Canadian Journal of Forest Research

OPEN ACCESS | Article

A correction for serial nonindependence in mountain pine beetle aerial survey data to reduce overestimation of cumulative damage

Devin W. Goodsman and **Jim Weber**

Canadian Forest Service — Natural Resources Canada, Northern Forestry Centre, 5320 122 Street NW, Edmonton, AB T6H 3S5, Canada

Corresponding author: **Devin W. Goodsman** (email: devin.goodsman@nrcan-rncan.gc.ca)

Abstract

At the beginning of this century, an unprecedented mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak expanded across western North America. Methods are needed to quantify its impact on forests. Although aerial survey data of infestation are readily available, their translation to tree mortality estimates is challenging because the polygons sketched by aerial surveyors to delimit infested regions often overlap from one year to the next, which produces a lack of independence in yearly estimates of infested landscape. This problem manifests when annual proportions of infested land are summed over years because the sum often exceeds one. Cumulative proportions of infested land larger than one implies overestimation of tree mortality if initial densities of available host trees are multiplied by annual proportions of infested landscape. To address this problem, we developed a probabilistic method for correcting for nonindependence in proportions of infested land recorded using overlapping polygons. We demonstrate our approach in Jasper National Park where the probabilistic correction reduces overestimation bias in regions that were infested multiple years when predictions are compared to validation data. The approach and the spatial raster data sets we generated will be useful to forest managers seeking to quantify the impact of mountain pine beetles.

Key words: aerial survey, remote sensing, tree mortality, insect, climate change, pine

Résumé

Au début de ce siècle, une épidémie sans précédent de dendroctone du pin (*Dendroctonus ponderosae* Hopkins) s'est étendue à l'ouest de l'Amérique du Nord. Il est nécessaire de trouver des méthodes pour quantifier son impact sur les forêts. Même si les données d'enquête aérienne sur l'infestation sont facilement disponibles, leur traduction en estimations de la mortalité des arbres est difficile car les polygones esquissés par les enquêteurs aériens pour délimiter les régions infestées se chevauchent souvent d'une année à l'autre, ce qui produit un manque d'indépendance dans les estimations annuelles du paysage infesté. Ce problème se manifeste lorsque les proportions annuelles de terrains infestés sont additionnées sur plusieurs années, car la somme est souvent supérieure à un. Les proportions cumulées de terrains infestés supérieures à un impliquent une surestimation de la mortalité des arbres si les densités initiales d'arbres hôtes disponibles sont multipliées par les proportions annuelles de terrains infestés. Pour aborder ce problème, nous avons développé une méthode probabiliste pour corriger la non-indépendance des proportions de terres infestées enregistrées à l'aide de polygones se chevauchant. Nous démontrons notre approche dans le parc national de Jasper où la correction probabiliste réduit le biais de surestimation dans les régions qui ont été infestées plusieurs années lorsque les prédictions sont comparées aux données de validation. L'approche et les ensembles de données matricielles spatiales que nous avons générés seront utiles aux gestionnaires forestiers qui cherchent à quantifier l'impact du dendroctone du pin ponderosa. [Traduit par la Rédaction]

Mots-clés : relevé aérien, télédétection, mortalité des arbres, insecte, changements climatiques, pin

