Studying the probability of spruce beetle caused mortality in Colorado's spruce forests using Bayesian hierarchical models

Warong Suksavate\textsuperscript{1,2} | Yu Wei\textsuperscript{3} | John Lundquist\textsuperscript{4}

\textsuperscript{1}Department of Forest Biology, Faculty of Forestry, Kasetsart University, Bangkok, Thailand
\textsuperscript{2}Special Research Unit for Wildlife Genomics (SRUWG), Faculty of Forestry, Kasetsart University, Bangkok, Thailand
\textsuperscript{3}Department of Forest and Rangeland Stewardship, Colorado State University, Fort Collins, Colorado
\textsuperscript{4}Region 10 Forest Health Protection and Pacific Northwest Research Station, United States Forest Service, Anchorage, Alaska

Abstract
The association between environmental factors and Engelmann spruce mortality in western Colorado was investigated using Bayesian hierarchical zero-and-one inflated beta regressive model. The results indicated that the probability of mortality occurrence rate was positively associated with moderate to warm temperature zones, moderate precipitation, and single canopy stands with the smaller size class of spruce within stands. The probability of full mortality rate was positively associated with warm and humid climate zones with spruce as sub-dominant species in the vertically complex stands. On the other hand, higher mortality rate was more associated with local stand characteristics than climate factors that low mortality rate associated with stands dominated by spruce of a large size classes. The correlation model offers an approach to create predicted risk map for probability and rate of spruce mortality that could help manager in precedent planning for the upcoming outbreak events.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. © 2020 The Authors. Natural Resource Modeling published by Wiley Periodicals LLC
Recommendations for Resource Managers

- This study aims to investigate the environmental association of spruce bark beetle caused mortality in landscape context with binomial-beta hierarchical framework.
- Both regional-scale climate and stand-level characteristics should be considered to assess the risk and severity of outbreak caused mortality at a large spatial scale.
- The mortality risk is higher and more severe in warmer zones, especially the stands with spruce as subdominant tree species should be more concerned.
- Predicted map could improve the ability to spatially assess the risk of spruce mortality assisting managers to be better prepared for the future outbreaks, particularly in the risky area.

KEYWORDS
bark beetle, forest health, landscape, risk mapping, zero-and-one inflated beta regression

1 | INTRODUCTION

Spruce bark beetle (*Dendroctonus Rufipennis* Kirby) is native to temperate coniferous forests of western North America where they have periodic large scale outbreaks (Massey & Wygant, 1954; Raffa et al., 2008). In the region of southern Rocky Mountains, Engelmann spruce (*Picea engelmannii* Parry ex Engelm.) is the main host of this species of bark beetle (Schmid & Frye, 1977). Spruce bark beetle outbreaks impact many services and products provided by forest ecosystems in this region (DeRose & Long, 2009; Griffin, Turner, & Simard, 2011; Jenkins, Page, Hebertson, & Alexander, 2012; Kaiser, McGlynn, & Emanuel, 2013; Kurz et al., 2008). Understanding the effects of environmental factors on the spatial distribution and severity of mortality is a major goal for forest entomologists and land managers. Such information can provide valuable insights for managing forested landscapes (Lundquist & Hamelin, 2005; Lundquist & Reich, 2014).

Risk predictive models have been developed to gain understanding of the driving forces of spruce beetle outbreak and to predict when and where the outbreak would occur. The previous studies indicated that Engelmann spruce mortality is associated with climate covariates and vegetation characteristics (Hart, Veblen, Eisenhart, Jarvis, & Kulakowski, 2014; Huberty & Denno, 2004; Inbar, Doostdar, & Mayer, 2001; Joern & Mole, 2005; Reich, Lundquist, & Hughes, 2016). For instance, Schmid and Frye (1977) developed a hazard rating system for spruce beetle in Colorado based on average size of spruce, basal area of stand, and physiographic location of the stand. Reynolds and Holsten
(1996) used a tree-based classification approach to model spruce beetle hazard in Lutz spruce in Alaska and showed that beetle incidence was associated with total stand basal area, dominance of spruce, spruce age classes, elevation, and aspect. Hebertson and Jenkins (2008) used a similar tree-based approach in Utah to show that temperature in winter and drought severity index precedently associated with outbreak years. DeRose and Long (2012) found that latitude, slope, aspect, stand structure, species composition, site productivity, winter minimum temperature, and summer maximum temperature were strongly related to spruce mortality in southern Utah in the early stages of an infestation, but eventually be weakened with time. A detailed examination of the complex mechanisms of outbreak were presented by Raffa et al. (2008) illustrated that outbreaks were results of a series of excessive internal and external factors over the thresholds, driving cross-scale interactions, feedback, and nonlinear relationships.

Probabilistic models are common statistics tools used to study ecological patterns those characterize spatially explicit occupancy of organisms (Hoeting, Leecaster, & Bowden, 2000; MacKenzie et al., 2002). These models can assess the influences of hypothesized environmental covariates on the presence, absence, or abundance of targeted taxa and can spatially address the probability of occurrence across the study area. For instance, generalized linear model (GLM) was used to describe the association between spatially explicit responses of organisms and associated environmental covariates (Guisan & Thuiller, 2005). Bayesian-likelihood models, in the other hand, allow multiple problems within a hierarchical structure to be linked via a normalized likelihood function and the results, the posteriors, were sampled with the Markov chain Monte Carlo (MCMC) approaches (Gelman & Hill, 2006; Hooten & Hobbs, 2015). Structure of spatial dependency can also be integrated into the probabilistic models to address the autocorrelated errors in which could increase the efficiency of the model, with Gaussian process being often used (Banerjee, Carlin, & Gelfand, 2014; Banerjee, Yandell, & Yi, 2008; Chelgren, Michael, Bailey, & Bury, 2011).

