

A Bayesian Space-Time Model for Discrete Spread Processes on a Lattice

Long, Jed A.¹ jlong@uvic.ca*

Robertson, Colin² 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:

Long, JA, C Robertson, FS Nathoo, and TA Nelson. 2012. A Bayesian Space-Time Model for Discrete Spread Processes on a Lattice. *Spatial and Spatial-Temporal Epidemiology*. 3. 151-162.

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).

1 **Abstract**

2 In this article we present a Bayesian Markov model for investigating environmental spread
3 processes. We formulate a model where the spread of a disease over a heterogeneous landscape
4 through time is represented as a probabilistic function of two processes: local diffusion and
5 random-jump dispersal which allows the model to represent the leptokurtic spread pattern typical
6 of many infectious diseases and biological invasions. We demonstrate the properties of this
7 model using a simulation experiment and an empirical case study – the spread of mountain pine
8 beetle in western Canada. Posterior predictive checking was used to validate the number of
9 newly inhabited regions in each time period. Map comparison analysis was used to measure
10 spatial agreement of spatially distributed model parameter estimates and observed values. The
11 model performed well in the simulation study in which a goodness-of-fit statistic measuring the
12 number of newly inhabited regions in each time interval fell within the 95% posterior predictive
13 credible interval in over 97% of simulations. The map comparison analysis revealed that in some
14 cases the magnitude of estimated parameter values differed markedly from the true values, but in
15 all cases an adequate recovery of the spatial structure was obtained, indicating good spatial
16 agreement. The case study of a mountain pine beetle infestation in Western Canada (1999 to
17 2009) extended the base model in two ways. First, spatial covariates thought to impact the local
18 diffusion parameters, elevation and forest cover, were included in the model. Second, a refined
19 definition for translocation or jump-dispersal based on mountain pine beetle ecology was
20 incorporated improving the fit of the model. Posterior predictive checks on the mountain pine
21 beetle model found that the observed goodness-of-fit test statistic fell within the 95% posterior
22 predictive credible interval for 8 out of 10 years. The simulation study and case study provide
23 evidence that the model presented here is both robust and flexible; and is therefore appropriate
24 for a wide array of spread processes in epidemiology and ecology.

25

26 **HIGHLIGHTS**

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

32

33 **KEYWORDS:** space-time binary data, spread process, spatial random effects, mountain pine
34 beetle, map comparison

35

36

37 **1. Introduction**

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

54 Due to the nature of disease surveillance systems which are the primary data sources for
55 disease modeling studies, data are often only available at discrete temporal intervals (e.g.,
56 weeks). Similarly, many pathogens spread via fomites at discrete time periods. For example,
57 marine invasive species such as the zebra mussel (*Dreissena polymorpha*) spread primarily due
58 to recreational boating, which peaks during the summer months (Schneider et al., 1998). Other
59 species disperse naturally at discrete temporal intervals. For instance, bark beetles emerge and

60 disperse on an annual basis (Safranyik & Carroll, 2006). As a result, there is considerable interest
61 in developing discrete-time representations of spread in ecological models.

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

76 The spread of disease is often the result of multiple mechanisms. For example, foot and
77 mouth disease typically spreads among animals and herds via airborne transmission, and among
78 farms and regions by animal movement networks and fomites (Green et al., 2006). Similarly, in
79 many ecological invasions, the resultant pattern of invasion is often multi-causal: locally through
80 diffusion or movement and over greater distances via intermediate species or translocation
81 vectors. Smith et al. (2002) provide a mathematical model for spread processes characterized by
82 both local and random-jump dispersing mechanisms, using the spread of raccoon rabies across a

83 lattice in their example. However, inference in Smith et al. (2002) was based on a stochastic
84 estimation algorithm rather than a formal statistical framework for inference. The spread of
85 raccoon rabies was also examined by Wheeler and Waller (2008) who link spatial variation in
86 patterns of spread (i.e., deviations from a travelling wave) to landscape heterogeneity using
87 spatially-varying regression, and adopt a Bayesian framework for inference.

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

96 Despite the inherent link between the spread of species of interest to ecologists, and
97 disease spread in human populations studied by epidemiologists, models have been developed
98 largely independently in these fields until very recently (Smith et al., 2002). Given that the
99 majority of emerging diseases are zoonotic in nature (Jones et al., 2008), and often of wildlife
100 origin, there exists a need for integrated ecological-epidemiological modeling. In this article we
101 present a hierarchical Bayes approach appropriate for modeling either disease or organism spread
102 across a landscape, and allow for landscape heterogeneity using spatially-varying parameters.
103 Although we investigate a specific ecological application (i.e., mountain pine beetle in western
104 Canada), the methods employed are generic and appropriate for a wide range of spread problems
105 in ecology and epidemiology..

106 In this article, we present a discrete-time Markov model for ecological spread processes
 107 allowing for spatially-varying spread rates and random-jump dispersal. In the next section we
 108 present the model. A simulation study follows, which evaluates model performance. Simulated
 109 datasets at multiple spatial scales, under various spread scenarios, are used to illustrate model
 110 strengths and weaknesses. Next, we report on an empirical case study investigating the spread of
 111 mountain pine beetle (*Dendroctonus ponderosae*) in western Canada. Finally, we discuss
 112 remaining challenges and practical issues related to the process of model development and
 113 conclude by linking this research with potential applications in epidemiology.

114 2. Model Development

115 Given n regions comprising a study area, we formulated a logistic model for a binary
 116 spread process (defined here as inhabited-uninhabited) where it was assumed that newly
 117 inhabited regions do not revert to being uninhabited (Mollison, 1977), and we
 118 let $Z_i(t) \in \{0, 1\}$ indicate the presence of an organism or disease in region i , $i=1, \dots, n$, at time t ,
 119 $t=1, \dots, T$. Here, $t = 1$ corresponds to the initial map of organism or disease presence, and the
 120 vector $\mathbf{Z}(t) = (Z_1(t), \dots, Z_n(t))'$ represents a binary map describing the progression of the
 121 organism or disease at time t . Spread is described through a stochastic process model for $\mathbf{Z}(t)$,
 122 which conditional on model parameters $\boldsymbol{\theta}$, is assumed to follow a first-order Markov
 123 assumption, so that

124 $Pr\{\mathbf{Z}(t)|\mathbf{Z}(t-1), \mathbf{Z}(t-2), \dots, \mathbf{Z}(1), \boldsymbol{\theta}\} = Pr\{\mathbf{Z}(t)|\mathbf{Z}(t-1), \boldsymbol{\theta}\}$ where the Markov transition kernel is
 125 indexed by $\boldsymbol{\theta}$. The Markov model is further simplified by assuming conditional independence
 126 across regions, so that

127 $Pr\{\mathbf{Z}(t)|\mathbf{Z}(t-1), \boldsymbol{\theta}\} = \prod_i Pr\{Z_i(t)|\mathbf{Z}(t-1), \boldsymbol{\theta}\}$. Spatial dependence is accommodated at the second
 128 level of the hierarchical specification with random effects incorporated into $\boldsymbol{\theta}$. The term $p_{it} =$

129 $Pr\{Z_i(t)|Z(t-1), \Theta\}$ represents the probability that the organism or disease is present in region i at
 130 time t , given the presence map $Z(t-1)$ and conditional on Θ . We assumed that regions where the
 131 organism or disease is present remain inhabited, so that $p_{it} = 1$ if $Z_i(t-1) = 1$; whereas, if region
 132 i is free of organism or disease at time $t-1$, so that $Z_i(t-1) = 0$, we assumed a logistic specification
 133 with space varying coefficients

$$134 \quad \log\{p_{it}/(1-p_{it})\} = \mu_t + \lambda_i NN_{[i,t-1]} \quad [1]$$

135 where $NN_{[i,t-1]}$ is the number of inhabited neighbors of region i at time $t-1$; μ_t is a time varying
 136 parameter representing a baseline probability of becoming inhabited; and λ_i is a spatially-varying
 137 parameter quantifying the local impact of inhabited regions on their uninhabited neighbors.
 138 Neighbors for $NN_{[i,t-1]}$ are defined using a Queen's case (Moore neighborhood – i.e., edge or
 139 corner in contact signifies a neighbor) definition of spatial neighbors; however, alternate
 140 neighborhood configurations could easily be explored. The baseline rate μ_t is common to all
 141 regions, and for an uninhabited region with no inhabited neighbors at time $t-1$ (isolated from the
 142 spread wave) we have

$$143 \quad \log\{p_{it}/(1-p_{it})\} = \mu_t \quad [2]$$

144 so that μ_t can be thought of as representing the time-varying probability of translocation events,
 145 describing random-jump movements by a species. Inclusion of terms for both diffusion and
 146 random-jump movements is important when spread occurs via separate mechanisms or at
 147 differing spatial scales. Mountain pine beetles, for example, spread via two independent
 148 mechanisms, actively over short distances (e.g., within forest stands), and passively via
 149 convective wind currents capable of transporting small populations for hundreds of kilometers
 150 (Safranyik & Carroll, 2006).