Introduction

The mountain pine beetle (*Dendroctonus ponderosae* Hopkins) is one of the most destructive forest insects in North America (Safranyik and Carroll 2006). The most recent outbreak in western Canada and the western United States exemplifies the destructive potential of the mountain pine beetle. In western North America, the recent outbreak was the most severe and expansive ever recorded (Kurz et al. 2008). The destructiveness of mountain pine beetle outbreaks is a direct result of the process of host tree colonization by adult beetles seeking to reproduce, which nearly always results in the death of colonized trees (Safranyik and Carroll 2006). After pine trees are colonized in the mid to late summer, their foliage begins to fade. However, foliage of attacked trees does not turn red until approximately 1 year after attack (Wulder et al. 2004, 2006). At this point, trees are typically called red-topped trees or red attack trees. An ever-diminishing proportion of red pine needles can remain on the tree in subsequent years and it is not unusual for trees to retain red foliage for 3 years (Wulder et al. 2004, 2006). One and a half to 5 years after mountain pine beetles kill trees, their needles begin to drop, and they gradually turn grey (Safranyik and Carroll 2006). Grey trees are called grey because they appear grey when viewed from a distance due to their lack of foliage. Aerial surveyors in fixed or rotary-wing aircraft delineate regions of pine forest in which tree crowns have recently turned red by recording clusters of trees as georeferenced points or by encompassing larger infestations using sketched polygons to quantify the extent and severity of mortality caused by mountain pine beetles across landscapes. Detailed records of individual infested trees or clusters of infested trees collected from rotary-wing aircraft are typically called detailed aerial surveys to distinguish them from polygons sketched from fixed wing aircraft, which are called aerial overview surveys. However, in the mountainous environment of Jasper National Park where our study is based, detailed aerial surveys and their polygon-based analogue were both collected from rotary-wing aircraft. At the time of writing, polygon-based aerial surveys remain the main method that land managers track outbreak size and severity for large outbreaks of forest insects.

Aerial overview surveys were originally conceived to assess the area impacted by various forest disturbance agents rather than to estimate tree mortality. Nevertheless, several studies have proposed methods to convert aerial survey coverage to mortality estimates (Coops et al. 2006; Robertson et al. 2009; Meddens et al. 2012; Meigs et al. 2015). Modern approaches for translating aerial survey data to mountain pine beetle impacts on forests often rely on remotely sensed data in combination with aerial survey data to estimate damage (Coops et al. 2006). Some specific examples include estimation of mortality area of killed trees based on their crown areas (Meddens et al. 2012), percentages of available pine hosts killed (Robertson et al. 2009), or mortality totals measured in basal area or stem density (Meigs et al. 2015). Regardless of the data integration approach and damage metric, it is important to recognize that aerial overview surveys can overestimate or underestimate the damage caused by mountain pine beetles and the extent of overestimation or underestimation depends on how severity classifiers are translated to proportions of infested land (Wulder et al. 2009; Egan et al. 2019, 2020). Data collected by aerial surveyors may overestimate damage caused by mountain pine beetles because as infestations grow, it becomes impossible to record individual clusters of infested trees and aerial surveyors switch to the polygon data format. Polygons will often overlap from one year to the next and in large and heavily impacted regions, it is unlikely that aerial surveyors remember which trees were affected in previous years — even when the same surveyor conducted previous surveys. Although with modern electronic tablets, surveyors can access previous year's surveys when conducting aerial surveys, this does not necessarily facilitate exclusion of previously infested regions. The reason is that polygon classifiers are not binary and so it would only make sense to completely exclude an infested region if all trees within it were infested. This is not how outbreaks typically progress. Instead, a given polygon classified as light in 1 year may have an overlapping proportion with a polygon recorded as moderate in the next year. It is unrealistic to expect aerial surveyors to do the mental arithmetic necessary to adjust the severity classes of overlapping polygons as they are surveying — even if they can view surveys from previous years while conducting the current year's survey. To distinguish recently attacked red trees from trees that were attacked more than 1 year ago that have retained red needles, surveyors rely on subtle changes in the hue of red foliage. In this context, persistent red needles, variation in light intensity between survey days, direction of survey flight in relation to the sun, and time of day are all subtle challenges to interpreting recent pine tree mortality. As a result of overlapping coverage of aerial survey polygons from one year to the next and the high likelihood that some proportion of area with red-crowned trees was recorded as red in previous years, records of infested trees delineated using polygons are not independent from year to year.