Hierarchical modeling approaches have been applied in the studying the impact of disturbances, and management interventions on species richness; effects of community composition on diseases of individual trees; effects of logging and sanitation thinning on density of bark beetle infestations (Giovanini, Andrew, Jones, Altman, & Arnett, 2013; Haas, Hooten, Rizzo, & Meentemeyer, 2011; Stadelmann, Bugmann, Wermelinger, Meier, & Bigler, 2013), and so forth. To mitigate the excessive zero data, which detection of presence is uncommon, the probabilistic models often be depended on zero-inflated modeling structure (Ridout, Demétrio, & Hinde, 1998). The inordinate absence in data are separately modeled using binomial likelihood. Sometimes, model was decomposed into multiple likelihood parts to account for both extreme and intermediate values in which each part is separately modeled using the logistic regression (Nishii & Tanaka, 2012). The general class of models with binomial-beta framework was proposed by Ospina and Ferrari (2010, 2012) to address the extreme values at both ends of the data by applying multiple stages logistic regression for proportional data that was separated into three processes consist of modeling of presence-absence, occurrence of 100-percent rate, and intermediate proportional values.

In this study, a spatially explicit Bayesian multistage model was developed to investigate the mortality of *P. engelmannii* infested by *D. rufipennis*. The relationships between mortality, environmental factors at local and landscape scale, and spatial dependency were statistically examined. A zero-and-one inflated beta regression model was formulated using three consecutive likelihood structures based on the framework presented by Ospina and Ferrari (2012) to account for presence-absence and rate of Engelmann spruce mortality
associated with climate, stand-level covariates, and spatial dependency. The aim of this study is to develop an approach to assess the effects of combined multi-scale environmental variables composed of climate data, stand structure, and composition on Engelmann spruce mortality, and to create a predicted risk map based on the associations described by the best model.

2 | METHODS

2.1 | Study area and data collection

Field data were collected from 55 sites in 8 National Forests in western Colorado during July 2013 and August 2014 by cluster sampling (Figure 1). A sampling site was composed of three variable size plots approximately 50 m separated from each other, and each site was randomly located at least 50 m from the edge of forest road with 1.5 km minimum distance between sites. *P. engelmannii* distribution map obtained from part of the landcover data (Gap Analysis Program; https://gapanalysis.usgs.gov/gaplandcover/) was used for study sites placement conditional on spruce-fir forest type. In each plot, all individual trees, both alive and dead, designated by relascope to be within the plot were tallied for species and measured diameter at breast height. The presence of *D. rufipennis* larval feeding galleries was searched and identified on every suspected tree. Stand characteristic data were summarized for average basal area per tree, species richness, species dominance, vertical stand complexity (number of canopy layers), and mortality rate as proportion of total basal area.

**FIGURE 1** Study area in western Colorado and the locations of 55 study sites (black dots). The study area was delineated by the forested areas in Colorado.
Numeric plot data were discretized to account for the possible nonlinearity without adding excessive complexity to the model function. Class intervals were determined to attain equal frequencies among classes. Relative dominance (RD) of *P. engelmannii* was divided into four classes: RD1 for [0%, 25%], RD2 for (25%, 50%), RD3 for (50%, 75%), and RD4 for (75%, 100%). Average basal area per tree (BT) was classified into three categories: BT1 for between [0, 1] square feet/tree, BT2 for between (1, 2] square feet/tree, and BT3 for >2 square feet/tree. The vertical structure (S) of stand was categorized into three classes: S1 for stands with one story, S2 for stands with two stories, and S3 for stands with at least three stories. Climate data were obtained from a previous study (Aquirre-Bravo & Reich, 2006) by spatially intersecting locations of study plots with the predicted climate map. Precipitation (P, from P1 to P5) and average temperature (T, from T1 to T5) data were categorically combined into 25 climate zones (5-by-5 combinations) based on geostatistic interpolation obtained from previous studies (Figure 2; Aquirre-Bravo & Reich, 2006; Reich et al., 2014). Delineation of climate zones provide an opportunity to examine the nonlinear effect of climatic factors and avoiding the use of highly-bias numeric climate data (Pongpattananurak, Reich, Khosla, & Aguirre-Bravo, 2012; Reich et al., 2014).

### 2.2 Model specification

A hierarchical Bayesian framework was implemented to accommodate multiple steps of GLMs. Definition of model parameters were shown in Table 1. The first step in the formulation of model is zero-inflated component, which describe the binary response of spruce mortality occurrence ($y_i \geq 0$), given by $z_i$. The second binary response was then modeled to inflate full mortality ($y_i = 1$), given by $v_i$. Occurrence of mortality and full mortality among sample plots can be formulated as zero-and-one inflated model based on hierarchical binomial process. The estimated parameters $p_i$ and $\phi_i$ represent the probability of occurrence of mortality ($z_i = 1$) and
probability of full mortality \((v_i = 1)\) at plot \(i\), respectively. The following equations capture the relationships within zero-and-one inflated structures:

\[
\begin{align*}
\Pr(z_i = 1) &= p_i, \quad (1) \\
y_i &= \begin{cases} 
(0, 1); & z_i = 1, \\
0; & z_i = 0,
\end{cases} \quad (2)
\end{align*}
\]

\[
z_i \sim \text{binomial}(p_i, 1), \quad (3)
\]

\[
\Pr(v_i = 1) = \Pr(y_i = 1 | z_i = 1) = \phi_i, \quad (4)
\]

\[
v_i \sim \text{binomial}(\phi_i, 1), \quad (5)
\]

\[
\Pr(y_i) = \begin{cases} 
\phi_i p_i; & y_i = 1, \\
(1 - \phi_i)p_i; & 0 < y_i < 1, \\
1 - p_i; & y_i = 0.
\end{cases} \quad (6)
\]

