A Bayesian Space-Time Model for Discrete Spread Processes on a Lattice
Long, Jed A. 1 jlong@uvic.ca*
Robertson, Colin 2 crobertson@wlu.ca
Nathoo, Farouk S. 3 nathoo@math.uvic.ca
Nelson, Trisalyn A. 1 trisalyn@uvic.ca

1 Spatial Pattern Analysis & Research (SPAR) laboratory, Department of Geography, University of Victoria, Victoria, British Columbia, Canada
2 Department of Geography & Environmental Studies, Wilfrid Laurier University, Waterloo, Ontario, Canada
3 Department of Mathematics & Statistics, University of Victoria, Victoria, British Columbia, Canada

*Corresponding Author
Jed A. Long
Department of Geography, University of Victoria
PO Box 3060 STN CSC,
Victoria, BC, Canada, V8W 3R4
jlong@uvic.ca
+1 250 853 3271

Pre-print of published version.

Reference:

DOI:
http://dx.doi.org/10.1016/j.sste.2012.04.008

Disclaimer:
The PDF document is a copy of the final version of this manuscript that was subsequently accepted by the journal for publication. The paper has been through peer review, but it has not been subject to any additional copy-editing or journal specific formatting (so will look different from the final version of record, which may be accessed following the DOI above depending on your access situation).
Abstract

In this article we present a Bayesian Markov model for investigating environmental spread processes. We formulate a model where the spread of a disease over a heterogeneous landscape through time is represented as a probabilistic function of two processes: local diffusion and random-jump dispersal which allows the model to represent the leptokurtic spread pattern typical of many infectious diseases and biological invasions. We demonstrate the properties of this model using a simulation experiment and an empirical case study – the spread of mountain pine beetle in western Canada. Posterior predictive checking was used to validate the number of newly inhabited regions in each time period. Map comparison analysis was used to measure spatial agreement of spatially distributed model parameter estimates and observed values. The model performed well in the simulation study in which a goodness-of-fit statistic measuring the number of newly inhabited regions in each time interval fell within the 95% posterior predictive credible interval in over 97% of simulations. The map comparison analysis revealed that in some cases the magnitude of estimated parameter values differed markedly from the true values, but in all cases an adequate recovery of the spatial structure was obtained, indicating good spatial agreement. The case study of a mountain pine beetle infestation in Western Canada (1999 to 2009) extended the base model in two ways. First, spatial covariates thought to impact the local diffusion parameters, elevation and forest cover, were included in the model. Second, a refined definition for translocation or jump-dispersal based on mountain pine beetle ecology was incorporated improving the fit of the model. Posterior predictive checks on the mountain pine beetle model found that the observed goodness-of-fit test statistic fell within the 95% posterior predictive credible interval for 8 out of 10 years. The simulation study and case study provide evidence that the model presented here is both robust and flexible; and is therefore appropriate for a wide array of spread processes in epidemiology and ecology.

HIGHLIGHTS
- Develop and implement a hierarchical Bayes Markov model for spread processes.
- Case study describing the spread of mountain pine beetle in Western Canada, 1999-2009.
- Model assessment uses posterior predictive simulation and map comparison statistics.
- East of Rocky Mountains spread is dominated by translocation events.
- Model is flexible at handling complex spread processes across heterogeneous landscapes.

KEYWORDS: space-time binary data, spread process, spatial random effects, mountain pine beetle, map comparison
1. Introduction

Understanding the emergence and spread of infectious diseases is of increasing concern for promoting global health (Jones et al., 2008). While the reasons for changes in disease patterns over time and space are complex and multidimensional (Morse, 1995), there is a growing need for models capable of describing variation in the spread pattern once cases are being reported (Riley, 2007). Further, understanding underlying risk factors associated with disease amplification is needed to establish appropriate control measures. For example, animal movement and network structure often have an important role in how zoonotic disease epidemics or epizootics develop and spread (Kiss et al., 2006). Similarly in ecology, the spread of non-native species are routinely linked to anthropogenic vectors (e.g., Coetzee et al., 2009) or climate change (e.g., Cudmore et al., 2010). As a result, the study of spread processes, defined here as the ability of an organism or disease to expand its current range, is receiving considerable attention in both the epidemiological and ecological literature, and increasingly detailed spatial-temporal datasets are providing new opportunities to study the dynamics of spread (see for example, Hooten et al., 2010). Given the increased rate at which many organisms are spreading (Ricciardi, 2007), continued development of methods and tools capable of modeling complex spread processes are warranted.

Due to the nature of disease surveillance systems which are the primary data sources for disease modeling studies, data are often only available at discrete temporal intervals (e.g., weeks). Similarly, many pathogens spread via fomites at discrete time periods. For example, marine invasive species such as the zebra mussel (*Dreissena polymorpha*) spread primarily due to recreational boating, which peaks during the summer months (Schneider et al., 1998). Other species disperse naturally at discrete temporal intervals. For instance, bark beetles emerge and
disperse on an annual basis (Safranyik & Carroll, 2006). As a result, there is considerable interest in developing discrete-time representations of spread in ecological models.

Spatially, data can be represented as either an aggregated spatial unit (i.e., discrete) or as point-events (i.e., continuous). For most ecological models, aggregated data are used simply due to ease of field sampling. Units are analogous to quadrats in which the presence / absence or abundance of the species is measured. While continuous spatial data provide a high level of spatial detail, this is typically purchased at the expense of the spatial extent of the study. For ecological models at the landscape scale, point-event data are often not feasible. For both discrete and continuous spatial data, representation of landscape heterogeneity is often a major limitation in models of spread (Pitt et al., 2009). For example, physical barriers such as mountains and rivers are often poorly represented using traditional geographic data formats (Cova and Goodchild, 2002). This issue is exacerbated by the fact that processes at multiple spatial scales act in concert to produce spread patterns on the landscape, yet modeling is often carried out at a single spatial scale (Pitt et al., 2009). Pearson and Dawson (2003) have proposed hierarchical modeling as a potential solution which can incorporate ecological mechanisms at multiple spatial scales.

The spread of disease is often the result of multiple mechanisms. For example, foot and mouth disease typically spreads among animals and herds via airborne transmission, and among farms and regions by animal movement networks and fomites (Green et al., 2006). Similarly, in many ecological invasions, the resultant pattern of invasion is often multi-causal: locally through diffusion or movement and over greater distances via intermediate species or translocation vectors. Smith et al. (2002) provide a mathematical model for spread processes characterized by both local and random-jump dispersing mechanisms, using the spread of raccoon rabies across a
lattice in their example. However, inference in Smith et al. (2002) was based on a stochastic estimation algorithm rather than a formal statistical framework for inference. The spread of raccoon rabies was also examined by Wheeler and Waller (2008) who link spatial variation in patterns of spread (i.e., deviations from a travelling wave) to landscape heterogeneity using spatially-varying regression, and adopt a Bayesian framework for inference.

Bayesian space-time spread models have previously been formulated for investigating environmental spread processes with point-referenced count data, (e.g., Wikle, 2003; Hooten and Wikle, 2006). As well, Gibson et al. (2006) have developed a Bayesian space-time percolation model for contact-based (local) spread across a spatial lattice; however, the model developed there does not accommodate translocation events, where the disease process spreads across disconnected regions. This is particularly important considering that the broad scale outcomes of spread by many ecological organisms are dominated by random-jump translocation events, and not local diffusion (e.g., Suarez et al., 2001).