To our knowledge, no method for correcting lack of yearly independence of polygon data has been proposed. Our research in this context was focused on three objectives. The first objective was to develop a method for correcting for overestimation of cumulative mortality proportions based on aerial survey data that arise due to serial nonindependence. Our second objective was to translate corrected mortality proportions based on aerial surveys into estimates of annual killed tree density. Our third objective was to compare the killed tree density estimates with validation data to determine whether the correction improved estimates. To achieve our first objective, we propose a probabilistic correction for the lack of yearly independence in polygon data. To realize our second objective, we exploit data derived from forest inventory data and moderate resolution imaging spectroradiometer (MODIS) satellite imagery (Beaudoin et al. 2014) to estimate annual and cumulative densities of aerially visible trees with red crowns. We produced annual spatial raster data sets of densities and error bounds of aerially visible red-topped lodgepole pine trees that were infested by mountain pine beetle per 625 m^2 raster grid cell (25 m × 25 m) in Jasper National Park. These gridded data allowed us to accomplish our third objective of comparing our corrected and uncorrected predictions to data in which numbers of mountain pine beetle infested trees with red crowns within aerially surveyed polygons were counted by ground survey crews.

Materials and methods

We propose a method for estimating densities of infested trees from aerial overview survey data by first correcting the lack of yearly independence in infestation data derived from polygon data and then using forest inventory data and satellite-based data in combination with corrected aerial survey data to estimate annual densities of infested trees. We then harmonized estimates of infested trees in which clusters of infested trees were recorded as points with estimates of infested trees derived from polygon-based aerial overview surveys. Finally, we evaluated our approach by comparing data in which numbers of killed trees were counted by ground survey crews to tree mortality estimates we generated using the corrected mortality proportion in combination with remotely sensed data. To elucidate the challenges that our method addresses, we first describe our data before outlining the approach that allowed us to correct for lack of serial independence in aerial overview survey data.

Aerial survey data of mountain pine beetle infestation

The mountain pine beetle infestation data used in this study were collected in Jasper National Park during annual aerial surveys. The aerial survey data were collected by sketch mapping, in which aerial surveyors delineated infested regions with red-topped trees in the autumn of each year by sketching them on georeferenced tablet computers. The aerial survey data collected within Jasper National Park include both polygon and point data collected by observers in rotary-wing aircraft to better navigate the mountainous terrain in Jasper National Park. The point data are considered detailed aerial survey data as surveyors attempted to delineate individual trees or clusters of infested trees. Conversely, the polygon data are typically called aerial overview survey data even though they were collected using a rotary-wing aircraft (overview data are more often collected using fixed wing aircraft). Generally, whether point or polygon data were collected depends on the extent and severity of infestations as well as on whether land managers intend to cut and burn individual infested trees in which case, point data are preferred. In Jasper National Park, for all years up to but not including 2014, infestations were recorded as point data. Although mountain pine beetle infestation data were recorded as point data up to 2014, this was not because managers intended to cut and burn individual infested trees as no beetle infested trees were controlled by cutting and burning in this period. When the outbreak expanded in 2014, the new widespread infestations were more pragmatically represented using polygons that encompassed large patches of red-crowned trees that were infested by mountain pine beetles. However, even after 2013, small and isolated infestations were documented using points even though most infestations were recorded using polygon data (see Supplementary Fig. S1 for example). Both point and polygon aerial survey data represent the relative abundance of red-topped trees killed by mountain pine beetles, albeit in different ways.

To better quantify the relative abundance of red-topped trees represented using polygon or point data, aerial surveyors assign all the point and polygon data they collect a light, moderate, or severe classification. For point data, the severity classification depends on the number of infested trees in the cluster. For polygon data, the light classifier is assigned to polygons in which between 1% and 10% of trees have red crowns, the moderate classifier is assigned to polygons in which 10%–30% of trees have red crowns, and the severe classifier is assigned when more than 30% of trees have red crowns within a polygon. Note that this three-class severity rating is different from the five-class system used in British Columbia and the United States. In British Columbia, light, moderate, and severe classes are sandwiched between trace and very severe classes at the lowest and highest levels of infestation, respectively.