151 Comparison of equations [1] and [2] reveals that λ_i characterizes the change in the log-
 152 odds of habitation for region i , arising from the presence of organisms in neighboring regions.
 153 Larger values of λ_i correspond to an increasing probability of spread from inhabited regions to
 154 uninhabited neighbors. As such, λ_i controls the rate of organism mobility or diffusion into region
 155 i , and the vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$ characterizes spatial variability in diffusion across the entire
 156 study area.

157 Using a hierarchical modeling approach, we allowed for temporal variation in the
 158 translocation parameters $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)'$ and spatial variation in diffusion parameters $\boldsymbol{\lambda} = (\lambda_1, \dots,$
 159 $\lambda_n)'$ using mixed model random effect specifications. The translocation component $\boldsymbol{\mu}$, is
 160 composed of a constant μ_c , coupled with time-varying mean-zero effects θ_t , so that

$$161 \quad \mu_t = \mu_c + \theta_t \quad [3]$$

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

166 The spatially-varying diffusion parameters $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$ are modeled using a
 167 convolution prior

$$168 \quad \lambda_i = \alpha_0 + h_i + a_i \quad [4]$$

169 where α_0 represents the baseline level of spread across the study area; $h_i \sim N(0, \tau_h)$ are
 170 independent and identically distributed random effects representing spatially unstructured
 171 variation; and a_i is a spatially correlated random effect, with the vector $\boldsymbol{a} = (a_1, \dots, a_n)'$ modeled
 172 using an intrinsic conditional autoregressive model – CAR(τ_a). This random effect formulation
 173 follows Besag, York and Mollie (1991), who suggest including both independent and spatially

174 correlated random effects.. Additional terms corresponding to spatially varying covariates can be
175 easily incorporated in [4] to investigate relationships between covariates and the local rate of
176 diffusion. The spatial CAR model for a_i uses a Queen's case (i.e., adjacency) definition of
177 neighbors. In addition, a binary definition of weights is used, with neighbors coded as 1 and non-
178 neighbors as 0. Finally, our model specification is made complete by assigning a flat prior to α_0 ,
179 and weakly informative *inverse-gamma*(0.5,0.0005) hyper-priors for the variance components τ_h
180 and τ_a .

181 The vector Θ is the set of parameters in our model, with $\Theta = \{\mathbf{h}, \mathbf{a}, \boldsymbol{\theta}, \alpha_0, \mu_c, \tau_a, \tau_h, \tau_\theta\}$.
182 With this specification, Bayesian inference for Θ is based on the posterior distribution
183 $[\Theta|\mathbf{Z}(1), \dots, \mathbf{Z}(T)]$, where $\mathbf{Z}(1), \dots, \mathbf{Z}(T)$ are binary data vectors representing a realized ecological
184 spread process. The posterior distribution is computed using Markov chain Monte Carlo methods
185 implemented in the free software WinBUGS (Lunn et al., 2000). The model code used for fitting
186 this model can be obtained from the first author upon request.

187

188 **3. Model Evaluation Using a Simulation Study**

189 *3.1 Simulation Study Data*

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

196 simulation-estimation approach to investigate the sensitivity of the model to changes in model
197 parameter values and spatial scale.

198 Spread datasets were simulated from a set of patterns representing realistic scenarios of
199 spatial diffusion (Figure #a) and time-varying translocation (Figure #b), in combination with
200 three different spatial scales. Spatial diffusion represented as a linear trend (Λ_1) can be thought to
201 represent a simplified spread process across a more homogeneous landscape, this scenario might
202 be encountered, for example, in the presence of a latitudinal gradient. The Gaussian random field
203 scenario (Λ_2) is an example of a more complex, heterogeneous landscape, characteristic of a
204 wide range of spread processes. Constant translocation (M_1) represents a static level of random
205 translocation events. Linear decreasing translocation (M_2) represents a situation where a spread
206 mechanism (e.g., fomite transmission of foot and mouth disease) is decreasing over time as
207 control measures are put in place. Finally, oscillating translocation (M_3) represents a seasonal
208 cycle to translocation, as in the case of wind-driven transport which depends primarily on
209 seasons. Variation in spatial scale (20x20, $n = 400$; 40x40, $n = 1600$; 80x80, $n = 6400$) is used to
210 examine the impact of scale on model performance. Specifically, we are interested in how
211 changes in the grain (resolution) of the data may affect results. Extent, the other aspect of spatial
212 scale, is of less interest here, as we assume the extent of the disease is contained by the study
213 area.

214 < approximate location for Figure #a,b >The simulation-estimation procedure involved
215 three steps: 1) generate a realistic spread scenarios using combinations of diffusion (Λ),
216 translocation (M), and spatial scale (see Figure #); 2) simulate spread data from the scenarios;
217 and 3) estimate the model based on the simulated data. Each combination of spatial diffusion
218 pattern, time-varying translocation pattern, and spatial scale comprised one spread scenario, for a

219 total of 18 different scenarios (the simulated spread datasets for the 6 scenarios at the 40x40
220 scale are shown in Figure ##).

221 < approximate location for Figure ## >

222 3.2 Examining Model Fit

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

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

$$239 \quad Inh_t = \sum_{i=1}^n Z_i(t) - Z_i(t-1) \quad [5]$$

240 Two of the simulated datasets were selected for checking based on the results of the parameter
241 estimation recovery analysis so that one of the better performing scenarios, and one of the poorer

242 performing scenarios were evaluated, both using the 40x40 spatial scale to ensure comparability.
243 In each case the test statistic was evaluated at each time point to determine if a value of Inh_t
244 computed for the data falls outside the main mass of the corresponding posterior predictive
245 distribution, in which case we have evidence that the model does not fit this aspect of the data.

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

263
$$SSIM(i) = [l(i)]^\alpha \cdot [c(i)]^\beta \cdot [s(i)]^\gamma \quad [6]$$

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

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

279 3.3. Simulation Study Results

280 The global goodness of fit analysis for the 27 different simulated spread scenarios are
281 reported in Table #. These global chi-square tests reveal that in none of the scenarios were the
282 estimated parameters significantly different than the true values. The standardized residuals
283 demonstrate the effects of changes in the pattern of spread and spatial scale on diffusion
284 (Appendix B). Residuals tended to increase with larger study areas. For the 20x20 and 40x40
285 spatial scales, the Λ_2 spatial trend produced larger error than the Λ_1 spatial trend or the Λ_3
286 Gaussian random field; however error was largely similar for all three patterns for the 80x80

287 study area. For translocation, the opposite general pattern holds, with larger residuals for smaller
288 datasets. This is likely due to the lack of available cells for translocation to occur at smaller
289 scales. The interdependency between diffusion and translocation is illustrated in Appendix B.
290 Interdependency is clearly impacted by spatial scale, with fairly similar patterns in residuals at
291 the 20x20 scale, and less so at larger scales.