The first part (Equation 1–3) is zero-inflated binomial process created to describe the presence of mortality, \(z_i\), with proportion parameter \(p_i\). The subsequent structure (Equations 4 and 5) described the probability \(\phi_i\) to inflate extreme mortality rate \((y_i = 1)\). Since neither \(y_i = 0\) nor 1 are in the domain of beta distribution, absence and extreme value must be modeled separately leaving out the value of mortality rate in-between. Equation (6) shows the full probability of the zero-and-one inflated structure. The formulation of beta model follows after zero-and-one binomial process with conditional probability of \(1 - \phi_i\) (Equation 6), to address the value of mortality rate between 0 and 1 given the mortality occurs. Then, beta likelihood was incorporated in the last part of hierarchical structure to model mortality rate, \(y_i\), given \(0 < y < 1\). The model can be formulated as in Equation (7), parameterized by mean, \(\mu_{y_i}\), and variances, \(\sigma_{y_i}^2\).

\[
(y_i | v_i = 0) \sim \text{beta}(\mu_{y_i}, \sigma_{y_i}^2). \quad (7)
\]

To address the association between mortality and environmental factors, logistic regression was implemented to model the probability of mortality occurrence, \(p\), probability full mortality, \(\phi\),

| Parameters | Description |
|------------|-------------|
| \(z_i\)   | Presence of mortality |
| \(y_i\)   | Proportion of mortality |
| \(p_i\)   | Probability of mortality occurrence |
| \(\phi_i\) | Probability of full mortality occurrence |
| \(\mu\)   | Logit mean of regression |
| \(\alpha_i\) and \(\beta_i\) | Parameters of beta process |
| \(\sigma_w^2\) | Effect of spatial covariance |
| \(\lambda\) | Exponential decay rate |
and average mortality rate, $\mu_y$, as the responses in logit form (Banerjee et al., 2014; Finley, Banerjee, & McRoberts, 2007). Equation (8) represents the logistic regression with $X$ as design matrix and $\gamma$ as model coefficients. To account for the spatial dependence, a normal distribution of spatial errors with exponential decay covariance function was implemented in the following structure (Equations 9 and 10):

$$\text{logit}(\mu) = X\gamma + w,$$

$$w \sim \text{normal}(0, \Sigma(\lambda, \sigma^2_w)),$$

$$\Sigma = \sigma^2_w \exp(-\lambda D).$$

Given that the value of logit response ($\mu$) is a linear function of environmental covariates with spatial error terms ($w$) described by latent Gaussian process. Spatial covariance matrix ($\Sigma$) address spatial dependence between pairs of observed locations described by variance effect parameter ($\sigma^2_w$), exponential diminishing function of Euclidean distance, $D$, between pairs of points, and exponential decay parameter, $\lambda$. Probability of mortality occurrence, probability of full mortality, and average mortality rate were all modeled in the same approach. Directed acyclic graph of hierarchical structure was shown in Figure 3 and full likelihood functions of binomial zero-and-one inflated model and beta model could be, respectively, written as the following equations:

$$L(\gamma, \lambda, \sigma^2_w | z) \propto \prod \text{binomial}(z | g(\gamma, w)) \text{normal}(0, \Sigma(\lambda, \sigma^2_w)),$$

$$L(\gamma, \lambda, \sigma^2_w | y) \propto \prod \text{beta}(y | f_\alpha(g(\gamma, w)), f_\beta(\sigma^2_z)) \text{normal}(0, (\lambda, \sigma^2_w)).$$

The hierarchical model was developed under the Bayesian framework (Gelman & Hill, 2006). Given $f_\alpha$ and $f_\beta$ are the moment matching function for beta mean and variance of mortality rate. The coefficients, $\gamma$, were assigned by vague normal priors centered at zero with large variances. The spatial variance effect term, $\sigma^2_w$, received vague gamma priors centered at one with large variances. Centers of priors for regression parameters were specified by a preliminary analysis of fitted GLM model without spatial dependency. Then, the residuals obtained from the preliminary GLM model were used to fit spatial semi-variogram model with GeoR package (R Core Team, 2014; Ribeiro & Diggle, 2001). The estimated range parameter from preliminary semi-variogram model were used as center of the uniform prior for spatial decay parameter, $\lambda$, with value between zero and two times of center’s value. Model fitting was performed by Just another Gibbs sampler Software (JAGS; Plummer, 2004) through R via rjags package (Plummer, 2013). MCMC with Metropolis and Gibbs sampling algorithm implemented to update the posteriors. The sampling was performed on three independent MCMC chains of 100,000 iterations with 10,000 burn-in samples. Parameters for each chain were randomly initialized within a feasible range according to priors. Gelman–Rubin diagnostics were applied to assess the convergence of posteriors.

Model evaluation and selection was done using deviance information criterion (DIC; Spiegelhalter, Nicola, Carlin, & Van Der Linde, 2002). Candidate model with the lowest DIC value suggest to be the most optimal. Furthermore, the other evaluation methods were also used to address for accuracy of the model without involving in model selection. Mean squared prediction error (MSPE) was used in posterior predictive check to assess the accuracy of the
prediction. Besides, log predictive density (LPD) was estimated by averaging over the iteration of total likelihood conditional on parameters within MCMC chains.

### 2.3 Prediction of mortality at regional scale

The probability of mortality occurrence, probability of full mortality, and mortality rate were predicted from 0.5 quantile of posteriors across MCMC conditional on the presence of host species (20,547 of the 1-km grids). First, the regional climate covariates (temperature and precipitation zones) were known throughout whole study area, however, prediction for the whole study area required the covariates of stand-level characteristics to be estimated for the whole spatial prediction domain. The stand-level covariates were estimated by linear spatial interpolation using the inverse distance weight (IDW) method considering 12 nearest neighbors (Figure 4). The interpolation processes were applied using ArcMap (Environmental Systems

---

**FIGURE 3** Conceptual diagram of the hierarchical Bayesian model (Directed Acyclic Graph) of zero- and one-inflated beta model. Above is the model for $p$ and $\phi$, the zero-and-one inflated binomial process. Below is the beta model for continuous proportion of mortality, $y$. Solid lines represent stochastic relationship between parameters while dashed lines represent deterministic relationship.
The hierarchical Bayesian zero-and-one inflated beta regression models were fitted to examine the relationship between environmental factors and spruce mortality. All candidate models were validated by DIC, LPD, and MSPE posterior predictive checks. The best five candidates for separate models were selected based on the information criteria of DIC to incorporate tradeoff between model predictive power versus complexity result in the optimal simple model. The comparable of information criterion for the best models were shown in Table 2.

