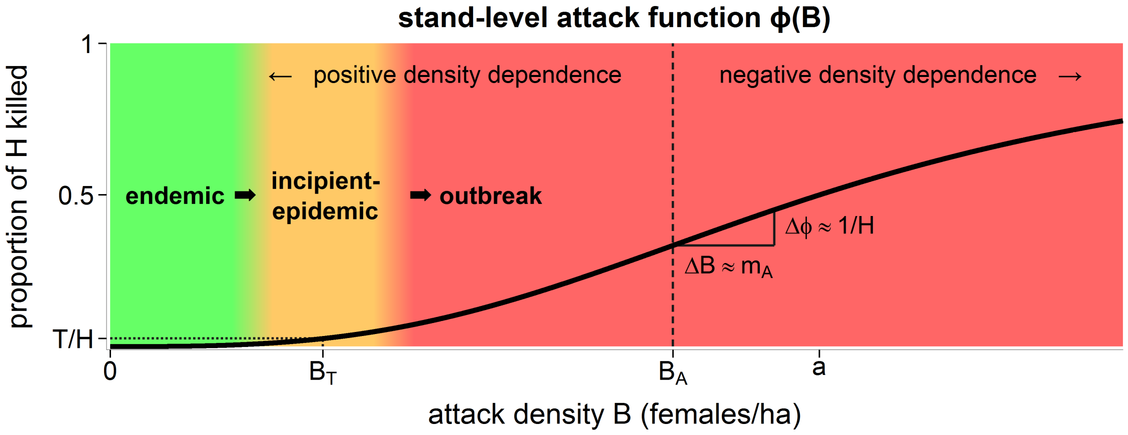

Figure 4.2: Host mortality as a function (4.1) of MPB attack density B for $\kappa = 3$. Below the inflection point (B_A) is a regime of negative density dependence. When $B \approx 0$, the endemic population is too small to mass-attack healthy pine. When B rises to the incipient-epidemic transition point B_T, mass attacks become feasible and the MPB are released from the endemic phase. At moderate densities, each attacked pine accounts for $\approx m_A$ beetles. At higher densities, intraspecific competition leads to diminishing returns and negative density dependence

B_T is of course scale dependent; Amman (1984) estimated a quite different transition point at the $1/T = 40.5$ ha sampling resolution. However, given any scale of interest (T), and given the values of κ and a, B_T can be estimated by setting $\phi(B_T) = T/H$ and inverting (4.1) to get:

$$B_T = \frac{a}{\sqrt[\kappa]{H/T - 1}}. \tag{4.3}$$

The number of beetles aggregating for a mass attack appears to be carefully moderated by pheromones (Safranyik and Carroll, 2006), suggesting that MPB attempt to find an optimal attack density that is sufficient to overwhelm the tree but low enough to avoid crowd competition. This optimum lies around 61 attackers/m^2 (Raffa and Berryman, 1983), so its stand-wide average would depend on the bark area of the trees. For example, a tree with 5.5 m^2 of bark available for attack (typical of the pine-leading stands in our study area) would have an optimum around 340 females/stem.

Under ideal conditions for MPB attack, this optimal density will presumably match the average attack density per attacked tree, which we call the *mass attack number* m_A (in females/stem). This average is approximated by the slope of $\phi(B)$ near its inflection point $B_A = a\sqrt[\kappa]{(\kappa-1)/(\kappa+1)}$ (where $\phi''(B_A) = 0$), since, at this intermediate density, $\phi(B)$ is nearly linear, and increases with B at rate $\phi'(B_A) \approx 1/(m_A H)$. From (4.1) we can therefore compute the approximation:

$$m_A \approx 1/(\phi'(B_A)H) = \frac{4B_A\kappa}{H(\kappa^2-1)}. \tag{4.4}$$

4.2.2 Stand susceptibility (a)

Though (4.3) and (4.4) have a complex nonlinear dependence on κ, both equations scale linearly with the half-saturation value a. Stands that are highly susceptible to MPB attack have lower values, and vice versa. It is therefore appropriate to view a as a susceptibility measure, which can vary with environmental factors such as weather and stand characteristics such as pine density.

However, lacking a clear biology-based model to connect half-saturation to these factors, we simply take the best linear approximation on the logit-log scale, writing $\boldsymbol{x\beta} = -\kappa\log(a)$ for a set of unknown regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots \beta_{n_\beta})$ and (covariate) data $\boldsymbol{x} = (x_1, \ldots x_{n_\beta})$. By embedding $\boldsymbol{x\beta}$ into (4.1) we obtain an absolute risk model in terms of mortality (ϕ) (Nelson et al., 2008), as opposed to a relative risk model such as the stand susceptibility index (SSI) of Shore, Safranyik, and Lemieux (2000).

Similar regression models such as in Aukema et al. (2008) have been useful for identifying environmental factors that have a significant ($\beta_k \neq 0$) effect on outbreak occurence. For our purposes the $\boldsymbol{\beta}$ simply serves as a (location-wise) correction of a through which to estimate MPB population sizes, so we do not focus on the β_k or their effect sizes in our analysis. However, interested readers will find the full set of linear regression covariates listed in Appendix 4.5.1.

4.2.3 Endemic populations (ϵ)

The (aspatial) red-top model of Heavilin and Powell (2008) has no endemic equilibrium: low density populations are viewed as unstable, tending to extinction, and occuring only by means of immigrations from a reservoir of distant outbreaks appearing stochastically across the landscape. However, empirical data (*eg*. Boone et al., 2011; Bleiker et al., 2014) suggest that stable resident endemic populations are widespread and persistent. These low-density populations subsist on a small number of defensively compromised pines and an assemblage of secondary bark beetle species that assist in the colonization of weakened trees.

We introduce this stable endemic equilibrium into the red-top model by adding the constant term $\epsilon > 0$ (in females/ha) to the post-dispersal MPB population B in the red-top model at all sites/years prior to attack. The number ϵ represents a constant and spatially uniform background level of emerging MPB. Should an in-flight from a neighbouring outbreak occur, its density is added to the endemic cohort ϵ, and the combined population attacks pines according to (4.1). The effect of the endemic term is therefore to boost the effective size of spreading populations, increasing the likelihood that an incipient-epidemic transition will succeed in sparking a local outbreak.

In the absence of immigrating MPB, the endemic population is too small to attack healthy pines, so it instead seeks out defensively weakened trees. Because this pool of suitable hosts is ephemeral and extremely small compared to H, endemic MPB incur a much higher flight-establishment mortality cost than do outbreaking populations: Safranyik and Carroll (2006) estimates the generation mortality of endemic MPB at 97.5%. Assuming most of this loss can be attributed to the search flight, the rate of attack on defensively weakened hosts under this model would be $(1 - 0.975)\epsilon$ females/ha (or slightly above), with the healthy pine population variable H unaffected.

However, if an endemic population joins with a cohort of immigrating outbreak-level MPB, suitable hosts suddenly become abundant, and the flight-establishment losses should drop accordingly. We assume that the generation mortality in populations capable of mass-attacks – thought to lie in the range 80-98.6% (Safranyik and Carroll, 2006; Amman, 1984) – mostly occurs as a result of tree defenses and crowd-competition. Unlike search flight losses, these effects are subsumed into $\phi(B)$ under the model (4.1). Therefore, we estimate the total number of attacking beetles at a given site as the sum of ϵ and any MPB (local or immigrant) originating from mass-attacked trees.

4.2.4 Reproduction (λ)

Reproduction connects subsequent years, so we must now make the dependence of our model variables on time and location explicit. In the red-top model, reproduction is summarized by $\tilde{B}_{i,t} = \lambda_{t-1}\phi_{i,t-1}H_{i,t-1}$. This expresses that $\tilde{B}_{i,t}$, the density of (non-endemic) mature MPB emerging in year t at location i, is proportional to the number of mass-attacked stems in year $t - 1$.

The scaling constant λ_{t-1} is a productivity parameter giving the (per-tree) average number of female MPB brood that survive to maturity and attack the following year. This kind of large scale averaging foregoes some precision, but simplifies the model considerably, summarizing in a single constant the many MPB within-tree growth and developement processes that cannot be observed in aerial surveys (Berryman, 1974).

Under this model, productivity λ_t is not identifiable from data on $\phi_{i,t}$ and $H_{i,t}$ without knowledge of $a_{i,t}$. So we instead fixed the value of $\lambda_t = \lambda$ in all years to a plug-in estimate of $\lambda = (2/3)(250) = 166.7$ (females/stem) suggested by empirical productivity data for epidemic phase MPB (Cole and Amman, 1969, Fig. 9), and assuming a 1:2 male-female sex ratio (Reid, 1962). This productivity value is consistent with a 90% generation mortality rate, calculated using the brood production regression in (Safranyik, 1988, eq. 14) on the mean diameters (Carroll et al., 2006) and heights (Safranyik and Linton, 1991) of pine in our study area.

Although a time (and space) dependent λ would be more realistic, it would complicate the model considerably. We do however allow all other process model parameters to vary with time (*eg.* ϵ_t, κ_t, $\boldsymbol{\beta}_t$, and the parameters of D_t), estimating them separately for each year in our analysis. Variations in productivity are therefore reflected in changing stand susceptibility $a_{i,t}$, which varies both spatially and temporally through $\boldsymbol{\beta}_t$ and the local covariates $\boldsymbol{x}_{i,t}$ (Table 4.2).

submodel	vector	symbols	definition	units
attack	$\boldsymbol{\theta}_{\phi_t}$	κ_t	density dependence shape value	unitless
		λ	beetle production per attacked host	females/stem
		ϵ_t	emerging endemic MPB population level	females/ha
		$a_{i,t}$	half-saturation / susceptibility value	females/ha
		$\boldsymbol{\beta}_t$	linear regression coefficients for $a_{i,t}$	-
dispersal	$\boldsymbol{\theta}_{D_t}$	$\boldsymbol{\Delta}_{k,t}$	pWMY kernel: angle, shape and range	-
error	$\boldsymbol{\theta}_{V_t}$	σ_t^2	marginal variance	unitless
		$\boldsymbol{\rho}_t$	Gaussian autocorrelation range (x and y)	km

Table 4.2: Parameters of the generalized red-top model. All except for λ are fitted to data separately by year (t). For dispersal, a 5-parameter product-WMY (pWMY) kernel (Appendix 4.5.2) is assigned to each of $m = 625$ data blocks, indexed by $k = 1, \ldots m$. A vector of 44 regression coefficients ($\boldsymbol{\beta}_t$) defines stand susceptibility through the linear model $\kappa_t \log(a_{i,t}) = \boldsymbol{x}_{i,t}\boldsymbol{\beta}_t$ for local covariates $\boldsymbol{x}_{i,t}$ (Appendix 4.5.1), where i indexes location.

4.2.5 Dispersal ($\tilde{B} \rightarrow B$)

We model dispersal using a redistribution kernel D_t. If the emerging MPB population $\tilde{B}_{t,i}$ is observed at n spatial locations, D_t specifies an $n \times n$ matrix ($\boldsymbol{D}_t$) whose i, j^{th} entry $[D_t]_{ij}$ is the expected proportion of the population $\tilde{B}_{t,j}$ that will move to cell i in the course of dispersal (Appendix 4.5.2). Thus, after adding the endemic MPB, the expected attack density is $\mathbb{E}\left(B_{i,t}\right) = \epsilon_t + \sum_j \left([D_t]_{ij}\tilde{B}_{t-1,j}\right)$. The equivalent matrix-vector equation $\mathbb{E}\left(\boldsymbol{B}_t\right) = \epsilon_t \boldsymbol{I} + \boldsymbol{D}_t \tilde{\boldsymbol{B}}_t$ allows us to drop the cumbersome location indices (i), so we will use this simpified notation whenever possible.

The i^{th} entry of $\mathbb{E}\left(\boldsymbol{B}_t\right)$ estimates the density of attackers, sometimes called beetle pressure, at stand location i in year t. Beetle pressure is a common feature of MPB outbreak risk models (*eg.* Wulder et al., 2006; Preisler et al., 2012), where it expresses proximity to infestations by a weighted sum of severity values or presence/absence indicators in a neighbourhood of the target stand. The choice of D_t reflects assumptions about how MPB redistribute in search of new hosts. Ad-hoc assignments of weights to $[D_t]_{ij}$ often suffices for predictive modeling (*eg.* Kärvemo et al., 2014; Kunegel-Lion, McIntosh, and Lewis, 2018), but when studying attack dynamics it is more desireable to use a model for the process of flight (Nelson et al., 2008) such as a biodiffusion-based redistribution kernel.

Our flight model approximates the Whittle-Matérn-Yasuda (WMY) kernel family (Yasuda, 1975), which describes diffusive movements through complex habitat (see Chapter 3). Included in this family are a number of distinct isotropic kernels that have been advocated in previous studies of similar datasets (*eg.* Turchin and Thoeny, 1993; Heavilin and Powell, 2008; Goodsman et al., 2016). Figure 4.3 (middle) is one example, arising from diffusion with constant settling.

We will calculate the $[D_t]_{ij}$ values using pWMY kernels (Appendix 4.5.2), which in addition to closely approximating the WMY, easily incorporate anisotropic (directed) movement patterns (Figure 4.3, right) as might be expected from the effect of local winds (Ainslie and Jackson, 2011) and patchy habitat (Powell et al., 2018).

Importantly, the pWMY can be computed far more quickly than the WMY. Computational simplicity allows different dispersal patterns to be quickly fitted at different sites within a dataset. In our study area, this revealed a complex pattern of directionality (nonstationarity) that varies depending on the position of the source population.

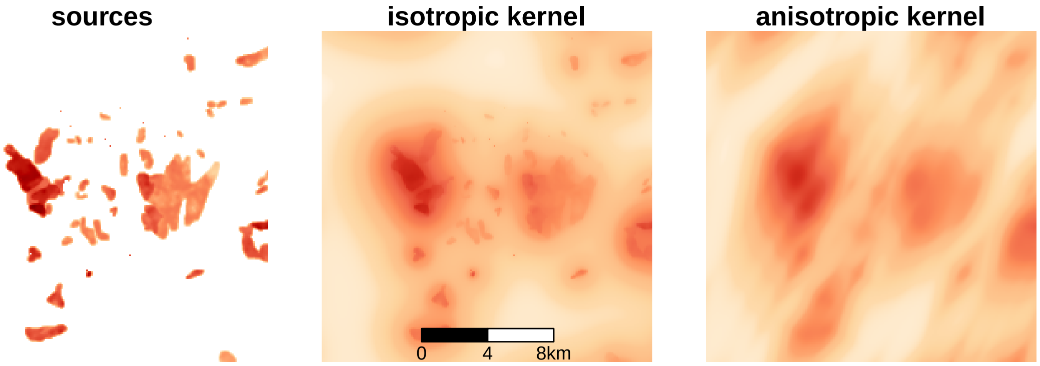

Figure 4.3: MPB density pre (left) and post-dispersal (middle and right) for two models of MPB flight patterns: an isotropic Bessel kernel (middle) with parameters from Goodsman et al. (2016), and an anisotropic pWMY kernel (right) parametrized to resemble it, but with the addition of a northeast-facing directionality

Nonstationarity in dispersal is unsurprising in light of work by Powell and Bentz (2014) on cues for direction and motility in MPB flights. However, lacking high resolution data on these cues, we opted for a phenomenological model that combines multiple stationary (pWMY) kernels to form a nonstationary one. Each stationary kernel is fitted separately to a 10×10 km block of the data, and a weighted average of their predictions is taken, with weights inversely related to distance from the block centroid (Figure 4.4). The resulting nonstationary dispersal model is itself a kernel, so we refer to it as D_t (with associated matrix $\boldsymbol{D}_t$). Its explicit mathematical form is derived in Appendix 4.5.2.

To construct D_t we used a total of 625 stationary kernels, positioned on a 25×25 grid of block centroids covering the study area. Each captures local flight patterns only; Thus we allow it to influence beetle pressure $\mathbb{E}\left(\boldsymbol{B}_t\right)$ only in a neighbourhood of the block over which it was fitted. The distance-weighting function was chosen such that zero weight is assigned beyond a distance of 7.1 km, corresponding to the distance from the centroid to the corners of the block.

This scheme tracks movements up to 14.2 km, a reasonable upper bound on self-powered dispersal given laboratory studies suggesting fewer than 10% of MPB are capable of flight beyond this distance (Shegelski, Evenden, and Sperling, 2019). To avoid overparameterizing

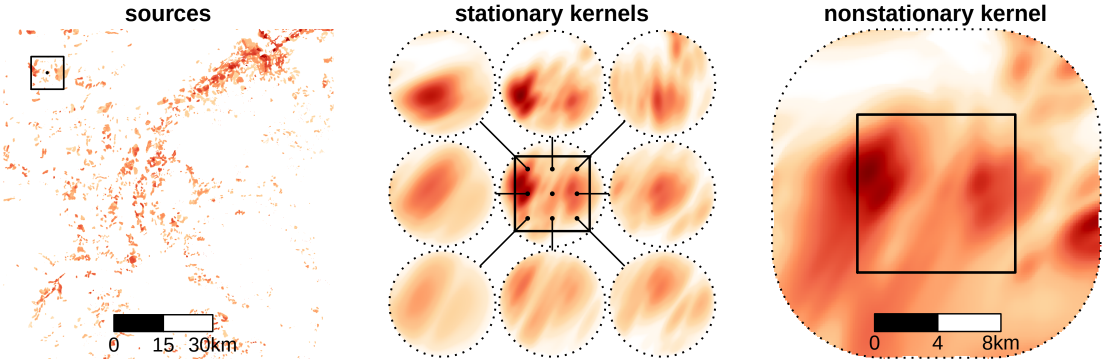

Figure 4.4: A nonstationary flight pattern estimation scheme: stationary kernels are separately fitted to small overlapping blocks of data (at left, a block and its centroid). Expected beetle pressure (detail, at right) is computed as the distance-weighted average of nearby kernel predictions. The middle panel shows the nearest 9 block centroids and their kernel predictions before averaging

an already complicated model – and lacking data on wind patterns – we assumed that atmospherically-driven flight events (such as those documented by Jackson et al., 2008) were rare enough to ignore.

