Running title: Mountain pine beetle management assessment
Number of words: ~4935
Number of tables: 2
Number of figures: 9
Number of references: 46

Management assessment of mountain pine beetle infestation in Cypress Hills, SK

Mélodie Kunegel-Lion[1,*], **Rory L. McIntosh**[2], **& Mark A. Lewis**[3].

1. *Department of Biological Sciences, University of Alberta, CW 405 Biological Sciences Bldg, Edmonton, AB T6G 2E9, Canada. kunegel@ualberta.ca*
2. *Forest Service Branch, Saskatchewan Ministry of Environment, Box 3003 McIntosh Mall, Prince Albert, SK S6V 6G1, Canada. rory.mcintosh@gov.sk.ca*
3. *Department of Biological Sciences, University of Alberta, CW 405 Biological Sciences Bldg, Edmonton, AB T6G 2E9, Canada and Department of Mathematical and Statistical Sciences, University of Alberta, 632 CAB, Edmonton, AB T6G 2G1, Canada. mark.lewis@ualberta.ca*

*Correspondence author.

Date of submission: 6th November 2018

Résumé

Insect epidemics such as the mountain pine beetle (MPB) outbreak have a major impact on forest dynamics. In Cypress Hills, Canada, the Forest Service Branch of the Saskatchewan Ministry of Environment aims to control as many new infested trees as possible by conducting ground-based surveys around trees infested in previous years. Given the risk posed by MPB, there is a need to evaluate how well such a control strategy performs. Therefore, the goal of this study is to assess the current detection strategy compared to competing strategies (random search and search based on model predictions via machine learning), while taking management costs into account. Our model predictions via machine learning used a generalized boosted classification tree to predict locations of new infestations from ecological and environmental variables. We then ran virtual experiments to determine control efficiency under the three detection strategies.

The classification tree predicts new infested locations with great accuracy (AUC = 0.93). Using model predictions for survey locations gives the highest control efficiency for larger survey areas. Overall, the current detection strategy performs well but control could be more efficient and cost-effective by increasing the survey area as well as adding locations given by model predictions.

Keywords : beetle pressure, control efficiency, detection, insect epidemics, management cost

Introduction

The mountain pine beetle (MPB ; *Dendroctonus ponderosae*, Hopkins 1902) epidemic has caused extensive mortality in North American pine forests, which is in conflict with human objectives in many places. At a large scale, the epidemic is linked to climate change as well as population dynamics that shift intermittently between endemic and epidemic states (Carroll *et al.*, 2004; Shore *et al.*, 2006; Raffa *et al.*, 2008; Preisler *et al.*, 2012). MPB's spread is unaffected by most environmental barriers such as low mountain ranges and fragmented forests thanks to its ability to disperse long distances (de la Giroday *et al.*, 2012; Bentz *et al.*, 2016). To better control MPB populations, we need to determine areas at risk and assess the efficiency of current detection strategies.

The MPB is a bark beetle that infests and kills various species of pines. In North America, lodgepole pine (*Pinus contorta*, Dougl. ex Loud. var. *latifolia* Engelm) is the primary MPB host although MPB is a threat to almost all pine species (Safranyik & Carroll, 2006). During an epidemic, MPB individuals coordinate their attacks, using aggregation pheromones, to form a "mass attack" and overwhelm the defences of large and healthy trees (Bordon, 1982). Therefore, an epidemic population of MPB presents a threat to healthy pine stands.

The MPB is primarily univoltine, meaning that each new generation is produced over a year (see Mitton & Ferrenberg, 2012; Bentz & Powell, 2014; Mitton & Ferrenberg, 2014). In summer, the beetles disperse and reproduce, and the females lay eggs in galleries they excavate under the bark. Individuals usually overwinter as larvae. In spring, they resume their development and finally emerge as adults later in the summer (Safranyik & Carroll, 2006). Trees are seriously injured by the gallery excavation process and the development of MPB larvae and their associated blue stain fungi, and generally die and turn red by the end of the MPB life-cycle. During the following years, attacked trees become grey. As a result, red-top trees, infested during the summer of the previous year are easily spotted during aerial surveys of stands, becoming good proxy for the status of the previous year's MPB infestation.

At a landscape level, two types of dispersal strategies have been observed for MPB (Safranyik & Carroll, 2006; Robertson *et al.*, 2007) : long-distance dispersal, passive downwind flight over the canopy, and short-distance dispersal, active flight a few meters above ground. Researchers estimate the short-distance dispersal range to be within a stand (Safranyik & Carroll, 2006) at the order of 20 to 50 meters, although some beetles can go as far as 100 meters (Robertson *et al.*, 2007). By way of contrast, long-distance dispersal range is tens to hundreds of kilometres (Safranyik & Carroll, 2006; Jackson *et al.*, 2008). While short-distance dispersal is much more common than long-distance dispersal (Safranyik *et al.*, 1989; Chen & Walton, 2011), the MPB's epidemic behaviour associated with outbreaks arising from long-distance dispersal can pose a threat to entire regions of pine forests.

In Canada, since 2006, a local MPB epidemic has emerged in the Cypress Hills area, located in the southwest of Saskatchewan and southeast of Alberta. The Cypress Hills inter-provincial park comprises the West Block, divided between Alberta (219 km^2) and Saskatchewan (126 km^2), and the

Center Block, in Saskatchewan (58 km^2). For the purpose of this paper, our study focuses on the Saskatchewan portion of the park. Therefore the use of "the park" and "Cypress Hills" in the text refers to the Saskatchewan portion. The local MPB population is endemic to the park and probably came from southern populations in Montana, USA (R. L. McIntosh, pers. comm.). It could have been partly sustained by beetle flights from the south and west. Indeed, during spring and summer, during MPB dispersal, the dominant wind comes from the southwest.

Studying and controlling MPB in the Cypress Hills area is essential for two reasons. First, as an inter-provincial park and national heritage, Cypress Hills has significant natural, economic and cultural values. Second, even though this park is somewhat isolated compared to lodgepole and jack pine ranges (Little, 1971; Cullingham *et al.*, 2012), the presence of a MPB epidemic, in association with the long-distance dispersal ability of the insect and the wind direction, makes the Cypress Hills area a possible stepping-stone facilitating the infestation of the remainder of Saskatchewan and regions further east. Therefore, there is an urgent need for analysis of management and for infestation prediction in Cypress Hills.

Aware of the need for management, the Forest Service Branch of the Saskatchewan Ministry of Environment has implemented a "zero-tolerance" policy designed to catch and control as many short-distance infestations as possible. This requires intensive surveillance to implement early detection and rapid aggressive response actions. The policy operates according to the following procedure. In early fall, after MPB have colonized new trees, an aerial survey of the park extent is conducted to collect geo-referenced data on potential red-top trees, which are dead or dying trees infested by MPB the previous year. These are later ground-truthed for MPB attacks. Then, 50 meter-radius circular survey plots are drawn around each of the red-top trees confirmed to have been killed by MPB. The survey plots are searched for green infested trees, which are trees recently infested by MPB during the summer. These are later controlled in late fall/winter which usually consists of felling and burning massively infested trees, ensuring that beetles are killed. The survey plot can be spatially extended if green infestations are spotted close to the plot's limits (Saskatchewan Ministry of Environment, 2016). In addition to these measures, areas presenting high densities of red-top trees are entirely surveyed and controlled. No detected infestations are left untreated. Such intensive control is expensive. Therefore, there is a need to determine how well this strategy is working.

Given this management strategy and the MPB context in Canada, our study aims to answer the question : Are there ways to improve detection strategies without increasing management costs ? If managers completely removed infested trees coming from MPB short-distance dispersal inside the park, the remaining source of infestation would be long-distance dispersal events from outside the park which are often considered spatially random when observed at a small scale (Long *et al.*, 2012; Powell *et al.*, 2018). Therefore, we hypothesize that a random search would be as efficient as a local search around red-top trees. Moreover, we hypothesize that, if other factors than distance to previous infestations influence the location of new infestations, then a search based on predictions from such factors would be more efficient than a local search around red-top trees. However, the management survey might not be big enough to include all infestations from short-distance dispersal events. Therefore, we make the third hypothesis that, as the search area increases, the detection efficiency will increase too.

Material and methods

MPB predictions

To predict MPB infestation a year ahead in Cypress Hills, we used the generalized boosted classification model which is a machine learning algorithm. Boosted classification trees generate results with an excellent fit for a binary response by successively fitting a tree to the previous tree's residuals to reduce significantly the final error variance (StatSoft, 2013).

Data

The covariates and response variable values were distributed discretely in space and time. We applied a grid of 18 317 cells of size 100m$\times$100m to the Cypress Hills park extent. For each cell for each year, the observation consisted of a set of environmental and ecological covariates plus the response variable. The response variable was the presence/absence of MPB derived from the presence/absence of green infested trees in each cell of the grid based on data from the Forest Service ground survey. From the Forest Service surveys, we got the locations of green infestations controlled by managers and we deduced which trees had been green infested in the previous year using the red-top trees.