The best zero-inflated model describing the environmental association with the probability of mortality occurrence was composed of temperature zone (T), precipitation zone (P), vertical structure of stand (S), and average basal area per tree (BT; Figure 5). For the spatial parameters,
The posterior of exponential decay rate ($\lambda$) and spatial dependence effect ($\sigma^2_w$) were median of 0.017 with 95% credible interval of [0.014, 0.021], and 7.028 [6.522, 7.555], respectively (Figure 6). The derived spatial effect was shown in Figure 6. Posterior density for coefficient of each temperature and precipitation zone was compared against middle baseline at temperature zone (T3) and middle precipitation zone (P3), respectively. The colder temperature zones T1, $-17.847 [-19.809, -15.895]$, and T2, $-2.033 [-3.225, -0.883]$, were negatively correlated with the occurrence of mortality compared with middle baseline, with an extremely large negative value of effect of T1, however, the coefficients of T4, $-0.534 [-1.922, 0.850]$, and T5, $-0.907 [-2.669, 0.825]$, were not significantly different from the middle baseline. For the effect of precipitation, moderately high precipitation zone P4, 2.307 [1.094, 3.555], was positively related to the occurrence of mortality against the middle baseline at P3, while the coefficients of other climate zones were not different from baseline. For the association between local stand characteristics and probability of mortality occurrence, relative effects of vertical stand structure and average

| Model parameters | DIC  | ΔDIC | LPD  | MSPE |
|------------------|------|------|------|------|
| Zero-inflated logistic regression model | | | | |
| T + P + S + BT | 90.020 | 0.000 | −41.205 | 0.115 |
| T + P + S + RD + BT | 91.040 | 1.020 | −40.683 | 0.111 |
| T + S + BT | 92.970 | 2.950 | −43.512 | 0.118 |
| T + S + RD + BT | 93.410 | 3.390 | −42.622 | 0.115 |
| P + S + BT | 93.740 | 3.720 | −46.294 | 0.124 |
| T + P + S | 94.370 | 4.350 | −43.944 | 0.122 |
| T + P + S + RD | 94.510 | 4.490 | −42.912 | 0.115 |
| One-inflated logistic regression model | | | | |
| T + P + S + RD | <0.001 | 0.000 | −0.001 | <0.001 |
| T + P + S + RD + BT | <0.001 | <0.001 | −0.001 | <0.001 |
| T + P + RD + BT | 12.710 | 12.710 | −5.536 | 0.065 |
| P + S + RD | 16.660 | 16.660 | −7.328 | 0.080 |
| T + P + S + BT | 17.770 | 17.770 | −7.704 | 0.090 |
| T + P + RD | 18.340 | 18.340 | −8.112 | 0.091 |
| T + S + RD + BT | 21.000 | 21.000 | −8.860 | 0.095 |
| Beta regression model | | | | |
| P + S + RD + BT | −17.920 | 0.000 | 26.111 | 0.019 |
| P + RD + BT | −15.010 | 2.910 | 22.684 | 0.022 |
| P + S + RD | −12.320 | 5.600 | 20.887 | 0.023 |
| T + P + ST + RD | −11.990 | 5.930 | 21.513 | 0.023 |
| T + P + RD + BT | −11.740 | 6.180 | 26.491 | 0.019 |
| T + P + S + RD + BT | −11.740 | 6.180 | 26.132 | 0.019 |
| T + P + RD | −11.550 | 6.370 | 24.970 | 0.025 |

Note: The posterior predictive check (MSPE and LPD) was listed. Three stages of modeling consist of zero-inflated model for dealing with the presence of spruce mortality, one-inflated model for the presence of full mortality conditional on the presence of mortality, and beta regression model estimating the proportion of partial mortality. The covariates include temperature zones (T), precipitation zones (P), number of stories (S), relative dominance (RD), and average basal area perper tree (BT).
FIGURE 5  Posterior distribution of coefficient parameters of the best-fitted zero-inflated (white boxes) and beta regression model (gray boxes), compared with baseline (gray solid lines). The covariates consist of temperature zone (T), precipitation zones (P), relative dominance (RD), number of stories (S), and average basal area per tree (BT) were included in the best-fitted model. Error bars showed 95% credible interval of posterior distributions.
basal area per tree were compared against the minimum baseline at the single level class (S1) and the smallest size class (BT1). The coefficients of size class were 2.006 [0.888, 3.159] and 1.186 [−0.074, 2.430] for BT2 and BT3, respectively. This appears to be a weak unimodal pattern with effect of BT2 was significant over BT1. Both stands with higher vertical complexity S2, −3.750 [−4.963, −2.550], and S3, −5.104 [−6.471, −3.799], were significantly negative related to the occurrence of mortality relative to S1, with a monotonic relationship.