292 < approximate location for Table # >

293 Analysis of model fit using posterior predictive checks based on the statistic Inh_t is
294 presented in Figure 5. For datasets simulated from scenario Λ_2M_3 and scenario Λ_3M_2 , the test
295 statistic fell within the 2.5 and 97.5 percentiles (95% credible interval), 97 and 98 times (out of
296 100) respectively, indicating a very good fit to the data. For the simulated data, the model
297 appears to capture the timing of newly inhabited cells well.

298 < approximate location for Figure 5 >

299 Map comparison of estimated (λ) and expected (Λ) diffusion maps (40x40 spatial scale)
300 revealed different trends from purely aspatial measures reported in Table ##. Estimated λ maps
301 associated with M_2 showed the lowest similarity in all three Λ scenarios. In the case of scenarios
302 Λ_1M_2 and Λ_2M_2 , mean SSIM values were extremely low (-0.026 and 0.064 respectively)
303 indicating poor model fit. These low SSIM values can be attributed to low scores in the
304 luminance component (0.231 and 0.214 respectively). Map similarity was considerably higher in
305 the other scenarios, with a maximum of 0.754 for scenario Λ_1M_1 .

306 < approximate location for Table ## >

307 Finally, model sensitivity to hyper-priors on variance parameters of the spatial CAR
308 component (τ_a) revealed the model inference to be robust to the prior forms considered. Effects
309 of changes in these hyper-priors on point estimates of diffusion and translocation are outlined in

310 Table 2. For both diffusion and translocation, changes in parameter estimates ranged from 0 to
311 0.03 for posterior means, and 0 to 0.02 for posterior standard deviations.

312 < approximate location for Table 2 >

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

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

328 *4.1. Background*

329 Mountain pine beetle is the most destructive biotic agent of mature pine forests in
330 western North America (Safranyik & Carroll, 2006). Endemic to this region, mountain pine
331 beetles typically attack weakened pine trees scattered throughout the forest. Periodically, when
332 favorable conditions manifest, mountain pine beetle populations escalate to epidemic levels,

333 causing mortality to mature pine trees covering thousands of hectares (Safranyik & Carroll,
334 2006). Originating around 1998, the current outbreak is the largest on record and has devastated
335 western Canada's forest industry, causing substantial timber losses (British Columbia Ministry of
336 Forests and Range, 2007). A warming climate combined with forest fire suppression has resulted
337 in an overabundance of mature lodgepole pine (*Pinus contorta*) trees on the landscape. As
338 lodgepole pine are the preferred host of mountain pine beetle, the combined effect of a warming
339 climate and forest fire suppression is listed as probable cause for the magnitude of the current
340 outbreak (Carroll et al., 2006).

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

352 < approximate location for Figure 3 >

353 The objective of this case study is to use the proposed Bayesian spread model to learn
354 about processes governing the spread of mountain pine beetle at the boundary of its historical
355 spatial range. Mountain pine beetle spread is of two types; active and passive spread, which

356 facilitate movement of beetle populations. Active spread represents spatially local dispersal
357 events where beetles fly within or between neighboring pine stands and is the principle mode of
358 spread (Safranyik et al., 1992). Mountain pine beetles are also capable of passive spread whereby
359 beetles are carried long distances via convective wind currents during periods of emergence
360 (Shepherd, 1966; Furniss and Furniss, 1972; Ainslie and Jackson, 2011) . The model we have
361 developed captures active spread through the spatially local diffusion parameters – λ , and passive
362 spread through the temporally stochastic translocation parameters – μ . We hope to gain insight
363 into mountain pine beetle spread during the current epidemic by interpreting spatial variation in
364 λ , and temporal variation in μ .

365 4.2. Data and Study Area

366 Mountain pine beetle infestation data were obtained from the British Columbia Ministry
367 of Forests and Range¹ and the Alberta Department of Sustainable Resource Development² for
368 each year of our study (1999 – 2009). These infestation data are primarily obtained through
369 aerial overview surveys, but also in situ measurements and remotely sensed data sources.

370 Mountain pine beetle emergence occurs during a short one month window during the summer in
371 the study area. As such, the spread process can be measured at discrete (i.e., annual) intervals.
372 Infestation events are represented as both points (indicating a small cluster of infested trees) and
373 polygons (a large area of infestation).

374 We selected a rectangular study area that covers the northern portion of the eastward
375 expansion by mountain pine beetle into the province of Alberta (inset Figure 3). A 12 km grid
376 was demarcated across the study area, generating $n = 2310$ contiguous spatial units. Similar 12
377 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>

378 beetle outbreak in British Columbia (Aukema et al., 2008; Zhu et al., 2008) and spatial
379 synchrony within the current outbreak (Aukema et al., 2006).

380 We were also interested in investigating relationships between mountain pine beetle
381 spread (Figure 4a) and environmental factors. Two spatial covariates, elevation and forest cover
382 (see Figure 4 b and c), were identified in the literature as important in governing local spread of
383 mountain pine beetle. Elevation was taken as the mean of elevation values within each spatial
384 unit using a fine grain elevation dataset (spatial resolution of 25 m). Percent forest cover for each
385 spatial unit was determined using a national land cover database (Wulder et al., 2008). These
386 spatial covariates were incorporated into the model for λ [4], and relate to local diffusion so that
387 [4] becomes

$$388 \lambda_i = \alpha_0 + X_i \beta + h_i + a_i \quad [7]$$

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

391 < approximate location for Figure 4 >

392 Exploratory spatial analysis revealed that mountain pine beetle translocation events
393 exhibited distance-dependence, whereby translocation events occur more frequently proximal to
394 previously infested regions. This phenomenon is commonly associated with the characteristic
395 leptokurtic pattern of spread (e.g., in animal-borne diseases #Ferguson, Lindstrom#, human
396 diseases #REF#, and with invading organisms, Lewis 1997) whereby translocation events are
397 distance-dependent relative to the spread wave. . In this scenario, extremely long distance
398 translocation events are rare, but still possible. To account for this effect, we considered a more
399 general model for the translocation component that more appropriately resembles this distance-
400 dependant spread process. The new distance-dependent translocation parameter is defined as:

401
$$\tilde{\mu}_i = \mu_c + \gamma d_{it} + \theta_t \quad [8]$$

402 where $\tilde{\mu}_i$ is a spatial and temporally varying translocation parameter, d_{it} is the distance (centroid
403 to centroid) from cell i to the nearest infested region at time t , with coefficient γ , and μ_c and θ_t are
404 as defined in [3]. This treats the distance values (d_{it}) as a space-time covariate modulating the
405 translocation component.

406 *4.3. Model Implementation*

407 Several variations of the model were implemented incorporating different parameters
408 (Table 1). For each model, two MCMC chains were run to fit the model. Convergence was
409 assessed following Brooks and Gelman (1998), and a conservative burn-in of 10 000 iterations
410 was selected. Following burn-in, 20 000 samples from each chain were retained for inference.
411 Model selection was based on the deviance information criterion (DIC), which combines the
412 deviance with a penalty for model complexity (Spiegelhalter et al., 2002). Posterior mean,
413 variance, and 95% equal-tailed credible intervals were used to summarize the posterior
414 distributions.

415 *4.4. Case Study Results*

416 Variation in DIC between model 1 and 2 was substantial, indicating that the inclusion of
417 a distance-dependent translocation term ($\tilde{\mu}_i$ from [8]) improved the model (see Table 1). As
418 such, the four subsequent model specifications all use the $\tilde{\mu}_i$ definition from [8]. The variation
419 in DIC between models that use the $\tilde{\mu}_i$ definition for translocation (models 2 – 6) was small
420 (Table 1). The addition of the aspatial random effect parameter (\mathbf{h}) had little effect on the DIC.
421 However we include \mathbf{h} in subsequent model specifications as it can be used to interpret variation
422 in mountain pine beetle spread not captured by the smoothing effect of the CAR model and/or λ
423 covariates; which may relate to barriers to spread. Model 6, the most complex model tested

424 (including both the elevation and forest cover covariates) resulted in the lowest DIC value, and
425 forms the basis for further discussion.