We used 14 covariates related to topography, weather, vegetation, and beetle pressure (Table 1). The weather variables were : the highest maximum daily temperature over the year, the overwinter survival probability of the larvae (Régnière & Bentz, 2007), and the average daily relative humidity in spring. Indeed, MPB dispersal is reduced with high temperatures (Safranyik & Carroll, 2006). The minimum temperatures in fall and winter impact MPB survival if the vulnerable stages–developing in the fall and at the end of the winter–are exposed to extreme temperatures (Cole, 1981; Safranyik & Carroll, 2006; Régnière & Bentz, 2007). Drought in the spring reduces pines' ability to defend themselves and increase MPB attacks' success rate (Safranyik, 1978; Creeden *et al.*, 2014; Sidder *et al.*, 2016). Additionally, MPB individuals need at least 833 degree-days above 5.5°C over a year to complete their growth (Safranyik *et al.*, 1975; Carroll *et al.*, 2006; Safranyik *et al.*, 2010). In the park, over the time period studied, the minimum number of degree-days above 5.5 °C was 923, which is above the threshold and so degree-days was not included in our model. Furthermore, high numbers of degree-days are not an issue as MPB rarely present multivoltinism (Bentz & Powell, 2014). We included the MPB presence at the same location and in the neighbourhood the year before in order to take into account the spatio-temporal autocorrelation of the data (Fig. 1). The beetle pressure from outside the park was represented by the distance to the park southern border (illustrated on Fig. 2) which was close to external infestations not managed by the Forest Service and potential sources of MPB. The rest of the variables included in the model were : pine cover, latitude, longitude, year, elevation, slope, and northerness and easterness derived from the aspect.

Topography data came from the Canadian Digital Elevation Map downloaded from the Geogratis website (geogratis.cgdi.gc.ca). We generated weather variables with the BioSIM software (Régnière *et al.*, 2014) at the location of each grid cell centroid. BioSIM uses data from surrounding weather stations and interpolates the weather variable values at each location of interest using a digital elevation map. The vegetation data came from Beaudoin *et al.* (2014). The authors computed these data from a 2001 MODIS imagery, and the vegetation parameters were assumed constant over our time period.

We used data from the years 2007 to 2015. Randomly, we chose 75% of these data, years combined, *i.e.* 149 278 observations, to train the model. The remaining 25%, 49 502 observations, were used to validate the model.

Generalized Boosted Model

We trained the generalized boosted classification model using the `gbm` function of the `R` package `gbm` (Ridgeway, 2015) on the 14 covariates in the training set. The process analyzed the performance of 50 000 classification trees and performed a 10-fold cross-validation in order to find the best classifier. The algorithm implemented in the `gbm` function consisted of reducing a loss function between the observed and the predicted response values using Friedman's Gradient Boosting Machine (Ridgeway, 2015). The loss function was represented by a Bernoulli error distribution, which is adapted to a binary response. The `gbm` function output provides the probability of MPB presence at each location. We tested the accuracy of the model's prediction using the area under the receiver operating characteristic curve (AUC ; Metz, 1978; Bradley, 1997), the false positive and false negative rates, and the misclassification rate which is the percentage of misclassified instances by the model. A receiver operating characteristic (ROC) curve (Metz, 1978) depicts, for a range of probability thresholds, the true positive rate (or 1 - false negative rate, also referred to as sensitivity) against the false positive rate (also referred to as 1 - specificity). We used Youden's method (Youden, 1950) to determine the probabilities threshold which selects the farthest point from the diagonal on the ROC curve. A high AUC ($0 \leq \text{AUC} \leq 1$) represents a good performance of a binary classifier in terms of correspondence between observed and predicted values.

Assessing management

Data

To assess the detection strategies, we needed the exact locations of red-top trees for a focus year and the following year. In 2011 to 2013, the data from the Forest Service included an exhaustive survey of red-top trees' locations and the number of green infestations controlled around each red-top tree. The other years included infested areas in which red-top trees' locations were not specified. For this reason, we only used data from 2011 and 2012 for this analysis. Furthermore, the years 2011 and 2012 happened to have a similar number of red-top trees/survey plots : 292 for 2011 and 284 for 2012, which made the two years comparable.

For controlled green infestations, we used the location of the circular plot centres (± 50 meters compared to the real locations of green infestations). For uncontrolled green infestations outside of survey plots, we used the location of red-top trees the year after. The total number of green infestations was

644 for 2011 and 936 for 2012.

Simulated detection strategies

To calculate the efficiency of the detection strategies, we simulated virtual experiments. For each year, we counted the number of green infestations in increasing virtual survey areas for three different strategies : 1) local search in circular plots of varying radius around red-top trees (similar to the current Forest Service strategy), 2) search in circular plots of varying radius randomly located in space, and 3) search in a varying number of 100×100m square plots placed at locations predicted by the boosted classification tree. In the predictions strategy, we used 100×100m square plots and not circular plots to match as much as possible the predicted locations from the classification tree. For the local and random searches, we used circular plots of increasing radius : from 50 to 100 meters by increment of 5, from 110 to 150 meters by increment of 10, 200, and 300 meters.

To be able to compare similar survey areas among detection strategies, we needed to be able to fix the number of search locations, and therefore the search area, from the classification tree output. We could simply select a certain number of locations with the highest probabilities. However, if the number of selected locations is small like it is the case here, some locations with relatively high probabilities might not be chosen whereas locations with slightly higher probabilities due to random noise will. To bypass this issue, we introduced some noise by randomly sampling the locations using the model probabilities to the power of 3 as weight. We investigated the impact of variation in this exponent value in Appendix A. For the random and prediction strategies, we performed 500 simulations for each year.

Control efficiency

We calculated control efficiency for each year for each survey area with the equation

$$\text{control efficiency} = \frac{\#\text{ green infestations controlled}}{\text{total }\#\text{ green infestations in the park}}. \tag{1}$$

From the area controlled (*i.e.* the sum of every survey plot area), we obtained the net survey area by removing the overlapping areas. For each year,

$$\text{net survey area} = \begin{cases} \#\text{ plots} \times \pi r^2 - \text{overlaps} & \text{for local/random} \\ \#\text{ square plots} \times 100^2 & \text{for predictions} \end{cases}. \tag{2}$$

We then determined the relationship between net survey area and control efficiency. This was achieved by fitting a non-linear function, using the `nls` function of the `R` package `stats`, to control efficiency versus net survey area in the two cases : local search around red-top trees, local control efficiency $= f_{\text{local}}(\text{net survey area})$, and model predictions strategy, prediction control efficiency $= f_{\text{prediction}}(\text{net survey area})$. For the random search case, we fitted a linear function using the `lm` function of the `R` package `stats` : random control efficiency $= f_{\text{random}}(\text{net survey area})$.

Management cost

To determine cost-effective recommendations for managers, we also examined the relationship between net survey area and management cost. The management cost variable included the cost of aerial survey, the cost of control, and the cost of surveying all non-overlapping 50 meter-radius circular plots. It was available for the years 2010 to 2015. Within each year, the cost per unit (control cost per tree and survey cost per plot) did not vary depending on the location. However, since the cost per unit varied among years due to economic fluctuations, we took the median cost per unit over the years 2010 to 2015 and multiplied it for each year by the number of units in each category (number of controlled trees and circular plots per year). Thus, for each year :

$$\begin{aligned}\text{management cost} = \;& \text{median aerial survey cost}\\ &+ \text{median control cost per tree} \times \# \text{ trees controlled}\\ &+ \text{median circular plot survey's cost} \times \# \text{ plots.} \qquad (3)\end{aligned}$$

The number of units in each category was available for the years 2006 to 2015. Therefore, we determined management cost values for 2006 to 2015. As a result, although total cost did vary year to year, the cost per plot and per tree did not. We fitted a linear regression line to the relationship between management cost and total area surveyed with circular plots (management cost = g(total area surveyed with circular plots) where $g(.)$ is a straight line function) using the `lm` function of the `R` package `stats`. The total area surveyed with circular plots does not contain overlaps (Saskatchewan Ministry of Environment, 2016) so this is equal to the net survey area with radius $= 50$ (equation (2)). To get to the next step, we assumed that the management cost increases proportionally with the plot area. Thus, the cost of the total area from several survey plots is equal to the cost of the area of a single much larger survey plot. Hence, management cost $= g($total area surveyed with circular plots$)$ became management cost $= g($net survey area$)$. We then defined the "management cost per controlled tree" which is the management cost divided by the control efficiency for one year. Note that this cost per controlled tree is scaled by the total number of infestations in the park for each year. We explored the relationship between management cost per controlled tree and net survey area using the two regression equations : control efficiency $= f($net survey area$)$ and management cost $= g($net survey area$)$:

$$\begin{aligned}\text{management cost per controlled tree} &= \frac{\text{management cost}}{\text{control efficiency}}\\ &= \frac{g(\text{net survey area})}{f(\text{net survey area})}.\end{aligned} \tag{4}$$

The net survey area value corresponding to the minimum management cost per controlled tree would be the optimal area to survey.