The best one-inflated model describing the probability of full mortality was composed of the coefficients of T, P, S, and RD. The posterior densities of model parameters according to the best models are shown in Figure 7. For the spatial parameters, posterior of exponential decay rate (λ) and spatial dependence effect (σw²) were 0.148 [0.078, 0.219], and 0.682 [0.533, 0.856], respectively (Figure 6). Due to nonexistence of full mortality sampling site within the coldest temperature zone (T1), the coefficient of T1 was unavailable. The coefficients of moderately cold, T2, and warm temperature zones, T4, were significantly different with the middle baseline, T3, with the posteriors −92.591 [−94.558, −90.645] and −23.455 [−25.413, −21.488], respectively. The coefficient of T2 was negative on T2 side, while the warmest temperature zone T5, 67.199 [65.247, 69.151], had a distinctively positive coefficient compare with both T4 and baseline. This showed the oscillatory pattern on the effects of temperature zones with T2 and T5 had a very distinct negative and positive coefficients, respectively. The coefficients of precipitation zones seem to had a milder oscillatory pattern than temperature zones. The moderately low and high precipitation zones, P2 and P4, have negative coefficients on the probability of full mortality with the posteriors of −47.155 [−49.116, −45.207] and −71.057 [−72.996, −69.096] respectively; while the wettest precipitation zone P5, 74.124 [72.172, 76.090], had a distinctly large positive coefficient compare with all others. For stand characteristics, the higher stand vertical complexity S2 and S3 had positive coefficients on full mortality with the posteriors of 138.796 [136.854, 140.745] and 93.642 [91.675, 95.615], respectively. This
showed a mild unimodal pattern when being compared with minimum baseline at S1. The coefficients of all classes with high relative dominance of *P. engelmannii* classes were negative compare with minimum baseline at RD1, however, RD4, $-47.262 \in [-49.217, -45.308]$, had weaker correlation with RD2, $-139.190 \in [-141.154, -137.206]$, and RD3, $-141.102 \in [-143.057, -139.134]$, showing the unimodal pattern with minimum.

The best beta model for mortality rate was a function of P, S, RD, and BT. The posterior density of parameters was shown in Figure 5. The posteriors of exponential decay parameter ($\lambda$) and spatial dependence effect ($\sigma_w^2$) were 0.007 [0.003, 0.009] and 1.094 [1.030, 1.159], respectively (Figure 6). Comparing to the middle baseline at P3, moderately high precipitation zone at P4, $1.056 \in [0.063, 1.967]$, had a positive coefficient on mortality rate, while the coefficients of colder precipitation zones P1 and P2 were not significantly different from middle baseline with posteriors $-0.731 \in [-2.177, 0.751]$ and $-0.949 \in [-2.183, 0.274]$, respectively. For the effect of stand structure, high vertical complexity classes of S2 and S3 had negative coefficients on mortality relative to minimum baseline at S1 with posteriors $-0.535 \in [-1.377, 0.348]$ and $-0.608 \in [-2.001, 0.810]$, respectively. The highest relative dominance class, RD4, was positively related to mortality rate, versus the minimum baseline at RD1, with posteriors $-1.084 \in [-2.110, -0.023]$.  

**FIGURE 7** Posterior distributions of coefficients of the best one-inflated model were plotted separately, due to the extremely high and low coefficient values. The model parameters consist of temperature zones (T), precipitation zones (P), number of stories (S), and relative dominance (RD).
There was no significant effect associated with relative dominant classes of RD2 and RD3 with posteriors $0.032 [-0.943, 1.044]$ and $-0.077 [-1.238, 1.092]$, respectively. Stands with the largest size class BT3 had a negative coefficient on mortality rate relative to minimum baseline at the smallest size class BT1 with posteriors $-1.258 [-2.52, -0.002]$, while the intermediate size class BT2 had a nonsignificant negative coefficient compare with BT1 with posteriors $-0.238 [-1.064, 0.574]$.

The predicted risk maps for probability of occurrence of mortality, probability of full mortality, and mortality rate were created conditional on MCMC chains of model coefficients

**FIGURE 8** Prediction map at 0.50 quantile of response values for the best-fitted zero-and-one inflated beta regression model. Each map represents the prediction of predicted probability and proportion for separated process of the model; zero-inflated (ZIM), one-inflated (OIM), and beta model (BRM). The study area was delineated by the occurrence of spruce-fir forest in Colorado.
following model with the lowest DIC (Figure 8). The regional prediction was done for median (0.50 quantile), lower bound of credible interval (0.025 quantile), and upper bound of credible interval (0.975 quantile). Table 3 showed the proportion of study area that the value fell in the 0 and 1 interval for each quantile predicted risk map. There was a significant difference of predicted value between quantile level of zero-inflated and beta model, while there was a less difference of prediction between quantile level of one-inflated model.

### 4 Discussion and Conclusions

The causes of bark beetle outbreaks are often difficult to be deciphered because vulnerable host populations and aggressive beetle populations must co-occur within a favorable spatially varying environments. Temperature and precipitation play especially significant roles in determining the occurrence, distributions, and severity of outbreaks across forested regions. This study showed that a Bayesian framework can be used to account for the occurrence and severity of beetle caused spruce mortality by applying zero-and-one inflated binomial and beta likelihood with spatial dependency of errors described by Gaussian process.

The results indicated that the occurrence and rate of spruce mortality were associated with regional-scale temperature and precipitation zones, and with community-level stand structure, relative dominance, and tree size class (Allen et al., 2010; Breshears et al., 2005; Hanson &
Low temperature was negatively related to both the probability of spruce mortality occurrence and full mortality rate, while warm temperature was positively associated with full mortality rate. For the relations with local stand characteristics, probability of occurrence and mortality were both higher in single-story stands with medium to large size class of stands, while full mortality rate was associated with higher stand structure complexity and low relative density of Engelmann spruce. These showed that relationships of environmental covariates at multiple scales with mortality occurrence and rate were necessary to predict the risk spruce mortality at landscape level in western Colorado. The spatial dependence of the model showed a very strong and mild spatial dependency in zero-inflated model and beta process model, respectively, with the range around 200 km. This showed a large scale spatial dependence of errors in predicting the occurrence and rate of mortality. Nevertheless, there was a very weak spatial dependency in one-inflated model addressing full mortality rate which was not common in observations and implied that local scale factors might be more important driving such extreme mortality rate. Nevertheless, the cluster observed locations in this study which based on sparse dispersion of available host and protected area cause the lack of the intermediate distance of around 30–80 km. This made the spatial dependency at intermediate distance to be underrated.