Despite the inherent link between the spread of species of interest to ecologists, and disease spread in human populations studied by epidemiologists, models have been developed largely independently in these fields until very recently (Smith et al., 2002). Given that the majority of emerging diseases are zoonotic in nature (Jones et al., 2008), and often of wildlife origin, there exists a need for integrated ecological-epidemiological modeling. In this article we present a hierarchical Bayes approach appropriate for modeling either disease or organism spread across a landscape, and allow for landscape heterogeneity using spatially-varying parameters. Although we investigate a specific ecological application (i.e., mountain pine beetle in western Canada), the methods employed are generic and appropriate for a wide range of spread problems in ecology and epidemiology.
In this article, we present a discrete-time Markov model for ecological spread processes allowing for spatially-varying spread rates and random-jump dispersal. In the next section we present the model. A simulation study follows, which evaluates model performance. Simulated datasets at multiple spatial scales, under various spread scenarios, are used to illustrate model strengths and weaknesses. Next, we report on an empirical case study investigating the spread of mountain pine beetle (*Dendroctonus ponderosae*) in western Canada. Finally, we discuss remaining challenges and practical issues related to the process of model development and conclude by linking this research with potential applications in epidemiology.

2. Model Development

Given *n* regions comprising a study area, we formulated a logistic model for a binary spread process (defined here as inhabited-uninhabited) where it was assumed that newly inhabited regions do not revert to being uninhabited (Mollison, 1977), and we let \( Z_i(t) \in \{0, 1\} \) indicate the presence of an organism or disease in region \( i, i=1,..,n \), at time \( t \), \( t=1,...,T \). Here, \( t = 1 \) corresponds to the initial map of organism or disease presence, and the vector \( Z(t) = (Z_1(t), ..., Z_n(t))' \) represents a binary map describing the progression of the organism or disease at time \( t \). Spread is described through a stochastic process model for \( Z(t) \), which conditional on model parameters \( \Theta \), is assumed to follow a first-order Markov assumption, so that \( Pr[Z(t)|Z(t-1),Z(t-2),...,Z(1),\Theta] = Pr[Z(t)|Z(t-1),\Theta] \) where the Markov transition kernel is indexed by \( \Theta \). The Markov model is further simplified by assuming conditional independence across regions, so that \( Pr[Z(t)|Z(t-1),\Theta] = \Pi_i Pr[Z_i(t)|Z(t-1),\Theta] \). Spatial dependence is accommodated at the second level of the hierarchical specification with random effects incorporated into \( \Theta \). The term \( p_{it} = \ldots \)
Pr[Z_i(t)|Z(t-1),Θ] represents the probability that the organism or disease is present in region i at time t, given the presence map Z(t-1) and conditional on Θ. We assumed that regions where the organism or disease is present remain inhabited, so that \( p_{it} = 1 \) if \( Z_i(t-1) = 1 \); whereas, if region i is free of organism or disease at time t-1, so that \( Z_i(t-1) = 0 \), we assumed a logistic specification with space varying coefficients

\[
\log\left(\frac{p_{it}}{1 - p_{it}}\right) = \mu_t + \lambda_i NN_{i,t-1} \tag{1}
\]

where \( NN_{i,t-1} \) is the number of inhabited neighbors of region i at time t-1; \( \mu_t \) is a time varying parameter representing a baseline probability of becoming inhabited; and \( \lambda_i \) is a spatially-varying parameter quantifying the local impact of inhabited regions on their uninhabited neighbors.

Neighbors for \( NN_{i,t-1} \) are defined using a Queen’s case (Moore neighborhood – i.e., edge or corner in contact signifies a neighbor) definition of spatial neighbors; however, alternate neighborhood configurations could easily be explored. The baseline rate \( \mu_t \) is common to all regions, and for an uninhabited region with no inhabited neighbors at time t-1 (isolated from the spread wave) we have

\[
\log\left(\frac{p_{it}}{1 - p_{it}}\right) = \mu_t \tag{2}
\]

so that \( \mu_t \) can be thought of as representing the time-varying probability of translocation events, describing random-jump movements by a species. Inclusion of terms for both diffusion and random-jump movements is important when spread occurs via separate mechanisms or at differing spatial scales. Mountain pine beetles, for example, spread via two independent mechanisms, actively over short distances (e.g., within forest stands), and passively via convective wind currents capable of transporting small populations for hundreds of kilometers (Safranyik & Carroll, 2006).
Comparison of equations [1] and [2] reveals that $\lambda_i$ characterizes the change in the log-odds of habitation for region $i$, arising from the presence of organisms in neighboring regions. Larger values of $\lambda_i$ correspond to an increasing probability of spread from inhabited regions to uninhabited neighbors. As such, $\lambda_i$ controls the rate of organism mobility or diffusion into region $i$, and the vector $\lambda = (\lambda_1, \ldots, \lambda_n)'$ characterizes spatial variability in diffusion across the entire study area.

Using a hierarchical modeling approach, we allowed for temporal variation in the translocation parameters $\mu = (\mu_1, \ldots, \mu_T)'$ and spatial variation in diffusion parameters $\lambda = (\lambda_1, \ldots, \lambda_n)'$ using mixed model random effect specifications. The translocation component $\mu_i$ is composed of a constant $\mu_c$, coupled with time-varying mean-zero effects $\theta_i$, so that

$$\mu_i = \mu_c + \theta_i \quad [3]$$

A weakly informative prior for $\mu_c$ was adopted, using a Normal distribution with mean 0 and a precision of $1/1000$, or $\mu_c \sim N(0,0.001)$. The $\theta_i$ represent year-to-year variation in translocation, and are modeled independently as $\theta_i \sim N(0,\tau_{\theta})$ with the variance $\tau_{\theta}$ assigned a conditionally conjugate inverse-Gamma hyper-prior $\tau_{\theta} \sim inverse-gamma(0.01,0.01)$.

The spatially-varying diffusion parameters $\lambda = (\lambda_1, \ldots, \lambda_n)'$ are modeled using a convolution prior

$$\lambda_i = \alpha_0 + h_i + a_i \quad [4]$$

where $\alpha_0$ represents the baseline level of spread across the study area; $h_i \sim N(0,\tau_h)$ are independent and identically distributed random effects representing spatially unstructured variation; and $a_i$ is a spatially correlated random effect, with the vector $a = (a_1, \ldots, a_n)'$ modeled using an intrinsic conditional autoregressive model – CAR($\tau_a$). This random effect formulation follows Besag, York and Mollie (1991), who suggest including both independent and spatially
correlated random effects. Additional terms corresponding to spatially varying covariates can be easily incorporated in [4] to investigate relationships between covariates and the local rate of diffusion. The spatial CAR model for $a_i$ uses a Queen’s case (i.e., adjacency) definition of neighbors. In addition, a binary definition of weights is used, with neighbors coded as 1 and non-neighbors as 0. Finally, our model specification is made complete by assigning a flat prior to $\alpha_0$, and weakly informative $inverse-gamma(0.5, 0.0005)$ hyper-priors for the variance components $\tau_h$ and $\tau_a$.

The vector $\Theta$ is the set of parameters in our model, with $\Theta = \{h, a, \theta, \alpha_0, \mu_c, \tau_a, \tau_h, \tau_\theta\}$. With this specification, Bayesian inference for $\Theta$ is based on the posterior distribution $\Theta|Z(1), \ldots, Z(T)$, where $Z(1), \ldots, Z(T)$ are binary data vectors representing a realized ecological spread process. The posterior distribution is computed using Markov chain Monte Carlo methods implemented in the free software WinBUGS (Lunn et al., 2000). The model code used for fitting this model can be obtained from the first author upon request.