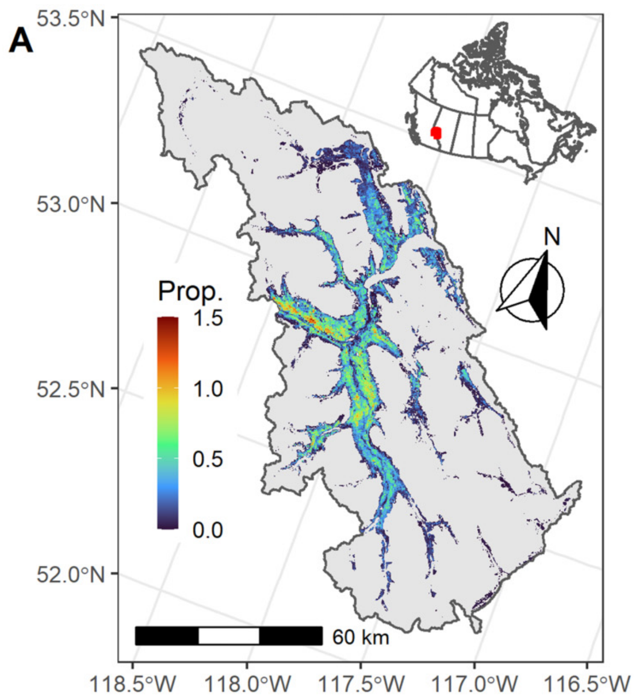
Regardless of severity or data type, surveyors attempted to distinguish recently red trees from red-topped trees that were infested more than 1 year prior when assigning severity classifiers, although this can be difficult when polygons are large and overlapping from one year to the next. It is important to note that aerial survey data are collected by observers looking down at forests. Therefore, the most visible trees to aerial surveyors are the dominant and subdominant overstory trees, which conveniently are the size classes that are preferentially attacked by mountain pine beetle populations at outbreak levels (Safranyik and Carroll 2006). However, one of the limitations of aerial overview survey approaches is their inability to detect infested trees that are below the canopy of codominant and dominant trees. Moreover, because the crowns of mountain pine beetle infested trees tend to be the largest crowns in the stand when mountain pine beetles are at outbreak levels, and because of visual biases towards red objects in humans, levels of infestation that are assessed by aerial surveyors may often be overestimates. In Jasper National Park, all data collected from 1999 to 2013 inclusive were point data. From 2014 to 2020 inclusive, when the outbreak in Jasper National Park accelerated, aerial overview survey data were primarily polygon data with occasional points to represent small and isolated infestations.

The uncorrected approach for calculating annual and cumulative impact from polygon data

To test whether our correction for serial nonindependence in polygon-based aerial survey data reduces overestimation of cumulative mountain pine beetle impact, we must compare it with an uncorrected equivalent. Therefore, before describing the correction in detail we will describe a simple uncorrected way of calculating annual proportions of the landscape affected by mountain pine beetles. Cumulative damage can be obtained by summing up annual proportions impacted area at each location in space. The first step in our calculation of annual impact is rasterizing the polygon data.

We rasterized aerial overview polygon data using a 25 m $\times$ 25 m raster grid for each severity class (light, moderate, and severe) by assigning proportions of infested land base to each grid cell that was covered by a polygon. Thus, we obtained three 25 m $\times$ 25 m raster grids per year: one each for light, moderate, and severe infestations. We chose a resolution of 25 m $\times$ 25 m because the validation polygons with infested stems counted by ground crews that we used to test our approach were small. The smallest validation polygon had an area of only 0.375 Ha, which is difficult to reconcile with predictions made at a coarser resolution such as 250 m $\times$ 250 m

Fig. 1. (A) Cumulative proportions of areas that were infested by mountain pine beetles (see eq. 2) estimated using aerial overview survey data from Jasper National Park between 2014 and 2020 often exceeded one and attained a maximum of 1.51 in the most severely impacted regions, indicating that they are not true proportions. (B) After applying a probabilistic correction (see eq. 7) for serial nonindependence, cumulative proportions were reduced and never exceeded one. Maps were produced using the ggplot2 package in R with aerial overview survey data collected by the Canadian Forest Service and National Parks Canada that were rasterized using GDAL software in Python.