4.2.6 Data

Pine density $\boldsymbol{H}_t$ was estimated using the model output of Beaudoin et al. (2014) for the year 2001, after adjusting for losses due to wildfire, logging, and pest damage incurred during the intervening years (Appendix 4.5.1). For simplicity we did not attempt to model regeneration, but rather assume that changes in density due to growth were small enough to ignore over the period 2001-2008.

Pine mortality data are drawn from the AOS of the Merritt TSA (Figure 4.1) for the attack years 2006-2008. These were rasterized by standard methods (Appendix 4.5.1) to produce a 1000×1000 grid of sample locations at a 1 ha resolution, matching the geometry of the pine density dataset. To avoid edge effects in dispersal calculations, we excluded a $\approx$10km buffer at the edge of this grid from the response data, forming the (logit-transformed) vector ϕ_t from the subgrid of dimensions 893×893 centered on this region (a within-year sample size of 797,449 points).

4.2.7 Errors and data-fitting

A redistribution kernel is a probabilistic model – it connects MPB damage patterns to the *expected* density of attackers arriving next year at each location $\mathbb{E}(\boldsymbol{B}_t)$. Variations of $\boldsymbol{B}_t$ about this mean should therefore be modeled as error. Investigations into ecological dispersal by Preston (1948) and Limpert, Stahel, and Abbt (2001) inform us these errors are likely to be lognormally distributed. Assuming, $\big(\mathbb{E}(B_{i,t}) - B_{i,t}\big) \overset{iid}{\sim} \text{lognormal}(0, \tilde{\sigma}_t^2)$, we can summarize Sections 4.2.1-4.2.4 in the equation:

$$\underbrace{\mathrm{logit}\,(\boldsymbol{\phi}_t)}_{\text{pine mortality log-odds}} = \underbrace{\boldsymbol{X}_t\boldsymbol{\beta}_t}_{\text{susceptibility}} + \kappa_t \log\underbrace{(\epsilon_t \boldsymbol{I} + \lambda \boldsymbol{D}_t\,(\boldsymbol{\phi}_{t-1} \odot \boldsymbol{H}_{t-1}))}_{\text{beetle pressure}} + \underbrace{\boldsymbol{Z}_t,}_{\text{error}} \tag{4.5}$$

where $\boldsymbol{X}_t = (\boldsymbol{x}'_{1,t}, \ldots \boldsymbol{x}'_{n,t})'$ is the (covariate) data matrix for year t, and $\boldsymbol{Z}_t$ is the vector of process errors arising from $\boldsymbol{B}_t$. The logit and log functions are applied elementwise, and the symbol $\odot$ denotes elementwise multiplication. This slight abuse of notation allows us to suppress the location indices i and write the complete model (4.5) in terms of length-n vector operations.

Under the lognormal assumption, $\boldsymbol{Z}_t$ is mean-zero multivariate normal (MVN), with a variance $\kappa_t\tilde{\sigma}_t^2$ that scales with the strength of the density dependence in $\phi(B)$. We assume that measurement error introduces an additional mean-zero MVN random vector appearing additively on the logit scale of (4.5). Since these errors are presumably independent of $\boldsymbol{B}_t$, their effect (by standard MVN theory) is to simply increase the variance of $\boldsymbol{Z}_t$. Thus, ignoring any autocorrelation, we could write $\boldsymbol{Z}_t \sim \text{MVN}\left(\boldsymbol{0}, \sigma_t^2\boldsymbol{I}\right)$, where σ_t^2 is the sum of the variances from process and measurement error.

For simplicity we ignored temporal autocorrelation by treating each year of data in the analysis as independent, as is commonly done in large-scale MPB outbreak analyses (*eg.* Heavilin and Powell, 2008; Goodsman et al., 2016). While this is not ideal, it avoids the difficulties associated with aligning subsequent years of raster data containing a large number of slight positional errors (Wulder et al., 2009), while simplifying the error model both mathematically and computationally.

Spatial autocorrelation, on the other hand, is more easily corrected using covariograms (Chiles and Delfiner, 2012). For computational efficiency we used the Gaussian covariogram, which generates a covariance matrix $\boldsymbol{V}_t$ (to replace $\sigma_t^2 \boldsymbol{I}$ above) based on σ_t^2 and a pair of correlation range parameters, $\boldsymbol{\rho}_t$. In this model, the logarithm of the likelihood function for observations of ϕ_t, given ϕ_{t-1} and $\boldsymbol{X}_t$ is proportional to:

$$\mathcal{L}\left(\boldsymbol{\theta}_t \mid \boldsymbol{Z}_t\right) = -\log\left(\det(\boldsymbol{V}_t)\right) - \boldsymbol{Z}_t' \boldsymbol{V}_t^{-1} \boldsymbol{Z}_t \quad \text{where } \boldsymbol{\theta}_t = \left(\boldsymbol{\theta}_{\phi_t}, \boldsymbol{\theta}_{D_t}, \boldsymbol{\theta}_{V_t}\right) \tag{4.6}$$

with $\boldsymbol{Z}_t$ as defined in (4.5), and model parameters $\boldsymbol{\theta}_t$ organized into components of attack dynamics ($\boldsymbol{\theta}_{\phi_t}$), dispersal ($\boldsymbol{\theta}_{D_t}$), and error ($\boldsymbol{\theta}_{V_t}$); as in Table 4.2. The model can now be fitted to data by maximum likelihood estimation (MLE), which finds the maximizer of (4.6), called $\hat{\boldsymbol{\theta}}_t = (\hat{\boldsymbol{\theta}}_{\phi_t}, \hat{\boldsymbol{\theta}}_{D_t}, \hat{\boldsymbol{\theta}}_{V_t})$.

Our estimation method for $\hat{\boldsymbol{\theta}}_t$ is based on the 2-step algorithm described in Crujeiras and Van Keilegom (2010), but with a blockwise approach to approximating the large number of parameters in $\hat{\boldsymbol{\theta}}_{D_t}$. Each of the 625 pWMY kernels is fitted indendently to the data in its block, before being combined to form the nonstationary kernel matrix $\hat{\boldsymbol{D}}_t$. By assuming $\boldsymbol{D}_t \approx \hat{\boldsymbol{D}}_t$, estimation of the remaining parameters $\hat{\boldsymbol{\theta}}_{\phi_t}$ and $\hat{\boldsymbol{\theta}}_{V_t}$ then becomes straightforward using generalized least squares (GLS) based methods (Chiles and Delfiner, 2012). Simulations indicated that our approach yields unbiased and reasonably precise estimates for $\hat{\boldsymbol{\theta}}_t$ (Appendix 4.5.3).

4.3 Results

The estimated endemic densities and attack curve shapes in all three years (Figure 4.5) matched closely with ground surveys of our study area during the period 2001-2005. The emerging endemic population (ϵ_t) was estimated at 388, 279, and 566 (females/ha), respectively for the years 2006 – 2008; After flight-establishment loss, this would indicate approximately 7 – 14 attacks/ha by the endemic phase, similar to the ranges reported in Boone et al. (2011) and Bleiker et al. (2014). A density dependence in attack was detected in all years, with κ estimated at 1.69, 1.32, and 1.67

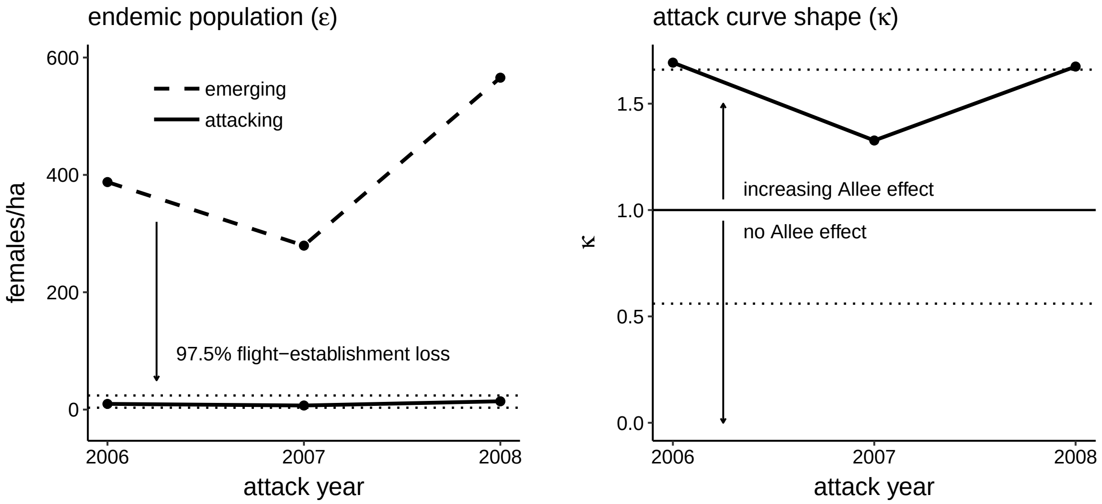

Figure 4.5: Fitted attack parameters. At left, estimates of the endemic population and expected attack rates lying within the range (dotted lines) reported in Boone et al. (2011). At right, estimates of the attack curve shape compared with reference levels from Cooke and Carroll (2017) (dotted lines)

Estimates of stand susceptibility $a_{i,t}$ varied across the landscape, being spatially dependent on $\boldsymbol{x}_{i,t}$. Locations unsuitable to MPB (such as unforested areas) tended to assume extremely large $a_{i,t}$ values whereas areas with optimal habitat for MPB assumed much smaller ones. Restricting our attention to optimal stands only – *ie.* those having a density of 800-1500 stems/ha and aged $>$ 80 yrs (Carroll et al., 2006), representing around 150,000 locations – the observed distribution of susceptibility values is more easily compared to empirical data from similar outbreaks (Figure 4.6).

For example the modes of the estimated $m_{A_{i,t}}$ values over these optimal stands were centered at 336, 932, and 480 females/stem, for the years 2006-2008 respectively. This is reasonably consistent with the 300-617 females/stem range observed in our study area by Safranyik and Linton (1991) during a previous outbreak in 1984. Using data on average diameters and attack heights for these optimal stands (23cm, Carroll et al., 2006; and

11.36m, Safranyik and Linton, 1991; respectively), we estimate a typical bark area of 5.5 m^2/stem (Safranyik, 1988, eq. 6). Our typical per-m^2 observed attack density ($m_{A_{i,t}}/5.5$) therefore lay in the range of 61-170 females/m^2. Note that the lower end of this range (observed in 2006) coincides exactly with the optimal attack density measured by Raffa and Berryman (1983) (Figure 4.6, right).

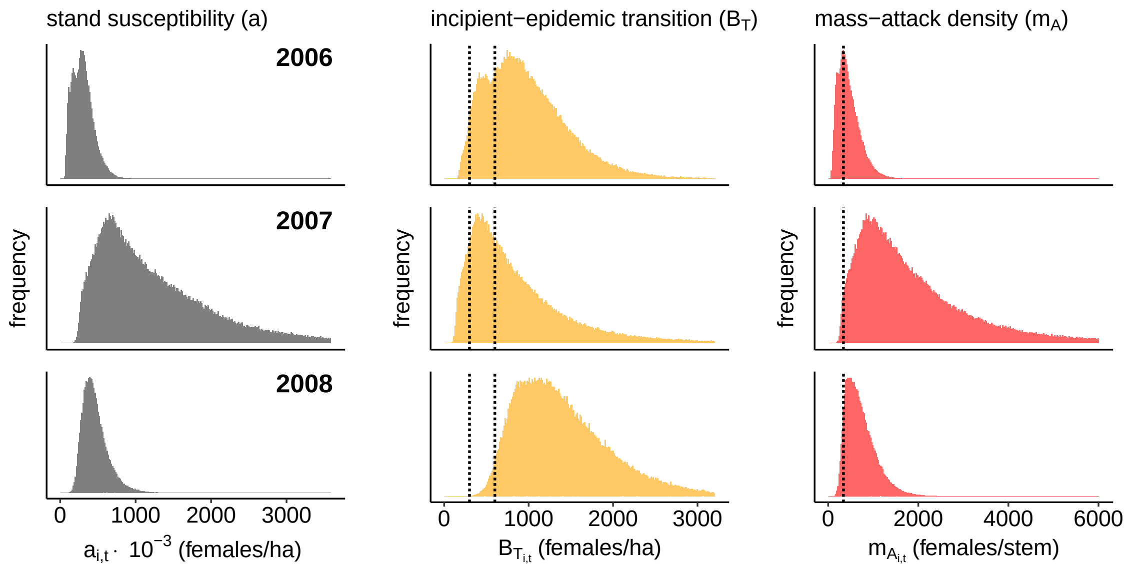

Figure 4.6: Histograms of estimated susceptibility ($a_{i,t}$ left) in stands optimal for MPB in the years 2006-2008, and two associated quantities: (middle) the beetle pressure required for one mass attack per 15 ha, with dotted lines indicating an empirical range (Cooke and Carroll, 2017); and the mass attack number (right), with a dotted line indicating the optimum of Raffa and Berryman (1983).

Flight events under the fitted model are summarized by the blockwise kernel estimates in Figure 4.7 and resulting beetle pressure heatmaps in Figure 4.8.

4.4 Discussion

The S-shaped recruitment curves that characterize the nonlinearity of MPB attack dynamics (*eg.* Raffa and Berryman, 1983; Boone et al., 2011) are usually fitted to field data on individual attacked trees, and so they relate attack density to the mortality among pines undergoing attack. This is a conditional probability model. For example the model of

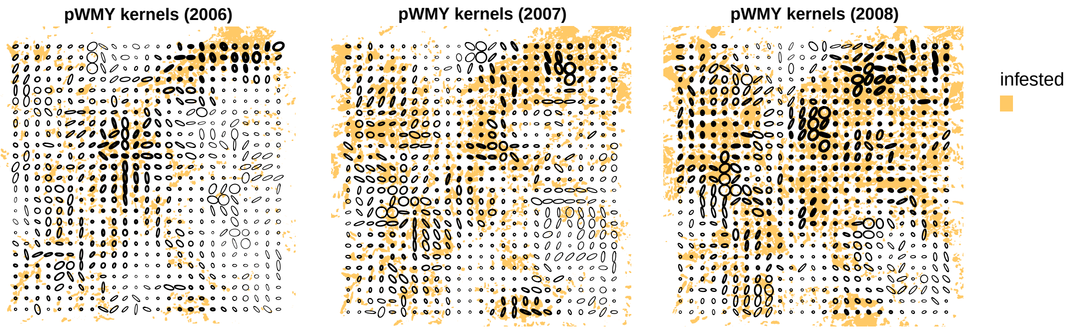

Figure 4.7: Diffusion ellipses summarizing the angle and effective range corresponding to each of the 625 fitted pWMY parameter sets used to construct $\hat{\boldsymbol{D}}_t$ for each year. Each ellipse inscribes a contour of constant density for dispersal from its center. Line thickness is scaled to match the estimated number of MPB displaced, emphasizing major outbreak centers. Infestations from the previous year are shaded to indicate the spatial distribution of source populations.

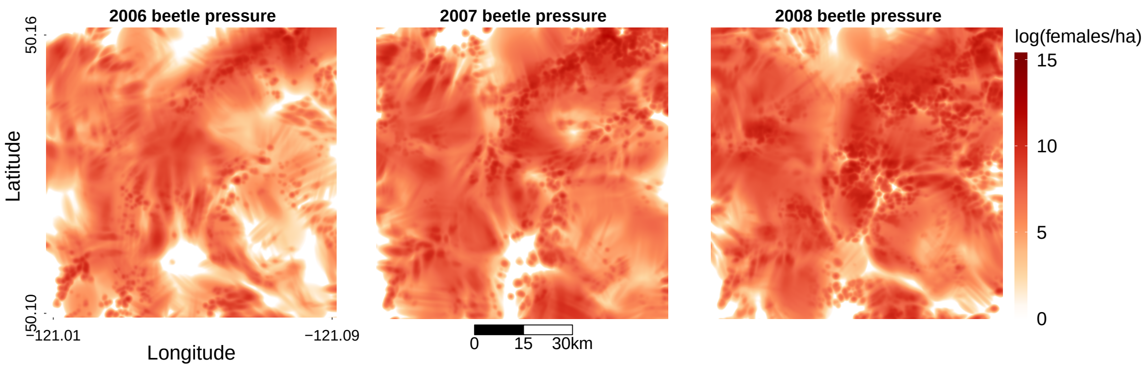

Figure 4.8: Heatmaps of $\log(\lambda \hat{\boldsymbol{D}}_t\,(\phi_{t-1} \odot \boldsymbol{H}_{t-1}))$, the fitted beetle pressure values arising from flight events in the years 2006-2008 (excluding endemic MPB). $\hat{\boldsymbol{D}}_t$ is the moving average of predictions from a 25×25 grid of local stationary models, each fitted to a local subset of the data

Cooke and Carroll (2017) has the form:

$$\text{logit}\left(\Pr(\text{pine mortality} \mid \text{attack})\right) = A + \kappa \log(N_a) = (A - \kappa \log(c)) + \kappa \log(B) \qquad (4.7)$$

where A is an dimensionless intercept; and N_a is the number of stems attacked within the study plot, which we expect to scale according to $cN_a \approx B$ with the attack density B (in females/ha).

Our model however is based on aerial data, from which failed attacks cannot be resolved.

In (4.2) we therefore related B to the *unconditional* probability of stand level mortality $\Pr(\text{pine mortality} \mid \text{attack}) \Pr(\text{attack})$, which we called ϕ. Notice that when $\Pr(\text{attack}) = 1$, both the red-top model of Heavilin and Powell (2008) and our generalization (4.1) coincide exactly with (4.7). In reality, attack rates will be much lower, so in the high-level description (4.1) we make the assumption that the logit-linear relationship (4.7) remains after aggregating mortality data at the 1 hectare scale.