However, one could also assign a cost θ to a missed green infestation as it would leads to several green infestations the following year. The cost of a missed green infestation θ times the number of missed green infestations is the avoided cost as it is the amount that would be saved in the future if these trees were actually controlled instead of being missed. In other words, θ is the marginal cost added to the following year cost if one green infestation is left and produce new infestations. Therefore, the total cost was defined as

$$\begin{aligned}\text{total cost} &= \text{management cost} + \text{avoided cost} \\ &= \text{management cost} + \theta \times \#\ \text{missed infestations}.\end{aligned} \tag{5}$$

Thus, the total cost per controlled tree is the management cost plus the avoided cost divided by the control efficiency. Again, note that this cost per controlled tree is scaled by the total number of green infestations for each year. We then compared the optimal survey area for the management cost and for the total cost depending on the strategy used. We also investigated the dependence of the optimal survey area on θ in Appendix B.

Results

MPB predictions

The generalized boosted classification model has a good predictive ability (Fig. 3) : the AUC value is 0.927. The probability threshold chosen from Youden's index is 0.003, which means that it is optimal in terms of misclassified instances to consider any probability value above this threshold as an infestation. Using this threshold, we calculated the confusion matrix (Table 2). The false negative and false positive rates calculated from it are, respectively, 0.187 and 0.118, which means that 18.7% of the infested locations are wrongly classified as non-infested and 11.8% of the non-infested locations are wrongly classified as infested. Additionally, the misclassification rate was 0.119 which means that 11.9% of the model results were misclassified compared to the observations.

We calculated the variables' impact on the classification tree output (*i.e.* relative importance). The MPB presence in the same location the year before is the most important variable (relative importance = 0.60), followed by the MPB pressure from neighbouring cells (0.26), the distance to the southern infested border of the park (0.10), and the overwinter survival (0.02). The remaining variables have each a relative importance below 0.01.

Assessing management

When increasing the radius of the circular plots or the number of square plots, and thus the area surveyed, the control efficiency increases and saturates for the local and predictions strategies (Fig. 4). The control efficiency of the search around random locations increases linearly with the net survey area. The local and predictions strategies are more efficient than the random search. For example, the local search reaches between 55.9% and 71.2% control efficiency at a 50-meters radius (current strategy), the predictions strategy between 54.3% and 63.3%, whereas it reaches only 0.01% control efficiency for the random search at the same survey area. For survey areas larger than those in the current strategy ($\sim$ 2 200 000 m^2), the predictions control efficiency is higher than the local control efficiency (Fig. 4). For example, for a survey area corresponding to 70-meters radius for the local search ($\sim$ 3 900 000 m^2), the control efficiency is 60.6% to 73.7% for the local search and 81.9% to 84.4% for the predictions strategy.

The management cost increases linearly with the net survey area (Fig. 5). We numerically obtain the net survey area values corresponding to the minimum management cost per controlled tree over the extent of net survey area values studied for the local and predictions strategies for 2011 and 2012 : 2 178 332 to 2 225 780 m^2 (Fig. 6a). We obtain the matching radius 50 meters using equation (2) for the local search. However, it is highly probable that the cost of missing a green infestation θ is non-negligible. As the management cost increases with the survey area and the avoided cost decreases, the total cost shows a minimum value larger than zero (Fig. 7 for $\theta = 1000$). Therefore, the minimum total cost per controlled tree with $\theta = 1000$ gives survey area values ranging from 3 010 378 to 5 062 968 m^2 and corresponding to the radius 60 to 65 meters using equation (2) for the local search (Fig. 6b).

Discussion

MPB infestations can be well predicted in space using a generalized boosted classification tree and variables related to the location of previous year infestations. A detailed analysis of the impact of survey areas on the control efficiency shows that combining an increase in survey area with a change in detection strategy leads to more cost-effective control.

MPB predictions

Generally, generalized boosted classification approaches often give better predictive accuracies than generalized linear approaches (Marmion *et al.*, 2009; Youssef *et al.*, 2016). Here, the percentage of correctly classified cells, $1-$ misclassification rate, is 84.9%. In comparison, Aukema *et al.* (2008) reported a predictive accuracy of 78% for a one-year ahead forecast using a spatial-temporal autologistic regression model on similar variables. At large scales (respectively 12x12 km and 1x1 km grid cell size in Aukema *et al.*, 2008; Preisler *et al.*, 2012), beetle pressure has a great impact on new infestations so it is not surprising to find indications that this is also the case in our results at a smaller scale.

While classification tree approaches can be used for prediction, they cannot be used to determine the actual impact of covariates on the response. Indeed, a classification approach, such as decision trees or boosted classification trees, often provide a relative importance index for each covariate, but this relative importance is an index of performance that depends highly on tree structures. A classification method does not test the impact of a covariate on the response like a traditional statistical method would, but rather attempts to explain the response by a sequence of binary choices among covariate values. However, it makes sense that environmental variables have less impact on the MPB presence than beetle pressure given that a small-size study area is usually relatively homogeneous.

Machine learning algorithms are widely used to detect/predict species locations (Marmion *et al.*, 2009) but few quantitatively compare the result to non-modelling/expert-knowledge methods like we did in this study (*e.g.* Boissard *et al.*, 2008).

Assessing management

The management assessment results show that the current detection strategy (searching in a 50 meter-radius plot around previous infestations) is efficient, but that using a larger survey area and a different strategy would improve efficiency. Robertson *et al.* (2007) found that 20 to 50 meters is the most common dispersal range but that MPB can go farther. These few individuals that go farther, and therefore are not removed during control, might be sufficient to sustain the population in the stand. MPB is subject to a strong Allee effect (Logan *et al.*, 1998; Goodsman *et al.*, 2016) : at low beetle densities, a certain number of individuals is needed for a successful mass attack. Below this threshold, the attack is unsuccessful and the beetles either do not survive or fall back into the endemic population phase. The transition between endemic and epidemic population phases highly depends on both intrinsic and extrinsic factors which are subjected to a lot of uncertainty, making the transition forecast problematic (Cooke & Carroll, 2017).

Because of the existence of this threshold, local densities of beetles are important to infestation success. For that reason, Strohm *et al.* (2016) found that increasing search radius is more important than increasing search effectiveness, which is the percentage of infestations found within a survey area. Indeed, search effectiveness does not need to be flawless to decrease the beetle number below the Allee threshold. However, if the search radius is too small, enough beetles can disperse from neighbouring locations and successfully infest trees. For a search effectiveness of approximatively 80%, Strohm *et al.* (2016) show that MPB population size would decrease only if the search radius increases despite increases in search effectiveness. In Cypress Hills, for 2011 and 2012, we estimated the search effectiveness at 89%. This supports our recommendation to increase the survey area. Overall, Strohm *et al.* (2016) show that the search plot size of the Alberta management strategy (similar to Saskatchewan's strategy) was not large enough to reach the desired goal of reducing MPB population by 80% (Alberta Sustainable Resource Development, 2007) and the present study shows results consistent with this conclusion.

Local search around red-top trees, associated with short-distance dispersal, is a more efficient method than the random search, associated with random events from long-distance dispersal. This suggests that, despite intensive management, short-distance dispersal is still the main MPB dispersal strategy in Cypress Hills. However, a mechanistic model, such as the ones developed in Heavilin & Powell (2008), Rodrigues *et al.* (2015) and Goodsman *et al.* (2016), or the method described in Chen & Walton (2011), adapted for this area could likely give more insights on the subject by, in particular, quantifying the importance of both dispersal strategies.

An alternative to the local search around red-top trees is to survey locations with high predicted infestation probabilities. For a survey area larger than the one corresponding to the current strategy, it becomes more efficient to use the predictions strategy rather than the local strategy. This could be explained by the spatial scale of our model predictions. One $100{\times}100$m grid cell area and one 50 meter-radius circular plot area have the same order of magnitude. For a similar number of plots, the previous infestation at the same location decides for half of the model predictions results according to the relative importance whereas a red-top tree is always at the center of a circular plot. As the survey area increases, more of the red-top trees are included in the predictions survey in addition to other susceptible locations whereas the number of red-top trees included in the local survey does not change. Therefore, while in the local survey fewer and fewer green infestations are present the further away from the red-top tree, the predictions survey focuses on additional high risk locations chosen according to other variables, mainly the distance to the southern infested border, increasing the chance of finding more green infestations. One could combine both strategies : surveying first around red-top trees than adding extra survey plots in predicted areas that were not already surveyed until the alloted budget is reached.

Introducing a management cost allows for more informed decisions upon which to choose survey area size and detection strategy. Indeed, there is a minimum cost per controlled tree that corresponds to an optimal survey area larger than zero. This optimal survey area varies with the cost of missing a green infestation which can be calculated, for example, by the cost of a circular survey plot plus the cost of removing a certain number of new green infestations due to this red-top tree.

Limitations

A potential limitation of this work is the assumption that the cost associated with several 50 meter-radius plots is equivalent to the cost of one much larger plot of the same total area, and that this relationship is linear, even for areas as large as 20% of the park surface. One could also assume that the relationship's slope would decrease as survey locations are closer in space and managers spend less money and time travelling between locations. These savings seem negligible, nonetheless, it would decrease the slope of the relationship between cost per controlled tree and survey area at larger survey areas. However, it would probably have little impact on the location of the minimum cost and thus the optimal survey area size.