(6.25 Ha). In assigning severity multipliers, we followed the standard, which is to use severity multipliers of 0.05, 0.1, and 0.3 for light, moderate, and severe classifications. These severity multipliers represent the proportion of aerially visible dominant and codominant pine trees within each polygon that were recently infested by mountain pine beetles and possessed red crowns. The severity multipliers were multiplied by the appropriate raster grids representing the proportion of each grid cell covered by the corresponding polygon severity class to obtain an overall proportion of lodgepole pine in the raster that were lightly, moderately, or severely infested. Then to obtain a total proportion of codominant and dominant lodgepole pines that were infested per grid cell, we summed overall proportions computed within light, moderate, and severe raster grids for the corresponding grid cell. In this way, our three raster grids were reduced to a single raster grid per year of the same dimensions as the three raster grids for each severity class.

Each grid cell of the annual raster grids described above represents the local proportion of susceptible pine hosts that were infested and recorded as red tops (P_t). The uncorrected ($D_{u,t}$) density of host pines can be estimated using

$$D_{u,t} = X_0\,(A)\,(D_0)\,P_t \tag{1}$$

wherein X_0 is the initial proportion of the landscape that comprises suitable host trees for mountain pine beetles, A is the area of a 25 m × 25 m grid cell ($A = 625$ m^2), and D_0 is the initial density of pine trees per m^2 before the outbreak. The initial proportion of the landscape that comprises suitable host trees (X_0) and the initial density of host pine trees per m^2 before the outbreak (D_0) need to be estimated using remotely sensed data as we describe in a later section.

Finally, to compute the cumulative proportion of codominant and dominant lodgepole pine trees within each grid cell that were impacted over the outbreak, we summed up the overall proportion of codominant and dominant lodgepole pine trees impacted per year for each grid cell over all years (n) in our data set to obtain a single raster with dimensions equivalent to the annual raster grids:

$$C_{u,n} = X_0 \sum_{i=1}^{n} P_i \tag{2}$$

To calculate the uncorrected cumulative density of host trees impacted, one must multiply eq. 2 by the area of a grid cell in m^2 (A) and the initial density of host pine per m^2 (D_0). As we have previously indicated, estimating annual and cumulative impact in this way is problematic because the cumulative summed proportion of trees killed in heavily infested locations based on aerial overview surveys often exceeds one (Fig. 1A). Note that the fact that the cumulative proportion of landscape that is impacted by mountain pine beetles exceeds one is not an artefact of the rasterization process described above. The same effect can be observed without first rasterizing the polygon data if the proportions of infested trees derived from polygon severity classes are summed at a location where numerous overlapping infestations occurred. When calculations are done this way using the raw polygon data, the cumulative proportions at the most severely impacted points on the map will frequently exceed one. Because many grid cells in our raster were only partially covered by polygons, rasterization of the polygon data reduced the probability that the cumulative proportion exceeded one or left it unchanged when grid cells were completely covered by polygons.

The nonsensical nature of a proportion that exceeds one can be more clearly understood if it is perceived in terms of a finite pool of available host trees; over the course of an outbreak, more trees are predicted to have been killed in a cell than there are trees in the cell. The lag in fade of red beetle infested trees to grey and the inclusion of some trees that were infested more than 1 year ago within overlapping polygons explains why cumulative proportions of stands with red-topped trees often sum to greater than one.

A probabilistic approach for correcting for temporal dependence in polygon data

After rasterizing the polygon aerial survey data as described in the previous section, they can be corrected for some of the serial nonindependence that arises from polygon overlap from year to year. The following description provides details of our probabilistic correction approach. This approach can easily by applied to the raster grids, as it requires only that simple arithmetic operations of multiplication and subtraction be applied to the raster data.