3. Model Evaluation Using a Simulation Study

3.1 Simulation Study Data

We carried out a simulation study to assess the model performance under different scenarios describing the spread of disease. While employing Bayesian inference to ‘borrow strength’ can help address the issue of inaccurate estimation due to infrequent sampling (i.e., big area; small numbers), the opposite effect may be true for pooled estimates that are pulled too much towards the mean (Gelman and Price, 1999). When applied with real data the diffusion (spatial) and translocation (temporal) parameters will be unknown, therefore we adopt a
simulation-estimation approach to investigate the sensitivity of the model to changes in model parameter values and spatial scale.

Spread datasets were simulated from a set of patterns representing realistic scenarios of spatial diffusion (Figure #a) and time-varying translocation (Figure #b), in combination with three different spatial scales. Spatial diffusion represented as a linear trend ($\Lambda_1$) can be thought to represent a simplified spread process across a more homogeneous landscape, this scenario might be encountered, for example, in the presence of a latitudinal gradient. The Gaussian random field scenario ($\Lambda_2$) is an example of a more complex, heterogeneous landscape, characteristic of a wide range of spread processes. Constant translocation ($M_1$) represents a static level of random translocation events. Linear decreasing translocation ($M_2$) represents a situation where a spread mechanism (e.g., fomite transmission of foot and mouth disease) is decreasing over time as control measures are put in place. Finally, oscillating translocation ($M_3$) represents a seasonal cycle to translocation, as in the case of wind-driven transport which depends primarily on seasons. Variation in spatial scale ($20\times20, n = 400; 40\times40, n = 1600; 80\times80, n = 6400$) is used to examine the impact of scale on model performance. Specifically, we are interested in how changes in the grain (resolution) of the data may affect results. Extent, the other aspect of spatial scale, is of less interest here, as we assume the extent of the disease is contained by the study area.

The simulation-estimation procedure involved three steps: 1) generate a realistic spread scenarios using combinations of diffusion ($\Lambda$), translocation ($M$), and spatial scale (see Figure #); 2) simulate spread data from the scenarios; and 3) estimate the model based on the simulated data. Each combination of spatial diffusion pattern, time-varying translocation pattern, and spatial scale comprised one spread scenario, for a
total of 18 different scenarios (the simulated spread datasets for the 6 scenarios at the 40x40 scale are shown in Figure ##).<br/>

3.2 Examining Model Fit

To compare the true and estimated diffusion values we used a global chi-square goodness of fit statistic where bins were set at intervals of 0.25. The test is a comparison of the number of observations in each bin for the known parameters and the number of estimates in each bin in the estimated parameters. In addition to chi-square tests, we also report standardized residuals for both diffusion and translocation, defined as the absolute value of the difference between the true and estimated parameter value, divided by the number of observations (n for diffusion, T-1 for translocation).

To evaluate model fit to the simulated data, we used posterior predictive checking (Gelman et al., 2004). To perform posterior predictive checks, 100 draws from the posterior distribution of all diffusion and translocation parameters were obtained. Data were simulated with the model using these parameter values to create 100 replicate datasets from the posterior predictive distribution. We assessed similarity between these replicates and the observed data for some test quantity of interest (Gelman et al., 2004). In our case the observed data are the original data used to describe a spread process, and were generated by the model using the chosen parameter values, while in practical applications this would be the observed spread data. The test statistic we used was the number of new cells inhabited in each time period, Inh,

\[ Inh = \sum_{i=1}^{n} Z_i(t) - Z_i(t-1) \quad [5] \]

Two of the simulated datasets were selected for checking based on the results of the parameter estimation recovery analysis so that one of the better performing scenarios, and one of the poorer
performing scenarios were evaluated, both using the 40x40 spatial scale to ensure comparability. In each case the test statistic was evaluated at each time point to determine if a value of $Inh_t$ computed for the data falls outside the main mass of the corresponding posterior predictive distribution, in which case we have evidence that the model does not fit this aspect of the data.

We also examined the spatial structure of local diffusion using map comparison analysis. The objective of map comparison is to uncover similarities (or differences) between expected ($\Lambda$) and estimated ($\lambda$) diffusion maps, and evaluate whether two maps could have been generated by the same process. This is facilitated in the simulation examples as we have both an expected diffusion map (e.g., those $\Lambda$ in Figure 2 from which the spread process was generated) that we can compare to the mean posterior predictive estimates ($\bar{\lambda}$). In terms of model validation, considering spatial structure provides improved confidence in estimated $\lambda$ over purely aspatial comparisons. The structural similarity (SSIM) index was selected as an exploratory statistic for comparing maps (Wang et al., 2004). SSIM incorporates a Gaussian weighting function, to assess similarity across spatially local regions. This is in contrast to direct pixel to pixel comparisons, which ignore spatial structure in maps, often producing comparison measures highly sensitive to slight spatial misalignment (Pontius Jr., 2000). SSIM considers three components for map comparison: luminance, contrast, and structure, relating to local differences in mean, variance and covariance respectively (Wang et al., 2004). Note that these three components are relatively independent, and changes in one component will not necessarily affect others. SSIM takes the following spatially local form, computing a similarity statistic for each spatial unit:

$$SSIM(i) = (l(i))^{a} \cdot (c(i))^{b} \cdot (s(i))^{g}$$  \[6\]
where \( i \) denotes the \( i^{th} \) spatial unit, \( l \) the luminance component, \( c \) the contrast component, and \( s \) the structure component (see Wang et al., 2004 for further details). The exponents \( \alpha, \beta, \) and \( \gamma \) can be used to weight individual components, with default values taken as \( \alpha = \beta = \gamma = 1 \). The local components \( l(i) \) and \( c(i) \) are strictly positive while \( s(i) \) can take on negative values. We report only the mean global statistic for each of the three components and overall similarity, noting that although locally the product from [6] holds, due to summation rules the mean SSIM value does not equal the product of the means of each component. When two maps are identical, \( \text{SSIM} = 1 \), and values decrease from 1 as similarity decreases. SSIM was calculated for all 6 simulated scenarios at the 40x40 spatial scale. Given the map size (40x40), we selected a Gaussian weighting function with parameters \( h = 3 \) and \( sd = 0.5 \). SSIM results were insensitive to minor changes to \( h \) and \( sd \).

As a final check on model sensitivity, we varied the hyper-parameters corresponding to the prior variances for random effects governing variation in both diffusion and translocation to other suggested alternatives, \( \text{inverse-gamma}(0.001, 0.001) \) and \( \text{inverse-gamma}(0.1, 0.1) \). The effects of these prior adjustments on point estimates for diffusion and translocation are reported.

### 3.3. Simulation Study Results

The global goodness of fit analysis for the 27 different simulated spread scenarios are reported in Table #. These global chi-square tests reveal that in none of the scenarios were the estimated parameters significantly different than the true values. The standardized residuals demonstrate the effects of changes in the pattern of spread and spatial scale on diffusion (Appendix B). Residuals tended to increase with larger study areas. For the 20x20 and 40x40 spatial scales, the \( \Lambda_2 \) spatial trend produced larger error than the \( \Lambda_1 \) spatial trend or the \( \Lambda_3 \) Gaussian random field; however error was largely similar for all three patterns for the 80x80
study area. For translocation, the opposite general pattern holds, with larger residuals for smaller datasets. This is likely due to the lack of available cells for translocation to occur at smaller scales. The interdependency between diffusion and translocation is illustrated in Appendix B. Interdependency is clearly impacted by spatial scale, with fairly similar patterns in residuals at the 20x20 scale, and less so at larger scales.