Another limitation is that we only undertook the analysis for years with a number of red-top trees approximately equal to 300 as only data for these years were available. The survey area values are directly linked to the number of survey plots and, thus, the number of red-top trees for each year. Therefore, the survey area values are not directly applicable to years with a different number of red-top trees, although the curve patterns would be similar. The results also vary with the ratio total number of green infestations to number of red-top trees. This ratio was larger in 2012 than 2011. However, we scaled most of the results by the total number of green infestation to allow a fair comparison of both years.

Furthermore, the selection of only two consecutive years of data makes the analysis potentially susceptible to bias due, for example, to particular weather conditions or to the specific details of implementation of management work for these two years. To minimize the latter, however, a detailed survey protocol is implemented.

Conclusion

The control efficiency in Cypress Hills could be slightly increased for a smaller cost, which includes the future savings made by controlling an infested tree now rather than several ones the following year. This would be done by engaging more management resources, such as an increased survey area, in combination with using a search strategy that exploits criteria other than the location of red-top trees.

Acknowledgements

We thank Brian Poniatowski for the useful discussions around technical and surveillance aspect in Cypress Hills and Samuel Fischer for his valuable insights on the methodology and results. We also thank the Lewis Research Group for helpful feedback regarding modelling and analysis.

This research was supported by a grant to M.A.L. from the Natural Science and Engineering Research Council of Canada (grant no. NET GP 434810-12) to the TRIA Network, with contributions from Alberta Agriculture and Forestry, Foothills Research Institute, Manitoba Conservation and Water Stewardship, Natural Resources Canada-Canadian Forest Service, Northwest Territories Environment and Natural Resources, Ontario Ministry of Natural Resources and Forestry, Saskatchewan Ministry of Environment, West Fraser and Weyerhaeuser. M.A.L. is also grateful for the support through the NSERC Discovery and the Canada Research Chair programs.

Références

Alberta Sustainable Resource Development (2007) Mountain Pine Beetle Management Strategy.

Aukema, B.H., Carroll, A.L., Zheng, Y., Zhu, J., Raffa, K.F., Dan Moore, R., Stahl, K. & Taylor, S.W. (2008) Movement of outbreak populations of mountain pine beetle : influences of spatiotemporal patterns and climate. *Ecography*, **31**, 348–358. doi : 10.1111/j.0906-7590.2007.05453.x.

Beaudoin, A., Bernier, P., Guindon, L., Villemaire, P., Guo, X., Stinson, G., Bergeron, T., Magnussen, S. & Hall, R. (2014) Mapping attributes of Canada's forests at moderate resolution through kNN and MODIS imagery. *Canadian Journal of Forest Research*, **44**, 521–532. doi : 10.1139/cjfr-2013-0401.

Bentz, B.J., Duncan, J.P. & Powell, J.A. (2016) Elevational shifts in thermal suitability for mountain pine beetle population growth in a changing climate. *Forestry*, **89**, 271–283. doi : 10.1093/forestry/cpv054.

Bentz, B.J. & Powell, J.A. (2014) Mountain Pine Beetle Seasonal Timing and Constraints to Bivoltinism : A Comment on Mitton and Ferrenberg, "Mountain Pine Beetle Develops an Unprecedented Summer Generation in Response to Climate Warming.". *The American Naturalist*, **184**, 787–796. doi : 10.1086/678405.

Boissard, P., Martin, V. & Moisan, S. (2008) A cognitive vision approach to early pest detection in greenhouse crops. *Computers and Electronics in Agriculture*, **62**, 81–93. doi : 10.1016/j.compag.2007.11.009.

Bordon, J.H. (1982) Aggregation pheromones. *Bark Beetles in North American Conifers A System for the Study of Evolutionary Biology* (eds. J.B. Mitton & K.B. Sturgeon), pp. 74–139. University of Texas Press, Austin, TX.

Bradley, A.P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30**, 1145–1159.

Carroll, A.L., Shore, T.L. & Safranyik, L. (2006) Direct control : theory and practice. *The mountain pine beetle : a synthesis of biology, management, and impacts on lodgepole pine.* (eds. L. Safranyik & B. Wilson), pp. 155–172. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia.

Carroll, A.L., Taylor, S.W., Régnière, J. & Safranyik, L. (2004) Effects of climate change on range expansion by the mountain pine beetle in British Columbia. Information Report BC-X-399, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Kelowna, British Columbia, Canada.

Chen, H. & Walton, A. (2011) Mountain pine beetle dispersal : spatiotemporal patterns and role in the spread and expansion of the present outbreak. *Ecosphere*, **2**, 1–17. doi : 10.1890/ES10-00172.1.

Cole, W.E. (1981) Some risks and causes of mortality in mountain pine beetle populations : A long-term analysis. *Researches on Population Ecology*, **23**, 116–144. doi : 10.1007/BF02514096.

Cooke, B.J. & Carroll, A.L. (2017) Predicting the risk of mountain pine beetle spread to eastern pine forests : Considering uncertainty in uncertain times. *Forest Ecology and Management*, **396**, 11–25. doi : 10.1016/j.foreco.2017.04.008.

Creeden, E.P., Hicke, J.A. & Buotte, P.C. (2014) Climate, weather, and recent mountain pine beetle outbreaks in the western United States. *Forest Ecology and Management*, **312**, 239–251. doi : 10.1016/j.foreco.2013.09.051.

Cullingham, C.I., James, P.M.A., Cooke, J.E.K. & Coltman, D.W. (2012) Characterizing the physical and genetic structure of the lodgepole pine x jack pine hybrid zone : mosaic structure and differential introgression. *Evolutionary Applications*, **5**, 879–891. doi : 10.1111/j.1752-4571.2012.00266.x.

de la Giroday, H.M.C., Carroll, A.L. & Aukema, B.H. (2012) Breach of the northern Rocky Mountain geoclimatic barrier : initiation of range expansion by the mountain pine beetle. *Journal of Biogeography*, **39**, 1112–1123. doi : 10.1111/j.1365-2699.2011.02673.x.

Goodsman, D.W., Koch, D., Whitehouse, C., Evenden, M.L., Cooke, B.J. & Lewis, M.A. (2016) Aggregation and a strong Allee effect in a cooperative outbreak insect. *Ecological Applications*, **26**, 2623–2636. doi : 10.1002/eap.1404.

Heavilin, J. & Powell, J. (2008) A novel method of fitting spatio-temporal models to data, with applications to the dynamics of Mountain Pine Beetles. *Natural Resource Modeling*, **21**, 489–524. doi : 10.1111/j.1939-7445.2008.00021.x.

Jackson, P.L., Straussfogel, D., Lindgren, B.S., Mitchell, S. & Murphy, B. (2008) Radar observation and aerial capture of mountain pine beetle, Dendroctonus ponderosae Hopk. (Coleoptera : Scolytidae) in flight above the forest canopy. *Canadian Journal of Forest Research*, **38**, 2313–2327. doi : 10.1139/X08-066.

Little, E.L.J. (1971) Atlas of United States trees. Vol 1, conifers and important hardwoods, U.S. Department of Agriculture Miscellaneous Publication 1146.

Logan, J.A., White, P., Bentz, B.J. & Powell, J.A. (1998) Model Analysis of Spatial Patterns in Mountain Pine Beetle Outbreaks. *Theoretical Population Biology*, **53**, 236–255. doi : 10.1006/tpbi.1997.1350.

Long, J.A., Robertson, C., Nathoo, F.S. & Nelson, T.A. (2012) A Bayesian space–time model for discrete spread processes on a lattice. *Spatial and Spatio-temporal Epidemiology*, **3**, 151–162. doi : 10.1016/j.sste.2012.04.008.

Marmion, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecological Modelling*, **220**, 3512–3520. doi : 10.1016/j.ecolmodel.2008.10.019.

Metz, C.E. (1978) Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, **8**, 283–298. doi : 10.1016/S0001-2998(78)80014-2.

Mitton, J.B. & Ferrenberg, S. (2014) Field Studies Demonstrate Bivoltinism in the Mountain Pine Beetle : A Reply to Bentz and Powell. *The American Naturalist*, **184**, 797–801. doi : 10.1086/678404.

Mitton, J.B. & Ferrenberg, S.M. (2012) Mountain Pine Beetle Develops an Unprecedented Summer Generation in Response to Climate Warming. *The American Naturalist*, **179**, E163–E171. doi : 10.1086/665007.

Powell, J.A., Garlick, M.J., Bentz, B.J. & Friedenberg, N. (2018) Differential dispersal and the Allee effect create power-law behaviour : Distribution of spot infestations during mountain pine beetle outbreaks. *Journal of Animal Ecology*, **87**, 73–86. doi : 10.1111/1365-2656.12700.

Preisler, H.K., Hicke, J.A., Ager, A.A. & Hayes, J.L. (2012) Climate and weather influences on spatial temporal patterns of mountain pine beetle populations in Washington and Oregon. *Ecology*, **93**, 2421–2434. doi : 10.1890/11-1412.1.

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**, 501–517. doi : 10.1641/B580607.