Let X_t be the proportion of landscape comprising susceptible host pine trees that are not yet infested by mountain pine beetles. Then, over time the remaining proportion of the landscape within a raster grid cell, for example, that contains susceptible host trees can be represented as

$$X_t = X_0 \prod_{i=0}^{t} (1.0 - P_i) \tag{3}$$

where X_0 is the initial proportion of the landscape that comprises suitable host trees for mountain pine beetles and P_t is the yearly varying probability of infestation by mountain pine beetle. It follows then that the proportion of susceptible pine trees infested by mountain pine beetles (Y_t) is given by the product

$$Y_t = P_t X_{t-1} \tag{4}$$

which can be rewritten using eq. 3 as follows:

$$Y_t = P_t X_0 \prod_{i=0}^{t-1} (1.0 - P_i) \tag{5}$$

Equation 5 can be understood as X_0 multiplied by the conditional probability of a region becoming infested in year t and not being infested in any previous year. Equation 5 and the use of conditional probabilities in this way reduce the nonindependence of proportions in regions where infestations were recorded using overlapping polygons because it prevents previously infested landscape from being reinfested. This makes biological sense for mountain pine beetles because trees die once they are infested, and thus they are no longer suitable for subsequent infestation. When the probabilities in eq. 5 are replaced with the estimated proportions of the region with red crowns (from the rasterized polygon data, for example), it can be used to correct proportions of landscape comprising trees that were recently killed by mountain pine beetles. Equation 5 can be applied to each grid cell of the annual raster grids of the proportion of landscape infested per year. Equivalently, the equation can be applied to the whole raster for each year using raster arithmetic, which leads to much more efficient calculations. To estimate corrected annual infested tree densities per raster grid cell ($D_{c,t}$), eq. 5 must be multiplied by the initial density of host pines per m^2 (D_0) and the area of an individual grid cell (A):

$$D_{c,t} = P_t X_0 (A)(D_0) \prod_{i=0}^{t-1} (1.0 - P_i) \tag{6}$$

Recall that A is the area of a 25 m × 25 m grid cell (A = 625 m^2). Equation 6 is the corrected equivalent of eq. 1. In eq. 6, (and eq. 1), the initial density of pine trees per unit area before the outbreak (D_0) and the initial proportion of the landscape comprising suitable pine hosts for mountain pine beetle (X_0) are all that is required to produce estimates of annual infested tree density. In the section that follows, we describe how we estimated the initial proportion of the landscape comprising suitable pine hosts for mountain pine beetle (X_0) and the initial density of host pines per m^2 (D_0) across our gridded study landscape using remotely sensed data in tandem with forest inventory data.

Using the same probabilistic framework starting with eq. 3, it follows that the corrected cumulative proportion of the landscape that is infested over all n years in the outbreak is

$$C_{u,n} = X_0 \left(1 - \prod_{i=0}^{n} (1.0 - P_i)\right) \tag{7}$$

Equation 7 is the corrected equivalent of eq. 2. Simple mathematical logic can be applied to understand that provided both X_0 and P_i are always less than one and greater than zero, $C_{u,n}$ will never exceed one.

Forest inventory and satellite data

We obtained spatial data that characterize Canadian forests based on Canadian Forest Inventory and MODIS satellite data (Beaudoin et al. 2014) to estimate the density of dominant lodgepole pine trees across Jasper National Park as well as the proportion of each of our raster grid cells that comprised suitable host trees for mountain pine beetle infestation. These data do not already contain estimates of densities of dominant lodgepole pine trees or of the proportion of area with susceptible lodgepole pine trees and so we derived our estimates from other variables present in the data set. The remotely sensed data were particularly appropriate for estimation of the initial density of host trees before mountain pine beetle infestation and the initial proportion of raster grids that contained susceptible hosts before infestation because they relied on remotely sensed imagery and forest inventory data from 2011, before the mountain pine beetle outbreak in Jasper National Park accelerated.