Our results (Figure 4.5) supported this assumption. Indeed the estimated density dependence parameter $\hat{\kappa}$ in 2006 and 2008 very nearly matched the value of 1.66 reported by Cooke and Carroll (2017) for pooled recruitment curve data from the preceeding years 2002-2003 and 2005. This indicates that not only is density dependence detectable from stand-level AOS data (in the absence of failed attack counts) – supporting the findings of Goodsman et al. (2016) on Allee effects – but also that the precise shape of the attack curve in (4.7) can be estimated from aerial data on $\boldsymbol{\phi}_t$ and $\boldsymbol{H}_t$ alone. This includes both the Allee and compensatory (crowd competition) effects (Figure 4.2).

In Section 4.2.1 we showed how, via stand-susceptibility (a), this κ value is mathematically linked to the mass attack number and the incipient-epidemic transition point. A comparison of our point estimates for these parameters with empirical data from previous years showed reasonably good agreement, supporting the theory behind formulae (4.3) and (4.4). From yearly plots of the distribution of these values over locations of optimal habitat (Figure 4.6), we see that 2006 was a year of strong population growth for MPB, with a relatively low threshold for outbreak emergence (B_T), and mass attack numbers (m_A) centered at or near the optimum for brood production.

MPB populations continued to expand through the next two years, before collapsing in 2009. This expansion was accompanied by a large number incipient epidemic transition events. Our model indicates that in optimal habitat these events typically happened when MPB attack densities increased through the range 427-1114 of females/ha (the modes of the estimated $B_{T_{i,t}}$ by year; Figure 4.6, middle). This agrees with empirical observations by

(Cooke and Carroll, 2017) of a transition point in the 300-600 range during the five years leading up to 2006, and indicates that B_T values spiked as the epidemic neared collapse in 2008

These increases, along with the elevated m_A levels in 2007-2008, can be attributed to host depletion. MBP tend to prefer pine of a certain phloem, size, and vigour class (Shrimpton and Thomson, 1985; Cole and McGregor, 1983; Raffa and Berryman, 1983). As the preferred hosts become scarce, MPB likely balance increasing fitness costs by first intensifying mass attacks on the few that remain (Lewis, Nelson, and Xu, 2010), thus effectively increasing m_A above its optimal level. Similarly, a scarcity of suitable mass-attack targets can be expected to make spontaneous eruptions from the endemic phase less likely.

These findings support the observation of Carroll et al. (2006) that the incipient-epidemic transition point seems to occur at a level slightly above the density required to mass-attack a single pine. Our model expresses this quantity by the ratio $B_{T_{i,t}}/m_{A_{i,t}}$, whose median values (in optimal MPB habitat) were 2.2, 0.5, and 2.0 in the years 2006-2008, respectively.

Our estimates of the endemic population size ϵ_t in each year (Figure 4.5) were in remarkably close agreement with ranges expected from field studies of this cryptic population phase (Boone et al., 2011; Bleiker et al., 2014). On dividing the $B_{T_{i,t}}$ values in Figure 4.6 by these estimated ϵ_t and taking medians, we find that a $2.5 - 3.2$X increase in the endemic population was typically sufficient to initiate an outbreak. This illustrates how outbreaks might sporadically arise across the landscape – if environmental conditions were to double or triple the number of injured/weakened pines available to the endemic population, this could allow it to grow to the point of exceeding B_T in the absence immigrating MPB – in accordance with the theory of Berryman (1978), and the explanation of Cooke and Carroll (2017) as to the origin of the outbreaks analysed in Section 4.3.

In-flights of MPB are of course crucial to understanding MPB outbreak dynamics – allowing them to spread through space. This is clear from the large number of spatial regression studies pointing to beetle pressure as the single most significant factor in outbreak development (*eg*. Aukema et al., 2008; Preisler et al., 2012; Sambaraju et al., 2012). As we explained in Section 4.2.5, beetle pressure simply expresses our modeling assumptions about MPB dispersal; Different modelling approaches handle this problem in different ways.

With few exceptions (such as Powell and Bentz, 2014; Powell et al., 2018) forecasting models often reconstruct beetle pressure in a heuristic way, by defining infestation indicator variables that are summed over local spatial neighbourhoods (see *eg*. Shore, Safranyik, and Lemieux, 2000; Aukema et al., 2008; Robertson et al., 2009; Kunegel-Lion, McIntosh, and Lewis, 2018). Many attack dynamics regression models also employ this trick (*eg*. Zhu, Huang, and Reyes, 2010; Preisler et al., 2012; Sambaraju et al., 2012; Kärvemo et al., 2014), and indeed a stationary and isotropic kernel-based representation (as in Heavilin and Powell, 2008; Goodsman et al., 2016) is simply a refinement that finds a biology-based shape (and range) for the filter. Our method refined this idea further, in a novel way, by introducing directedness and location-dependence by means of a weighted combination of stationary kernels.

These fitted pWMY kernels identified a large number of highly directed (anisotropic) dispersal events in all years. Their combination to form a nonstationary kernel ($\hat{D}_t$) brings into focus a complex landscape of MPB movement patterns (Figure 4.8), illustrating how detailed information on beetle pressure can be recovered from AOS data by rethinking the usual modeling assumptions about dispersal. The fitted grid of dispersal kernel parameters ($\hat{\boldsymbol{\theta}}_{D_{k,t}}$) that generate $\hat{D}_t$ (Figure 4.7) resembles a smooth vector field, suggestive of a connection with local wind patterns or habitat characteristics. Future work might connect environmental drivers to $\boldsymbol{\theta}_{D_{k,t}}$ (similar to Powell and Bentz, 2014) as a means of studying the dispersal process itself.

Though we do not analyse the kernel parameters ($\boldsymbol{\theta}_{D_t}$) here, it is worth remarking that in most of the pWMY kernels a leptokurtic pattern of dispersal was favoured over the simpler Gaussian model of biodiffusion. This highlights the versatility of the pWMY in modeling different flight mechanisms (see also Chapter 3), and suggests that a wide range of MPB flight behaviours are realized across the landscape: including both the fat-tailed patterns, proposed by Goodsman et al. (2016) and Turchin and Thoeny (1993); and the Gaussian, suggested by Heavilin and Powell (2008).

The model-fitting procedure of Section 4.2.7 is intended to study attack dynamics (at least) one year after they occur, not to predict them in future summers. Nor is our estimate of stand susceptibility $a_{i,t}$ (as a log-linear function of local covariates) intended for extrapolation; A more judicious choice of covariates whose values can be projected in time (combined with a significance-based feature selection) would be needed in a predictive risk model for MPB damage. Nonetheless we think the framework in (4.5) – and in particular the nonstationary approach to dispersal – will be helpful in building model-based solutions to management and forecasting problems.

For example, given observations (or projections) of pre-dispersal MPB density ($\tilde{\boldsymbol{B}}_t$), the fitted dispersal matrix and endemic level (say $\hat{\boldsymbol{D}}$ and $\hat{\epsilon}$) from a previous year could be used to predict beetle pressure using the equation $\boldsymbol{B}_t = \epsilon \boldsymbol{I} + \hat{\boldsymbol{D}} \tilde{\boldsymbol{B}}_t$ from Section 4.2.5. This in turn could be compared with a given incipient-epidemic transition point B_T to classify stands as potentially infested ($B_{i,t} > B_T$) or not. We illustrate the idea in Figure 4.9 using the empirical threshold of $B_T = 450$ (the midpoint of the range reported in Cooke and Carroll, 2017), and comparing to estimates that assume $\epsilon_t = 0$. Notice that neither $a_{i,t}$, κ_t nor $\boldsymbol{\theta}_{V_t}$ is needed for this classification.

The true positive rate in the training year 2006 was 93.5%, and in the forecast for 2007 it improved to 98.0%. Note that these high detection rates lie near the level mentioned in Six, Biber, and Long (2014) for stabilizing outbreaks by mitigation measures (such as cut and burn). However with high recall comes a high false positive rate (low precision); Moreover the 2007 prediction required information on pre-dispersal density that is typically not available until *after* the attack summer being predicted – recall that $\tilde{\boldsymbol{B}}_t$ is derived from crown fade data with a one-year lag. One possible solution would be to iterate equation

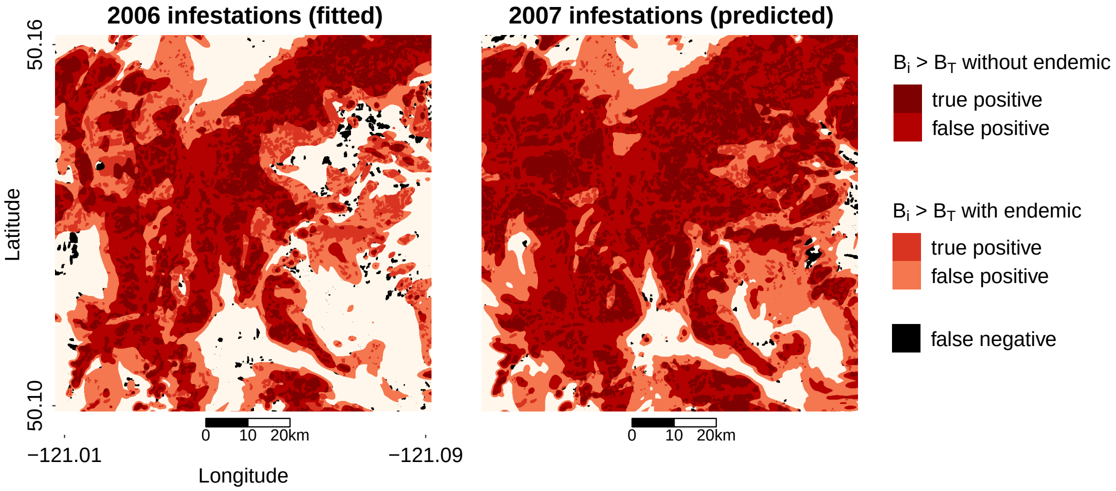

Figure 4.9: Infested locations identified in the training year and next-year forecasts. Using the fitted values of ϵ_t and $\boldsymbol{D}_t$ from the training year 2006 (left), locations were classified as infested (shaded) if the predicted beetle pressure exceeded $B_T = 450$. Using these same parameters along with the observed attack damage and pine density in 2006, we then predicted infestations in 2007 (right). For comparison, an endemic-free estimate is also plotted (darker shaded regions) by replacing $B_{i,t}$ with $B_{i,t} - \epsilon_t$. The effect is to withdraw the contours of infestation inward, limiting spread considerably.

(4.5) with simulated error to produce a suite of multi-year forecasts under various scenarios of stand susceptibility and process error, an idea we plan to explore in future work.

Figure 4.9 illustrates an important consequence of the ubiquity of endemic MPB in their natural range: it increases the potential for outbreaks to spread into new areas. By including the endemic population in our beetle pressure estimates, the contours of the infestation predictions broadened, sometimes by several kilometers. This improved detection rates (true positive rate in 2006 without the endemic component: 71%; and in 2007: 84%), a finding that may be important to modelers building outbreak spread forecasts.

It also suggests that the potential for range expansion may be underestimated if the endemic contribution to MPB outbreaks is ignored. This will be of particular relevance in contemporary areas of concern, such as the Boreal forest in Alberta (Safranyik et al., 2010). The establishment of endemic populations should be monitored as it has the potential to accelerate the spread of outbreaks and thus speed the range expansion of the MPB.

4.5 Appendices to Chapter 4

4.5.1 Datasets

Host density H

Host density $H_{i,t}$, or the number of susceptible overstory pine trees within the 1 ha square of land in cell i in year t, was approximated using variable $h_{i,t}$, the combined above-ground biomass at cell i attributed to *Pinus* species. We calculated this from a 2001 baseline estimate supplied in Beaudoin et al. (2014), making adjustments in each subsequent year by subtracting losses from wildfire and logging (using data from http://www.hectaresbc.org), as well as MPB.

While raw data in units of live mature stems/ha would be both more realistic and convenient, they are typically not available at the scale and resolution that we are interested in. We therefore simply rescaled h_i to match empirical distributions of H_i based on ground surveys (Figure 4.10).

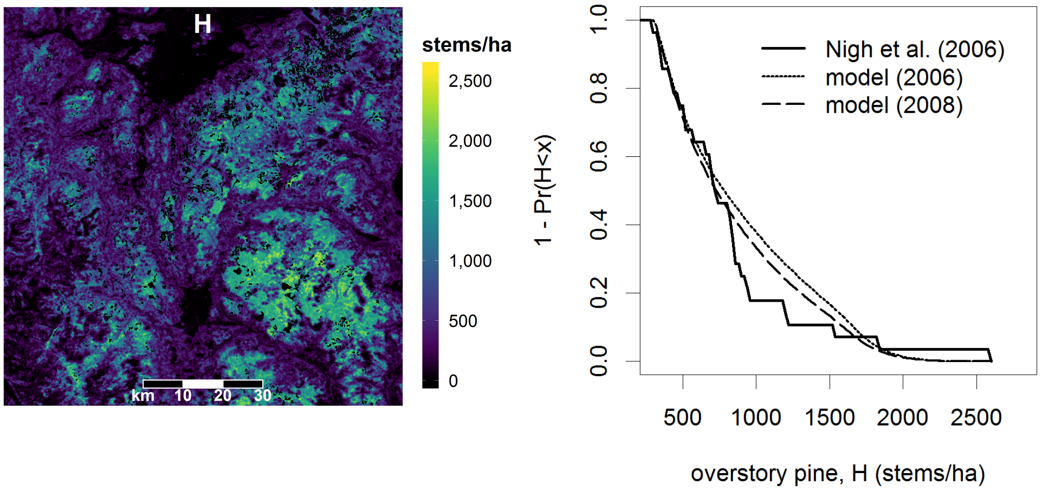

Figure 4.10: Heatmap of host density $H_{i,t} = s_h h_{i,t}$, estimated from rescaled pine volume data $h_{i,t}$. At right, the empirical CDF of pre-attack host density in 2006 and 2008 (dashed and dotted lines) are compared against an overstory pine density survey from 2006 by Nigh, Antos, and Parish (2008) (solid line).

In a 2006 survey of 28 high-density stands in the Merritt TSA by Nigh, Antos, and Parish (2008), the highest observed density was 2810 stems/ha, 92% of which was pine. Based on that maximum we assigned a scaling factor of $s_h = 0.92 \times 2810/\max_i(h_{i,2006}) = 9.1$ and fixed $H_{i,t} = s_h h_{i,t}$. A more cautious approach, for example using mensurational projections for the stands in our study area, is possible but was avoided for the sake of simplicity. However our linear rescaling produced a reasonably close agreement in empirical cumulative distribution (CDF) functions (Figure 4.10), and the scaling factor of $\approx 10X$ is essentially a linearization of the more carefully constructed nonlinear volume-density curve derived in Goodsman et al. (2016) (Appendix S1).

Pine mortality ϕ

Our response variable $\phi_{i,t}$ is the percent of $H_{i,t}$ killed by pine beetle attack in the summer of year t. We derived these values from AOS data comprising two types of GIS information collected annually by BC's provincial forest management agency: polygons with categorical damage severity attributes (digitized sketch maps) indicating large contiguous areas of infestation; and spot data indicating a small cluster of infested stems at a particular location. Because crown-fade typically happens with a one-year delay, we refer to year $t + 1$ in the AOS dataset as the *attack year* t.

To convert polygons to raster format we followed a protocol introduced by Chen and Walton (2011); the five AOS damage severity categories (corresponding to intervals of percent mortality: trace <1%, light 1-10%, moderate 11-29%, severe 30-49%, very severe >50%) were interpreted by multiplying the midpoint of each interval with the percent area of overlap with each cell. Spots were interpreted by defining a quarter-hectare circle centered at the point coordinates and assigning it a 30% mortality value (reflecting AOS-wide average stand loading and spot infestation levels).

Some minor modifications of these mortality data were needed to correct obvious positional errors and to make our analysis approach feasible: Attack rates >1 (due to multiple overlapping damage observations) were truncated to one, and values at locations unsuitable for MPB (water bodies, non-treed areas, *etc.*) were set to zero. We then added the small constant $\xi = 4 \times 10^{-6}$ (equal to one half the minimum finite logit value) to each cell before dividing by $1 + \xi$. This ensured that $0 < \phi_i < 1$, so that $\text{logit}(\phi_i)$ is defineable all sites. It is also consistent with premise of ubiquitous endemic MPB populations, undetectable by the AOS (Wulder et al., 2006).

Stand susceptibility covariates

Beetle pressure is only part of the equation in MPB attack dynamics. Environmental conditions before and during an attack, as well as the density, composition, and health of the stand influence the ability of a given pine to resist bark beetle attack (Safranyik and Carroll, 2006; Nelson et al., 2008). These local conditions are often summarized as *stand susceptibility*, a ranking of relative risk (to MPB attack) computed from local covariates. The model of Shore, Safranyik, and Lemieux (2000), for example, uses the product of four covariates relating to: pine dominance; stand density; stand age; and elevation.

Our model uses a similar product of stand characteristics along with a suite of additional microclimate and topography-related covariates, similar to those found in Aukema et al. (2008). Local stand characteristics, such as H_i, were derived from the Beaudoin et al. (2014) model, and topographical features were drawn from provincial government datasets (http://www.hectaresbc.org). Local weather variables, such as temperature and precipitation highs/lows were constructed using the *climateBC* model, via elevation-adjusted extrapolations from weather station measurements and climatic norms (Wang et al., 2007). In total, we compiled 43 such covariates for each of the n sites in the study area ($n_\beta = 44$, including an intercept). These are the rows of the $n \times n_\beta$ matrices $\boldsymbol{X}_t$. They are summarized in Table 4.3.