Ridgeway, G. (2015) *gbm : Generalized Boosted Regression Models.*

Robertson, C., Nelson, T.A. & Boots, B. (2007) Mountain Pine Beetle Dispersal : The Spatial–Temporal Interaction of Infestations. *Forest Science*, **53**, 395–405.

Rodrigues, L.A.D., Mistro, D.C., Cara, E.R., Petrovskaya, N. & Petrovskii, S. (2015) Patchy Invasion of Stage-Structured Alien Species with Short-Distance and Long-Distance Dispersal. *Bulletin of Mathematical Biology*, **77**, 1583–1619. doi : 10.1007/s11538-015-0097-1.

Régnière, J., Saint-Amant, R. & Béchard, A. (2014) BioSIM 10 – User's manual. , Natural Resources Canada, Canadian Forest Service, Laurentian Forest Center, Québec (QC).

Régnière, J. & Bentz, B. (2007) Modeling cold tolerance in the mountain pine beetle, Dendroctonus ponderosae. *Journal of Insect Physiology*, **53**, 559–572. doi : 10.1016/j.jinsphys.2007.02.007.

Safranyik, L. (1978) Effects of climate and weather on mountain pine beetle populations. *Proceedings of Symposium on Theory and Practice of Mountain Pine Beetle Management in Lodgepole Pine Forests, April 25-27, 1978, Washington State University, Pullman, Washington* (eds. A.A. Berryman, G. Amman & R. Stark), pp. 77–84. College of Forestry, Wildlife and Range Sciences, University of Idaho, Moscow, Idaho.

Safranyik, L., Carroll, A., Régnière, J., Langor, D., Riel, W., Shore, T., Peter, B., Cooke, B., Nealis, V. & Taylor, S. (2010) Potential for range expansion of mountain pine beetle into the boreal forest of North America. *The Canadian Entomologist*, **142**, 415–442. doi : 10.4039/n08-CPA01.

Safranyik, L., Shrimpton, D.M. & Whitney, H.S. (1975) An interpretation of the interaction between lodgepole pine, the mountain pine beetle, and its associated blue stain fungi in western Canada. *Management of Lodgepole Pine Ecosystems Sympo sium Proceedings*, pp. 406 – 428. Washington State University Coop. Extension Service, Pullman, WA, Pullman, Washington, USA.

Safranyik, L., Silversides, R., McMullen, L.H. & Linton, D.A. (1989) An empirical approach to modeling the local dispersal of the mountain pine beetle (Dendroctonus ponderosae Hopk.) (Col., Scolytidae) in relation to sources of attraction, wind direction and speed. *Journal of Applied Entomology*, **108**, 498–511. doi : 10.1111/j.1439-0418.1989.tb00484.x.

Safranyik, L. & Carroll, A.L. (2006) The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. *The mountain pine beetle : a synthesis of biology, management and impacts on lodgepole pine* (eds. L. Safranyik & B. Wilson), pp. 3–66. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, Canada.

Saskatchewan Ministry of Environment (2016) Mountain Pine Beetle operational procedures document.

Shore, T.L., Safranyik, L. & Whitehead, R.J. (2006) Principles and Concepts of Management. *The Mountain Pine Beetle : A Synthesis of Biology, Management, and Impacts on Lodgepole Pine* (eds. L. Safranyik & B. Wilson). Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, Canada.

Sidder, A.M., Kumar, S., Laituri, M. & Sibold, J.S. (2016) Using spatiotemporal correlative niche models for evaluating the effects of climate change on mountain pine beetle. *Ecosphere*, **7**, 1–22. doi : 10.1002/ecs2.1396.

StatSoft, I. (2013) Introduction to boosting trees for regression and classification. *Electronic Statistics Textbook*. StatSoft, Tulsa, OK.

Strohm, S., Reid, M.L. & Tyson, R.C. (2016) Impacts of management on Mountain Pine Beetle spread and damage : A process-rich model. *Ecological Modelling*, **337**, 241–252. doi : 10.1016/j.ecolmodel.2016.07.010.

Youden, W.J. (1950) Index for rating diagnostic tests. *Cancer*, **3**, 32–35. doi : 10.1002/1097-0142(1950)3 :1<32 : :AID-CNCR2820030106>3.0.CO ;2-3.

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S. & Al-Katheeri, M.M. (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, **13**, 839–856. doi : 10.1007/s10346-015-0614-1.

668 # Tables

Table 1 Description and range of the covariates used in the generalized boosted classification model.

Name	Description	Range	Unit
PineCover	Coverage of *Pinus albicaulis* (Whitebark Pine), *Pinus banksiana* (jack pine) and *Pinus contorta* (includes subspecies lodgepole pine and shore pine)	0 – 76.1	%
TMax	The highest maximum daily temperature from September of the previous year to August	27.3 – 36.7	°C
OWS	The overwinter survival probabilities of larvae (Régnière & Bentz, 2007) using a 5-year lookback	0.23 – 0.50	–
RH	Average daily relative humidity in spring	56.9 – 73.8	%
BP0	Presence of previous year mountain pine beetle infestation in the focus cell	0/1	–
BPn	Previous year mountain pine beetle pressure in the neighbouring cells : BPn = $\sum$ BP0 in adjacent cells of radius 1 + 0.5 × $\sum$ BP0 in adjacent cells of radius 2 + 0.25 × $\sum$ BP0 in adjacent cells of radius 3 (Fig. 1)	0 – 9.25	–
DistSouth	Distance from the grid cell centroid to the South infested border of the park	5 – 36660	m
Latitude	Latitude of the grid cell centroid	49.55 – 49.61	dec. °
Longitude	Longitude of the grid cell centroid	-110.01 – -109.43	dec. °
Year	Year of the survey	2007 – 2015	–
Elevation	Elevation at the grid cell centroid	1055 – 1386	m
Slope	Slope at the grid cell centroid	0 – 20.31	°
Northerness	Tendency of the slope to face North	+1 – -1	–
Easterness	Tendency of the slope to face East	+1 – -1	–

Table 2 Confusion matrix showing the results of the model classification on the validation dataset ($n = 49502$) using the threshold 0.003 chosen using the Youden's index.

		Observed	
		absence	presence
Predicted	absence	43 059	129
	presence	5 752	562

List of figure captions

Fig. 1 :

Representation of the adjacent cells taken into account in the covariates (*cf.* Table 1). Striped blue : focus cell, dark grey : 4 adjacent cells (radius 1), light grey : next 8 adjacent cells (radius 2), medium grey : next 16 adjacent cells (radius 3).

Fig. 2 :

Cypress Hills park boundaries in Saskatchewan (grey). The dotted red line represents the park border close to outside infestations in the South. The dashed blue line represents the park border with Alberta.

Fig. 3 :

Observations (a) versus predictions (b) of the mountain pine beetle infestation in Cypress Hills, Saskatchewan, for 2011. On a), a dark red color represents cells with infested trees whereas a bright green color represents cells without infested trees. For b), the risk of infestation per cell ranges from bright green (low risk) to dark red (high risk).

Fig. 4 :

Management control efficiency (= number of infested trees controlled in the park divided by the total number of infested trees) in relation to the net survey area (= total area controlled without overlaps). Solid lines and circles represent the local search around red-top trees for each 2011 and 2012. Dashed lines and crosses represent the search at locations chosen from predictions for each 2011 and 2012. Dotted lines and pluses represent the search around random locations for 2011 and 2012 combined. Each year, the random and prediction strategies data are each the mean of 500 random simulations. The lines represent the fitted values for the local and prediction strategy using a non-linear least square model : control efficiency $_{\text{local}}$ = $1 - \exp(-a * \text{net survey area}^b)$ and control efficiency $_{\text{predictions}}$ $= 1 - \exp(-c * \text{net survey area}^d)$, where $a_{2011} = 0.004$, $b_{2011} = 0.358$, $a_{2012} = 0.018$ and $b_{2012} = 0.287$ (P-values < 0.001 for the null hypotheses $a = 0$ and $b = 1$, df $= 17$) for the local search, $c_{2011} = 2.25^{-6}$, $d_{2011} = 0.884$, $c_{2012} = 3.65^{-5}$ and $d_{2012} = 0.709$ (P-values $= 0.309$ and 0.164 respectively for the null hypotheses c_{2011} and $c_{2012} = 0$, and P-values < 0.001 for the null hypotheses d_{2011} and $d_{2012} = 1$, df $= 17$) for the predictions strategies. For the random search, we used a linear regression : control efficiency $_{\text{random}}$ $= e * \text{net survey area}$, if net survey area $\leq$ park area or 1 if net survey area $>$ park area, where $e = 5.31^{-9}$ (P-value < 0.001 for the null hypothesis $e = 0$, $R^2 = 0.999$, df $= 37$). The striped bars represent the percentage of park area covered by the survey.

Fig. 5 :

Cost of aerial survey, control and circular survey plots in relation to the total area surveyed using circular survey plots from 2006 to 2015. The line represent the fitted values using a linear regression : management cost $= k + l * \text{net survey area}$, where $k = 54\,540.00$ and $l = 0.057$ (P-values < 0.001 for the null hypotheses $k = 0$ and $l = 0$, $R^2 = 0.961$, df $= 8$).