The proportion of landscape that contains susceptible host pine is a function of the percentage of landcover that is treed and the percentage of treed landcover that is pine. Both of these variables have been estimated (Beaudoin et al. 2014). Thus, the proportion of a raster grid that contains susceptible pine (X_0), which is a variable in eqs. 1 and 6 that requires estimation, can be estimated from these variables as

$$X_0 = \left(\frac{L_P}{100}\right)\left(\frac{T_c}{100}\right) \tag{8}$$

in which L_P represents the percentage of lodgepole pine in a raster cell and T_c represents the percentage of the raster grid that is covered by trees.

To estimate susceptible lodgepole pine density in Jasper National Park prior to the mountain pine beetle outbreak, we relied on estimates of stand age based on forest inventory and MODIS data (Beaudoin et al. 2014). Our estimate of initial lodgepole pine host density also relied on an assessment of the site index in regions of Jasper National Park affected by mountain pine beetles. Site index is a measure of site productivity used by forest managers (Monserud et al. 2006) in which site productivity is quantified based on the height of an average tree that is 50 years old when its age is measured at breast height (1.3 m). Based on data in Monserud et al. (2006), we determined that the site index in mountain pine beetle affected areas of Jasper National Park was between 12 and 15 m at 50-year breast height age. These site indices correspond most closely to a medium site as defined by the Alberta Phase 3 Forest Inventory (Alberta Forest, Lands and Wildlife 1985), which is a site index of 14 m at 50-year breast height age. Having determined an appropriate site index, we accessed data (Supplementary Table S1) on the relationship between lodgepole pine stand age and the density of stems of diameter greater than 10 cm for a medium site from the Alberta Phase 3 Forest Inventory (Alberta Forest, Lands and Wildlife 1985). We fitted a nonlinear regression to data from the Alberta Phase 3 Forest Inventory, which allowed us to translate estimates of stand age across our gridded study area to estimates of density of susceptible lodgepole pine trees (see Supplementary Fig. S2). The fitted nonlinear regression allowed us to write the following expression for the initial density of lodgepole pine stems per m^2 (D_0 in eqs. 1 and 6):

$$D_0 = \frac{74}{10\,000}(Z-30)\,e^{-0.02493\times(Z-30)+0.0000492(Z-30)^2}, \quad z>30 \text{ years} \tag{9}$$

in which Z represents stand age in years. Notice that we only consider stands older than 30 years since stands younger than this age in the Canadian Rockies will not have stems of diameter equal to or larger than 10 cm. A similar approach was used to translate total tree volume at a stand level to densities of trees (Goodsman et al. 2016), but this prior approach differs because the current method finds an expression for tree density in terms of stand age rather than volume.

Using the raster data on forest attributes (Beaudoin et al. 2014), and eqs. 8 and 9, we were able to produce a spatial raster data set at 25 m × 25 m resolution of the density of lodgepole pine stems of diameter greater than 10 cm per 625 m^2 grid cell across Jasper National Park prior to the mountain pine beetle outbreak, which is the product $X_0(A)(D_0)$ in eqs. 1 and 6. Then, using eqs. 1 and 6, we were able to produce uncorrected (eq. 1) and corrected (eq. 6) annual raster grids in which each grid cell contained an estimate of the number of recently killed red-topped trees per 625 m^2 grid cell.

Validation of predicted densities of infested trees derived from polygon data

Our aerial survey data set contained only eight small validation polygons in which numbers of red infested trees were recorded by ground survey crews. Observed numbers of infested trees per validation polygon ranged only from two to six infested trees. Given this extremely narrow range of observed values, a linear regression validation approach did not make sense because, assuming infested trees were Poisson distributed, the uncertainty around each estimate would be as large or larger than the variation across the whole validation data set. Moreover, only two of the eight validation polygons with numbers of infested trees counted by ground crews were from polygons that overlapped infestations in previous years. To predict the number of infested trees we summed up estimates of densities of infested trees per 25 m × 25 m grid cell that were intersected by each of the validation polygons for estimates based on the corrected and uncorrected approaches. Predicted densities of infested trees were obtained by multiplying the density of susceptible host trees by the corrected and uncorrected proportions of each cell that were infested. To test predicted values against observed values, we assumed that the counts of infestations within polygons were Poisson distributed random variables. For all validation plots, we used a Poisson test (Garwood 1936) to test the null hypothesis that the true λ parameter of the Poisson distributed count of infested trees in each validation polygon was equal to the expected density of infested trees predicted using our corrected and uncorrected approaches. p-values greater than 0.05 led us to fail to reject our null hypothesis (a positive validation result), whereas a p-value less than 0.05 led us to reject our null hypothesis (a negative validation result).