4.5.2 Redistribution kernels for 2-dimensional space

Redistribution kernels view dispersal events as moving individuals from a fixed source to a random destination. If the coordinates of the i^{th} possible destination are $\boldsymbol{s}_i = (x_i, y_i)'$, then we write the *movement vector* from source (j) to destination (i) as $\boldsymbol{\delta}_{ij} = \boldsymbol{s}_i - \boldsymbol{s}_j = (\delta^x_{ij}, \delta^y_{ij})'$, where $\delta^x_{ij} = x_i - x_j$ and $\delta^y_{ij} = y_i - y_j$ are the components of the movement along the x and y axes. Direction (angle α_{ij}) and distance (d_{ij}) are then given by the identities $d^2_{ij} = |\boldsymbol{\delta}_{ij}|^2 = (\delta^x_{ij})^2 + (\delta^y_{ij})^2$, and $\tan(\alpha_{ij}) = \delta^y_{ij}/\delta^x_{ij}$.

We define the redistribution kernel $D(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\Delta})$ to be the probability mass function (PMF) for possible destinations, with parameters $\boldsymbol{\Delta}$. For simplicity modelers usually choose kernels that are spatially *stationary* (invariant to location), and *isotropic* (invariant to direction). Stationarity means movement probabilities depend only on direction and distance, so D can be written $D(\boldsymbol{\delta}_{ij}; \boldsymbol{\Delta})$. With the additional assumption of isotropy, D becomes a function of distance d_{ij} only, or $D = D(d_{ij}; \boldsymbol{\Delta})$. In general, we will write D for the function and $[D]_{ij}$ for its value with source j and destination i.

The (isotropic and stationary) Gaussian is the most common kernel in applications:

$$D_G(d_{ij}; \rho) = c \exp\left(-d^2_{ij}/\rho\right), \text{ where } d_{ij} = |\boldsymbol{s}_i - \boldsymbol{s}_j|. \tag{4.8}$$

c is a normalization constant, chosen such that with $\boldsymbol{s}_j$ fixed, the summation of (4.8) over all destinations is equal to one. This normalization is a general requirement of any PMF, but in the context of redistribution kernels it ensures that total population counts are conserved. More precisely, if we start from local source populations of size $\tilde{B}_j$, with individuals at each source dispersing independently and according to D, then the expected number to arrive at destination i is $B_i = \sum_j [D]_{ij}\tilde{B}_j$, and the sum of the B_i is equal to the sum of the source populations.

Thus D is sometimes chosen by selecting a function that matches the profile of empirical data on B_i. Other times, hypotheses about the movement mechanism lead to mathematical derivations. For example, under a quite general set of circumstances, diffusion through

2D space gives rise to the WMY kernel family (Yasuda, 1975; Yamamura, 2002; Hapca, Crawford, and Young, 2008):

$$D_W\left(d_{ij};\boldsymbol{\Delta}_W\right) = c\left(d_{ij}/\rho\right)^{\nu} K_{\nu}\left(d_{ij}/\rho\right), \text{ where } \boldsymbol{\Delta}_W = (\nu,\rho)', \tag{4.9}$$

with shape parameter $\nu > -1$; range parameter $\rho > 0$; normalization constant c (computed as above); and with K_{ν} to denote the ν^{th} order modified Bessel function of the second kind.

The Gaussian (4.8) and 2D Laplace kernels used in Heavilin and Powell (2008), are limiting/special cases of the WMY ($\nu \to \infty$, and $\nu = 1/2$, respectively). The Bessel kernel appearing in the bark beetle models of Turchin and Thoeny (1993) and Goodsman et al. (2016) is another special case ($\nu = 0$). In this sense (4.9) is robust with respect to hypotheses about movement. We use an approximation to (4.9) that is somewhat more flexible, the geometrically anisotropic product-WMY:

$$D_{\otimes}\left(\boldsymbol{\delta}_{ij};\boldsymbol{\Delta}\right) = cD_W\left(d^x_{ij};\boldsymbol{\Delta}^x\right)D_W\left(d^y_{ij};\boldsymbol{\Delta}^y\right), \tag{4.10}$$
$$\text{with } \boldsymbol{\Delta} = (\alpha,\boldsymbol{\Delta}^x,\boldsymbol{\Delta}^y)', \text{ and } (d^x_{ij},d^y_{ij})' = \boldsymbol{R}_{\alpha}\boldsymbol{\delta}_{ij},$$

where $\boldsymbol{R}_{\alpha}$ is the standard 2D rotation matrix for angle α and c the normalization constant. This kernel is similar to the WMY, closely approximating it over much of its parameter range, yet it can be computed far more quickly because, like the Gaussian, it is spatially separable (as discussed in Chapter 3). Moreover it better captures directed movements, by means of angle α and the independent shape/range parameter sets, $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$, representing two orthogonal directions. Thus unlike an isotropic kernel, (4.10) captures ellipsoid patterns of redistribution (Figure 4.3).

Our nonstationary formulation of D uses a weighted combination of $m = 625$ stationary kernels $D_{\otimes_k}$ ($k = 1\ldots m$), each of the form (4.10), and each with its own parameter set $\boldsymbol{\Delta}_k$. Each is spatially referenced, with coordinates $\boldsymbol{r}_k$ to denote the centroid of a 10×10 km block over which $D_{\otimes_k}$ is assumed to reasonably approximate local flight patterns. The predictions of these local kernels are combined by weighted averaging, with weights inversely related to distance from the centroid $\boldsymbol{r}_k$ to the prediction site $\boldsymbol{s}_i$. For a given weighting function $\omega(d)$, we define the nonstationary kernel:

$$D(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}_D) = c \sum_{k=1}^{m} \omega\left(|\boldsymbol{s}_i - \boldsymbol{r}_k|\right) D_{\otimes_k}\left(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Delta}_k\right), \text{ where } \boldsymbol{\theta}_D = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_m)' \quad (4.11)$$

with normalization constant c computed in the usual way. We used a bisquare weighting function $\omega(d) = \left\lceil 1 - (d/r)^2 \right\rceil^2$, with the ceiling function $\lceil x \rceil$ enforcing a cutoff distance of $r = 7.1$ km beyond which zero weight is assigned. Centroids $\boldsymbol{r}_k$ were arranged in a 25×25 grid of overlapping blocks, with a spacing of 3.3 km between centroids. This balanced a need for large samples within each block (10×10 km $= 10^4$ points) and a desire for high resolution estimates of $\mathbb{E}\left(\boldsymbol{B}_t\right)$.

4.5.3 Model-fitting and simulations

Covariograms are in many ways similar to redistribution kenels. We use a geometrically anisotropic Gaussian covariogram, which defines the covariance between errors at $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ to be:

$$[\boldsymbol{V}_t]_{ij} = \mathrm{Cov}\left(Z_{i,t}, Z_{j,t}\right) = \sigma_t^2 \exp\left((d_x)^2/\rho_x\right) \exp\left((d_{ij}^y)^2/\rho_y\right), \quad (4.12)$$

$$\text{where } (d_x, d_y)' = \boldsymbol{R}_\alpha |\boldsymbol{s}_i - \boldsymbol{s}_j|$$

with $\rho_x, \rho_y > 0$ the range parameters, and α the angle of orientation. For reasons of computational efficiency we fixed $\alpha = 0$ so that (4.12) remains spatially separable (for more detail on separable covariograms see Chapter 2).

4.5.4 Estimation

Supposing beetle pressure is known – either by direct measurement, or by fixing biologically reasonable values for ϵ_t and $\boldsymbol{\theta}_{D_t}$ – (4.5) will become linear in the remaining attack dynamics parameters κ_t and $\boldsymbol{\beta}_t$. The maximization problem (4.6) then becomes a spatial linear regression on stand susceptibility, much like in Aukema et al. (2008) and Zhu, Huang, and Reyes (2010) except with an explicit (rather than implicit) error model. In this situation, using generalized least squares (GLS), it is straightforward to find $\hat{\boldsymbol{\theta}}_{\phi_t}$ and $\hat{\boldsymbol{\theta}}_{V_t}$ numerically using a 2-step estimator (Chiles and Delfiner, 2012).

Similarly if $\boldsymbol{\theta}_{D_t}$, but not $\boldsymbol{\theta}_{\phi_t}$, is known, it remains a relatively straightforward 1-dimensional optimization problem to find $\hat{\boldsymbol{\theta}}_{\phi_t}$ and $\hat{\boldsymbol{\theta}}_{V_t}$ by profile likelihood on ϵ using GLS as above (Crujeiras and Van Keilegom, 2010). However with all three components unknown, the inference problem is far more involved. Our solution is three stage algorithm that requires an initial estimate of beetle pressure. We used $\epsilon = 0$ and the stationary Bessel kernel reported in Goodsman et al. (2016):

1. Assume $\boldsymbol{V}_t \propto \boldsymbol{I}$. Estimate $\hat{\boldsymbol{\beta}}_t$ and $\hat{\sigma}_t$ by OLS given the initial beetle pressure values. Estimate $\hat{D}_t$ by blockwise MLE given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_t$. Estimate $\hat{\boldsymbol{\theta}}_{\phi_t}$ by profile likelihood on ϵ given $D_t = \hat{D}_t$.

2. Estimate $\hat{\boldsymbol{\theta}}_{V_t}$ by MLE on the model residuals from stage 1. Refine the estimate of $\hat{\boldsymbol{\theta}}_{\phi_t}$ by profile likelihood on ϵ given $D_t = \hat{D}_t$.

3. Assume $\boldsymbol{\theta}_{V_t} = \hat{\boldsymbol{\theta}}_{V_t}$. Refine the estimate of $\hat{D}_t$ by blockwise MLE given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_t$ from stage 2. Refine the estimate of $\hat{\boldsymbol{\theta}}_{\phi_t}$ by profile likelihood on ϵ given $D_t = \hat{D}_t$.

In stages 1 and 3, "blockwise MLE" for $\hat{D}_t$ means the following: we split the dataset into square blocks (each containing 10^4 locations) centered over the 625 points of a 25×25 evenly spaced grid covering the study area, and assigned a pWMY kernel (with parameters $\boldsymbol{\Delta}_{k,t}$) to each one. Fixing $\boldsymbol{\beta}_t$ and $\boldsymbol{V}_t$ to their most current estimates as specified in stages 1/3, for each block we jointly estimated the seven remaining unknown parameters (ϵ, κ, and $\boldsymbol{\Delta}_{k,t}$) by numerically maximimizing the likelihood function (4.6), under the assumption that D_t is the stationary kernel (4.10) with parameters $\boldsymbol{\Delta}_{k,t}$. We then used the $\hat{\boldsymbol{\Delta}}_{k,t}$ to construct D_t using (4.11) (discarding the local estimates of ϵ and κ).

Simulations

Our 3-stage algorithm is similar to one described by Crujeiras and Van Keilegom (2010), where $\hat{\boldsymbol{\theta}}_{\phi_t}$ and $\hat{\boldsymbol{\theta}}_{D_t}$ are jointly estimated and $\hat{\boldsymbol{\theta}}_t$ is known to be asymptotically normal and unbiased. However since ours estimates $\hat{\boldsymbol{\theta}}_{\phi_t}$ and $\hat{\boldsymbol{\theta}}_{D_t}$ separately, we have no theoretical guarantees of its large-sample properties. Instead we investigated the properties of our estimators in simulations.

Since our model is computationally intensive, we conducted simulations on a smaller spatial scale – 33×33 km, covered by a 5×5 layout of blocks – and generated D_t using anisotropic Gaussian (instead of pWMY) kernels. The fitted model, however, was as described in the main text, using pWMY kernels to fit D_t. In each of 100 repetitions, we assigned values to the parameters in Table 4.2 uniformly at random within a biologically reasonable range – *eg.* $0.1 < \kappa < 25$, $1 < \epsilon < 500$, and $25 < \lambda < 1000$ – and used covariate data pulled from a randomly located subset of the full 2008 data. We then used (4.5) to compute the true response values $\text{logit}(\phi)$, adding them to randomly generated MVN errors $\boldsymbol{Z}_t$ to produce a (simulated) observed response.

The response and associated covariates were fed into the algorithm of Appendix 4.5.4 to yield estimates $\hat{\boldsymbol{\theta}}_t$ separately for each repetition. Errors in estimation for the attack parameters $\boldsymbol{\theta}_{\phi_t}$ and the angles of anisotropy $\alpha_{k,t}$ are summarized in Figure 4.11. For comparison we also report the errors after stage 1, where the $\boldsymbol{Z}_t$ are assumed to be indepedendent in space (a model misspecification).

Raster plots of the estimated post-flight MPB density closely approximated the true (simulated) ones. Individual fitted stationary kernels also closely resembled the true ones, favouring large shape values (and thus approximating the Gaussian closely), and estimating the dispersal orientation angles with remarkable precision. Interestingly the autocorrelation correction (stages 2-3) had little impact on these angle estimates, so although beetle density estimates differed slightly between stages 1 and 3, the error distribution of the $\alpha_{k,t}$ appears largely unchanged (Figure 4.11).

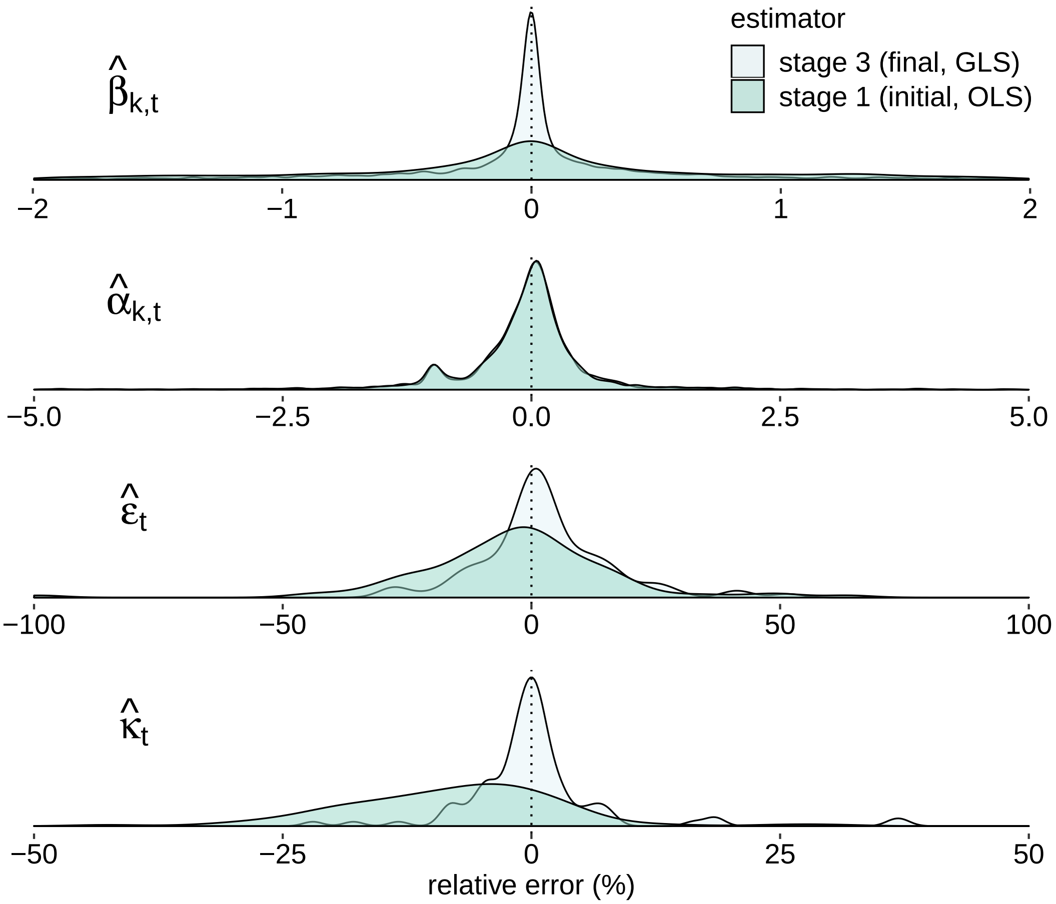

Figure 4.11: Kernel density plots (smoothed histograms) of the relative errors in parameter estimates using the 3-stage algorithm (Appendix 4.5.4), in 100 independent simulations (indexed by t). Results on 44 regression parameters ($\beta_{1,t} \dots \beta_{44,t}$), and 25 angles of dispersal anisotropy ($\alpha_{1,t} \dots \alpha_{25,t}$) are pooled. Stage 1 estimators ignore autocorrelation. Stage 3 estimators correct for it.

However the results for the other parameters highlight some of the reasons we must not ignore spatial autocorrelation: uncertainty is underestimated under an incorrect independence assumption, leading to a wider than expected spread of errors and more frequent misspecifications of $\hat{\alpha}_{i,t}$. In our case, this imprecision appeared to introduce bias in the more sensitive components of the model; Both ϵ and κ tended to be underestimated in stage 1. The stage 3 autocorrelation correction appears to largely eliminate this bias and improve precision (peakedness of the density plots).

The error distributions of the individual $\hat{\beta}_{k,t}$ showed good agreement with the large sample asymptoptic theory in Crujeiras and Van Keilegom (2010), from which confidence intervals can be computed by inverting the Fisher information matrix corresponding to (4.6). Uncertainty in the distributions for ϵ and κ, however, was underestimated by this theory; with only 38% (and 42%, respectively) of estimates lying inside their nominal 95% intervals. This may be due to an inadequate sample size, or a failure to find the values of $\hat{\boldsymbol{\theta}}_{\phi_t}$ and $\hat{\boldsymbol{\theta}}_{D_t}$ that *jointly* maximize (4.6) in stages 1 and/or 3. We therefore omit confidence intervals for these parameters in the main text, reporting the $\hat{\epsilon}$ and $\hat{\kappa}$ simply as point estimates.