Fig. 6 :

Management cost per controlled tree (a ; from management cost per controlled tree $_{\text{local}} = \frac{k+l*\text{net survey area}}{1-\exp(-a*\text{net survey area}^b)}$ and management cost per controlled tree $_{\text{pred.}} = \frac{k+l*\text{net survey area}}{1-\exp(-c*\text{net survey area}^d)}$) and total cost per controlled tree (b ; from equation (5) using $\theta = 1000$) in relation to the net survey area. Solid lines represent the local search around red-top trees for each 2011 and 2012. Dashed lines represent the search at locations chosen from model predictions for each 2011 and 2012. Black circles correspond to the minimum cost for the local search whereas white circles correspond to the minimum cost for the model predictions strategy.

Fig. 7 :

Management cost (dashed line), avoided cost with $\theta = 1000$ (dotted line) and management plus avoided costs (= total cost ; solid line) in relation to the net survey area for the model predictions strategy. The local search values, not presented here, display similar patterns.

Fig. A1 :

Control efficiency in relation to the classification tree probabilities exponent. Increasing the classification tree probabilities exponent gives more weight to locations with high predicted risks of infestation. Solid lines represent the local search around red-top trees for 2011. Dashed lines represent the search at locations chosen from model predictions for 2011. Dotted lines represent the search around random locations for 2011. Thin lines correspond to a survey area equivalent to the current Forest Service strategy (50 meter-radius circular plot ; 2 200 000 m^2). Thick lines correspond to a survey area of 6 000 000 m^2 which correspond to the circular plot radius 90 m for the local search. The data for 2012, not presented here, display similar patterns.

Fig. B1 :

Optimal net survey area (a) and minimum total cost per controlled tree (b) in relation to the cost of missing a green infestation θ. Solid lines represent the values for the local search whereas dashed lines represent the values for the model predictions strategy for each 2011 and 2012.

Figures

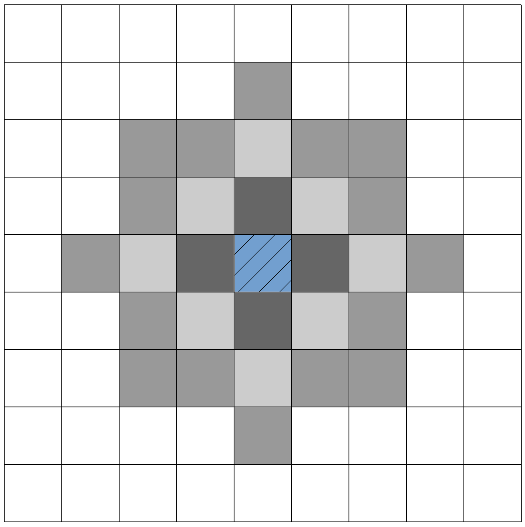

Figure 1 Representation of the adjacent cells taken into account in the covariates (*cf.* Table 1). Striped blue : focus cell, dark grey : 4 adjacent cells (radius 1), light grey : next 8 adjacent cells (radius 2), medium grey : next 16 adjacent cells (radius 3).

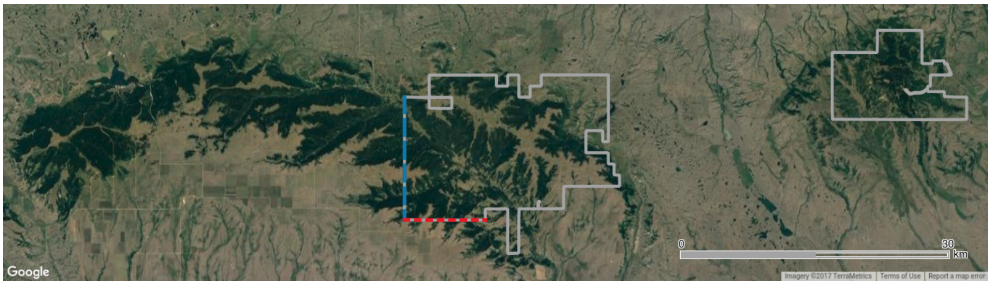

Figure 2 Cypress Hills park boundaries in Saskatchewan (grey). The dotted red line represents the park border close to outside infestations in the South. The dashed blue line represents the park border with Alberta.

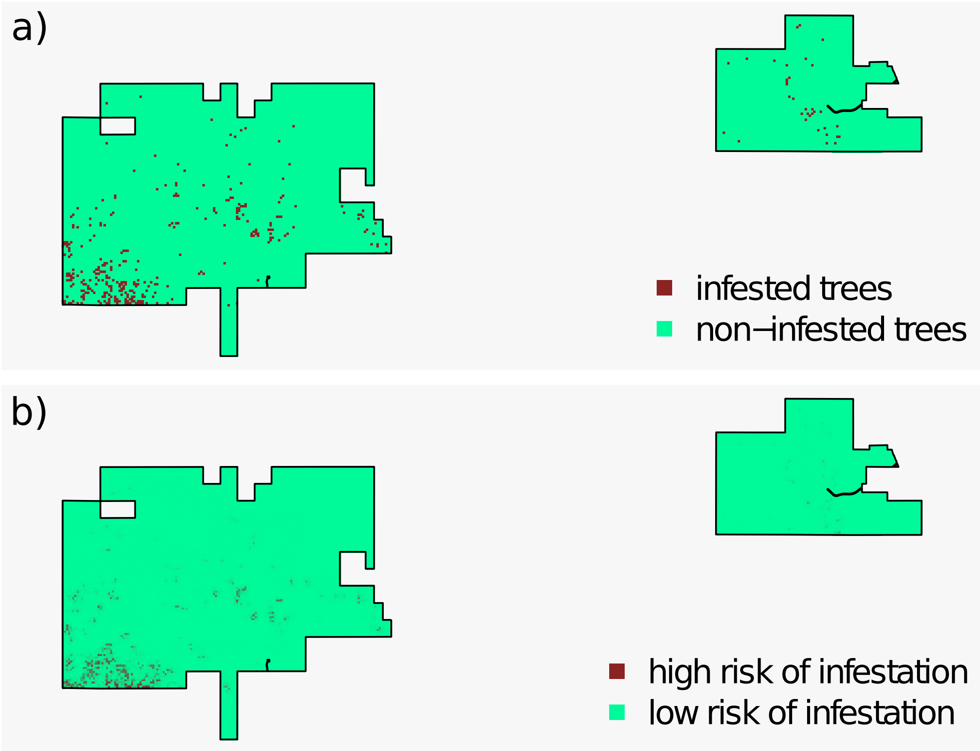

Figure 3 Observations (a) versus predictions (b) of the mountain pine beetle infestation in Cypress Hills, Saskatchewan, for 2011. On a), a dark red color represents cells with infested trees whereas a bright green color represents cells without infested trees. For b), the risk of infestation per cell ranges from bright green (low risk) to dark red (high risk).