Analysis of model fit using posterior predictive checks based on the statistic \( Inh \) is presented in Figure 5. For datasets simulated from scenario \( \Lambda_2M_3 \) and scenario \( \Lambda_3M_2 \), the test statistic fell within the 2.5 and 97.5 percentiles (95% credible interval), 97 and 98 times (out of 100) respectively, indicating a very good fit to the data. For the simulated data, the model appears to capture the timing of newly inhabited cells well.

Map comparison of estimated (\( \lambda \)) and expected (\( \Lambda \)) diffusion maps (40x40 spatial scale) revealed different trends from purely aspatial measures reported in Table ##. Estimated \( \lambda \) maps associated with \( M_2 \) showed the lowest similarity in all three \( \Lambda \) scenarios. In the case of scenarios \( \Lambda_1M_2 \) and \( \Lambda_2M_2 \), mean SSIM values were extremely low (-0.026 and 0.064 respectively) indicating poor model fit. These low SSIM values can be attributed to low scores in the luminance component (0.231 and 0.214 respectively). Map similarity was considerably higher in the other scenarios, with a maximum of 0.754 for scenario \( \Lambda_1M_1 \).

Finally, model sensitivity to hyper-priors on variance parameters of the spatial CAR component \( (\tau_a) \) revealed the model inference to be robust to the prior forms considered. Effects of changes in these hyper-priors on point estimates of diffusion and translocation are outlined in
Table 2. For both diffusion and translocation, changes in parameter estimates ranged from 0 to 0.03 for posterior means, and 0 to 0.02 for posterior standard deviations.

Overall, the simulation study provides convincing evidence that the model and the corresponding Bayesian inference are able to recover the parameter values used to simulate the data. Changes to variance hyper-parameters for diffusion and translocation have little impact on estimation in the settings we considered. Further, the effect of spatial scale has also been shown to be an important consideration. Highlighted by map comparison analysis, spatial structure does indeed play an important role in assessment of maps of true vs. estimated output parameters. Map comparison revealed that when maps of true (Λ) vs. estimated (λ) diffusion parameters were dissimilar the bulk of this difference can be attributed to the magnitude of the values (luminance), and that our model does effectively reveal expected spatial structure. This means that in some cases interpretation should be limited to spatial patterns observed, taking the magnitude of reported λ values as potentially misleading. In many cases modeling efforts primarily investigate spatial patterns of output parameters (e.g., high areas vs. low areas), and less so the magnitudes of output values. The model is effective at identifying such spatial variation in parameter estimates.

4. Empirical Case Study – Mountain Pine Beetle in Western Canada

4.1. Background

Mountain pine beetle is the most destructive biotic agent of mature pine forests in western North America (Safranyik & Carroll, 2006). Endemic to this region, mountain pine beetles typically attack weakened pine trees scattered throughout the forest. Periodically, when favorable conditions manifest, mountain pine beetle populations escalate to epidemic levels,
causing mortality to mature pine trees covering thousands of hectares (Safranyik & Carroll, 2006). Originating around 1998, the current outbreak is the largest on record and has devastated western Canada’s forest industry, causing substantial timber losses (British Columbia Ministry of Forests and Range, 2007). A warming climate combined with forest fire suppression has resulted in an overabundance of mature lodgepole pine (*Pinus contorta*) trees on the landscape. As lodgepole pine are the preferred host of mountain pine beetle, the combined effect of a warming climate and forest fire suppression is listed as probable cause for the magnitude of the current outbreak (Carroll et al., 2006).

The historical range of mountain pine beetle in Canada is predominantly within the province of British Columbia (Figure 3). Current epidemic mountain pine beetle populations have breached historic physiographic (e.g., Rocky Mountains) and climatic barriers to spread (Safranyik et al., 2010). Substantial beetle populations now exist in the province of Alberta, where the range of the beetles preferred host – lodgepole pine, meets the range of jack pine (*Pinus banksiana*) (Figure 3). Empirical evidence has found that jack pine is an alternative suitable host for mountain pine beetle (Furniss and Schenk, 1969; Cerezke, 1995). In the absence of climatic factors inhibiting beetle populations, jack pine, present throughout the boreal forest, could provide continuous habitat facilitating further eastward expansion by mountain pine beetle and negative economic and ecological consequences in Canada’s boreal region (Logan and Powell, 2001; Carroll et al., 2006; Safranyik et al., 2010).

The objective of this case study is to use the proposed Bayesian spread model to learn about processes governing the spread of mountain pine beetle at the boundary of its historical spatial range. Mountain pine beetle spread is of two types; active and passive spread, which
facilitate movement of beetle populations. Active spread represents spatially local dispersal events where beetles fly within or between neighboring pine stands and is the principle mode of spread (Safranyik et al., 1992). Mountain pine beetles are also capable of passive spread whereby beetles are carried long distances via convective wind currents during periods of emergence (Shepherd, 1966; Furniss and Furniss, 1972; Ainslie and Jackson, 2011). The model we have developed captures active spread through the spatially local diffusion parameters – $\lambda$, and passive spread through the temporally stochastic translocation parameters – $\mu$. We hope to gain insight into mountain pine beetle spread during the current epidemic by interpreting spatial variation in $\lambda$, and temporal variation in $\mu$.

4.2. Data and Study Area

Mountain pine beetle infestation data were obtained from the British Columbia Ministry of Forests and Range\(^1\) and the Alberta Department of Sustainable Resource Development\(^2\) for each year of our study (1999 – 2009). These infestation data are primarily obtained through aerial overview surveys, but also in situ measurements and remotely sensed data sources. Mountain pine beetle emergence occurs during a short one month window during the summer in the study area. As such, the spread process can be measured at discrete (i.e., annual) intervals. Infestation events are represented as both points (indicating a small cluster of infested trees) and polygons (a large area of infestation).

We selected a rectangular study area that covers the northern portion of the eastward expansion by mountain pine beetle into the province of Alberta (inset Figure 3). A 12 km grid was demarcated across the study area, generating $n = 2310$ contiguous spatial units. Similar 12 km spatial units have been used for investigating characteristics of a previous mountain pine

\(^1\) More Info at: http://www.for.gov.bc.ca/hfp/health/overview/overview.htm

\(^2\) More Info at: http://www.srd.alberta.ca/ManagingPrograms/ForestPests/ForestPestSurveyData.aspx
beetle outbreak in British Columbia (Aukema et al., 2008; Zhu et al., 2008) and spatial
synchrony within the current outbreak (Aukema et al., 2006).

We were also interested in investigating relationships between mountain pine beetle
spread (Figure 4a) and environmental factors. Two spatial covariates, elevation and forest cover
(see Figure 4 b and c), were identified in the literature as important in governing local spread of
mountain pine beetle. Elevation was taken as the mean of elevation values within each spatial
unit using a fine grain elevation dataset (spatial resolution of 25 m). Percent forest cover for each
spatial unit was determined using a national land cover database (Wulder et al., 2008). These
spatial covariates were incorporated into the model for \( \lambda \) [4], and relate to local diffusion so that
[4] becomes

\[
\lambda_i = \alpha_0 + X_i \beta + h_i + a_i \quad [7]
\]

where \( X_i \) is a vector of spatial covariates (e.g., elevation, percent forest cover) at location \( i \), and \( \beta \)
are the associated coefficients.