Category	Name	Units	Source
topography	altitude	m above sea level	provincial topography layers from hectaresbc.org (accessed 06/2019)
	slope	° above horizontal	
	aspect	° from true north	
	lakes indicator	binary	
stand inventory	treed area	%	damage-adjusted estimates from Beaudoin et al. (2014) based on remotely-sensed data from 2001 (see Section 4.5.1)
	stand age	years	
	pine density	stems/ha	
	log pine density	log(stems/ha)	
beetle activity	lagged pine mortality	%	ϕ_t and $\phi_t H_t$ lagged by one and two years (see Sections 4.5.1-4.5.1)
	lagged infested stems	stems/ha	
temperature	minima	°C	All climatic variables are seasonal, with separate covariates for: autumn of year $t-1$; winter, spring, and summer of year t. These are estimated using climateBC software from (Wang et al., 2007).
	averages	°C	
	maxima	°C	
cooling days	days below 0°C	°C · days	
	days below 18°C	°C · days	
warming days	days above 5°C	°C · days	
	days above 18°C	°C · days	
precipitation	totals	mm / 4 months	

Table 4.3: The 43 covariates included in the linear regression model for stand susceptibility. 31 of these are climatic (four seasons × 8 factors, with the exclusion of degree days above 18° to avoid collinearity problems); Four are lagged state variables (pine mortality and infested stem counts, lagged by one and two years); Four describe the local host population; and four are topographical.

Chapter 5

Discussion

In this thesis, I have contributed new methodologies for modelling the correlation structure of ecological data and the patterns that result from flight-based dispersal events. Chapters 2-3 developed these ideas separately, but they are alike in emphasizing how practical issues of computationally efficiency can be resolved using the mathematical concept of separability on spatial lattices. Chapter 4 showed how, in combination, these ideas become powerful tools for extracting information on subtle aspects of MPB attack behaviour from large-scale crown fade surveys.

I conclude here with a brief summary of my results in the context of established MPB ecology, and a discussion of the directions where these lines of thought may lead us in the future.

5.1 Separable models for spatial autocorrelation

The findings of Legendre (1993), Beale et al. (2010), Hawkins (2012) (and others) caution ecologists to be wary of issues that can arise if spatial autocorrelation (SAC) is ignored in spatial data. These issues were borne out, for example, in my simulation experiments in Chapter 4, where the sampling distributions of three attack dynamics parameters became far more precise (and for one of the nonlinear parameters, less biased) after repeating the inference using covariograms.

However, the MPB ecology literature remains split, with some authors correcting for

SAC and many others ignoring it. In part this may be due the mathematical complexity of the popular autoregressive approach, which can be daunting to ecologists unfamiliar with multivariate statistical methods. The covariogram on the other hand is more immediately intuitive, but suffers from issues of computational complexity that can make large-sample analyses infeasible.

I proposed a family of covariograms in Chapter 2 that exploit the mathematical properties of lattices to simplify computations involved in Gaussian likelihood-based data analyses. Covariance models for lattice data have a long history in spatial statistics, with influential theoretical contributions from researchers such as Besag (1974) and Cressie (1993) popularizing the use of Gaussian Markov random field models (GMRFs) such as the CAR and SAR (Fortin and Dale, 2005; Ver Hoef et al., 2018) that are now prevalent in the contemporary MPB modelling literature. My formulation also draws on random field theory. However, it is not a Markov model – it requires no Markovian assumptions about neighbourhoods of dependence, nor does it obscure the implied correlation structure by representing it through a matrix inverse, as do the CAR and SAR models (Wall, 2004). Instead it appeals to the more intuitive geostatistical paradigm (Matheron, 1962; Stein, 1999), in which covariances are given explicitly as a function of separation distance.

Chapter 2 assembled a number of logical next-steps in the discourse set forth by Zimmerman (1989), on separable lattice covariograms and anisotropy, combining them with some new results, in a context more accessible to ecologists who may not be well-versed in random field theory. My hope is that this work may breathe new life into the field of geostatistics, which is often dismissed out-of-hand by applied researchers for its computational issues (Simpson, Lindgren, and Rue, 2012).

Following Zimmerman (1989), I addressed computability by defining spatially separable covariograms based on the Whittle-Matérn (WM) model (Guttorp and Gneiting, 2006), and demonstrated through simulations their suitability as surrogates for non-separable covariograms. This showed how the "big n problem" in spatial statistics (Simpson, Lindgren, and

Rue, 2012) can be avoided without abandoning the intuitive appeal of covariograms. In that sense my research is similar to that of Lindgren, Rue, and Lindström (2011), who discussed situations in which the pervasive WM covariogram is well-approximated by a certain class of GMRFs for which computations are dramatically faster.

Along the way I discovered some less-obvious features of separable lattice models, including a data partitioning scheme that leads to a fast and accurate estimator of the angle of range anisotropy. In spatial ecology, such departures from isotropy often reflect subtle environmental conditions driving the data-generating process, such as wind and topography (Fortin and Dale, 2005). For example Zhu, Huang, and Reyes (2010) considered anisotropy in their SAR model as a result of directed dispersal flights by MPB. My AOS data analysis in Chapter 2 demonstrated how, using covariograms, these directions can be inferred over large nonstationary landscapes, revealing a smooth field of directionality in forest disturbance patterns. This warrants further research into the environmental origins of this directionality, as they might be used to improve the accuracy of outbreak forecasts.

5.1.1 Relating autocorrelation and dispersal

I described in Chapter 3 a very general set of ecological circumstances in which the expected pattern of redistribution for dispersing inviduals is a WMY kernel. In spatial statistics the same family of functions (the WM kernel) is well-known for its versatility as a covariogram (Stein, 1999). It is interesting that such different fields should converge on the same mathematical form, and the reader may wonder what (if any) meaningful connections are suggested between the concepts of redistribution and autocorrelation.

One way to connect the two concepts is to examine the correlation structure that is introduced into a random variable when it is convolved with a redistribution kernel. Recall that if $\mu(\boldsymbol{x})$ counts a population at location $\boldsymbol{x}$, and all individuals disperse according to the stationary kernel D, then the expected post-dispersal population is given by $\mu_D(\boldsymbol{x}) = D * \mu(\boldsymbol{x})$. Now suppose process error introduces variation about $\mu(\boldsymbol{x})$, and denote this pre-dispersal population by random variable $Z(\boldsymbol{x})$ with mean $\mu(\boldsymbol{x})$ and variance σ^2. The post-dispersal population is then $Z_D(\boldsymbol{x}) = D * Z(\boldsymbol{x})$. How would we characterize its distribution?

If $Z(\boldsymbol{x})$ is a Gaussian random field (RF), this is a well-known problem from sampling design in geostatistics, where $Z_D(\boldsymbol{x})$ is called a *regularization* of $Z(\boldsymbol{x})$ by the *sampling function* $\check{D}(\boldsymbol{r}) := D(-\boldsymbol{r})$ (Chiles and Delfiner, 2012, section 2.4). From standard theory, it follows that if D is square integrable then $Z_D(\boldsymbol{x})$ is a Gaussian random field with mean $D * \mu_Z(\boldsymbol{x})$ and its covariance kernel C_D is proportional to the convolution of D and $\check{D}$. This is sometimes called the *auto-correlation* of D. For example if we assume $Z(\boldsymbol{x})$ has covariance kernel $C_Z(r) = \sigma^2\delta(r)$ (a delta function, implying all finite samples of $Z(\boldsymbol{x})$ are uncorrelated and have equal variance σ^2), then the correlation induced in Z_D by D is:

$$\underbrace{C_D(\boldsymbol{r})/\sigma^2 = \left(C_Z * D * \check{D}\right)(\boldsymbol{r})/\sigma^2}_{\text{correlation function of } Z_D} = \underbrace{\left(D * \check{D}\right)(\boldsymbol{r}) = \int_{\mathbb{R}^d} D(\boldsymbol{s})D(\boldsymbol{r}+\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}}_{\text{auto-correlation of } D}. \tag{5.1}$$

When D is isotropic, (5.1) becomes $C_D/\sigma^2 = D * D$, the autoconvolution of D. In the moving average model of Matérn (1986) a similar equation was derived starting from a Poisson point process $Z(\boldsymbol{x})$ with intensity σ^2. In that context, the convolution with D was viewed as a disturbance produced by the measuring device. Here it is an ecological process, such as a flight-based dispersal.

I showed in Chapter 3 that the WMY kernel family ($\mathcal{D}$) has the desireable property of closure under convolutions: The autoconvolution of $D = \mathcal{D}\left(r;\kappa,\rho\right)$ is yet another WMY kernel; $D * D = \mathcal{D}\left(r; 1+2\kappa,\rho\right)$. This is unusual among 2-dimensional redistribution kernels (Schlägel and Lewis, 2016). In the context of equation (5.1), it shows that in a process where (uncorrelated) errors have been redistributed by a WMY kernel, the resulting correlation structure is described by the WM covariogram.

More generally, equation (5.1) defines a plausible covariance structure for post-dispersal measurements when the dominant source of unexplained variability in a data analysis is uncorellated pre-dispersal process error. This suggests a natural grouping of covariance and redistribution kernels into pairs, with one (D) giving rise to the other (C_D) via the spatial autocorrelation mechanism described above. A tabulation of these pairs (*eg*. Table 5.1) might be useful to modellers tasked with selecting the appropriate covariance kernel for a spatial analysis, given mechanistic assumptions about the physical dispersal of error sources. Some examples are derived in Appendix 5.5.2, in the hope that they might have pedagogical value and/or serve as a useful reference.

redistribution (D)	further reading on D	covariance ($C_D = D * D$)
spherical pulse	Chiles and Delfiner (2012)	spherical
2Dt (shape $\beta = 1/2$)	Clark et al. (1999)	Cauchy
Bessel (WMY, $\kappa = 0$)	Broadbent and Kendall (1953)	Whittle (WM $\kappa_D = 1$)
2D Laplace (WMY, $\kappa = 1/2$)	Joseph and Sendner (1958)	WM ($\kappa_D = 2$)
general WMY ($\kappa > -1/2$)	Yasuda (1975)	WM ($\kappa_D = 1 + 2\kappa$)
Gaussian	Skellam (1951)	Gaussian

Table 5.1: A reference list of 2-dimensional stationary isotropic covariance kernels C_D that can be written as the autoconvolution of a redistribution kernel D. Note that, up to normalization constants, the WMY and WM (Whittle-Matérn) kernels are mathematically identical, as are the 2Dt and rational quadratic kernels. κ_D denotes the shape parameter of the WM covariance kernel.

In my analyses in Chapters 3-4, I used the pWMY kernel as a model for redistribution, and, by adjusting attack density to account for dispersal flights, I attempted to remove as much autocorrelation as possible. Much of it remained however, in part because the adjustment itself was based on noisy AOS data. Although the autoconvolution of the pWMY is again a pWMY kernel, I modelled covariance using a Gaussian kernel for simplicity. It would be interesting to investigate whether a pWMY covariance kernel might have improved model fit. Further research is needed to investigate whether the autoconvolutions in Table 5.1 could be used more generally to improve error models in situations where dispersal plays a central role in the data-generating process.

5.2 The WMY and pWMY redistribution kernels

It is unfortunate that the redistribution kernel developed by Yasuda (1975) should have such a low profile in the ecological literature today. After all, it generalizes three of the more commonly used kernels in quantitative ecology (the Gaussian, 2D Laplace, and Bessel), and was independently derived and published at least twice in prominent journals (the second time by Hapca, Crawford, and Young, 2008). Yet this kernel family is referenced only indirectly, for example, in the comprehensive theoretical reviews of Okubo and Levin (2001) and Lewis, Petrovskii, and Potts (2016), and is altogether absent from the extensive table of phenomenological 2-dimensional kernels in Clobert et al. (2012).

Given the popularity of the Bessel kernel (Broadbent and Kendall, 1953; Williams, 1961; Awerbuch, Samson, and Sinskey, 1979; Turchin and Thoeny, 1993; Goodsman et al., 2016), it appears that a more parsimonious hypothesis of constant hazard settling is preferred by modellers over the gamma-distributed settling times of Yasuda (1975). Thus I started from the constant hazard assumption in my mechanistic derivation of the WMY kernel in Chapter 3. However, recognizing the spatial heterogeneity of the habitat through which forest insect flight takes place (Hapca, Crawford, and Young, 2008), I replaced the usual PDE for Fickian diffusion through homogeneous media (Okubo and Levin, 2001) with a PDE developed in statistical physics by O'Shaughnessy and Procaccia (1985) that better describes diffusion through disordered media.

In deriving the Bessel kernel, Broadbent and Kendall (1953) used Fickian diffusion to describe the movement of a parasitic larva wandering at random through a field until it is trapped at the end of a blade of grass, where it waits to be eaten by its host. We are reminded of the famous "ant in a labyrinth" problem from percolation theory (deGennes, 1976) that inspired a generation of research into anomolous diffusion through complex media (Metzler, Glöckle, and Nonnenmacher, 1994). De Gennes' labyrinth seems a better analogy for insect dispersal than does (unimpeded) Fickian diffusion. An insect moving through vegetation – be it a larva in a field of grass or a MPB in flight through the forest canopy – must navigate a mazelike network of obstacles and passages.

It can be mathematically convenient to describe the geometry of such disordered habitats as fractal media (Zeide, 1991). By doing so in Chapter 3, I obtained a PDE with a modified Laplace operator to represent random walks through disorded media – and by adding a constant settling term, I defined a subset of the WMY kernel family that includes the Bessel kernel as a special case. Thus I generalized the model of Broadbent and Kendall (1953) by introducing a more realistic depiction of movement through the available space.

I extended this model to include the full WMY kernel family by considering what happens when the movement process is iterated (independently) over multiple stages of dispersal. This result used the property that the WMY kernel is closed under convolutions, which itself is a highly desireable characteristic from a sampling design standpoint (Schlägel and Lewis, 2016). However it also provides a new interpretation for the WMY kernel shape parameter: Under the model of Yasuda (1975), it corresponds to the shape of the gamma-distributed stopping time; In my model it is determined by both the number of stages and the Hausdorf dimension d_f of the habitat.

Radar-based surveys such as Ainslie and Jackson (2011) and laboratory experiments by Shegelski, Evenden, and Sperling (2019) suggested the possibility of multi-stage dispersal events by MPB. In future work it would be interesting to examine whether the number of dispersal stages could be inferred from MPB data by fitting the WMY kernel to data and estimating its shape parameter. Alternatively, one might fix the number of stages and estimate d_f locally based on habitat covariates such as the spatial distribution of forestland (similar to Hargis, Bissonette, and Turner, 1999), to parametrize a kernel for movement ahead of model-fitting. This is not unlike the approach of Powell and Bentz (2014), who used a diffusion model to describe MPB movements with diffusivity a decreasing function of host density.

The second half of Chapter 3 turned to phenomenological extensions of the WMY kernel that characterize aspects of movement where the mechanism is impractical to model.

For example, data on local wind patterns during MPB dispersal would be difficult (if not impossible) to obtain at high resolution in a retrospective study. However wind likely introduces a strong directionality into flight patterns (Jackson et al., 2008). Flying MPB have been observed orienting themselves towards the upwind and downwind directions (Safranyik et al., 1992), as is common among flying insect taxa (Hu et al., 2016). Thus I proposed two simple extensions to introduce anisotropy into general isotropic and/or separable redistribution kernels. These substantially improved model fit on AOS data, supporting the position of Powell and Bentz (2014) that the common assumption of isotropic and stationary dispersal is far too simplistic for the predictive modelling of MPB damage patterns.

Chapters 2 and 3 are not only similar in both using variants of the same kernel (WM and WMY), but also in that they both address computational bottlenecks using the mathematics of separability on the lattice. In Chapter 2, a matrix inversion problem of $O(n^3)$ arithmetic complexity was reduced to one of $O(n^{3/2})$ by means of the product covariogram; and in Chapter 3, this same trick was used to reduce a $O(n^2)$ matrix multiplication to $O(n)$ complexity by approximating the WMY kernel with the pWMY. Through numerical simulations, I examined the parameter space of the WMY to identify situations where this approximation is adequate. Future researchers may use these numerical results as a guide in deciding when the computationally superior pWMY may be used as a drop-in replacement for the more realistic WMY kernel. However as I was unable to establish an analytical result on approximability, I leave this problem to future investigators.

5.3 Inferring MPB attack behaviour from AOS data

My data analysis in Chapter 4 demonstrated how the ideas developed in the foregoing chapters can be synthesized to create a MPB outbreak model suitable for AOS data. I addressed some of the shortcomings of the red-top model introduced in Chapter 1 by using the product-covariograms of Chapter 2 to correct for SAC, and the pWMY redistribution kernels of Chapter 3 to couple together sampling locations through a more realistic depiction of dispersal flights.

One of the more novel ideas presented in Chapter 4 is the construction of a nonstationary redistribution pattern as a distance-weighted average of the predictions of numerous locally fitted stationary pWMY kernels. This nonstationary construction was motivated by the results of Chapters 2-3, which indicated a high degree of nonstationarity in dispersal patterns at the one-hectare resolution, combined with the need for a computationally simple model. My scheme has the flexibility to capture nonstationarity, yet remains computationally efficient and scalable, as I demonstrated in fitting to an unusually large dataset encompassing nearly 800,000 locations.

The rationale for using a weighted average is that when the pWMY kernel fitted to the block centered at $\boldsymbol{y}$ is a good model for flight events near $\boldsymbol{y}$, then predictions $\hat{B}_{\boldsymbol{y}}(\boldsymbol{x})$ of beetle pressure based on this kernel are most informative at points $\boldsymbol{x}$ that are close to $\boldsymbol{y}$. This same principle underlies the theory of geographically weighted regression (Fotheringham, Brunsdon, and Charlton, 2003): viewing $\hat{B}_{\boldsymbol{y}}(\boldsymbol{x})$ as a measurement of the true beetle pressure $B(\boldsymbol{x})$ from the vantage point of location $\boldsymbol{y}$, my weighted average would become the *kernel estimator* of $B(\boldsymbol{x})$ (Staniswalis, 1989). Of course $\hat{B}_{\boldsymbol{y}}(\boldsymbol{x})$ is a model prediction, not a measurement, so we cannot appeal to the unbiasedness results from kernel estimator theory. In future research I hope to develop a rigorous theory to support this heuristic reasoning.

Similar to Powell and Bentz (2014), my model captures sinuous patterns of nonstationary and anisotropic movement across the landscape. However it differs in using a phenomenological model for dispersal for which the correction for SAC in parameter inference is made easier. There are advantages and disadvantages to this approach; On one hand, by addressing SAC we may expect to improve the precision of our estimators; On the other hand, the mechanistic PDE-based model of Powell and Bentz (2014) is more readily extended to predictions in novel habitat.