Past studies have explained some of the correlations between susceptibility of host tree, population dynamics of bark beetle, and environmental factors. This study provides an empirical evidence that deviation from optimal climatic conditions could either alter physiological processes of hosts and increase their susceptibility to bark beetles, or increase population of bark beetle that would overcome even the non-susceptible trees (Huberty & Denno, 2004; Park Williams et al., 2013). Several hypotheses have explained these relationships, for example, Plant Stress Hypothesis (PSH) in which tree under stress generate higher concentrations of amino acids providing more palatable food source for mortality causing agents (Mattson & Haack, 1987; White, 1993). Moreover, high physiological stress also decreases hosts’ ability to synthesize defensive secondary chemicals (Rhoades, 1983). Higher palatability and weakened defensive mechanisms could make plants to become more susceptible to insects and diseases. Our study found that mortality rate was lower in cold places and the other way around. Higher evaporation in warmer climate zones could cause higher level of water vapor deficit leading to physiological stress of host species. Reich et al. (2016) referred warm and dry climate as the marginal sites in which the both host population and site suitability is low, on the other hand, host has higher population and be less susceptible in niche center. Spatial distribution of marginal sites across could have impact to the landscape heterogeneity of mortality occurrence and rate, especially in regime of climate shifting (Mietkiewicz, Kulakowski, & Veblen, 2018). The results can also be conjectured by Climate Release Hypothesis (CRH). The CRH states that population dynamics of insects and other disease causing agents are directly related to abiotic factors (Price, 1991), for example, the increasing of dietary substances due to climate induced stress of host species; or warmer climate accelerates life cycle of agents (Burnett, 1949; Ayres & Lombardero, 2000; Mattson & Addy, 1975; Bentz & Six, 2006). This increase suitability for insect populations that eventually respond to the conditions by altering assemblage composition, range extension, ecological interactions, and timing of life history. These ecological shifting could enhance the aggressiveness or virulence of mortality causing agents to overcome defensive mechanism of host regardless of susceptibility (Burnett, 1949; Danks, 1992; Weed, Ayres, & Hicke, 2013).

Spatial data and statistical modeling could be applied to investigate the influence of the environmental factors on the spatial extent, distribution, and severity of mortality related to forest
health. The insight of host-agent-environment interactions could help manager made a spatially explicit prediction of risk, severity, and spreading of mortality in the regional level based on existence of information at multiple scales. The predicted map created from the best model showed that predicted probability of mortality occurrence and the mortality rate are higher in the northern part of study area, while the southern part had higher probability of full mortality. According to field data, there is a gradual change in pattern of forest community structure and composition across the study area in north–south direction (see Figures 2 and 4). This also implied the pattern of spatial distribution of local environmental factors affect the heterogeneity of risk and severity of spruce mortality across the region conditional on the suitable climate conditions. The predicted risk of this study is also congruent with the several historical major outbreak events in central Rocky Mountain in the mid-20th century. This field data of this study captured the rising period of spruce beetle outbreak which was at its peak in 2017 that impacted more 67,000 acres of Engelmann spruce forest, and revealed the spreading of outbreak into historically undamaged stands those strongly link to proximity to the extent of historic outbreak, and shifting in climate regime (Hart, Veblen, Schneider, & Molotch, 2017). Moreover, there were clusters of severe mortality patch of spruce forest in the southern part of Colorado, predicted with high probability of full mortality while probability of mortality occurrence is low. According to the outbreak event in 2017, the Southern Colorado forest suffered very high rate of mortality. Therefore, understanding the importance of environmental drivers at both local and regional-scale could provide us an ability to predict occurrence and severity of bark beetle epidemic.

Spruce beetle outbreaks have been one of the most damaging disturbances of forests in Western North America. Understanding what drivers cause the onset of an outbreak, sustain it, and cause it to subside would greatly help managers to be better prepared for the future outbreaks. This study helps improve our ability to predict spruce mortality by using hierarchical models to address for the interactions between spruce bark beetles, hosts, and environmental factors that could lead to useful prediction of risk occurrence and severity of the devastating outbreaks.

ACKNOWLEDGMENTS
I would like to express my very great appreciation to late Dr. Robin Reich for his invaluable guidance during my years at Colorado State University and for suggestions in planning, development, and conducting fieldwork of this study. I would also like to thank my colleagues at the Department of Forest and Rangeland Stewardship, Kristina Hughes and Ryan Davy, who gave valuable support during fieldwork of this study that related to their projects.

AUTHOR CONTRIBUTIONS
W. S.: Writing original draft, review, and editing. Y. W.: Leading supervision. J. L.: Supporting supervision.

ORCID
Warong Suksavate  http://orcid.org/0000-0003-1879-4905
Yu Wei  http://orcid.org/0000-0001-8473-6856
John Lundquist  https://orcid.org/0000-0003-1935-7520

REFERENCES
Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., ... Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, 259, 660–684.
Aquirre-Bravo, C., & Reich, R. M. (2006). Spatial statistical modeling and classification of climate for the state of Colorado and adjacent lands of neighboring states. Rocky Mountain Research Station, USDA Forest Service, Fort Collins, CO.

Ayres, M. P., & Lombardero, M. J. (2000). Assessing the consequences of global change for forest disturbance from herbivores and pathogens. *Science of the Total Environment, 262*(3), 263–286. https://doi.org/10.1016/s0048-9697(00)00528-3

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). Hierarchical modeling and analysis for spatial data. CRC Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=c9xLKzZWoZ4C&oi=fnd&pg=PP1&dq=sudipto&ots=Anx3yiQM5&sig=nQfBvL6u8L_k377OzdrBWli_mok

Banerjee, S., Yandell, B. S., & Yi, N. (2008). Bayesian Quantitative Trait Loci Mapping for Multiple Traits. *Genetics, 179*(4), 2275–2289.