< approximate location for Figure 4 >

Exploratory spatial analysis revealed that mountain pine beetle translocation events
exhibited distance-dependence, whereby translocation events occur more frequently proximal to
previously infested regions. This phenomenon is commonly associated with the characteristic
leptokurtic pattern of spread (e.g., in animal-borne diseases #Fergusan, Lindstrom#, human
diseases #REF#, and with invading organisms, Lewis 1997) whereby translocation events are
distance-dependent relative to the spread wave. In this scenario, extremely long distance
translocation events are rare, but still possible. To account for this effect, we considered a more
general model for the translocation component that more appropriately resembles this distance-
dependant spread process. The new distance-dependent translocation parameter is defined as:
\[
\tilde{\mu}_s = \mu_c + \gamma d_t + \theta_t, \quad [8]
\]

where \(\tilde{\mu}_s\) is a spatial and temporally varying translocation parameter, \(d_t\) is the distance (centroid to centroid) from cell \(i\) to the nearest infested region at time \(t\), with coefficient \(\gamma\), and \(\mu_c\) and \(\theta_t\) are as defined in [3]. This treats the distance values \((d_t)\) as a space-time covariate modulating the translocation component.

4.3. Model Implementation

Several variations of the model were implemented incorporating different parameters (Table 1). For each model, two MCMC chains were run to fit the model. Convergence was assessed following Brooks and Gelman (1998), and a conservative burn-in of 10,000 iterations was selected. Following burn-in, 20,000 samples from each chain were retained for inference.

Model selection was based on the deviance information criterion (DIC), which combines the deviance with a penalty for model complexity (Spiegelhalter et al., 2002). Posterior mean, variance, and 95% equal-tailed credible intervals were used to summarize the posterior distributions.

Variation in DIC between model 1 and 2 was substantial, indicating that the inclusion of a distance-dependent translocation term (\(\tilde{\mu}_s\) from [8]) improved the model (see Table 1). As such, the four subsequent model specifications all use the \(\tilde{\mu}_s\) definition from [8]. The variation in DIC between models that use the \(\tilde{\mu}_s\) definition for translocation (models 2–6) was small (Table 1). The addition of the aspatial random effect parameter \((h)\) had little effect on the DIC. However we include \(h\) in subsequent model specifications as it can be used to interpret variation in mountain pine beetle spread not captured by the smoothing effect of the CAR model and/or \(\lambda\) covariates; which may relate to barriers to spread. Model 6, the most complex model tested
(including both the elevation and forest cover covariates) resulted in the lowest DIC value, and forms the basis for further discussion.

To evaluate model fit we performed a posterior predictive check similar to that described in the simulation experiment. For the case study data, we drew 1000 samples from the posterior distribution of each parameter and these were then used to draw 1000 replicate datasets ($Z_{rep}$) from the posterior predictive distribution. The test statistic ($inh_i – see [5]$) of the true infestation data fell within the 2.5 and 97.5 percentiles (95% posterior predictive credible interval) of the simulated $Z_{rep}$ data in all but two years (8 out of 10), indicating a reasonable model fit (Figure 6). The two anomalies were associated with the first year after initial infestation (2000) and an extreme peak in infestations that occurred in 2007.

A negative relationship (posterior mean = -0.153, 95% C.I. = [-0.337, 0.019]) was observed between local diffusion rate and the elevation covariate. A negative relationship between mountain pine beetle and elevation in British Columbia has been previously reported, and is believed to be linked to elevational constraints on pine species, the beetles preferred host (Aukema et al., 2008; Zhu et al., 2008). This relationship does not necessarily apply east of the Rocky Mountains and may be reason that this relationship is rather weak (e.g., the 95% credible interval covers zero). The rugged topography of the Rocky Mountains historically provided a physical barrier to eastward expansion by mountain pine beetle, with only a few small cases of infestation observed east of the Rockies (see Cerezke, 1989). The current epidemic has breached this barrier, and continued eastward expansion by mountain pine beetle through the boreal will not be hindered by topography as it is comparably flat.
A positive relationship (posterior mean = 0.2681, 95% C.I. = [0.130, 0.406]) was identified between local diffusion rate and forest cover. In British Columbia, elevation had previously been used as a surrogate for forest cover information, as here lodgepole pine is found primarily at lower elevations. As previously mentioned, east of British Columbia topography is far less variable, and pine species are found throughout the range of elevations within the boreal forest. As mountain pine beetle continues its eastward expansion, variables that more appropriately represent the availability of suitable pine hosts directly will be most useful for predicting infestation.

Using a map of the spatially-varying diffusion parameter (λ) we can highlight regional variability in the rate of local spread (Figure 7a). Comparing to the time of infestation map (Figure 4a) we can clearly see that high λ values are found immediately east of the originally infested regions. Here due to the rugged topography, mountain pine beetle spread quickly along linear forest tracts in valley bottoms as has been previously demonstrated (Robertson et al., 2009). Caution should be taken interpreting λ values in regions where no infestation has occurred (such as in the most eastern portion of our study area). Here the model infers a continuous λ surface from few relevant λ measurements and posterior variance is highest (Figure 7d).

The map of the aspatial random effect parameters (h), can be used to identify regions where the smoothing of the CAR effect (Figure 7b) over-estimates (negative values) or under-estimates (positive values) the rate of local diffusion (Figure 7c). Regions over-estimated (negative values in Figure 7c) by the CAR portion of the model are likely due to abrupt changes in land cover resistant to mountain pine beetle (such as large lakes or mountain peaks). This interpretation may provide a valuable tool for examining landscape barriers on spread where...
representation of barriers is discrete rather than continuous. Reason for areas under-estimated
(positive values in Figure 7c) may be due to the topographic effect mentioned earlier. However,
the magnitude of the $h$ effects are quite small (-0.007 to 0.005). Given that the $h$ effect is quite
small, the $\lambda$ maps are therefore predominantly associated with a combination of the CAR effect
($a$) and the environmental covariates ($X\beta$).

We also examine annual changes in translocation ($\tilde{\mu}_t$) through time, by way of the
parameter $\theta_t$. Mountain pine beetle translocation is highest in 2007, identified by the sharp peak
in $\theta_t$ values in that year (Figure 8). In this example, interpretation of temporal trends in $\theta_t$
requires consideration of original mountain pine beetle infestation data sources. Aerial overview
surveys and remotely sensed data rely on visual cues of tree mortality (i.e., foliage turning red),
which occurs 1-2 years after beetle presence. Thus, the peak value of $\theta_t$ observed in 2007
actually corresponds to increased translocation events by mountain pine beetle in 2005 or 2006.
This is in line with reports of extensive beetle activity in 2006 (Carroll, 2010). Although
temporal climate covariates were not investigated here, factors such as uncharacteristically warm
summers or cold winters influence beetle populations, and the success of the beetles passive
spread mechanism (Stahl et al., 2006).