Moreover, more research is needed to develop a rigourous statistical theory to justify the model-fitting methodology described in the appendix to Chapter 4. Although numerical simulations indicated that estimators of both the attack function parameters and the angles of anisotropy in redistribution are reasonably unbiased, it would preferable to support this evidence with a robust statistical theory (*eg.* similar to Crujeiras and Van Keilegom, 2010).

5.3.1 Strengths and weaknesses of the generalized red-top model

I introduced the generalized red-top model in detail in Chapter 4, explaining how it is connected to the log-logit attack curve of Cooke and Carroll (2017). My analysis showed how its parameters can reveal information on attack dynamics that have been overlooked in other analyses based on the red-top model: In particular, my estimates of the mass-attack number (m_A) and the incipient-epidemic transition point (B_T) were in reasonable agreement with published data on ground surveys in my study area, which lends support to my interpretation of the attack curve in these terms.

AOS-based estimates of these attack curve properties can be useful in validating ecological hypotheses on MPB behaviour. For example my results indicated that for older, high-density stands, the number of MPB aggregating to mass-attack a single tree (m_A) tended to lie near (but often above) the optimum suggested in Raffa and Berryman (1983). Further, the transition point at which MPB behaviour switched from endemic to epidemic (B_T) happened when per-hectare attack densities lay close to this optimum, in support of the hypothesis by Carroll et al. (2006).

One of the biggest advantages in my modelling approach is that an analysis similar to that of Chapter 4 could be carried out over any region covered by the AOS. I chose the Merritt region for my study area in part so that field data such as Carroll et al. (2006) and Boone et al. (2011) could be compared against my results. However since there are many remote areas of the province for which ground surveys are lacking, the AOS in combination with my modelling techniques would allow quantities such as m_A and B_T to estimated virtually anywhere. Thus it would be interesting in future work to fit the model to other areas of the province. One could, for example, use estimates of B_{T_i} to map out areas where eruptions in MPB populations are most likely (for a given in-flight density), similar to the risk-ranking system of Shore and Safranyik (1992).

I must, however, acknowledge some limitations shared by the red-top model and its generalization. For example, neither can explicitly account for within-host scramble competition since, whenever an attack is successful, the model assumes a fixed number (β) of adults will emerge from the infested tree in the following summer. By contrast, empirical attack data reveals a strong negative feedback effect on productivity at high densities – mated females must compete under the bark for a limited amount of egg gallery space, and their offspring must compete for a limited food supply (Raffa and Berryman, 1983). The resulting compensatory effect is sometimes described by exponentially decaying productivity curves, $\beta(B_t)$, such as the Ricker-like model of Berryman (1974). In a more mechanistic analysis of Goodsman, Cooke, and Lewis (2017), this type of overcompensation was shown to drive outbreaks to collapse well ahead of host depletion.

The generalized red-top model does indeed exhibit negative density dependence with large MPB populations (B_{t+1}/B_t decreases as B_t increases due to scramble competition for the limited number of hosts). However since host mortality is nondecreasing in B_t, the model cannot describe early collapse due to intra-specific competition and overcompensation. Thus where the generalized red-top model is fitted to data on outbreak damage at very high beetle population levels, we may expect it to overestimate pine mortality.

The likely result is a bias towards lower estimates of stand susceptibility (*ie.* higher $a_{i,t}$, in the notation of Chapter 4) as the model attempts to match high B_t values with lower-than-expected ϕ_{t+1}. This in part explains why my estimators of the incipient-epidemic transition ($B_{T_{i,t}}$) and mass-attack density ($m_{A_{i,t}}$) – both of which scale with $\hat{a}_{i,t}$ – were on the high-end of the range of values reported in the literature. For example Goodsman, Cooke, and Lewis (2017) estimated the minimum number of attackers needed to colonize a host at 289, whereas my estimates of the incipient-epidemic transition point – which is expected to lie close this minimum (Carroll et al., 2006) – ranged from 427–1114.

Nevertheless, the generalized red-top model remains a good model for outbreak onset, since in that early stage, MPB populations are far from the point of host depletion. The years fitted in Chapter 4 represented the leadup to the collapse of the MPB epidemic in my study area, so the effect of overcompensation likely played a minor role in my attack function parameter estimates. However future investigators will need to consider this effect more carefully if the model is to be fitted to a MPB population undergoing collapse at the landscape level.

5.3.2 Endemic MPB populations

The dynamics of endemic MPB populations are rarely investigated in the modelling literature. To my knowledge, model predictions of year-to-year dynamics at these low levels have never been confronted with data, simply because it is very difficult to collect a sufficiently long time series. However, in theoretical work by Berryman (1974) and Cooke and Carroll (2017) it is hypothesized that the absence of an Allee effect creates a stable equilibrium (ϵ) for low MPB densities. Weakened hosts therefore provide a refugium for populations that would otherwise collapse.

Favourable climatic conditions, such as drought, may release the endemic MPB from this stable state by providing access to higher quality hosts, which in subsequent years allows attack densities to rise far above ϵ. This is the "outbreak epicenter" hypothesis of Royama (2012), which is supported by empirical studies such as in Chapman, Veblen, and Schoennagel (2012), Bleiker et al. (2014) and Cooke and Carroll (2017), for example. Heavilin and Powell (2008) offered a different explanation in which there is no endemic equilibrium, suggesting that low MPB populations are sustained by in-flights from distant outbreaks across the landscape.

The AOS contains no data on endemic populations, so in estimating ϵ in Chapter 4, I used a far simpler representation than the multi-equilibrium models of Berryman (1974) and Cooke and Carroll (2017) – one that represents endemic MPB implicitly through their effect on outbreaking populations. By assuming that non-outbreak MPB populations are all at the spatially uniform equilibrium level ϵ, I was able to reconcile the mathematically simple attack dynamics model of Heavilin and Powell (2008) with the outbreak epicenter hypothesis. My approach models attack densities at levels far below the detection threshold of the AOS; Yet the estimator $\hat{\epsilon}$ was in remarkably close agreement with ground surveys of my study area by Boone et al. (2011) and Carroll et al. (2006) (in the range of 7-14 attackers/ha). Had there been no stable endemic population in the area, we would expect $\hat{\epsilon}$ to be near zero.

Thus while the endemic population is small, it is not necessarily small enough to ignore for the purpose of modelling MPB spread. Local endemic MPB inflate the attack density in areas experiencing in-flights from nearby outbreaks, effectively lowering the Allee threshold. This threshold was conceptualized in Goodsman and Lewis (2016) as a plane whose intersection with the 2-dimensional MPB density function inscribes areas of positive MPB population growth. A fixed background endemic population has the effect of lowering this plane by a fixed amount – contours of positive growth get larger and the minimum founding population becomes smaller.

In this way the endemic population appears to prime its habitat for outbreaks to arise more easily, by requiring fewer immigrating MPB to meet the incipient-epidemic threshold. This effect could have important consequences for the management of MPB spread in novel habitats east of the Rocky Mountains, as the establishment of endemic populations may foreshadow the onset of outbreak-level populations that might otherwise fail to take hold.

5.4 Conclusion

My work here is just one small piece of a larger multi-disciplinary effort in recent decades to anticipate and manage the effects of the MPB on the pine forest ecosystem. However I expect the ideas presented in this thesis to assist future MPB researchers in calibrating and refining their models for outbreak spread, by providing a more feasible means of accessing the information contained in large, high-resolution spatial datasets such as the AOS. Through such improvements we get closer to the goal of understanding, managing, and forecasting MPB outbreaks effectively.

In the course of constructing the outbreak model of Chapter 4, I developed a suite of modelling tools with applications in spatial ecology that go far beyond the MPB. As survey datasets grow more expansive, the covariograms of Chapter 2 and the redistribution kernels of Chapter 3 will be useful in guiding quantitative ecologists to better account for both the error and dispersal processes that drive their models.

5.5 Appendices to Discussion

5.5.1 The Fourier and Hankel transforms

Denote by $\hat{D}(\boldsymbol{u})$ the Fourier transform of the integrable function $D(\boldsymbol{r}) \geq 0$, where $\boldsymbol{r} \in \mathbb{R}^d$:

$$\hat{D}(\boldsymbol{u}) = \int_{\mathbb{R}^d} D(\boldsymbol{r}) \exp(-2\pi i \boldsymbol{u} \cdot \boldsymbol{r}) \, \mathrm{d}\boldsymbol{r}, \tag{5.2a}$$

$$D(\boldsymbol{r}) = \int_{\mathbb{R}^d} \hat{D}(\boldsymbol{u}) \exp(2\pi i \boldsymbol{u} \cdot \boldsymbol{r}) \, \mathrm{d}\boldsymbol{u}\,. \tag{5.2b}$$

When D is radially symmetric, its spectral density $\hat{D}$ is also radially symmetric in d-dimensional Fourier space. Thus, upon switching to a spherical coordinate system, (5.2a) becomes an integration over $r = |\boldsymbol{r}|$, and $\hat{D}$ becomes a function of frequency radius $u = |\boldsymbol{u}|$ only. It is therefore often notationally convenient to treat D and $\hat{D}$ as single-variable functions. This will be indicated by an unbolded argument, *ie.* $D(r)$ and $\hat{D}(u)$, respectively. Note that $\hat{D}(u)$ refers to the *d-dimensional* transform of $D(\boldsymbol{r})$ – not to be confused with the one-dimensional Fourier transform of $D(r)$.

As functions of d variables, D and $\hat{D}$ are related by the Fourier transform. However, as functions of a single variable, they can be more simply related using Hankel transforms, as for example in the spectral density inversion formula of Stein (1999) (pp. 42–46) for isotropic kernels. This useful identity is restated for reference below, along with the definition of the ν^{th} order Hankel transform $\mathcal{H}_\nu$, which is written in terms of the ν^{th} order Bessel function of the first kind, J_ν:

$$\hat{D}(u) = (2\pi u)^{-d/2} u \mathcal{H}_{d/2-1} \left(r^{d/2-1} D(r) \right)(u), \tag{5.3}$$
$$\text{where } \mathcal{H}_\nu \left(f(r) \right)(u) = \int_0^\infty J_\nu(ur) r f(r) \, \mathrm{d}r$$

When deriving the spectral density of a d-dimensional isotropic covariance function $C(r)$ (where $r = |\boldsymbol{r}|$), equation (5.3) is often simpler to work with than (5.2a) (see Schlägel and Lewis, 2016). The inverse transformation, $\hat{C} \rightarrow C$, is equally simple thanks to the self-reciprocal nature of the Hankel transform – $\mathcal{H}_\nu$ is its own inverse (Birkinshaw, 1994). For example when $d = 2$,

$$\hat{D}(u) = (1/2\pi)\mathcal{H}_0 \left(D(r) \right)(u), \text{ with inverse } D(r) = 2\pi \mathcal{H}_0 \left(\hat{D}(u) \right)(r). \tag{5.4}$$

The integral transform in (5.3) is generally not trivial to solve analytically. However because many popular covariance models are derived by first building a spectral representation (see *eg.* Cressie, 1993, section 2.5), their Fourier and Hankel transforms often have uncomplicated forms, as we will see in several examples in the next section.

5.5.2 Examples of kernel autoconvolution pairs

Convolutions become products in Fourier space. This means that when D is isotropic the spectral density of the auto-correlation function $C_D/\sigma^2 = D * D$ is simply the square of the spectral density of D. Thus in the simple case of 2-dimensional space ($d = 2$) equation (5.4) implies that:

$$C_D(r) = \sigma^2 2\pi \mathcal{H}_0 \left(\left(\mathcal{H}_0 \left(D \right) \right)^2 \right). \tag{5.5}$$

Note that while a proper redistribution kernel must be scaled to satisfy $\int_{\mathbb{R}^d} D(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} = 1$, its normalization constant is unimportant to establishing functional relationships between D and C_D. This is because once the functional form of C_D is determined, we may derive its scaling constant using the identity $\lim_{r\to 0} C_D(r) = \sigma^2$. Thus in the following examples I will omit the normalization constants for D and C_D whenever doing so simplifies the presentation.

The rectangular pulse and triangle functions

The *rectangular pulse* function $\Pi_d(\boldsymbol{r};\rho)$ is an indicator variable taking value 1 over the d-dimensional cube of side length ρ centered over the origin, and zero elsewhere. It is separable, since its dependence on spatial coordinates $\boldsymbol{r} = (r_1, r_2, \ldots r_d)^T$ has the form $\Pi_d(\boldsymbol{r},\rho) = \Pi_1(r_1;\rho)\ldots\Pi_1(r_d;\rho)$. Separability makes it simple to derive the spectral density directly from (5.2a) and a table of integral transforms (*eg*. Gradshteyn and Ryzhik, 1965), from which we find:

$$\hat{\Pi}_d\left(\boldsymbol{u}\right) = \prod_{k=1}^{d} \left(\operatorname{sinc}(u_k/\rho)/\rho\right), \text{ where } \operatorname{sinc}(x) = \begin{cases} 1 & \text{for } x = 0, \\ \sin(x)/x & \text{otherwise,} \end{cases} \tag{5.6}$$

where $\boldsymbol{u} = (u_1, u_2, \ldots u_d)^T$. The effect of the pulse function on a point mass is to simply spread the density evenly within the surrounding hypercube. From equations (5.1) and (5.6) one finds the spectral density of its auto-correlation, $\hat{C}_D(\boldsymbol{u})/\sigma^2 = \Pi_{k=1}^d \left(\operatorname{sinc}(u_k/\rho)/\rho\right)^2$, which is easily inverted:

$$C_D(\boldsymbol{r};\rho) \propto \Pi_{k=1}^d \Lambda\left(r_k;\rho\right)/\rho^2, \quad \text{where } \Lambda\left(r;\rho\right) = \begin{cases} 1 - r/\rho & \text{for } 0 \le r < \rho, \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

In the general d-dimensional case, equation (5.7) shows that C_D is proportional to a separable product of *triangle* functions (Λ). In the one-dimensional case ($d = 1$), this reduces to an isotropic kernel known as the *tent* (or triangle) covariogram (Cressie, 1993, p.84). Though rarely used in practice, the tent covariogram is a useful textbook example; It is a one-dimensional covariance function that becomes invalid in higher dimensions (*eg.* Λ is not positive definite when r is distance in $\mathbb{R}^2$). This serves to remind that when considering extensions of a given covariance model to higher dimensions, it is not enough to simply redefine the measure by the obvious relabeling $r = |\boldsymbol{r}|$.

On the other hand $C_D(\boldsymbol{r};\rho)$ in (5.7) is positive definite in any dimension. In fact, any product of kernels that are valid in one-dimensional space (with respect to the component distances $r_1, \ldots r_d$) is a valid model in $\mathbb{R}^d$ (Rasmussen and Williams, 2006). This idea can be used to construct computationally efficient extensions of existing models in higher dimensional spaces, as I did with the pWMY kernel (a product of one-dimensional WMY kernels).

The spherical pulse and covariance models

Consider the redistribution kernel $D(\boldsymbol{r};\rho) \propto \mathbb{1}_{|\boldsymbol{r}|<\rho}$ that spreads point mass evenly in the surrounding d-dimensional sphere of radius ρ. This produces the *spherical* covariance family (*eg.* as derived in Chiles and Delfiner, 2012, pp. 85–88). I showed in the previous section that the $\mathbb{R}^1$ case produces a tent covariogram. In $\mathbb{R}^3$ the spherical kernel takes the form:

$$C_D(r;\rho) \propto \begin{cases} 1 - (3/2)(r/\rho) + (1/2)(r/\rho)^3 & \text{for } 0 \leq r < \rho \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

Despite some theoretical and practical drawbacks (*eg.* as discussed in Stein, 1999, pp. 52-53), the spherical covariance kernel in equation (5.8) remains in popular use (both in $\mathbb{R}^3$, and in $\mathbb{R}^2$ where C_D remains positive definite) – in large part because of computational advantages related to its compact support.

The WMY and and Whittle-Matérn kernel families

Recall from Chapter 3 that a WMY redistribution kernel has the form:

$$\mathcal{D}(r;\rho,\kappa) \propto (r/\rho)^{\kappa} K_{\kappa}(r/\rho), \quad \text{for } \rho > 0, \text{ and } \kappa > -1, \tag{5.9}$$

where K_κ denotes the order ν modified Bessel function of the second kind. Consulting a table of integrals (Gradshteyn and Ryzhik, 1965, equation 6.565.4), we find the Hankel transform of $\mathcal{D}(r;\rho,\kappa)$ to be a simple rational function of u, which is easily squared and back-transformed:

$$\begin{aligned}
&\mathcal{H}_0\left(\mathcal{D}(r;\rho,\kappa)\right) \propto \left(u^2+\rho^{-2}\right)^{-(\kappa+1)} \quad \text{for } \kappa > -3/4 \qquad (5.10)\\
\implies &\left(\mathcal{H}_0\left(\mathcal{D}(r;\rho,\kappa)\right)\right)^2 \propto \left(u^2+\rho^{-2}\right)^{-(2\kappa+2)} = \left(u^2+\rho^{-2}\right)^{-((1+2\kappa)+1)}\\
\implies &\mathcal{D}(r;\rho,\kappa) * \mathcal{D}(r;\rho,\kappa) \propto \mathcal{H}_0\left(\left(\mathcal{H}_0\left(\mathcal{D}(r;\rho,\kappa)\right)\right)^2\right) \propto \mathcal{D}(r;\rho,1+2\kappa).
\end{aligned}$$

Note that $C_{\mathcal{D}}(r)/\sigma^2 = \mathcal{D} * \mathcal{D}$ is a valid *covariance* kernel only if its shape parameter is strictly positive (Guttorp and Gneiting, 2006), hence the restriction $\kappa > -1/2$ on the shape parameter of $\mathcal{D}$ in Table 5.1. Notable examples include Whittle's (1954) correlation model (a WM with shape $\kappa = 1$), which arises as the autoconvolution of a Bessel redistribution kernel ($\kappa = 0$) (Williams, 1961); and the WM with shape $\kappa = 2$, which corresponds to the autoconvolution of a 2D Laplace kernel (Joseph and Sendner, 1958). Interestingly, the discretizations of these two covariance kernels closely approximate second and third-order Markov random field models (Lindgren, Rue, and Lindström, 2011).