Bentz, B. J., & Six, D. L. (2006). Ergosterol content of fungi associated with Dendroctonus ponderosae and dendroctonus rufipennis (Coleoptera: Curculionidae, Scolytinae). *Annals of the Entomological Society of America, 99*(2), 189–194. http://doi.org/10.1603/0013-8746(2006)99[0189:ecoaw]2.0.co;2

Breshears, D. D., Cobb, N. S., Rich, P. M., Price, K. P., Allen, C. D., Balice, R. G., ... Meyer, C. W. (2005). Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Sciences of the United States of America, 102*(42), 15144–15148.

Burnett, T. (1949). The effect of temperature on insect host-parasite population. *Ecology, 30*, 113–134.

Chelgren, N. D., Michael, J. A., Bailey, L. L., & Bury, R. B. (2011). Using multilevel spatial models to understand salmoner site occupancy patterns after wildfire. *Ecology, 92*(2), 408–421. https://doi.org/10.1890/10-0322.1

Danks, H. V. (1992). Arctic insects as indicators of environmental change. *Arctic, 45*, 159–166.

DeRose, R. J., & Long, J. N. (2009). Wildfire and spruce beetle outbreak: Simulation of interacting disturbances in the central rocky mountains. *Ecoscience, 16*(1), 28–38. https://doi.org/10.2980/16-1-3160

DeRose, R. J., & Long, J. N. (2012). Factors influencing the spatial and temporal dynamics of Engelmann spruce mortality during a spruce beetle outbreak on the Markagunt Plateau, Utah. *Forest Science, 58*(1), 1–14.

Environmental Systems Research Institute. (2011). *ArcGIS Desktop: Release 10*. Redlands, CA: Author.

Finley, A. O., Banerjee, S., & McRoberts, R. E. (2007). A Bayesian approach to multi-source forest area estimation. *Environmental and Ecological Statistics, 15*(2), 241–258. https://doi.org/10.1007/s10651-007-0049-5

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=c9xLKzZWoZ4C&oi=fnd&pg=PP1&dq=Gelman+and+Hill,+2006&ots=baV4M0Omm9&sig=O0ccUvDnGKSvo4eGocywECyn0k

Giovanini, J., Andrew, J. K., Jones, J. E., Altman, B., & Arnett, E. B. (2013). Effects of management intervention on post-disturbance community composition: An experimental analysis using Bayesian hierarchical models. *PLOS One, 8*(3), e59900. https://doi.org/10.1371/journal.pone.005990

Griffin, J. M., Turner, M. G., & Simard, M. (2011). Nitrogen cycling following mountain pine beetle disturbance in Lodgepole pine forests of greater yellowstone. *Forest Ecology and Management, 261*(6), 1077–1089. https://doi.org/10.1016/j.foreco.2010.12.031

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters, 8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Haas, S. E., Hooten, M. B., Rizzo, D. M., & Meentemeyer, R. K. (2011). Forest species diversity reduces disease risk in a generalist plant pathogen invasion. *Ecology Letters, 14*(11), 1108–1116. https://doi.org/10.1111/j.1461-0248.2011.01679.x

Hanson, P. J., & Weltzin, J. F. (2000). Drought disturbance from climate change: Response of United States forests. *The Science of the Total Environment, 262*, 205–220.

Hart, S. J., Veblen, T. T., Eisenhart, K. S., Jarvis, D., & Kulakowski, D. (2014). Drought induces spruce beetle (*Dendroctonus rufipennis*) Outbreaks across Northwestern Colorado. *Ecology, 95*(4), 930–939. https://doi.org/10.1890/13-0230.1

Hart, S. J., Veblen, T. T., Schneider, D., & Molotch, N. P. (2017). Summer and winter drought drive the initiation and spread of spruce beetle outbreak. *Ecology, 98*(10), 2698–2707. https://doi.org/10.1002/ecy.1963

Hebertson, E. G., & Jenkins, M. J. (2008). Climate factors associated with historic spruce beetle (Coleoptera: Curculionidae) outbreaks in Utah and Colorado. *Environmental Entomology, 37*(2), 281–292.
Hoeting, J. A., Leecaster, M., & Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics, 5*(1), 102–114. https://doi.org/10.2307/1400634

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs, 85*(1), 3–28. https://doi.org/10.1890/14-0661.1

Huberty, A. F., & Deno, R. F. (2004). Plant water stress and its consequences for herbivorous insects: A new synthesis. *Ecology, 85*(5), 1383–1398. https://doi.org/10.1890/03-0352

Inbar, M., Doostdar, H., & Mayer, R. T. (2001). Suitability of stressed and vigorous plants to various insect herbivores. *Oikos, 94*, 228–235.

Jenkins, M. J., Page, W. G., Hebertson, E. G., & Alexander, M. E. (2012). Fuels and fire behavior dynamics in bark beetle–attacked forests in Western North America and implications for fire management. *Forest Ecology and Management, 275*(July), 23–34. https://doi.org/10.1016/j.foreco.2012.02.036

Joern, A., & Mole, S. (2005). The plant stress hypothesis and variable responses by blue grama grass (*Bouteloua Gracilis*) to water, mineral nitrogen, and insect herbivory. *Journal of Chemical Ecology, 31*(9), 2069–2090.