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

434 < approximate location for Figure 6 >

435 A negative relationship (posterior mean = -0.153, 95% C.I. = [-0.337, 0.019]) was
436 observed between local diffusion rate and the elevation covariate. A negative relationship
437 between mountain pine beetle and elevation in British Columbia has been previously reported,
438 and is believed to be linked to elevational constraints on pine species, the beetles preferred host
439 (Aukema et al., 2008; Zhu et al., 2008). This relationship does not necessarily apply east of the
440 Rocky Mountains and may be reason that this relationship is rather weak (e.g., the 95% credible
441 interval covers zero). The rugged topography of the Rocky Mountains historically provided a
442 physical barrier to eastward expansion by mountain pine beetle, with only a few small cases of
443 infestation observed east of the Rockies (see Cerezke, 1989). The current epidemic has breached
444 this barrier, and continued eastward expansion by mountain pine beetle through the boreal will
445 not be hindered by topography as it is comparably flat.

446 A positive relationship (posterior mean = 0.2681, 95% C.I. = [0.130, 0.406]) was
447 identified between local diffusion rate and forest cover. In British Columbia, elevation had
448 previously been used as a surrogate for forest cover information, as here lodgepole pine is found
449 primarily at lower elevations. As previously mentioned, east of British Columbia topography is
450 far less variable, and pine species are found throughout the range of elevations within the boreal
451 forest. As mountain pine beetle continues its eastward expansion, variables that more
452 appropriately represent the availability of suitable pine hosts directly will be most useful for
453 predicting infestation.

454 Using a map of the spatially-varying diffusion parameter (λ) we can highlight regional
455 variability in the rate of local spread (Figure 7a). Comparing to the time of infestation map
456 (Figure 4a) we can clearly see that high λ values are found immediately east of the originally
457 infested regions. Here due to the rugged topography, mountain pine beetle spread quickly along
458 linear forest tracts in valley bottoms as has been previously demonstrated (Robertson et al.,
459 2009). Caution should be taken interpreting λ values in regions where no infestation has occurred
460 (such as in the most eastern portion of our study area). Here the model infers a continuous λ
461 surface from few relevant λ measurements and posterior variance is highest (Figure 7d).

462 < approximate location for Figure 7 >

463 The map of the aspatial random effect parameters (h), can be used to identify regions
464 where the smoothing of the CAR effect (Figure 7b) over-estimates (negative values) or under-
465 estimates (positive values) the rate of local diffusion (Figure 7c). Regions over-estimated
466 (negative values in Figure 7c) by the CAR portion of the model are likely due to abrupt changes
467 in land cover resistant to mountain pine beetle (such as large lakes or mountain peaks). This
468 interpretation may provide a valuable tool for examining landscape barriers on spread where

469 representation of barriers is discrete rather than continuous. Reason for areas under-estimated
470 (positive values in Figure 7c) may be due to the topographic effect mentioned earlier. However,
471 the magnitude of the h effects are quite small (-0.007 to 0.005). Given that the h effect is quite
472 small, the λ maps are therefore predominantly associated with a combination of the CAR effect
473 (α) and the environmental covariates ($X\beta$).

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

485 < approximate location for Figure 8 >

486 With some ecological invasions, barriers may be introduced as a management tactic,
487 effective at slowing the spread of an invading species (Sharov and Liebhold, 1998). In Western
488 Canada, clear-cut harvests and controlled forest fires have been implemented as barriers to
489 mountain pine beetle spread through the removal of large, contiguous sections of potential host
490 trees. In British Columbia, the voracity of the current epidemic has circumvented any mitigation
491 efforts (Wilson 2004), however in Alberta there is still hope that these, and other preventative

492 measures will prove successful at stopping future eastward beetle spread. Preventing continued
493 expansion by mountain pine beetle will be challenging given evidence that the beetles
494 reproductive success *improves* in lodgepole pine stands outside of its historical range (Cudmore
495 et al., 2010). Further, from our analysis it is clear that east of the Rocky mountains, mountain
496 pine beetle spread, like many other spread processes (e.g., Suarez et al., 2001), is dominated by
497 translocation events. When translocation dominates ecological invasions, the organism is often
498 able to jump spread barriers rendering them ineffective. In such cases, it is necessary to carefully
499 evaluate whether the introduction of barriers will provide the intended ecosystem and economic
500 benefits (Sharov and Liebhold, 1998).

501

502 **5. Discussion & Conclusions**

503

504 The validation process we adopt is an example of how both aspatial and spatial indices
505 can be incorporated into the validation procedure. Following guidelines of Gelman et al. (2004),
506 we use posterior predictive checking as an aspatial measure of model fit that can be used even
507 when true parameter values are unknown. This posterior predictive check revealed that our
508 model is sufficient at recovering the timing of new infections. In the simulation study, we were
509 able to complement this aspatial technique with a map comparison analysis (SSIM - Wang et al.,
510 2004) to assess spatial structure of λ values. Map comparison analysis revealed that in some
511 cases estimated λ values were different from expected in magnitude, but that in general the
512 spatial pattern of λ values was retrieved. The SSIM method enables creation of maps of local
513 differences in mean, variance, and covariance, providing information on the spatial structure and
514 differences in each (although we did not include such maps in this presentation). This spatial

515 approach to model evaluation represents a relatively simple procedure that can be easily
516 implemented with existing models providing valuable and unique insight on how spatial
517 structure of parameters relate to model performance. The SSIM measure was originally designed
518 for evaluating image compression algorithms, and only recently has been proposed as a useful
519 measure for quantitative comparison of continuous-value maps #Hagen-Zanker#. Thus,
520 improving our understanding of how the SSIM (or other similar statistics) can be used as spatial
521 measures of model evaluation remains an ongoing endeavor. .

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

536 Working within a hierarchical Bayesian framework allows for data at multiple
537 scales/representations to be incorporated. Often the accommodating nature of a hierarchical

538 Bayes framework is used as a security blanket when tasked with modeling erroneous or sparse
539 datasets. However, inferences resulting from such analyses are still a product of limited datasets
540 (i.e., garbage in – garbage out). Thus care must be taken to utilize a hierarchical Bayes
541 framework in such a way as to maximize the potential learning from available data, while
542 recognizing the limitations of a given dataset. The flexibility of the model presented here, from
543 the basic structure introduced in section 2 to the more complex variations used in the mountain
544 pine beetle example, and further proposed in the discussion, can provide an accommodating
545 framework for modeling many characteristic dual-mechanism (diffusion-translocation) spread
546 processes.. A key feature of the proposed model is the incorporation of spatially-varying
547 diffusion parameters, which allow for local differences in rates of spread across the study region,
548 accommodating diffusion across heterogeneous landscapes. Incorporating spatially varying
549 parameter values into a model framework (for example using geographically weighted regression
550 #Fotheringham# or, more broadly, any spatially varying coefficient model #e.g., Waller et al.
551 2007#) is becoming increasingly popular for examining spatial heterogeneity in a wide range of
552 applications, for example disease mapping (Best et al. 2005), crime rates, (Wheeler and Waller
553 2009), and housing values (Bitter et al. 2007). The additional complexity of varying parameters
554 over space is no longer a computational burden given modern computing capabilities, and
555 resulting maps of parameter estimates, such as those in Figure 7, can provide material for
556 interesting spatially-specific inference. However, it can be easy to attempt increasingly complex
557 models beyond what is capable of being learned from the data (i.e., over-fitting). This can lead to
558 a variety of problems including high parameter variance and sensitivity #REF#, and poor overall
559 fit. It is up to the researcher to determine what can be realistically learned from the data with
560 respect to spatially varying parameters,

561 Here, using a convolution model for the diffusion process, we examined spatially
562 structured (a) and non-structured (h) error terms (e.g., following the BYM model, Besag, York,
563 and Mollie 1991). In theory, maps of the non-structured term could be linked to spread barriers,
564 although in this example the effect was relatively small in magnitude. In ecological examples,
565 barriers are often a function of physical properties of the landscape (Sharov and Liebhold 1998).
566 However in epidemiology, where infections are commonly transferred along networks, the
567 identification of barriers will be more complex as they are related to the connectivity of infected
568 and susceptible nodes (Eubank et al. 2004, Keeling 2005). The identification and interpretation
569 of various anomalies (e.g., barriers) within maps of spatially varying parameter estimates can
570 provide valuable insight into a given process or limitations of a given model. However,
571 quantifying barriers (whether they are physical objects or properties of the underlying data)
572 remains a challenging endeavor in various facets of spatial data analysis (Cova and Goodchild
573 2002).