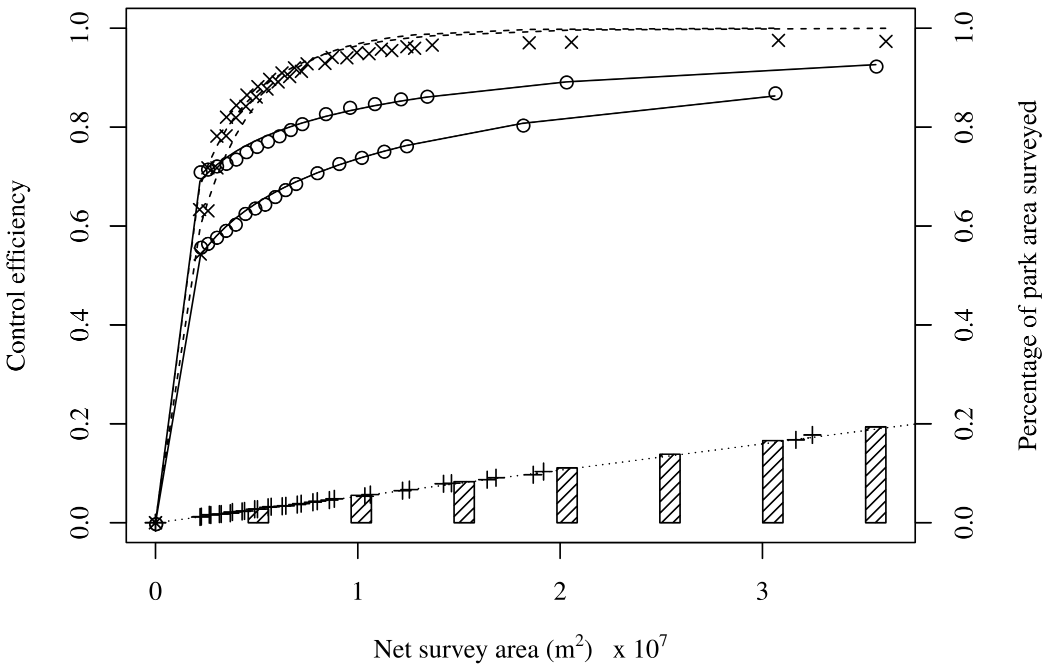

FIGURE 4 Management control efficiency (= number of infested trees controlled in the park divided by the total number of infested trees) in relation to the net survey area (= total area controlled without overlaps). Solid lines and circles represent the local search around red-top trees for each 2011 and 2012. Dashed lines and crosses represent the search at locations chosen from predictions for each 2011 and 2012. Dotted lines and pluses represent the search around random locations for 2011 and 2012 combined. Each year, the random and prediction strategies data are each the mean of 500 random simulations. The lines represent the fitted values for the local and prediction strategy using a non-linear least square model : control efficiency $_{\text{local}} = 1 - \exp(-a * \text{net survey area}^b)$ and control efficiency $_{\text{predictions}} = 1 - \exp(-c * \text{net survey area}^d)$, where $a_{2011} = 0.004$, $b_{2011} = 0.358$, $a_{2012} = 0.018$ and $b_{2012} = 0.287$ (P-values < 0.001 for the null hypotheses $a = 0$ and $b = 1$, df $= 17$) for the local search, $c_{2011} = 2.25^{-6}$, $d_{2011} = 0.884$, $c_{2012} = 3.65^{-5}$ and $d_{2012} = 0.709$ (P-values $= 0.309$ and 0.164 respectively for the null hypotheses c_{2011} and $c_{2012} = 0$, and P-values < 0.001 for the null hypotheses d_{2011} and $d_{2012} = 1$, df $= 17$) for the predictions strategies. For the random search, we used a linear regression : control efficiency $_{\text{random}} = e * \text{net survey area}$, if net survey area $\leq$ park area or 1 if net survey area $>$ park area, where $e = 5.31^{-9}$ (P-value < 0.001 for the null hypothesis $e = 0$, $R^2 = 0.999$, df $= 37$). The striped bars represent the percentage of park area covered by the survey.

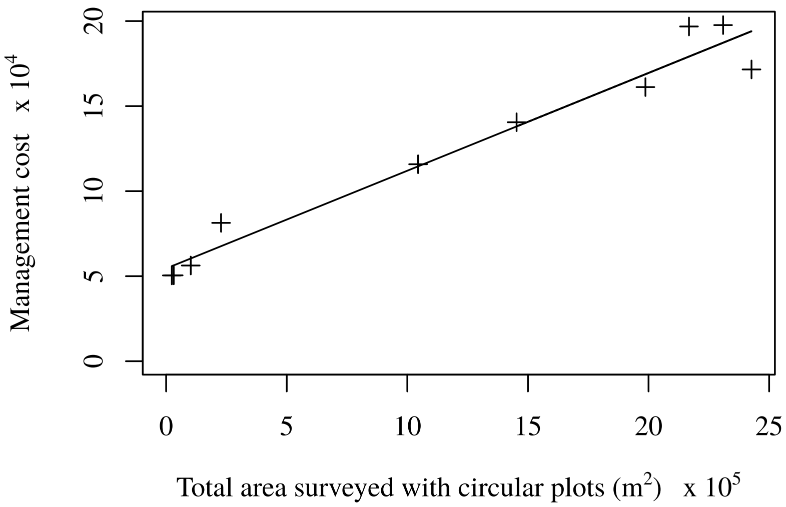

FIGURE 5 Cost of aerial survey, control and circular survey plots in relation to the total area surveyed using circular survey plots from 2006 to 2015. The line represent the fitted values using a linear regression : management cost = $k+l*$net survey area, where $k = 54\,540.00$ and $l = 0.057$ (P-values < 0.001 for the null hypotheses $k = 0$ and $l = 0$, $R^2 = 0.961$, df $= 8$).

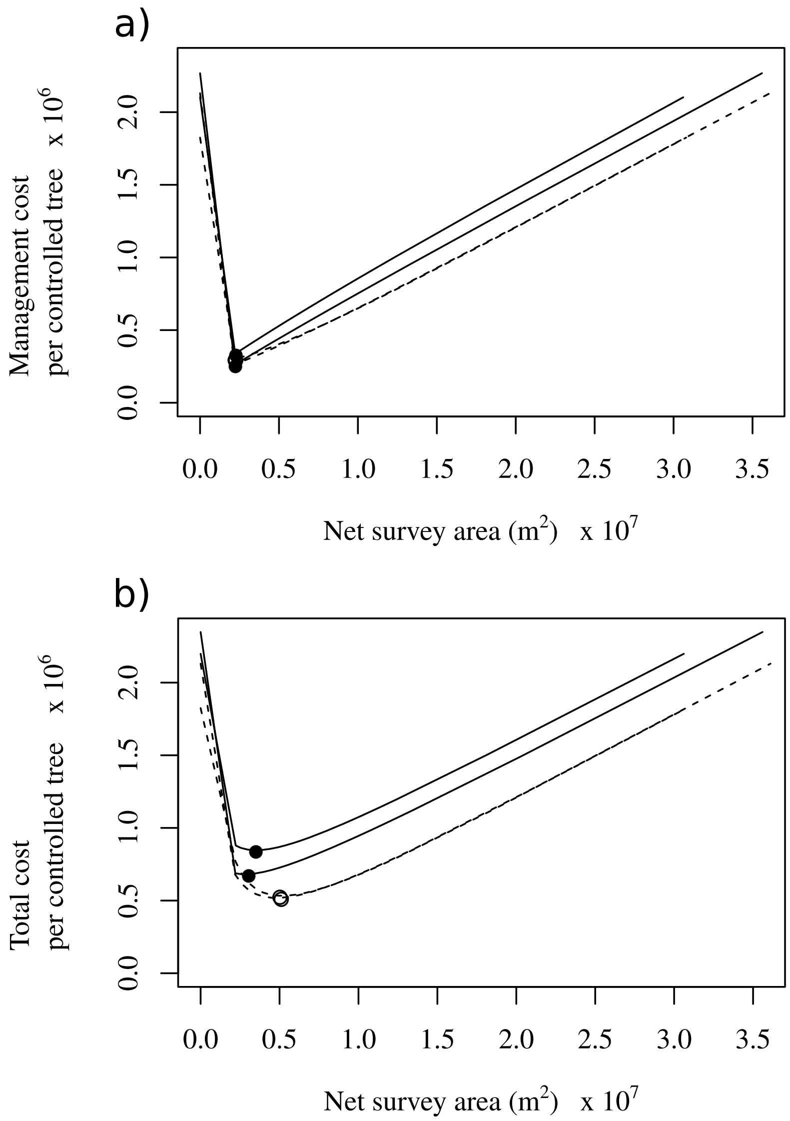

FIGURE 6 Management cost per controlled tree (a; from management cost per controlled tree $_{\text{local}} = \frac{k+l*\text{net survey area}}{1-\exp(-a*\text{net survey area}^b)}$ and management cost per controlled tree $_{\text{pred.}} = \frac{k+l*\text{net survey area}}{1-\exp(-c*\text{net survey area}^d)}$) and total cost per controlled tree (b; from equation (5) using $\theta = 1000$) in relation to the net survey area. Solid lines represent the local search around red-top trees for each 2011 and 2012. Dashed lines represent the search at locations chosen from model predictions for each 2011 and 2012. Black circles correspond to the minimum cost for the local search whereas white circles correspond to the minimum cost for the model predictions strategy.

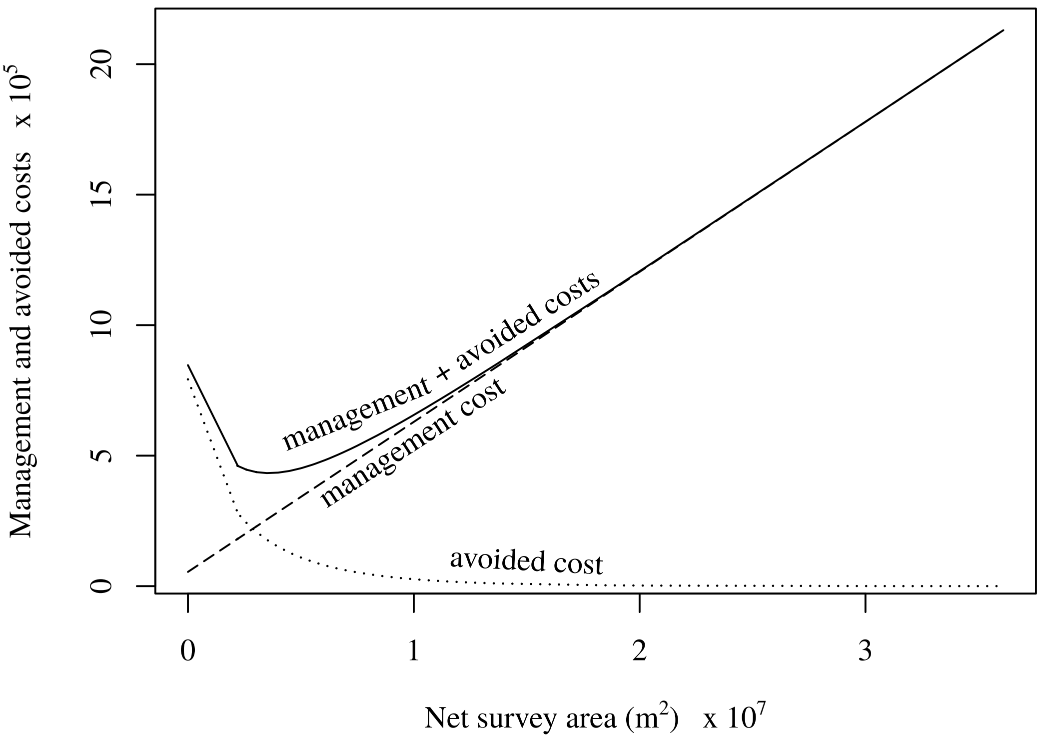

Figure 7 Management cost (dashed line), avoided cost with $\theta = 1000$ (dotted line) and management plus avoided costs (= total cost ; solid line) in relation to the net survey area for the model predictions strategy. The local search values, not presented here, display similar patterns.