With some ecological invasions, barriers may be introduced as a management tactic,
effective at slowing the spread of an invading species (Sharov and Liebhold, 1998). In Western
Canada, clear-cut harvests and controlled forest fires have been implemented as barriers to
mountain pine beetle spread through the removal of large, contiguous sections of potential host
trees. In British Columbia, the voracity of the current epidemic has circumvented any mitigation
efforts (Wilson 2004), however in Alberta there is still hope that these, and other preventative
measures will prove successful at stopping future eastward beetle spread. Preventing continued expansion by mountain pine beetle will be challenging given evidence that the beetles reproductive success improves in lodgepole pine stands outside of its historical range (Cudmore et al., 2010). Further, from our analysis it is clear that east of the Rocky mountains, mountain pine beetle spread, like many other spread processes (e.g., Suarez et al., 2001), is dominated by translocation events. When translocation dominates ecological invasions, the organism is often able to jump spread barriers rendering them ineffective. In such cases, it is necessary to carefully evaluate whether the introduction of barriers will provide the intended ecosystem and economic benefits (Sharov and Liebhold, 1998).

5. Discussion & Conclusions

The validation process we adopt is an example of how both aspatial and spatial indices can be incorporated into the validation procedure. Following guidelines of Gelman et al. (2004), we use posterior predictive checking as an aspatial measure of model fit that can be used even when true parameter values are unknown. This posterior predictive check revealed that our model is sufficient at recovering the timing of new infections. In the simulation study, we were able to complement this aspatial technique with a map comparison analysis (SSIM - Wang et al., 2004) to assess spatial structure of $\lambda$ values. Map comparison analysis revealed that in some cases estimated $\lambda$ values were different from expected in magnitude, but that in general the spatial pattern of $\lambda$ values was retrieved. The SSIM method enables creation of maps of local differences in mean, variance, and covariance, providing information on the spatial structure and differences in each (although we did not include such maps in this presentation). This spatial
approach to model evaluation represents a relatively simple procedure that can be easily implemented with existing models providing valuable and unique insight on how spatial structure of parameters relate to model performance. The SSIM measure was originally designed for evaluating image compression algorithms, and only recently has been proposed as a useful measure for quantitative comparison of continuous-value maps #Hagen-Zanker#. Thus, improving our understanding of how the SSIM (or other similar statistics) can be used as spatial measures of model evaluation remains an ongoing endeavor.

Both in the simulation examples and the case study we model spread across a regular tessellation (grid). In the mountain pine beetle example, we selected 12 km units as spatial unit for which to model spread. Our analysis was undoubtedly impacted by this selection, but also, by the scales at which the infestation and covariate data were collected. In the context of epidemiological spread models, how the at-risk population, the environment, , and population-environment interaction are represented will undoubtedly impact results. The use of regular and/or square spatial units are not required, and as in Smith et al. (2002) an irregular lattice (such as counties) could be appropriately used with this model. The implementation of an irregular lattice map structure would require careful consideration to the definition of spatial weights. For example, it may be useful to consider the proportion of the boundary associated with infected polygons surrounding an uninfected region as a way to accommodate the spatial structure of infected neighbors in [4]. Alternatively, higher-order spatial weights functions (e.g., using a distance-decay effect) may be useful for quantifying disease pressure in uninfected regions #REF#.

Working within a hierarchical Bayesian framework allows for data at multiple scales/representations to be incorporated. Often the accommodating nature of a hierarchical
Bayes framework is used as a security blanket when tasked with modeling erroneous or sparse datasets. However, inferences resulting from such analyses are still a product of limited datasets (i.e., garbage in – garbage out). Thus care must be taken to utilize a hierarchical Bayes framework in such a way as to maximize the potential learning from available data, while recognizing the limitations of a given dataset. The flexibility of the model presented here, from the basic structure introduced in section 2 to the more complex variations used in the mountain pine beetle example, and further proposed in the discussion, can provide an accommodating framework for modeling many characteristic dual-mechanism (diffusion-translocation) spread processes. A key feature of the proposed model is the incorporation of spatially-varying diffusion parameters, which allow for local differences in rates of spread across the study region, accommodating diffusion across heterogeneous landscapes. Incorporating spatially varying parameter values into a model framework (for example using geographically weighted regression #Fotheringham# or, more broadly, any spatially varying coefficient model #e.g., Waller et al. 2007#) is becoming increasingly popular for examining spatial heterogeneity in a wide range of applications, for example disease mapping (Best et al. 2005), crime rates, (Wheeler and Waller 2009), and housing values (Bitter et al. 2007). The additional complexity of varying parameters over space is no longer a computational burden given modern computing capabilities, and resulting maps of parameter estimates, such as those in Figure 7, can provide material for interesting spatially-specific inference. However, it can be easy to attempt increasingly complex models beyond what is capable of being learned from the data (i.e., over-fitting). This can lead to a variety of problems including high parameter variance and sensitivity #REF#, and poor overall fit. It is up to the researcher to determine what can be realistically learned from the data with respect to spatially varying parameters.
Here, using a convolution model for the diffusion process, we examined spatially
structured ($a$) and non-structured ($h$) error terms (e.g., following the BYM model, Besag, York,
and Mollie 1991). In theory, maps of the non-structured term could be linked to spread barriers,
although in this example the effect was relatively small in magnitude. In ecological examples,
barriers are often a function of physical properties of the landscape (Sharov and Liebhold 1998).
However in epidemiology, where infections are commonly transferred along networks, the
identification of barriers will be more complex as they are related to the connectivity of infected
and susceptible nodes (Eubank et al. 2004, Keeling 2005). The identification and interpretation
of various anomalies (e.g., barriers) within maps of spatially varying parameter estimates can
provide valuable insight into a given process or limitations of a given model. However,
quantifying barriers (whether they are physical objects or properties of the underlying data)
remains a challenging endeavor in various facets of spatial data analysis (Cova and Goodchild
2002).

In many applications, spread processes are impacted by factors varying across space and
time (e.g., environmental, socio-economic). In the area considered in our study, mountain pine
beetle are sensitive to warm august temperatures (Logan and Bentz, 1999), which trigger
emergence and local dispersal. Given sufficient climate data for each spatial and temporal unit, a
spatially- and temporally-varying climate covariate ($c_{ij}$) could be included through simple
modifications to [4] to help characterize diffusion rates associated with beetle sensitivity to
temperature in summer months. In epidemiological problems, a similar term could be associated
with a dynamic space-time covariate associated with the spatial diffusion process. Alternatively
factors associated with the baseline probability of infection (related to translocation here) can be
incorporated. We used a distance-dependent covariate to incorporate the fact that mountain pine
beetle translocation tended to occur proximal to existing infestations. In human disease spread, population mobility has been identified as an important factor in the underlying probability of disease spread (Viboud et al. 2006). Population mobility is dynamic in both space (regional differences) and time (seasonal mobility patterns), and could be represented using a space-time covariate for the baseline probability of infection in [3].

In conclusion, there is a growing demand, in epidemiology as well as ecology, for tools to incorporate a variety of spatially and temporally explicit data sources in a flexible statistical modeling framework in order to study spread processes. As we have demonstrated using simulated datasets along with the case study investigating mountain pine beetle spread in Western Canada, the framework we have proposed affords the ability to generate a finer understanding of how landscape features might affect dispersal mechanisms, while also allowing for unpredictable translocation events. This dual-mechanism (diffusion-translocation) process of spread is characteristic of a wide array of diseases (Green et al., 2006), as well as biological invasions (Andow et al., 1990; Lewis, 1997; Bossenbroek et al., 2001). Finally, the approach taken here presented a novel and insightful method for model-checking; specifically, the use of map comparison for evaluating spatially varying parameter estimates. The development of spatial measures for model evaluation remains an ongoing research problem, and is one area for future work being pursued by the authors.