Kaiser, K. E., McGlynn, B. L., & Emanuel, R. E. (2013). Ecohydrology of an outbreak: Mountain pine beetle impacts trees in drier landscape positions first. *Ecohydrology, 6*(3), 444–454. https://doi.org/10.1002/eco.1286

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., … Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature, 452*(7190), 987–990. https://doi.org/10.1038/nature06777

Lundquist, J., & Hamelin, R. (2005). Landscape pathology-forest pathology in the era of landscape ecology. Lundquist, J. E., & Reich, R. M. (2014). Landscape dynamics of mountain pine beetles. *Forest Science, 60*(3), 464–475. https://doi.org/10.5849/forsci.13-064

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology, 83*(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Massey, C., & Wygant, N. (1954, January). Biology and control of the Engelmann spruce beetle in Colorado. *The Bark Beetles, Fuels, and Fire Bibliography*. Retrieved from http://digitalcommons.usu.edu/barkbeetles/193

Mattson, W. J., & Addy, N. D. (1975). Phytophagous insects as regulators of forest primary production. *Science, 190*(4214), 515–522. http://doi.org/10.1126/science.190.4214.515

Mattson, W. J., & Haack, R. A. (1987). The role of drought in outbreaks of plant-eating insects. *BioScience, 37*(2), 110–118. http://doi.org/10.2307/1310365

Mietkiewicz, N., Kulakowski, D., & Veblen, T. T. (2018). Pre-outbreak forest conditions mediate the effects of spruce beetle outbreaks on fuels in Subalpine forests of Colorado. *Ecological Applications, 28*(2), 457–472. https://doi.org/10.1002/eap.1661

Nishii, R., & Tanaka, S. (2012). Modeling and inference of forest coverage ratio using zero-one inflated distributions with spatial dependence. *Environmental and Ecological Statistics, 20*(2), 315–336. https://doi.org/10.1007/s10651-012-0227-y

Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers, 51*(1), 111–126. https://doi.org/10.1007/s00362-008-0125-4

Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis, 56*(6), 1609–1623. https://doi.org/10.1016/j.csda.2011.10.005

Park Williams, A., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C. A., Meko, D. M., … McDowell, N. (2013). Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change, 3*(3), 292–297. https://doi.org/10.1038/nclimate1693

Plummer, M. (2004). *JAGS: Just Another Gibbs Sampler*. Retrieved from http://www3.jouy.inra.fr/miaj/public/matrisq/Contacts/applibugs.07_11_08.plummer.pdf

Plummer, M. (2013). *Rjags: Bayesian graphical models using MCMC*. R Package Version 3.

Pongpattananurak, N., Reich, R. M., Khasla, R., & Aguirre-Bravo, C. (2012). Modeling the spatial distribution of soil texture in the state of Jalisco, Mexico. *Soil Science Society of America Journal, 76*(1), 199–209. https://doi.org/10.2136/sssaj2011.0180

Price, P. W. (1991). Patterns in communities along latitudinal gradients. In P. W. Price, T. M. Lewinson, G. W. Fernandes & W. W. Benson (Eds.), *Plant animal interactions: Evolutionary ecology of tropical and temperate region* (pp. 51–68). New York: John Wiley and Sons.
R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raffa, K. F., Aukema, B. H., Bentz, B. J., Carroll, A. L., Hicke, J. A., Turner, M. G., & Romme, W. H. (2008). Cross-scale drivers of natural disturbances prone to anthropogenic amplification: The dynamics of bark beetle eruptions. *BioScience*, 58(6), 501–517. https://doi.org/10.1641/B580607

Reich, R. M., John, E. L., & Hughes, K. (2016). Host-environment mismatches associated with subalpine fir decline in Colorado. *Journal of Forestry Research*, 27(April), 1177–1189. https://doi.org/10.1007/s11676-016-0234-1

Reich, R. M., Lundquist, J. E., & Acciavatti, R. E. (2014). Influence of climatic conditions and elevation on the spatial distribution and abundance of Trypodendron Ambrosia beetles (Coleoptera: Curculionidae: Scolytinae) in Alaska. *Forest Science*, 60(2), 308–316. https://doi.org/10.5849/forsci.12-092

Rennenberg, H., Loreto, F., Polle, A., Brilli, F., Fares, S., Beniwal, R. S., & Gessler, A. (2006). Physiological responses of forest trees to heat and drought. *Plant Biology (Stuttgart)*, 8(5), 556–571.

Reynolds, K. M., & Holsten, E. H. (1996). Classification of spruce beetle hazard in Lutz and Sitka spruce stands on the Kenai Peninsula, Alaska. *Forest Ecology and Management*, 84(1-3), 251–262.

Rhoades, D. F. (1983). Herbivore population dynamics and plant chemistry. In R. F. Denno & M. S. McClure (Eds.), *Herbivore population dynamics and plant chemistry* (pp. 155–200). Orlando, FL: Academic Press.

Ribeiro, P. J., Jr., & Diggle, P. J. (2001). geoR: A package for geostatistical analysis. *R News*, 1(2), 14–18.

Ridout, M., Demétrio, C. G. B., & Hinde, J. (1998). Models for count data with many zeros. In *Proceedings of the XIXth International Biometric Conference*, 19:179–192. Retrieved from http://sada2013.sciencesconf.org/conference/sada2013/pages/ridout98ZIPexcesszerosmodelsreview.pdf

Schmid, J., & Frye, R. (1977, January). Spruce beetle in the rockies. *The Bark Beetles, Fuels, and Fire Bibliography*. Retrieved from http://digitalcommons.usu.edu/barkbeetles/86

Spiegelhalter, D. J., Nicola, G. B., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Stadelmann, G., Bugmann, H., Wermelinger, B., Meier, F., & Bigler, C. (2013). A predictive framework to assess spatio-temporal variability of infestations by the European spruce bark beetle. *Ecography*, 36(11), 1208–1217. https://doi.org/10.1111/j.1600-0587.2013.00177.x

Weed, A. S., Ayres, M. P., & Hicke, J. A. (2013). Consequences of climate change for biotic disturbances in North American forests. *Ecological Monographs*, 83(4), 441–470.

White, T. C. (1993). *The inadequate environment: Nitrogen and the abundance of animals* (Vol. 1). Berlin Heidelberg: Springer-Verlag.

---

**How to cite this article:** Suksavate W, Wei Y, Lundquist J. Studying the probability of spruce beetle caused mortality in Colorado’s spruce forests using Bayesian hierarchical models. *Natural Resource Modeling*. 2021;34:e12290. https://doi.org/10.1111/nrm.12290