Appendices

APPENDIX A : Varying the probability exponent

To vary the amount of noise that we introduced in the random sampling of locations from the model probabilities, we raised the model probabilities to an exponent ranging from 0 to 5. We then sampled the locations without replacement using the new probabilities as weight. The exponent 0 gives the same weight to all locations and, therefore, would give results equivalent to the random strategy. In opposition, a high exponent value increases the differences between low and high probabilities and eventually leads to a deterministic situation where the same locations with the highest probabilities are always chosen.

When we fixed the net survey area and varied the exponent, the predictions control efficiency varies from values similar to the random search at exponent 0 to values similar to the local search at high exponent (Fig. A1). When the fixed survey area is equivalent to the one used in the current strategy (2 200 000 m^2), we can see that the local control efficiency is always higher than the predictions control efficiency no matter the exponent value. However, for a net survey area of 5 000 000 m^2, the prediction control efficiency is larger than the local control efficiency for an exponent value from about 1-1.5 to 5.

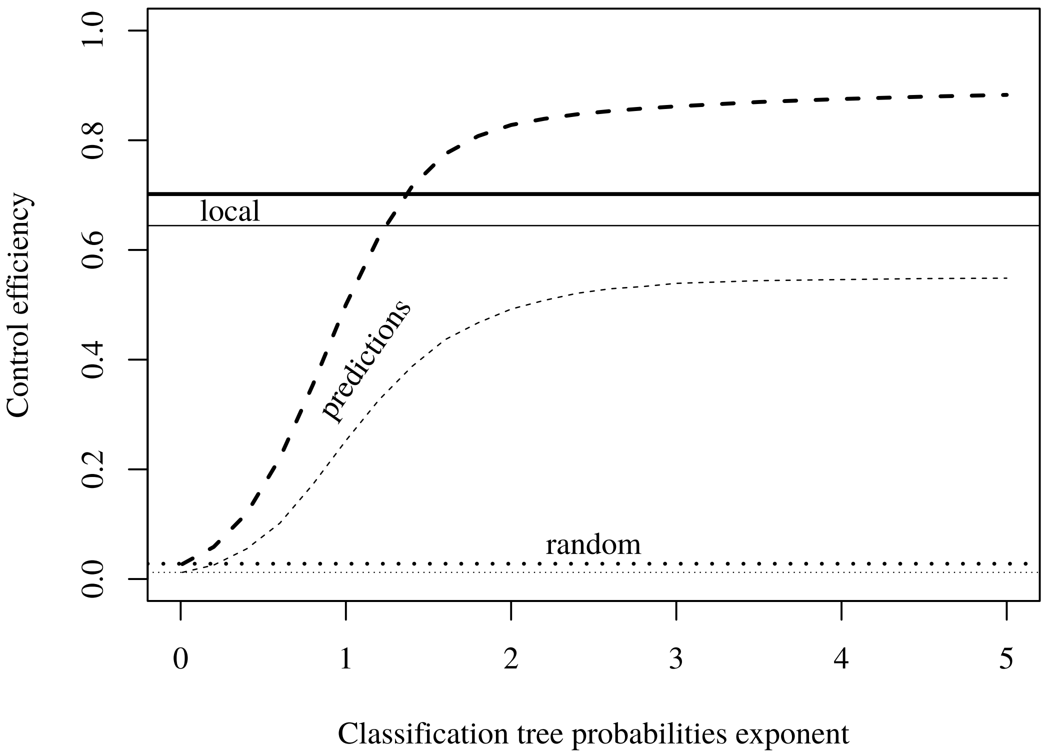

Figure A1 Control efficiency in relation to the classification tree probabilities exponent. Increasing the classification tree probabilities exponent gives more weight to locations with high predicted risks of infestation. Solid lines represent the local search around red-top trees for 2011. Dashed lines represent the search at locations chosen from model predictions for 2011. Dotted lines represent the search around random locations for 2011. Thin lines correspond to a survey area equivalent to the current Forest Service strategy (50 meter-radius circular plot ; 2 200 000 m^2). Thick lines correspond to a survey area of 6 000 000 m^2 which correspond to the circular plot radius 90 m for the local search. The data for 2012, not presented here, display similar patterns.

APPENDIX B : Varying the cost of a missed green infestation

We varied the cost of a missed green infestation θ from 0 to 2000 and investigated its impact on the optimal survey area and the minimum cost per controlled tree depending on the detection strategy.

The optimal net survey area increases with θ for both the local and predictions strategies, although the optimal area is consistently larger using the predictions strategy (Fig. B1a). However, the minimum total cost per controlled tree associated with the optimal survey area is lower for the predictions strategy than the local strategy for $\theta \geqslant 500$ (Fig. B1b).

This means that the more expensive a green infestation, *i.e.* the more new infestations produced by one infested tree, the better in term of costs it is to increase the management effort now rather than controlling the additional new infestations in the future.

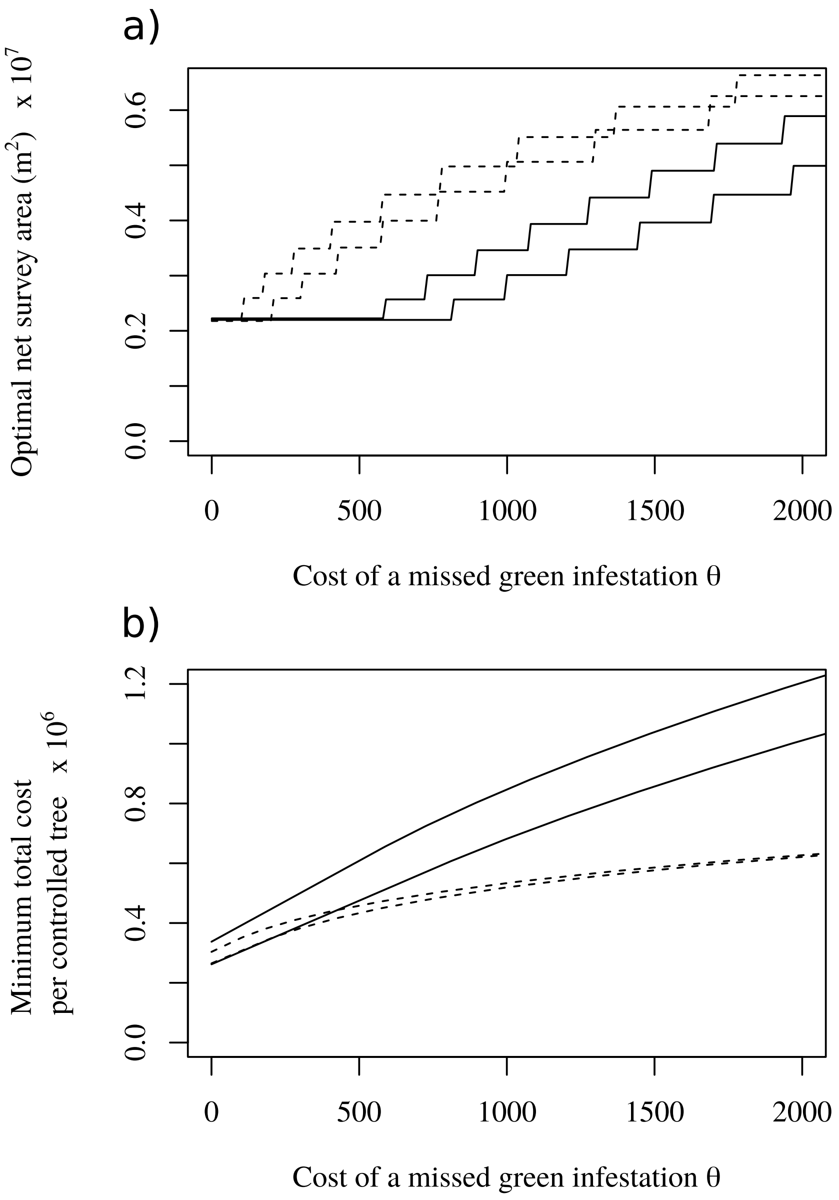

FIGURE B1 Optimal net survey area (a) and minimum total cost per controlled tree (b) in relation to the cost of missing a green infestation θ. Solid lines represent the values for the local search whereas dashed lines represent the values for the model predictions strategy for each 2011 and 2012.