Clark's 2Dt redistribution kernel and the Cauchy covariance model

Clark et al. (1999) developed a model for seed dispersal by replacing the range parameter in the Gaussian redistribution kernel by a gamma-distributed random variable. This represents a diffusion process where diffusivity varies at random among the propagules. The resulting density function is called the *2Dt*, so named because it extends a Student's *t* distribution to 2-dimensional space:

$$D(r;\beta,\rho) = \left(r^2+\rho^2\right)^{-(\beta+1)}. \qquad (5.11)$$

where $\rho > 0$ is a range parameter and $\beta > 0$ a shape parameter. Note that when $\beta = 0$ equation (5.11) is known as the *Cauchy kernel*, which arises from Lévy walks in 2-dimensional space (Kot, Lewis, and Driessche, 1996). Note also that equation (5.11) has the same form as the spectral density of the WMY kernel; The 2Dt and WMY models are Fourier duals.

When using likelihood-based methods, Clark et al. (1999) recommended fixing parameter β to a low value (suggesting 1/2 or 1) for reasons of numerical stability. Let us consider the case $\beta = 1/2$. Equation (5.10) shows that the Hankel transform of (5.11) is proportional to $\mathcal{D}(u; \beta, 1/\rho)$, which for $\kappa = 1/2$ is proportional to $\exp(-\rho r)$ (see Appendix 3.7.2). This implies that:

$$C_D(r)/\sigma^2 = \mathcal{H}_0\left(\left(\mathcal{H}_0\left(D\left(r; 1/2, \rho \right) \right) \right)^2 \right) \propto \mathcal{H}_0\left(\exp(-2\rho r) \right) \propto D(r; 1/2, 2\rho).$$

Therefore the autoconvolution of a 2Dt kernel with $\beta = 1/2$ is again a 2Dt, but with a doubled scale parameter. As a covariance kernel the function in (5.11) is known to the statistics community as the *Cauchy model* (Chiles and Delfiner, 2012, pp. 89-90) and in the machine learning community, as the *rational quadratic* kernel (Rasmussen and Williams, 2006, pp. 86-87)

Bibliography

Abu-Jeib, I (2002). “Centrosymmetric Matrices: Properties and an Alternative Approach”. In: *Can Appl Math Q* 10.42, pp. 9–445.

Ainslie, Bruce and Peter L Jackson (2011). “Investigation into mountain pine beetle above-canopy dispersion using weather radar and an atmospheric dispersion model”. In: *Aerobiologia* 27.1, pp. 51–65.

Ambikasaran, S. et al. (2016). “Fast Direct Methods for Gaussian Processes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.2, pp. 252–265. DOI: 10.1109/TPAMI.2015.2448083.

Amman, Gene D (1984). “Mountain pine beetle (Coleoptera: Scolytidae) mortality in three types of infestations”. In: *Environmental Entomology* 13.1, pp. 184–191.

Andersen, Mark (1991). “Properties of some density-dependent integrodifference equation population models”. In: *Mathematical Biosciences* 104.1, pp. 135–157.

Aukema, Brian H et al. (2006). “Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak”. In: *Ecography* 29.3, pp. 427–441.

Aukema, Brian H et al. (2008). “Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate”. In: *Ecography* 31.3, pp. 348–358.

Awerbuch, TE, R Samson, and AJ Sinskey (1979). “A quantitative model of diffusion bioassays”. In: *Journal of Theoretical Biology* 79.3, pp. 333–340.

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton, FL, USA: CRC Press.

Beale, Colin M et al. (2010). “Regression analysis of spatial data”. In: *Ecology Letters* 13.2, pp. 246–264.

Beaudoin, A et al. (2014). “Mapping attributes of Canada’s forests at moderate resolution through kNN and MODIS imagery”. In: *Can J For Res* 44.5, pp. 521–532.

Berryman, AA (1979). “Dynamics of bark beetle populations: analysis of dispersal and redistribution”. In: *Mitt Schweiz Entomol Ges* 52, pp. 227–234.

Berryman, Alan A (1974). “Dynamics of bark beetle populations: towards a general productivity model”. In: *Environmental Entomology* 3.4, pp. 579–585.

Berryman, Alan A (1978). “Towards a theory of insect epidemiology”. In: *Researches on Population Ecology* 19.2, pp. 181–196.

Berryman, Alan A et al. (1985). “Evolution of optimal group attack, with particular reference to bark beetles (Coleoptera: Scolytidae)”. In: *Ecology* 66.3, pp. 898–903.

Besag, Julian (1974). “Spatial interaction and the statistical analysis of lattice systems”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236.

Birkinshaw, Mark (1994). “Radially-symmetric Fourier transforms”. In: *Astronomical Data Analysis Software and Systems III*. Vol. 61, p. 249.

Bleiker, KP et al. (2014). “Characterisation of attacks made by the mountain pine beetle (Coleoptera: Curculionidae) during its endemic population phase”. In: *The Canadian Entomologist* 146.3, pp. 271–284.

Boone, Celia K et al. (2011). “Efficacy of tree defense physiology varies with bark beetle population density: a basis for positive feedback in eruptive species”. In: *Canadian Journal of Forest Research* 41.6, pp. 1174–1188.

Broadbent, SR and David G Kendall (1953). “The random walk of *Trichostrongylus retortaeformis*”. In: *Biometrics* 9.4, pp. 460–466.

Burnham, Kenneth P and David R Anderson (2004). “Multimodel inference: understanding AIC and BIC in model selection”. In: *Sociological Methods & Research* 33.2, pp. 261–304.

Candau, Jean-Noël, Richard A Fleming, and Anthony Hopkin (1998). “Spatiotemporal patterns of large-scale defoliation caused by the spruce budworm in Ontario since 1941”. In: *Canadian Journal of Forest Research* 28.11, pp. 1733–1741.

Carroll, AL et al. (2006). “Mountain pine beetle outbreak development: the endemic-incipient epidemic transition”. In: *Canadian Forest Service, Mountain Pine Beetle Initiative Project* 1, p. 22.

Carroll, Allan L et al. (2003). “Effect of climate change on range expansion by the mountain pine beetle in British Columbia”. In: *Shore, TL et al. (editors) Mountain Pine Beetle Symposium: Challenges and Solutions. Victoria (Canada), Natural Resources Canada,* *Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-399, pages 223–232*.

Chapman, Teresa B, Thomas T Veblen, and Tania Schoennagel (2012). “Spatiotemporal patterns of mountain pine beetle activity in the southern Rocky Mountains”. In: *Ecology* 93.10, pp. 2175–2185.

Chen, Huapeng (2014). “A spatiotemporal pattern analysis of historical mountain pine beetle outbreaks in British Columbia, Canada”. In: *Ecography* 37.4, pp. 344–356.

Chen, Huapeng and Adrian Walton (2011). “Mountain pine beetle dispersal: spatiotemporal patterns and role in the spread and expansion of the present outbreak”. In: *Ecosphere* 2.6, pp. 1–17.

Chen, Huapeng et al. (2015). “A spatiotemporal pattern analysis of potential mountain pine beetle emergence in British Columbia, Canada”. In: *Forest Ecology and Management* 337, pp. 11–19.

Chesson, Peter and Charlotte T Lee (2005). “Families of discrete kernels for modeling dispersal”. In: *Theoretical Population Biology* 67.4, pp. 241–256.

Chiles, JP and P Delfiner (2012). *Geostatistics: Modeling Spatial Uncertainty*. 2nd. Hoboken, NJ: John Wiley & Sons.

Chipperfield, Joseph D et al. (2011). “On the approximation of continuous dispersal kernels in discrete-space models”. In: *Methods Ecol Evol* 2.6, pp. 668–681.

Clark, James S, Eric Macklin, and Leslie Wood (1998). “Stages and spatial scales of recruitment limitation in southern Appalachian forests”. In: *Ecological monographs* 68.2, pp. 213–235.

Clark, James S et al. (1999). “Seed dispersal near and far: patterns across temperate and tropical forests”. In: *Ecology* 80.5, pp. 1475–1494.

Clobert, Jean et al. (2012). *Dispersal Ecology and Evolution*. Oxford, UK: Oxford University Press.

Cole, Walter E and Gene D Amman (1969). *Mountain pine beetle infestations in relation to lodgepole pine diameters*. Vol. 95. US Dept. of Agriculture, Forest Service, Intermountain Forest & Range . . .

Cole, Walter E and Mark D McGregor (1983). *Estimating the rate and amount of tree loss from mountain pine beetle infestations*. Tech. rep. US Department of Agriculture, Forest Service, Intermountain Forest and Range . . .

Cooke, Barry J and Allan L Carroll (2017). “Predicting the risk of mountain pine beetle spread to eastern pine forests: considering uncertainty in uncertain times”. In: *Forest Ecology and Management* 396, pp. 11–25.

Corbett, LJ et al. (2015). “The economic impact of the mountain pine beetle infestation in British Columbia: provincial estimates from a CGE analysis”. In: *Forestry: An International Journal of Forest Research* 89.1, pp. 100–105.

Cressie, Noel (1993). *Statistics for Spatial Data*. New York, NY, USA: John Wiley & Sons.

Crujeiras, Rosa M and Ingrid Van Keilegom (2010). “Least squares estimation of nonlinear spatial trends”. In: *Computational Statistics & Data Analysis* 54.2, pp. 452–465.

Das, Abhik et al. (2002). “Modelling a discrete spatial response using generalized linear mixed models: application to Lyme disease vectors”. In: *Int J Geogr Inf Sci* 16.2, pp. 151–166.

deGennes, PG (1976). “La percolation: Un concept unificateur”. In: *La Recherche* 7, pp. 921–926.

Dhar, Amalesh, Lael Parrott, and Scott Heckbert (2016). “Consequences of mountain pine beetle outbreak on forest ecosystem services in western Canada”. In: *Can J Forest Res* 46.8, pp. 987–999.

Dhar, Amalesh et al. (2015). “Impact of a mountain pine beetle outbreak on young lodgepole pine stands in central British Columbia”. In: *Forests* 6.10, pp. 3483–3500.

Dietrich, CR (1993). “Computationally efficient Cholesky factorization of a covariance matrix with block Toeplitz structure”. In: *J Stat Comput Sim* 45.3-4, pp. 203–218.

Dormann, Carsten et al. (2007). “Methods to account for spatial autocorrelation in the analysis of species distributional data: a review”. In: *Ecography* 30.5, pp. 609–628.

Evenden, Maya L, CM Whitehouse, and J Sykes (2014). “Factors influencing flight capacity of the mountain pine beetle (Coleoptera: Curculionidae: Scolytinae)”. In: *Environmental Entomology* 43.1, pp. 187–196.

Fortin, Marie-Josée and Mark RT Dale (2005). *Spatial Analysis: A Guide for Ecologists*. Cambridge, UK: Cambridge University Press.

Fotheringham, A Stewart, Chris Brunsdon, and Martin Charlton (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Garlick, Martha J et al. (2011). “Homogenization of large-scale movement models in ecology”. In: *Bulletin of Mathematical Biology* 73.9, pp. 2088–2108.

Genton, Marc G (2007). “Separable approximations of space-time covariance matrices”. In: *Environmetrics* 18.7, pp. 681–695.

Gilbert, Mark A et al. (2017). “Speeding up the simulation of population spread models”. In: *Methods Ecol Evol* 8.4, pp. 501–510.

Giroday, Honey-Marie C de la, Allan L Carroll, and Brian H Aukema (2012). “Breach of the northern Rocky Mountain geoclimatic barrier: initiation of range expansion by the mountain pine beetle”. In: *J Biogeogr* 39.6, pp. 1112–1123.

Golub, Gene H and Charles F Van Loan (2012). *Matrix computations*. Vol. 3. Baltimore, MD, USA: JHU Press.

Goodbody, Tristan RH et al. (2018). “Digital aerial photogrammetry for assessing cumulative spruce budworm defoliation and enhancing forest inventories at a landscape-level”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 142, pp. 1–11.

Goodchild, Michael F and David M Mark (1987). “The fractal nature of geographic phenomena”. In: *Annals of the Association of American Geographers* 77.2, pp. 265–278.

Goodsman, Devin W and Mark A Lewis (2016). “The minimum founding population in dispersing organisms subject to strong Allee effects”. In: *Methods in Ecology and Evolution* 7.9, pp. 1100–1109.

Goodsman, DW, BJ Cooke, and MA Lewis (2017). “Positive and negative density-dependence and boom-bust dynamics in enemy-victim populations: a mountain pine beetle case study”. In: *Theoretical Ecology* 10.2, pp. 255–267.

Goodsman, DW et al. (2016). “Aggregation and a strong Allee effect in a cooperative outbreak insect”. In: *Ecological Applications* 26.8, pp. 2623–2636.

Gradshteyn, Izrail Solomonovich and Iosif Moiseevich Ryzhik (1965). *Table of Integrals, Series, and Products*. 4th. New York, NY: Academic press.

Guan, Yongtao, Michael Sherman, and James A Calvin (2004). “A nonparametric test for spatial isotropy using subsampling”. In: *JASA* 99.467, pp. 810–821.

Guillot, Gilles et al. (2014). “Validity of covariance models for the analysis of geographical variation”. In: *Methods Ecol Evol* 5.4, pp. 329–335.

Guttorp, Peter and Tilmann Gneiting (2006). “Studies in the history of probability and statistics XLIX: On the Matérn correlation family”. In: *Biometrika* 93.4, pp. 989–995.

Hall, RJ et al. (2016). “Remote sensing of forest pest damage: a review and lessons learned from a Canadian perspective”. In: *The Canadian Entomologist* 148.S1, S296–S356.

Hapca, Simona, John W Crawford, and Iain M Young (2008). “Anomalous diffusion of heterogeneous populations characterized by normal diffusion at the individual level”. In: *Journal of the Royal Society Interface* 6.30, pp. 111–122.

Hargis, Christina D, John A Bissonette, and David L Turner (1999). “The influence of forest fragmentation and landscape pattern on American martens”. In: *Journal of Applied Ecology* 36.1, pp. 157–172.

Hassell, Michael Patrick (1978). *The Dynamics of Arthropod Predator-Prey Systems*. Princeton University Press.

Hawkins, Bradford A (2012). “Eight (and a half) deadly sins of spatial analysis”. In: *Journal of Biogeography* 39.1, pp. 1–9.

Heagerty, Patrick J and Subhash R Lele (1998). “A composite likelihood approach to binary spatial data”. In: *JASA* 93.443, pp. 1099–1111.

Heavilin, Justin and James Powell (2008). “A novel method of fitting spatio-temporal models to data, with applications to the dynamics of mountain pine beetles”. In: *Natural Resource Modeling* 21.4, pp. 489–524.

Hilborn, Ray and Marc Mangel (1997). *The ecological detective: confronting models with data*. Vol. 28. Princeton, NJ, USA: Princeton University Press.

Hirano, Toshihiro (2014). “Pseudo best estimator by a separable approximation of spatial covariance structures”. In: *Journal of the Japan Statistical Society* 44.1, pp. 43–71.

Hohn, Michael E, Andrew M Liebhold, and Linda S Gribko (1993). “Geostatistical model for forecasting spatial dynamics of defoliation caused by the gypsy moth (Lepidoptera: Lymantriidae)”. In: *Environmental Entomology* 22.5, pp. 1066–1075.

Holling, Crawford S (1959). “Some characteristics of simple types of predation and parasitism”. In: *The Canadian Entomologist* 91.7, pp. 385–398.

Hooke, Robert and Terry A Jeeves (1961). “Direct Search Solution of Numerical and Statistical Problems”. In: *Journal of the ACM* 8.2, pp. 212–229.

Horn, Roger A and Charles R Johnson (2013). *Matrix analysis*. New York, NY, USA: Cambridge University Press.

Hu, Gao et al. (2016). “Wind-related orientation patterns in diurnal, crepuscular and nocturnal high-altitude insect migrants”. In: *Frontiers in Behavioral Neuroscience* 10.32, pp. 1–8.

Jackson, Peter L et al. (2008). “Radar observation and aerial capture of mountain pine beetle, *Dendroctonus ponderosae* Hopk. (Coleoptera: Scolytidae) in flight above the forest canopy”. In: *Canadian Journal of Forest Research* 38.8, pp. 2313–2327.

Jonckheere, Inge et al. (2006). “A fractal dimension-based modelling approach for studying the effect of leaf distribution on LAI retrieval in forest canopies”. In: *Ecological Modelling* 197.1, pp. 179–195.

Joseph, J and H Sendner (1958). “Über die horizontale Diffusion im Meere”. In: *Deutsche Hydrografische Zeitschrift* 11.2, pp. 49–77.

Kareiva, Peter (1994). “Space: the final frontier for ecological theory”. In: *Ecology* 75.1, pp. 1–1.

Kärvemo, Simon et al. (2014). “Large-scale risk mapping of an eruptive bark beetle–importance of forest susceptibility and beetle pressure”. In: *Forest Ecology and Management* 318, pp. 158–166.

Kautz, Markus (2014). “On correcting the time-lag bias in aerial-surveyed bark beetle infestation data”. In: *Forest ecology and management* 326, pp. 157–162.

Keitt, Timothy H et al. (2002). “Accounting for spatial pattern when modeling organism-environment interactions”. In: *Ecography* 25.5, pp. 616–625.

Klutsch, Jennifer G et al. (2009). “Stand characteristics and downed woody debris accumulations associated with a mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak in Colorado”. In: *Forest Ecology and Management* 258.5, pp. 641–649.

Kot, Mark (2001). *Elements of Mathematical Ecology*. Cambridge, UK: Cambridge University Press.

Kot, Mark, Mark A Lewis, and Pauline van den Driessche (1996). “Dispersal data and the spread of invading organisms”. In: *Ecology* 77.7, pp. 2027–2042.