Acknowledgements

The authors would like to thank the editor and an anonymous referee for comments that improved the presentation of this manuscript. The authors also thank Professor Lance Waller for helpful comments on an initial version of the model developed in this paper. Funding for this work was provided by GEOIDE through the Government of Canada’s Networks for Centres of Excellence program.
References


Logan JA, Bentz BJ. Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. Environ Entomol 1999; 28:924-34.


Riley S. Large-Scale Spatial-Transmission Models of Infectious Disease. Science 2007; 316:1298-301.


Schneider DW, Ellis CD, Cummings KS. A transportation model assessment of the risk to native mussel communities from zebra mussel spread. Conserv Biol 1998; 12: 788-800.


List of Table Captions

Table 1: Model parameters used in each of 6 implemented models, along with DIC results. Model 6 was identified as the best model based on DIC and selected for further analysis.

Table 2: Sensitivity analysis of the prior distribution of the variance parameter for the CAR model (τ_a) used in the model (*) with two alternate selections.
List of Figure Captions

Figure 1: Graph diagramming the modeled relationship between input data, parameters, and hyper-parameters of the Bayesian space-time model for spread processes.

Figure 2: Simulated-estimation approach used in the study. Data were simulated using the model using three patterns for translocation (M) over 100 time-steps and three patterns for diffusion (Λ) onto 20x20, 40x40, and 80x80 study areas. The model was fitted to simulated datasets to obtain estimates of the generating parameters.

Figure 3: Historical range of mountain pine beetle and its preferred host species (lodgepole pine) and a potential new host species (jack pine) with the extent of mountain pine beetle infestation in 2009. Our study area (inset) contains the northern portion of the eastward expansion of mountain pine beetle into the boreal forest. Historical range of mountain pine beetle adapted from Fig. 4 in Safranyik and Carroll (2006). Tree species range maps from Little (1971), available at: http://esp.cr.usgs.gov/data/atlas/little/.

Figure 4: Maps showing: a) year of initial mountain pine beetle infestation across the study area and two spatial covariates used in the model b) mean elevation, and c) percent forest cover.

Figure 5: Posterior predictive checking for scenario Λ2M1 (top) and Λ3M2 (bottom) evaluating the number of newly inhabited cells at each time period. Error bars generated from replicates simulated from 100 draws of the posterior distributions of model
parameters ($\lambda_i, \mu_t$), dots indicate the observed data. Number of times the test statistic (see equation [5]) of the observed data fell outside of the 95% C.I. for the $y_{rep}$ data was: 3 in scenario $\Lambda_2M_1$ and 2 in scenario $\Lambda_3M_2$, indicating excellent model fit.

Figure 6: Posterior predictive checking for mountain pine beetle case study evaluating the number of newly infested cells at each time period. Error bars generated from replicates simulated from 1000 draws of the posterior distributions of model parameters ($\lambda_i, \tilde{\mu}_t$), dots indicate observed data. Note: No Inf. is the number of cells that do not become infested over the study time period.

Figure 7: Maps of the posterior means for a) local diffusion parameter – $\lambda$, b) CAR model effect – $a$, and c) aspatial random effect – $h$; with maps of posterior variance shown below.

Figure 8: Graph of mean posterior of parameter $\theta_t$, with 95% C.I. error bars, depicting annual variation in sporadic translocations events, which relates to passive spread by mountain pine beetle. The peak observed in 2007 corresponds to extensive beetle activity documented in 2006.
Table 2: Sensitivity analysis of the prior distribution of the variance parameter for the CAR model ($\tau_a$) used in the model (*) with two alternate selections.

<table>
<thead>
<tr>
<th>gamma</th>
<th>(0.5, 0.0005)*</th>
<th>(0.001, 0.001)</th>
<th>(0.1, 0.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior Mean</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu[37]$</td>
<td>-4.88</td>
<td>-4.89</td>
<td>-4.90</td>
</tr>
<tr>
<td>$\mu[80]$</td>
<td>-6.76</td>
<td>-6.79</td>
<td>-6.78</td>
</tr>
<tr>
<td>$\lambda[119]$</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>$\lambda[794]$</td>
<td>0.79</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>Posterior Standard Deviation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu[37]$</td>
<td>0.24</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>$\mu[80]$</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$\lambda[119]$</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>$\lambda[794]$</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
</tr>
</tbody>
</table>
Table 1: Model parameters used in each of 6 implemented models, along with DIC results. Model 6 was identified as the best model based on DIC and selected for further analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Parameters</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \lambda )</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>x x x x x x x</td>
<td>6935.5</td>
</tr>
<tr>
<td>2</td>
<td>x x x x x x</td>
<td>5623.0</td>
</tr>
<tr>
<td>3</td>
<td>x x x x x x</td>
<td>5626.2</td>
</tr>
<tr>
<td>4</td>
<td>x x x x x x</td>
<td>5626.1</td>
</tr>
<tr>
<td>5</td>
<td>x x x x x x</td>
<td>5625.7</td>
</tr>
<tr>
<td>*6</td>
<td>x x x x x x x x</td>
<td>*5617.9</td>
</tr>
</tbody>
</table>

* - model with best fit, presented in case study

elev – elevation covariate, %for – forest cover covariate
Figure 1: Graph diagramming the modeled relationship between input data, parameters, and hyper-parameters of the Bayesian space-time model for spread processes.
Figure 2: Simulated estimation approach used in the study. Data were simulated using the model using three patterns for translocation (M) over 100 time-steps and three patterns for diffusion (Λ) onto 20x20, 40x40, and 80x80 study areas. The model was fitted to simulated datasets to obtain estimates of the generating parameters.
Figure 3: Historical range of mountain pine beetle and its preferred host species (lodgepole pine) and a potential new host species (jack pine) with the extent of mountain pine beetle infestation in 2009. Our study area (inset) contains the northern portion of the eastward expansion of mountain pine beetle into the boreal forest. 