574 In many applications, spread processes are impacted by factors varying across space and
575 time (e.g., environmental, socio-economic). In the area considered in our study, mountain pine
576 beetle are sensitive to warm august temperatures (Logan and Bentz, 1999), which trigger
577 emergence and local dispersal. Given sufficient climate data for each spatial and temporal unit, a
578 spatially- and temporally-varying climate covariate (c_{ij}) could be included through simple
579 modifications to [4] to help characterize diffusion rates associated with beetle sensitivity to
580 temperature in summer months. In epidemiological problems, a similar term could be associated
581 with a dynamic space-time covariate associated with the spatial diffusion process. Alternatively
582 factors associated with the baseline probability of infection (related to translocation here) can be
583 incorporated. We used a distance-dependent covariate to incorporate the fact that mountain pine

584 beetle translocation tended to occur proximal to existing infestations. In human disease spread,
585 population mobility has been identified as an important factor in the underlying probability of
586 disease spread (Viboud et al. 2006). Population mobility is dynamic in both space (regional
587 differences) and time (seasonal mobility patterns), and could be represented using a space-time
588 covariate for the baseline probability of infection in [3].

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

602

603 **Acknowledgements**

604 The authors would like to thank the editor and an anonymous referee for comments that
605 improved the presentation of this manuscript. The authors also thank Professor Lance Waller for
606 helpful comments on an initial version of the model developed in this paper. Funding for this
607 work was provided by GEOIDE through the Government of Canada's Networks for Centres of
608 Excellence program.
609

References

- Ainslie B, Jackson PL. Investigation into mountain pine beetle above-canopy dispersion using weather radar and an atmospheric dispersion model. *Aerobiol* 2011; 27:51-65.
- Andow DA, Kareiva PM, Levin SA, Okubo A. Spread of invading organisms. *Landsc Ecol* 1990; 4:177-88.
- Aukema BH, Carroll AL, Zheng Y, Zhu J, Raffa KF, Moore RD., et al. Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate. *Ecography* 2008; 31:348-58.
- Aukema BH, Carroll AL, Zhu J, Raffa KF, Sickley TA, Taylor SW. Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak. *Ecography* 2006; 29:427-41.
- Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; 43:1-59.
- Bossenbroek JM, Kraft CE, Nekola JC. Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecol Appl* 2001; 11:1778-88.
- British Columbia Ministry of Forests and Range. Timber supply and the mountain pine beetle infestation in British Columbia: 2007 update. British Columbia Ministry of Forests and Range, Forest Analysis and Inventory Branch, Victoria, BC 2007; 38p.
- Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J of Comp and Graph Stat* 1998; 7:434-55.
- Carroll AL. Personal Communication with T.A. Nelson. 2010.
- Carroll AL, Regniere J, Logan JA, Taylor SW, Bentz BJ, Powell JA. Impacts of climate change on range expansion by the mountain pine beetle. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Mountain Pine Beetle Initiative Working Paper 2006-14, Victoria, BC 2006. 27p.
- Cerezke HF. Mountain pine beetle aggregation semiochemical use in Alberta and Saskatchewan, 1983-1987. Proceedings-Symposium on the management of lodgepole pine to minimize losses to mountain pine beetle (ed.s G.D. Amman), 1989. USDA Forest Service, Intermountain Research Station, Ogden, UT. INT-GTR-262, July 12-14, 1988, Kalispell, Montana; 108-13.
- Cerezke HF. Egg gallery, brood production, and adult characteristics of mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Scolytidae), in three pine hosts. *Can Entomol* 1995; 127: 955-65.
- Clark JS. Why environmental scientists are becoming Bayesians. *Ecol Lett* 2005; 8: 2-14.
- Coetzee JA, Hill MP, Schlange D. Potential spread of the invasive plant *Hydrilla verticillata* in South Africa based on anthropogenic spread and climate suitability. *Biol Invasions* 2009; 11:801-12.
- Cova TJ, Goodchild MF. Extending geographical representation to include fields of spatial objects. *Int J Geogr Inf Sci* 2002; 16:509-32.

- Cressie N, Calder CA, Clark JS, Ver Hoef JM, Wikle CK. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol Appl* 2009; 19:553-70.
- Cudmore TJ, Bjorklund N, Carroll AL, Lindgren BS. Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naive host tree populations. *J Appl Ecol* 2010; 47:1036-43.
- Davis MB, Shaw RG. Range shifts and adaptive responses to Quaternary climate change. *Science* 2001; 292:673-9.
- Furniss MM, Furniss R. Scolytids (Coleoptera) on snowfields above timberline in Oregon and Washington. *Can Entomol* 1972; 104:1471-7.
- Furniss MM, Schenk JA. Sustained natural infestation by the mountain pine beetle in seven new *Pinus* and *Picea* hosts. *J Econ Entomol* 1969; 62:518-9.
- Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. New York, NY: Chapman & Hall, CRC Press; 2004.
- Gelman A, Price PN. All maps of parameter estimates are misleading. *Stat Med* 1999; 18:3221-34.
- Gibson GJ, Otten W, Filipe JA, Cook A, Marion G, Gilligan CA. Bayesian estimation for percolation models of disease spread in plant populations. *Stat Comp* 2006; 16:391-402.
- Green DM, Kiss IZ, Kao RR. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proc Roy Soc B* 2006; 273: 2729-35.
- Hooten MB, Anderson J, Waller LA. Assessing North American influenza dynamics with a statistical SIRS model. *Spat Spat-Temp Epidemiol* 2010; 1:177-85.
- Hooten MB, Wikle CK. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. *Environ Ecol Stat* 2006; 15:59-70.
- Kiss IZ, Green DM, Kao RR. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. *J Roy Soc Interface* 2006; 3:669-77.
- Lewis MA. Variability, patchiness, and jump dispersal in the spread of an invading population. In: Tilman D, Kareiva PM, editors. *Spatial ecology: The role of space in population dynamics and interspecific interactions*. Princeton, NJ: Princeton University Press; 1997. p. 46-74.
- Little EL. Atlas of United States Trees: Volume 1, Conifers and Important Hardwoods. U.S. Department of Agriculture, Miscellaneous Publication 1146 1971; 9p:200 maps.
- Logan JA, Bentz BJ. Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. *Environ Entomol* 1999; 28:924-34.
- Logan JA, Powell JA. Ghost forests, global warming, and the mountain pine beetle (Coleoptera: Scolytidae). *Am Entomol* 2001; 47:160-73.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comp* 2000; 10:325-37.
- McCarthy MA, Masters P. Profiting from prior information in Bayesian analyses of ecological data. *J Appl Ecol* 2005; 42:1012-9.
- Mollison D. Spatial contact models for ecological and epidemic spread. *J Roy Stat Soc B* 1977; 39:283-326.