Kot, Mark and William M Schaffer (1986). “Discrete-time growth-dispersal models”. In: *Mathematical Biosciences* 80.1, pp. 109–136.

Kunegel-Lion, Mélodie, Rory L McIntosh, and Mark A Lewis (2018). “Management assessment of mountain pine beetle infestation in Cypress Hills, SK”. In: *Canadian Journal of Forest Research* 49.2, pp. 154–163.

Kurz, Werner A et al. (2008). “Mountain pine beetle and forest carbon feedback to climate change”. In: *Nature* 452.7190, p. 987.

Legendre, Pierre (1993). “Spatial autocorrelation: trouble or new paradigm?” In: *Ecology* 74.6, pp. 1659–1673.

Lele, Subhash, Mark L Taper, and Stuart Gage (1998). “Statistical analysis of population dynamics in space and time using estimating functions”. In: *Ecology* 79.5, pp. 1489–1502.

Lewis, MA (2000). “Spread rate for a nonlinear stochastic invasion”. In: *Journal of Mathematical Biology* 41.5, pp. 430–454.

Lewis, Mark A, William Nelson, and Cailin Xu (2010). “A structured threshold model for mountain pine beetle outbreak”. In: *Bulletin of Mathematical Biology* 72.3, pp. 565–589.

Lewis, Mark A, Sergei V Petrovskii, and Jonathan R Potts (2016). *The Mathematics Behind Biological Invasions*. Vol. 44. Switzerland: Springer.

Limpert, Eckhard, Werner A Stahel, and Markus Abbt (2001). “Log-normal distributions across the sciences: keys and clues”. In: *BioScience* 51.5, pp. 341–352.

Lindgren, BS and KF Raffa (2013). “Evolution of tree killing in bark beetles (Coleoptera: Curculionidae): trade-offs between the maddening crowds and a sticky situation”. In: *Can Entomol* 145.05, pp. 471–495.

Lindgren, Finn, Håvard Rue, and Johan Lindström (2011). “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach”. In: *JRSS Series B* 73.4, pp. 423–498.

Lindsay, Bruce G (1988). “Composite likelihood methods”. In: *Contemporary Mathematics* 80.1, pp. 221–239.

Logan, Jesse A et al. (1998). “Model analysis of spatial patterns in mountain pine beetle outbreaks”. In: *Theoretical Population Biology* 53.3, pp. 236–255.

Ludwig, Donald, Dixon D Jones, and Crawford S Holling (1978). “Qualitative analysis of insect outbreak systems: the spruce budworm and forest”. In: *The Journal of Animal Ecology*, pp. 315–332.

Lutscher, Frithjof, Elizaveta Pachepsky, and Mark A Lewis (2005). “The effect of dispersal patterns on stream populations”. In: *Siam Review* 47.4, pp. 749–772.

Martin, RJ (1979). “A subclass of lattice processes applied to a problem in planar sampling”. In: *Biometrika* 66.2, pp. 209–217.

Matérn, Bertil (1986). *Spatial Variation*. 2nd. Berlin: Springer-Verlag.

Matheron, Georges (1962). *Traité de Géostatistique Appliquée*. Vol. 1. Editions Technip-Paris.

May, Robert M (1978). “Host-parasitoid systems in patchy environments: a phenomenological model”. In: *The Journal of Animal Ecology*, pp. 833–844.

Méndez, Vicenç, Daniel Campos, and Frederic Bartumeus (2014). *Stochastic Foundations in Movement Ecology*. New York: Springer.

Metzler, Ralf, Walter G Glöckle, and Theo F Nonnenmacher (1994). “Fractional model equation for anomalous diffusion”. In: *Physica A: Statistical Mechanics and its Applications* 211.1, pp. 13–24.

Myers, Donald E (1989). “To be or not to be... stationary? That is the question”. In: *Mathematical Geology* 21.3, pp. 347–362.

Nakagawa, Shinichi and Robert P Freckleton (2008). “Missing inaction: the dangers of ignoring missing data”. In: *Trends Ecol Evolut* 23.11, pp. 592–596.

Nelson, William A et al. (2008). “Balancing ecological complexity in predictive models: a reassessment of risk models in the mountain pine beetle system”. In: *Journal of Applied Ecology* 45.1, pp. 248–257.

Neubert, Michael G, Mark Kot, and Mark A Lewis (1995). “Dispersal and pattern formation in a discrete-time predator-prey model”. In: *Theoretical Population Biology* 48.1, pp. 7–43.

Neumaier, Arnold (1998). “Solving ill-conditioned and singular linear systems: A tutorial on regularization”. In: *SIAM Review* 40.3, pp. 636–666.

Nicholson, Alexander J and Victor A Bailey (1935). “The Balance of Animal Populations – Part I”. In: *Proceedings of the Zoological Society of London*. Vol. 105(3). Wiley Online Library, pp. 551–598.

Nigh, Gordon D, Joseph A Antos, and Roberta Parish (2008). “Density and distribution of advance regeneration in mountain pine beetle killed lodgepole pine stands of the Montane Spruce zone of southern British Columbia”. In: *Canadian Journal of Forest Research* 38.11, pp. 2826–2836.

Okubo, Akira and Simon A Levin (2001). *Diffusion and Ecological Problems: Modern Perspectives*. 2nd. New York, NY: Springer-Verlag.

O’Shaughnessy, Ben and Itamar Procaccia (1985). “Analytical solutions for diffusion on fractal objects”. In: *Physical Review Letters* 54.5, pp. 455–458.

Painter, Kevin J and Thomas Hillen (2018). “From random walks to fully anisotropic diffusion models for cell and animal movement”. In: *Cell Movement*. Ed. by Magdalena Stolarska and Nicoleta Tarfulea. Cham, Switzerland: Springer Nature, pp. 103–141.

Peters, Robert Henry and Robert H Peters (1991). *A critique for ecology*. Cambridge, England: Cambridge University Press.

Pick, Georg (1899). “Geometrisches zur zahlenlehre”. In: *Sitzenber. Lotos (Prague)* 19, pp. 311–319.

Powell, James A and Barbara J Bentz (2014). “Phenology and density-dependent dispersal predict patterns of mountain pine beetle (*Dendroctonus ponderosae*) impact”. In: *Ecological Modelling* 273, pp. 173–185.

Powell, James A et al. (2018). “Differential dispersal and the Allee effect create power-law behaviour: Distribution of spot infestations during mountain pine beetle outbreaks”. In: *Journal of Animal Ecology* 87.1, pp. 73–86.

Preisler, Haiganoush K et al. (2012). “Climate and weather influences on spatial temporal patterns of mountain pine beetle populations in Washington and Oregon”. In: *Ecology* 93.11, pp. 2421–2434.

Preston, Frank W (1948). “The commonness, and rarity, of species”. In: *Ecology* 29.3, pp. 254–283.

Raffa, Kenneth F and Alan A Berryman (1986). “A mechanistic computer model of mountain pine beetle populations interacting with lodgepole pine stands and its implications for forest managers”. In: *Forest Science* 32.3, pp. 789–805.

Raffa, Kenneth F et al. (2008). “Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions”. In: *Bioscience* 58.6, pp. 501–517.

Raffa, KF and AA Berryman (1983). “The role of host plant resistance in the colonization behavior and ecology of bark beetles (Coleoptera: Scolytidae)”. In: *Ecological Monographs* 53.1, pp. 27–49.

Rasmussen, Carl Edward and Christopher KI Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT press.

Rasmussen, Lynn A (1974). *Flight and attack behavior of mountain pine beetles in lodgepole pine of northern Utah and southern Idaho*. Vol. 180. US Department of Agriculture, Forest Service, Intermountain Forest & Range.

Reid, RW (1962). “Biology of the mountain pine beetle, *Dendroctonus monticolae* Hopkins, in the East Kootenay Region of British Columbia II. Behaviour in the host, fecundity, and internal changes in the female”. In: *The Canadian Entomologist* 94.6, pp. 605–613.

Reyes, Perla E, Jun Zhu, and Brian H Aukema (2012). “Selection of spatial-temporal lattice models: assessing the impact of climate conditions on a mountain pine beetle outbreak”. In: *Journal of Agricultural, Biological, and Environmental Statistics* 17.3, pp. 508–525.

Roberts, Stephen et al. (2013). “Gaussian processes for time-series modelling”. In: *Phil. Trans. R. Soc. A* 371.1984, p. 20110550.

Robertson, Colin et al. (2008). “Risk rating for mountain pine beetle infestation of lodgepole pine forests over large areas with ordinal regression modelling”. In: *Forest ecology and management* 256.5, pp. 900–912.

Robertson, Colin et al. (2009). “Determination of the compositional change (1999–2006) in the pine forests of British Columbia due to mountain pine beetle infestation”. In: *Environ Monit Assess* 158.1, pp. 593–608.

Royama, Tomoo (2012). *Analytical Population Dynamics*. Vol. 10. Springer Science & Business Media.

Safranyik, L (1988). “Estimating attack and brood totals and densities of the mountain pine beetle in individual lodgepole pine trees”. In: *The Canadian Entomologist* 120.4, pp. 323–331.

Safranyik, L and DA Linton (1991). “Unseasonably low fall and winter temperatures affecting mountain pine beetle and pine engraver beetle populations and damage in the British Columbia Chilcotin Region”. In: *Journal of the Entomological Society of British Columbia* 88, pp. 17–21.

Safranyik, L et al. (1992). “Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand”. In: *Journal of Applied Entomology* 113.1-5, pp. 441–450.

Safranyik, L et al. (2010). “Potential for range expansion of mountain pine beetle into the boreal forest of North America”. In: *The Canadian Entomologist* 142.5, pp. 415–442.

Safranyik, Les and Allan L Carroll (2006). “The biology and epidemiology of the mountain pine beetle in lodgepole pine forests”. In: *The Mountain Pine Beetle: A Synthesis of Biology, Management, and Impacts on Lodgepole Pine*. Ed. by Les Safranyik and Bill Wilson. Victoria, Canada: Canadian Forest Service, pp. 3–66.

Sambaraju, Kishan R et al. (2012). “Climate change could alter the distribution of mountain pine beetle outbreaks in western Canada”. In: *Ecography* 35.3, pp. 211–223.

Sampson, Paul D and Peter Guttorp (1992). “Nonparametric estimation of nonstationary spatial covariance structure”. In: *JASA* 87.417, pp. 108–119.

Schlägel, Ulrike E and Mark A Lewis (2016). “A framework for analyzing the robustness of movement models to variable step discretization”. In: *Journal of Mathematical Biology* 73.4, pp. 815–845.

Searle, Shayle R, George Casella, and Charles E McCulloch (2009). *Variance components*. Vol. 391. Hoboken, NJ, USA: John Wiley & Sons.

Seuront, Laurent (2010). *Fractals and Multifractals in Ecology and Aquatic Science*. Boca Raton, FL: CRC Press.

Shegelski, Victor A, Maya L Evenden, and Felix AH Sperling (2019). “Morphological variation associated with dispersal capacity in a tree-killing bark beetle *Dendroctonus ponderosae* Hopkins”. In: *Agricultural and Forest Entomology* 21.1, pp. 79–87.

Sherman, Michael (2011). *Spatial statistics and spatio-temporal data: covariance functions and directional properties*. West Sussex, UK: John Wiley & Sons.

Shore, Terence Leckie and L Safranyik (1992). *Susceptibility and risk rating systems for the mountain pine beetle in lodgepole pine stands*. Vol. 336. Forestry Canada, Pacific Forestry Centre, Victoria BC.

Shore, TL, L Safranyik, and JP Lemieux (2000). “Susceptibility of lodgepole pine stands to the mountain pine beetle: testing of a rating system”. In: *Canadian Journal of Forest Research* 30.1, pp. 44–49.

Shrimpton, DM and AJ Thomson (1985). “Relationship between phloem thickness and lodgepole pine growth characteristics”. In: *Canadian Journal of Forest Research* 15.5, pp. 1004–1008.

Simpson, Daniel, Finn Lindgren, and Håvard Rue (2012). “In order to make spatial statistics computationally feasible, we need to forget about the covariance function”. In: *Environmetrics* 23.1, pp. 65–74.

Six, Diana L, Eric Biber, and Elisabeth Long (2014). “Management for mountain pine beetle outbreak suppression: Does relevant science support current policy?” In: *Forests* 5.1, pp. 103–133.

Skellam, John Gordon (1951). “Random dispersal in theoretical populations”. In: *Biometrika* 38, pp. 196–218.

Staniswalis, Joan G (1989). “The kernel estimate of a regression function in likelihood-based models”. In: *JASA* 84.405, pp. 276–283.

Stein, Michael L (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY: Springer-Verlag.

Stein, Michael L (2005). "Space–time covariance functions". In: *JASA* 100.469, pp. 310–321.

Strohm, S, ML Reid, and RC Tyson (2016). "Impacts of management on Mountain Pine Beetle spread and damage: a process-rich model". In: *Ecological Modelling* 337, pp. 241–252.

Strohm, S, RC Tyson, and JA Powell (2013). "Pattern formation in a model for mountain pine beetle dispersal: linking model predictions to data". In: *Bulletin of mathematical biology* 75.10, pp. 1778–1797.

Taylor, RAJ (1978). "The relationship between density and distance of dispersing insects". In: *Ecological Entomology* 3.1, pp. 63–70.

Tobler, Waldo R (1970). "A computer movie simulating urban growth in the Detroit region". In: *Economic Geography* 46.supplement 1, pp. 234–240.

Turchin, Peter and William T Thoeny (1993). "Quantifying dispersal of southern pine beetles with mark-recapture experiments and a diffusion model". In: *Ecological Applications* 3.1, pp. 187–198.

Van Loan, Charles F (2000). "The ubiquitous Kronecker product". In: *J Comput Appl Math* 123.1, pp. 85–100.

Varin, Cristiano, Nancy Reid, and David Firth (2011). "An overview of composite likelihood methods". In: *Statistica Sinica*, pp. 5–42.

Ver Hoef, Jay M et al. (2018). "Spatial autoregressive models for statistical inference from ecological data". In: *Ecological Monographs* 88.1, pp. 36–59.

Wall, Melanie M (2004). "A close look at the spatial structure implied by the CAR and SAR models". In: *Journal of Statistical Planning and Inference* 121.2, pp. 311–324.

Wang, Tongli et al. (2007). "ClimateBC v3. 2: a program to generate climate normal, decade, annual, seasonal and monthly data for genecology and climate change studies in British Columbia". In: *UBC, Department of Forest Science and BC Ministry of Forests, Vancouver, Canada*.

Warton, David I and Francis KC Hui (2011). "The arcsine is asinine: the analysis of proportions in ecology". In: *Ecology* 92.1, pp. 3–10.

Watson, George Neville (1995). *A Treatise on the Theory of Bessel Functions*. London: Cambridge University Press.

Weller, Zachary D, Jennifer A Hoeting, et al. (2016). “A review of nonparametric hypothesis tests of isotropy properties in spatial data”. In: *Statistical Science* 31.3, pp. 305–324.

Westfall, J and T Ebata (2009). *Summary of forest health conditions in British Columbia*. Tech. rep. British Columbia Ministry of Forests and Range, Forest Practices Branch, Victoria.

Whittle, Peter (1954). “On stationary processes in the plane”. In: *Biometrika* 41.3/4, pp. 434–449.

Williams, EJ (1961). “The distribution of larvae of randomly moving insects”. In: *Australian Journal of Biological Sciences* 14.4, pp. 598–604.

Wilson, Andrew G et al. (2014). “Fast kernel learning for multidimensional pattern extrapolation”. In: *Adv Neural Inf Process Syst*, pp. 3626–3634.

Wulder, Michael A et al. (2006). “Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities”. In: *For Ecol Manag* 221.1-3, pp. 27–41.

Wulder, Michael A et al. (2009). “Aerial overview survey of the mountain pine beetle epidemic in British Columbia: Communication of impacts”. In: *Journal of Ecosystems and Management* 10.1.

Wulder, Michael A et al. (2010). “A provincial and regional assessment of the mountain pine beetle epidemic in British Columbia: 1999–2008”. In: *J. Environ. Inform* 15.1, pp. 1–13.

Yamamura, Kohji (2002). “Dispersal distance of heterogeneous populations”. In: *Population Ecology* 44.2, pp. 93–101.

Yasuda, Norikazu (1975). “The random walk model of human migration”. In: *Theoretical Population Biology* 7.2, pp. 156–167.

Zauderer, Erich (2006). *Partial Differential Equations of Applied Mathematics*. 3rd. Hoboken, NJ: John Wiley & Sons.

Zeide, Boris (1991). “Fractal geometry in forestry applications”. In: *Forest Ecol Manag* 46.3-4, pp. 179–188.

Zheng, Yanbing and Brian H Aukema (2010). “Hierarchical dynamic modeling of outbreaks of mountain pine beetle using partial differential equations”. In: *Environmetrics* 21.7-8, pp. 801–816.

Zhou, Guofa and Andrew M Liebhold (1995). “Forecasting gypsy moth defoliation with a geographical information system”. In: *Insect Science* 2.1, pp. 83–94.

Zhu, Jun, Hsin-Cheng Huang, and Perla E Reyes (2010). “On selection of spatial linear models for lattice data”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3, pp. 389–402.

Zhu, Jun, Hsin-Cheng Huang, and Jungpin Wu (2005). “Modeling spatial-temporal binary data using Markov random fields”. In: *Journal of Agricultural, Biological, and Environmental Statistics* 10.2, pp. 212–225.

Zhu, Jun et al. (2008). “Autologistic regression analysis of spatial-temporal binary data via Monte Carlo maximum likelihood”. In: *Journal of Agricultural, Biological, and Environmental Statistics* 13.1, pp. 84–98.

Zimmerman, Dale L (1989). “Computationally exploitable structure of covariance matrices and generalized convariance matrices in spatial models”. In: *J Stat Comput Sim* 32.1-2, pp. 1–15.

Zimmerman, Dale L (1993). “Another look at anisotropy in geostatistics”. In: *Mathematical Geology* 25.4, pp. 453–470.