1 Historical range of mountain pine beetle adapted from Fig. 4 in Safranyik and Carroll (2006).  

Figure 4: Maps showing: a) year of initial mountain pine beetle infestation across the study area and two spatial covariates used in the model b) mean elevation, and c) percent forest cover.
Figure 5: Posterior predictive checking for scenario $\Lambda_2M_1$ (top) and $\Lambda_3M_2$ (bottom) evaluating the number of newly inhabited cells at each time period. Error bars generated from replicates simulated from 100 draws of the posterior distributions of model parameters ($\lambda_i, \mu_t$), dots indicate the observed data. Number of times the test statistic (see equation [5]) of the observed data fell outside of the 95% C.I. for the $y_{rep}$ data was: 3/100 time periods in scenario $\Lambda_2M_1$ and 2/100 time periods in scenario $\Lambda_3M_2$, indicating excellent model fit.
Figure 6: Posterior predictive checking for mountain pine beetle case study evaluating the number of newly infested cells at each time period. Error bars generated from replicates simulated from 1000 draws of the posterior distributions of model parameters ($\lambda_i$, $\tilde{\mu}_i$), dots indicate observed data. Note: No Inf. is the number of cells that do not become infested over the study time period.
Figure 7: Maps of the posterior means for a) local diffusion parameter – $\lambda$, b) CAR model effect – $a$, and c) aspatial random effect – $h$; with maps of posterior variance shown below.
Figure 8: Graph of mean posterior of parameter $\theta_t$, with 95% C.I. error bars, depicting annual variation in sporadic translocations events, which relates to passive spread by mountain pine beetle. The peak observed in 2007, corresponds to extensive beetle activity documented in 2006.
Appendix A: Global goodness of fit analysis results comparing true and estimated diffusion and translocation parameters for different spread scenarios and spatial scales.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Diffusion</th>
<th></th>
<th></th>
<th>Translocation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std Res</td>
<td>Std Res / n</td>
<td>chi Sq</td>
<td>P-value</td>
<td>Std Res</td>
</tr>
<tr>
<td>Λ_1M_1</td>
<td>82.55</td>
<td>0.21</td>
<td>67.2</td>
<td>0.14</td>
<td>99.44</td>
</tr>
<tr>
<td>Λ_1M_2</td>
<td>117.96</td>
<td>0.29</td>
<td>37.1</td>
<td>0.24</td>
<td>65.98</td>
</tr>
<tr>
<td>Λ_1M_3</td>
<td>65.65</td>
<td>0.16</td>
<td>44</td>
<td>0.31</td>
<td>63.55</td>
</tr>
<tr>
<td>Λ_2M_1</td>
<td>389.77</td>
<td>0.97</td>
<td>60.5</td>
<td>0.13</td>
<td>222.72</td>
</tr>
<tr>
<td>Λ_2M_2</td>
<td>186.93</td>
<td>0.47</td>
<td>39.9</td>
<td>0.3</td>
<td>184.6</td>
</tr>
<tr>
<td>Λ_2M_3</td>
<td>313.87</td>
<td>0.78</td>
<td>31.5</td>
<td>0.49</td>
<td>208.35</td>
</tr>
<tr>
<td>Λ_3M_1</td>
<td>76.12</td>
<td>0.19</td>
<td>77</td>
<td>0.11</td>
<td>80.64</td>
</tr>
<tr>
<td>Λ_3M_2</td>
<td>157.8</td>
<td>0.39</td>
<td>39.9</td>
<td>0.3</td>
<td>2.18</td>
</tr>
<tr>
<td>Λ_3M_3</td>
<td>114.18</td>
<td>0.29</td>
<td>38.9</td>
<td>0.19</td>
<td>11.39</td>
</tr>
</tbody>
</table>

40x40

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Diffusion</th>
<th></th>
<th></th>
<th>Translocation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std Res</td>
<td>Std Res / n</td>
<td>chi Sq</td>
<td>P-value</td>
<td>Std Res</td>
</tr>
<tr>
<td>Λ_1M_1</td>
<td>551.69</td>
<td>0.34</td>
<td>99</td>
<td>0.24</td>
<td>121.69</td>
</tr>
<tr>
<td>Λ_1M_2</td>
<td>518.2</td>
<td>0.32</td>
<td>47.7</td>
<td>0.36</td>
<td>41.85</td>
</tr>
<tr>
<td>Λ_1M_3</td>
<td>580.08</td>
<td>0.36</td>
<td>58.7</td>
<td>0.31</td>
<td>48.65</td>
</tr>
<tr>
<td>Λ_2M_1</td>
<td>863.32</td>
<td>0.54</td>
<td>99</td>
<td>0.24</td>
<td>154.78</td>
</tr>
<tr>
<td>Λ_2M_2</td>
<td>1093.67</td>
<td>0.68</td>
<td>55</td>
<td>0.29</td>
<td>150.77</td>
</tr>
<tr>
<td>Λ_2M_3</td>
<td>1550.23</td>
<td>0.97</td>
<td>66</td>
<td>0.28</td>
<td>204.74</td>
</tr>
<tr>
<td>Λ_3M_1</td>
<td>624.96</td>
<td>0.39</td>
<td>99</td>
<td>0.08</td>
<td>62.18</td>
</tr>
<tr>
<td>Λ_3M_2</td>
<td>584.96</td>
<td>0.37</td>
<td>55</td>
<td>0.29</td>
<td>11.01</td>
</tr>
<tr>
<td>Λ_3M_3</td>
<td>597.65</td>
<td>0.37</td>
<td>58.7</td>
<td>0.31</td>
<td>27.35</td>
</tr>
</tbody>
</table>

80x80

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Diffusion</th>
<th></th>
<th></th>
<th>Translocation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std Res</td>
<td>Std Res / n</td>
<td>chi Sq</td>
<td>P-value</td>
<td>Std Res</td>
</tr>
<tr>
<td>Λ_1M_1</td>
<td>5932.75</td>
<td>0.93</td>
<td>99</td>
<td>0.24</td>
<td>134.11</td>
</tr>
<tr>
<td>Λ_1M_2</td>
<td>7100.85</td>
<td>1.11</td>
<td>66</td>
<td>0.28</td>
<td>44.98</td>
</tr>
<tr>
<td>Λ_1M_3</td>
<td>2111.98</td>
<td>0.33</td>
<td>66</td>
<td>0.28</td>
<td>52.6</td>
</tr>
<tr>
<td>Λ_2M_1</td>
<td>7224.69</td>
<td>1.13</td>
<td>99</td>
<td>0.24</td>
<td>151.21</td>
</tr>
<tr>
<td>Λ_2M_2</td>
<td>3658.6</td>
<td>0.57</td>
<td>66</td>
<td>0.28</td>
<td>88.71</td>
</tr>
<tr>
<td>Λ_2M_3</td>
<td>6446.2</td>
<td>1.01</td>
<td>57.7</td>
<td>0.34</td>
<td>92.83</td>
</tr>
<tr>
<td>Λ_3M_1</td>
<td>3494.29</td>
<td>0.55</td>
<td>99</td>
<td>0.24</td>
<td>130.61</td>
</tr>
<tr>
<td>Λ_3M_2</td>
<td>8238.91</td>
<td>1.29</td>
<td>66</td>
<td>0.28</td>
<td>22.76</td>
</tr>
<tr>
<td>Λ_3M_3</td>
<td>6543.47</td>
<td>1.02</td>
<td>66</td>
<td>0.28</td>
<td>33.26</td>
</tr>
</tbody>
</table>
Appendix B: Standardized residuals for 9 spread scenarios and 3 spatial scales.
Appendix C: Map comparison analysis results comparing estimated diffusion to the true diffusion used to simulate data.

<table>
<thead>
<tr>
<th></th>
<th>Luminance</th>
<th>Contrast</th>
<th>Structure</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_1$</td>
<td>0.951</td>
<td>0.881</td>
<td>0.899</td>
<td>0.754</td>
</tr>
<tr>
<td>$M_2$</td>
<td>0.231</td>
<td>0.739</td>
<td>0.890</td>
<td>-0.026</td>
</tr>
<tr>
<td>$M_3$</td>
<td>0.924</td>
<td>0.864</td>
<td>0.889</td>
<td>0.706</td>
</tr>
<tr>
<td>$\Lambda_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_1$</td>
<td>0.824</td>
<td>0.858</td>
<td>0.897</td>
<td>0.622</td>
</tr>
<tr>
<td>$M_2$</td>
<td>0.214</td>
<td>0.872</td>
<td>0.930</td>
<td>0.064</td>
</tr>
<tr>
<td>$M_3$</td>
<td>0.681</td>
<td>0.870</td>
<td>0.898</td>
<td>0.508</td>
</tr>
<tr>
<td>$\Lambda_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_1$</td>
<td>0.972</td>
<td>0.856</td>
<td>0.720</td>
<td>0.602</td>
</tr>
<tr>
<td>$M_2$</td>
<td>0.903</td>
<td>0.726</td>
<td>0.692</td>
<td>0.454</td>
</tr>
<tr>
<td>$M_3$</td>
<td>0.974</td>
<td>0.827</td>
<td>0.697</td>
<td>0.565</td>
</tr>
</tbody>
</table>