- Morse SS. Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1995; 1:7-15.
- Parmesan C, Yohe G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 2003; 421:37-42.
- Pearson RG, Dawson TP. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob Ecol Biogeogr* 2003; 12:361-71.
- Pitt JP, Worner SP, Suarez AV. Predicting Argentine ant spread over the heterogeneous landscape using a spatially explicit stochastic model. *Ecol Appl* 2009; 19:1176-86.
- Pontius Jr. RG. Quantification error versus location error in comparison of categorical maps. *Photogr Eng Remote Sens* 2000; 66:1011-6.
- Ricciardi A. Are modern biological invasions an unprecedented form of global change? *Conserv Biol* 2007; 21:329-36.
- Riley S. Large-Scale Spatial-Transmission Models of Infectious Disease. *Science* 2007; 316:1298-301.
- Robertson C, Nelson TA, Jelinski DE, Wulder MA, Boots B. Spatial-temporal analysis of species range expansion: the case of the mountain pine beetle, *Dendroctonus ponderosae*. *J of Biogeogr* 2009; 36:1446-58.
- Safranyik L, Carroll AL. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In: Safranyik L, Wilson B, editors. The mountain pine beetle, a synthesis of biology, management, and impacts on lodgepole pine. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada; 2006. p. 3-66.
- Safranyik L, Carroll AL, Regniere J, Langor DW, Riel WG, Shore TL, et al. Potential for range expansion of mountain pine beetle into the boreal forest of North America. *Can Entomol* 2010; 142:415-42.
- Safranyik L, Linton D, Silversides R, McMullen L. Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand. *J Appl Entomol* 1992; 113: 441-50.
- 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.
- Sharov AA, Liebhold AM. Bioeconomics of managing the spread of exotic pest species with barrier zones. *Ecol Appl* 1998; 8:833-45.
- Shepherd RF. Factors influencing the orientation and rates of activity of *Dendroctonus ponderosae* Hopkins (Coleoptera: Scolytidae). *Can Entomol* 1966; 98:507-18.
- Smith DL, Lucey B, Waller LA, Childs JE, Real LA. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc Nat Acad Sci* 2002; 99:3668-72.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J Roy Stat Soc B* 2002; 64:583-639.
- Stahl K, Moore RD, McKendry IG. Climatology of winter cold spells in relation to mountain pine beetle mortality in British Columbia, Canada. *Clim Res* 2006; 32: 13-23.

- Suarez AV, Holway DA, Case TJ. Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proc Nat Acad of Sci* 2001; 98:1095-100.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 2004; 13:600-11.
- Wheeler DC, Waller LA. Mountains, valleys, and rivers: The transmission of raccoon rabies over a heterogeneous landscape. *J Agric Biol Env Stat* 2008; 13:388-406.
- Wikle CK. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* 2003; 84:1382-94.
- Wilson B. An overview of the Mountain Pine Beetle Initiative. In: Shore T, Brooks JE, Stone JE, editors. *Proceedings of the mountain pine beetle symposium: challenges and solutions*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre. Information Report BC-X-399, October 30-31, 2003, Kelowna, BC; 2004.
- Wulder MA, White JC, Cranny M, Hall RJ, Luther JE, Beaudoin A, et al. Monitoring Canada's forests - Part 1: Completion of the EOSD land cover project. *Can J Rem Sens* 2008; 34:549-62.
- Zhu J, Zheng Y, Carroll AL, Aukema BH. Autologistic regression analysis of spatial-temporal binary data via Monte Carlo maximum likelihood. *J Agric Biol EnvStat* 2008; 13:84-98.

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. ¹ Historical range of mountain pine beetle adapted from Fig. 4 in Safranyik and Carroll (2006). ² 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 Λ_2M_1 (top) and Λ_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 (λ_i, μ_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 Λ_2M_1 and 2 in scenario Λ_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 – λ , 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 θ_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 (τ_a) used in the model (*) with two alternate selections.

gamma	(0.5, 0.0005)*	(0.001, 0.001)	(0.1, 0.1)
Posterior Mean			
$\mu[37]$	-4.88	-4.89	-4.90
$\mu[80]$	-6.76	-6.79	-6.78
$\lambda[119]$	-0.02	0.01	0.00
$\lambda[794]$	0.79	0.80	0.81
Posterior Standard Deviation			
$\mu[37]$	0.24	0.24	0.25
$\mu[80]$	0.70	0.70	0.70
$\lambda[119]$	0.30	0.30	0.30
$\lambda[794]$	0.27	0.27	0.27

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.

Model	Model Parameters								DIC
	λ					μ			
	α_0	a	h	$elev$	$\%for$	μ_{const}	θ	d_{ij}	
1	x	x				x	x		6935.5
2	x	x				x	x	x	5623.0
3	x	x	x			x	x	x	5626.2
4	x	x	x	x		x	x	x	5626.1
5	x	x	x		x	x	x	x	5625.7
*6	x	x	x	x	x	x	x	x	*5617.9

* - model with best fit, presented in case study

elev – elevation covariate, *%for* – forest cover covariate

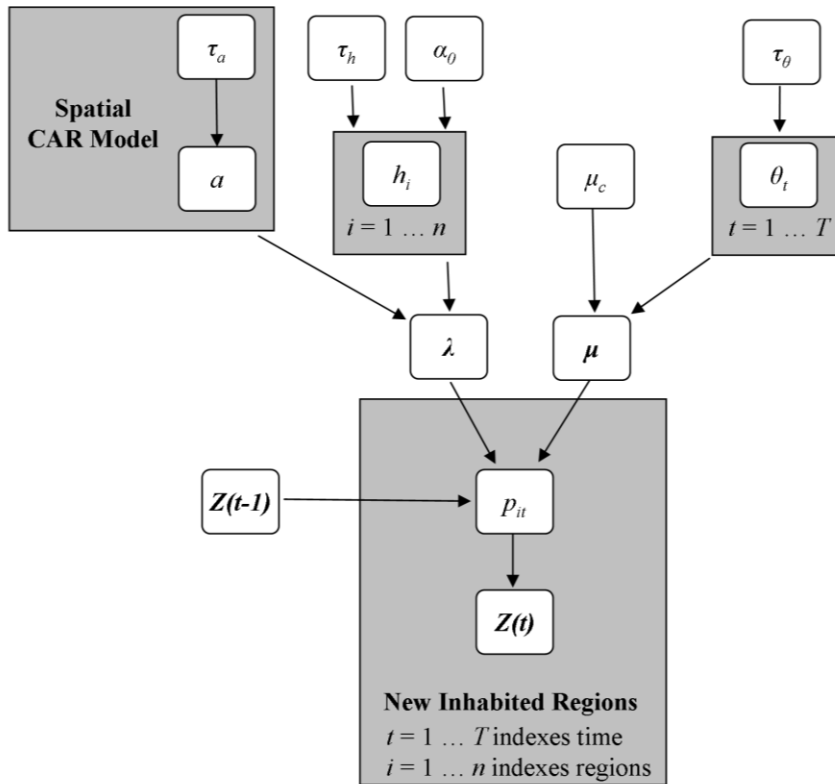
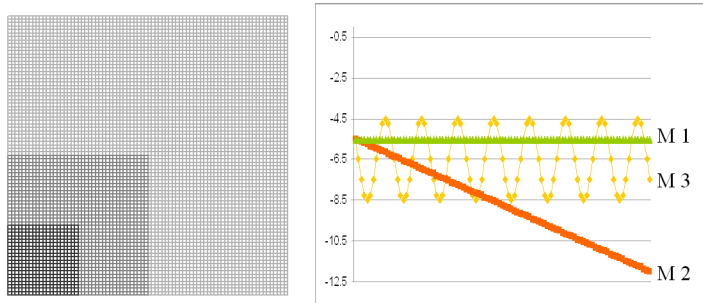


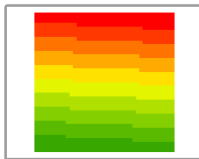
Figure 1: Graph diagramming the modeled relationship between input data, parameters, and hyper-parameters of the Bayesian space-time model for spread processes.

A) Simulation Scales and Parameters

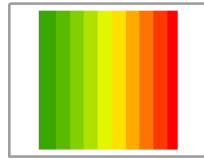
Spatial Scales 20x, 40x, 80x M: constant, linear, oscillating



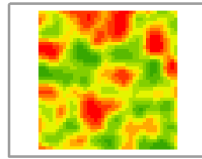
Λ 1: 0.09 - 2.25



Λ 2: 0.04 - 2.25

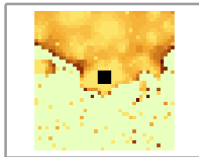


Λ 3: 0.46 - 2.66

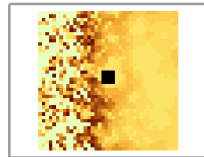


B) Simulated Data: Time of Infection

Λ 1, M 2



Λ 2, M 1



Λ 3, M 3

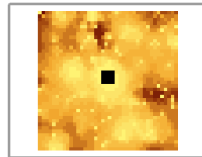


Figure 2: Simulated-estimation approach used in the study. Data were simulated using the model using three patterns for translocation (M) over 100 time-steps and three patterns for diffusion (Λ) onto 20x20, 40x40, and 80x80 study areas. The model was fitted to simulated datasets to obtain estimates of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

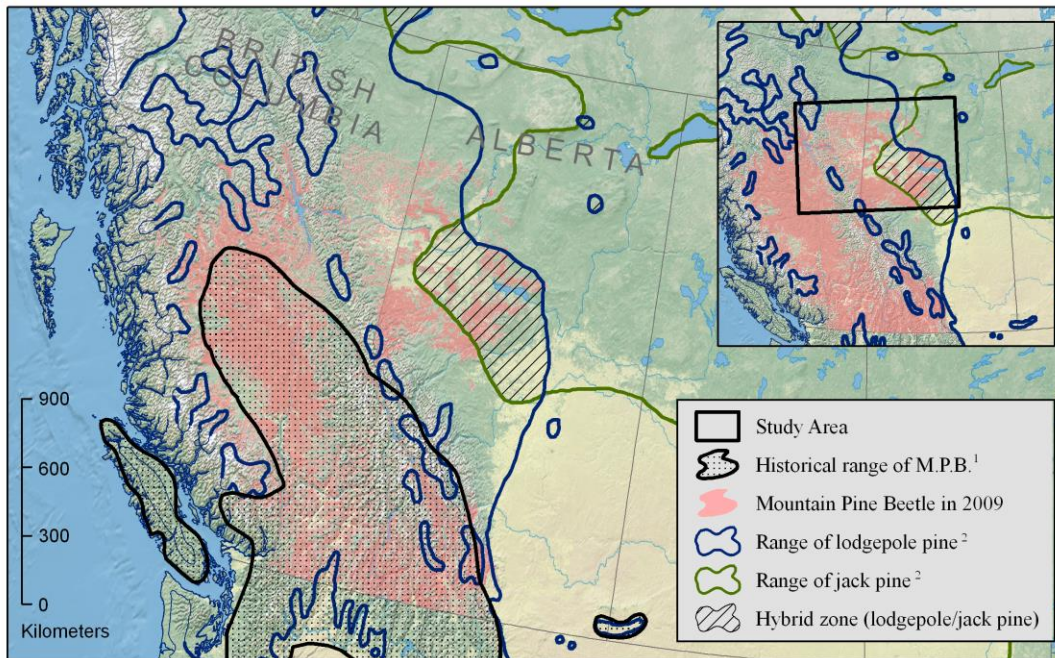


Figure 3: Historical range of mountain pine beetle and its preferred host species (lodgepole pine) and a potential new host species (jack pine) with the extent of mountain pine beetle infestation in 2009. Our study area (inset) contains the northern portion of the eastward expansion of mountain pine beetle into the boreal forest. ¹ Historical range of mountain pine beetle adapted from Fig. 4 in Safranyik and Carroll (2006). ² Tree species range maps from Little (1971), available at: <http://esp.cr.usgs.gov/data/atlas/little/>.

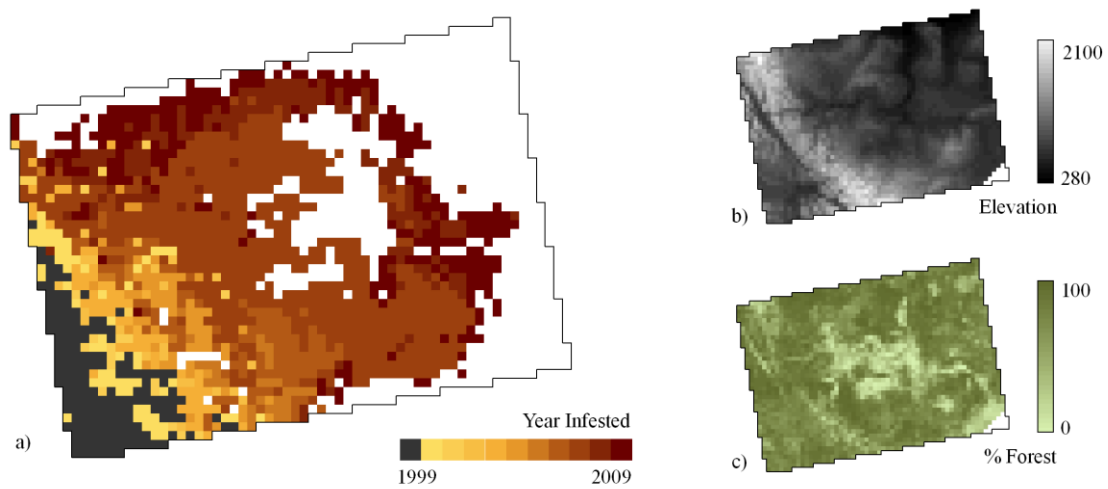


Figure 4: Maps showing: a) year of initial mountain pine beetle infestation across the study area and two spatial covariates used in the model b) mean elevation, and c) percent forest cover.

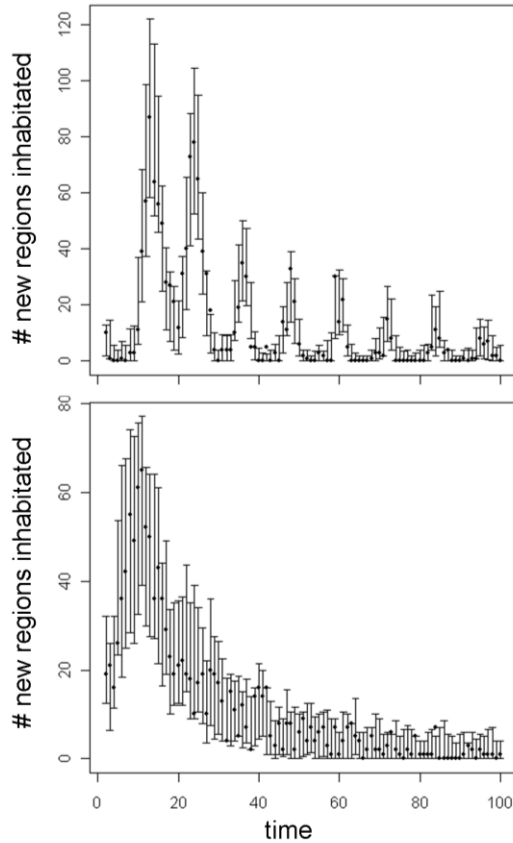


Figure 5: Posterior predictive checking for scenario Λ_2M_1 (top) and Λ_3M_2 (bottom) evaluating the number of newly inhabited cells at each time period. Error bars generated from replicates simulated from 100 draws of the posterior distributions of model parameters (λ_i, μ_t), dots indicate the observed 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 Λ_2M_1 and 2/100 time periods in scenario Λ_3M_2 , indicating excellent model fit.

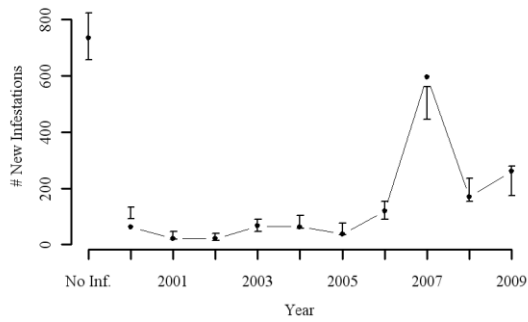


Figure 6: Posterior predictive checking for mountain pine beetle case study evaluating the number of newly infested cells at each time period. Error bars generated from replicates simulated from 1000 draws of the posterior distributions of model parameters $(\lambda_i, \tilde{\mu}_i)$, dots indicate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

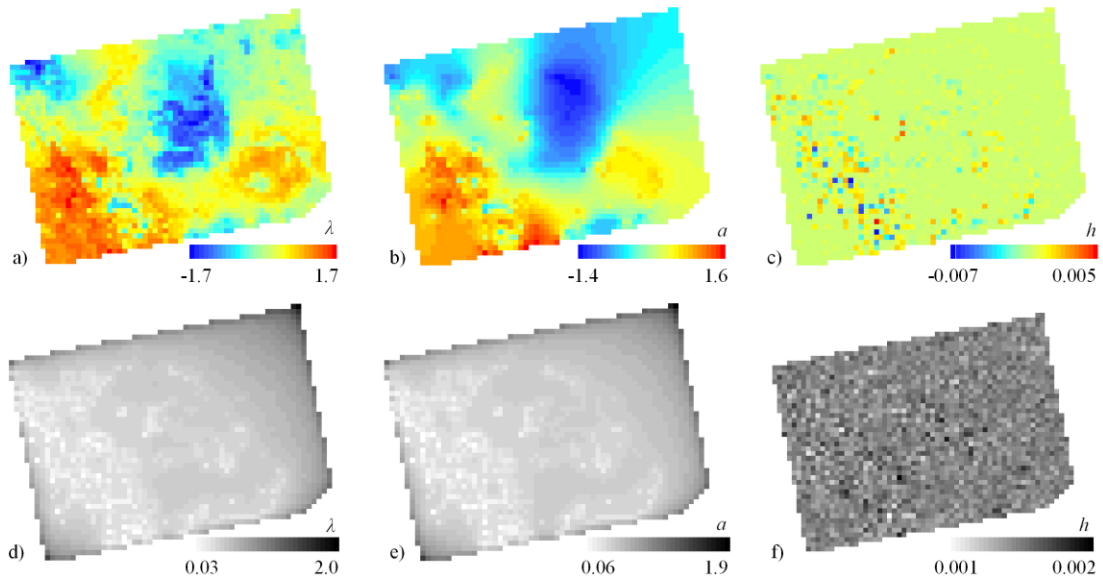


Figure 7: Maps of the posterior means for a) local diffusion parameter – λ , b) CAR model effect – a , and c) aspatial random effect – h ; with maps of posterior variance shown below.

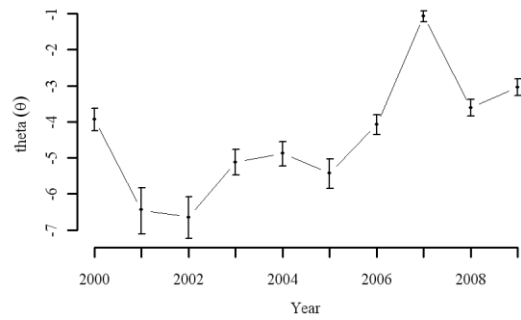
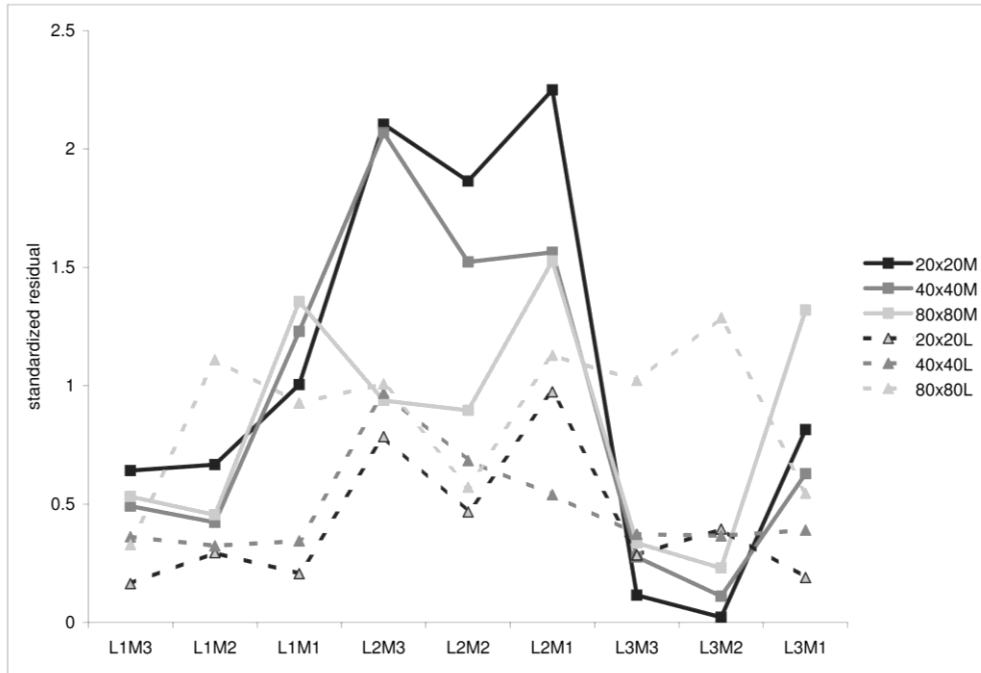


Figure 8: Graph of mean posterior of parameter θ_t , with 95% C.I. error bars, depicting annual variation in sporadic translocations events, which relates to passive spread by mountain pine beetle. The peak observed in 2007, corresponds to extensive beetle activity documented in 2006.

Appendix A: Global goodness of fit analysis results comparing true and estimated diffusion and translocation parameters for different spread scenarios and spatial scales.

Dataset	Diffusion				Translocation		
	20x20	Std Res	Std Res / n	chi Sq	P-value	Std Res	Std Res / t-1
Λ_1M_1		82.55	0.21	67.2	0.14	99.44	1
Λ_1M_2		117.96	0.29	37.1	0.24	65.98	0.67
Λ_1M_3		65.65	0.16	44	0.31	63.55	0.64
Λ_2M_1		389.77	0.97	60.5	0.13	222.72	2.25
Λ_2M_2		186.93	0.47	39.9	0.3	184.6	1.86
Λ_2M_3		313.87	0.78	31.5	0.49	208.35	2.1
Λ_3M_1		76.12	0.19	77	0.11	80.64	0.81
Λ_3M_2		157.8	0.39	39.9	0.3	2.18	0.02
Λ_3M_3		114.18	0.29	38.9	0.19	11.39	0.12
<hr/>							
40x40							
Λ_1M_1		551.69	0.34	99	0.24	121.69	1.23
Λ_1M_2		518.2	0.32	47.7	0.36	41.85	0.42
Λ_1M_3		580.08	0.36	58.7	0.31	48.65	0.49
Λ_2M_1		863.32	0.54	99	0.24	154.78	1.56
Λ_2M_2		1093.67	0.68	55	0.29	150.77	1.52
Λ_2M_3		1550.23	0.97	66	0.28	204.74	2.07
Λ_3M_1		624.96	0.39	99	0.08	62.18	0.63
Λ_3M_2		584.96	0.37	55	0.29	11.01	0.11
Λ_3M_3		597.65	0.37	58.7	0.31	27.35	0.28
<hr/>							
80x80							
Λ_1M_1		5932.75	0.93	99	0.24	134.11	1.35
Λ_1M_2		7100.85	1.11	66	0.28	44.98	0.45
Λ_1M_3		2111.98	0.33	66	0.28	52.6	0.53
Λ_2M_1		7224.69	1.13	99	0.24	151.21	1.53
Λ_2M_2		3658.6	0.57	66	0.28	88.71	0.9
Λ_2M_3		6446.2	1.01	57.7	0.34	92.83	0.94
Λ_3M_1		3494.29	0.55	99	0.24	130.61	1.32
Λ_3M_2		8238.91	1.29	66	0.28	22.76	0.23
Λ_3M_3		6543.47	1.02	66	0.28	33.26	0.34

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.

		Luminance	Contrast	Structure	SSIM
Λ_1	M_1	0.951	0.881	0.899	0.754
	M_2	0.231	0.739	0.890	-0.026
	M_3	0.924	0.864	0.889	0.706
Λ_2	M_1	0.824	0.858	0.897	0.622
	M_2	0.214	0.872	0.930	0.064
	M_3	0.681	0.870	0.898	0.508
Λ_3	M_1	0.972	0.856	0.720	0.602
	M_2	0.903	0.726	0.692	0.454
	M_3	0.974	0.827	0.697	0.565