Rasterizing point aerial survey data

Point data were rasterized differently from how polygon data were rasterized. This difference is an inevitable consequence of the nature of the data. From 2013 until 2016, point aerial overview survey data collected in Jasper National Park often included both estimates of numbers of red-topped trees and severity classifiers. Conversely, from 1999 to 2012, point data were not accompanied by estimates of numbers of red-topped trees but each point featured a severity classifier of light, moderate or severe. To assign mean and error estimates to numbers of trees infested, we first converted point data in each severity class to spatial raster data on a 25 m × 25 m raster grid. Thus, for each year we produced up to three raster data sets, depending on whether all three severity classes were present in the data or not. The value assigned to each cell was the count of the numbers of points of each severity class per 25 m × 25 m cell. This resolution of grid was used to facilitate harmonization of raster grids derived from polygon data with those derived from point data. Note that at this fine resolution, no raster cell contained more than one infestation point, but each point could represent a cluster of multiple trees. Then to estimate mean numbers of red-crowned trees per cell, we drew 10 000 bootstrap samples per cell in the raster grid from the numbers of red-topped trees per point in the corresponding severity class in the data that contained both types of information (i.e., from a subset of the data collected from 2013 to 2016). For severe point data for example, in a cell that contained a point that was assigned the severe classifier, we sampled 10 000 times with replacement from the data set that contained both the severe classifier and associated numbers of red-topped trees. To assign appropriate expected value to the cell, we averaged the number of red-topped trees that were randomly drawn in the 10 000 samples. We then computed 95% CI on expected density of pines in each cell using the quantile approach (Efron and Tibshirani 1994).

Evidently, bootstrapping means for the raster data derived from point infestation data was unnecessary when point data were assigned a severity classification and an estimate of the numbers of red-topped trees in the cluster. However, we repeated this process for locations and years where point data included estimates of numbers of red-topped trees because it allowed us to assign upper and lower confidence limits (95% CI) using the quantiles associates with probabilities of observation of 0.025 and 0.975. In addition, bootstrapping statistics in this way allowed us to conduct an internal validation of the approach. To assess the bootstrap approach, we selected raster grid cells in which only point data with estimates of the numbers of red-topped trees were present and regressed observed counts of red-topped trees against the expected numbers of red-topped trees per cell as predicted by the bootstrapped average. To evaluate our approach, we then calculated the coefficient of determination for a linear regression of predicted versus observed (R^2).

Producing landscape scale summaries

To estimate total annual numbers of aerially visible lodgepole pine trees killed by mountain pine beetles across Jasper National Park, we summed estimates in raster cells within park boundaries over each year. To produce confidence intervals on predictions we first computed the variance of our predictions versus observations for the validation polygons. There was a power-law relationship between the predicted mean based on the probabilistic correction and the variance (Supplementary Fig. S3). We used this relationship to extrapolate the variances for each grid cell in each year. Then, we calculated standard deviations from the variances and constructed Wald-type upper and lower 95% CI on corrected predictions of killed tree density across each cell in our raster grids for each year. Summing up the upper and lower 95% CI raster data sets for each year provided annual lower and upper 95% CI for landscape summaries for the entire study region.

Spatial data coordinate reference systems, projections, and reprojections

The aerial survey data used the Canada Albers Equal Area Conic coordinate reference system (EPSG: 102001), whereas the forest attribute data used the NAD 83/LCC Canada AVHRR-2 reference system (EPSG: 42309). We reprojected the aerial survey data to the coordinate reference system of the forest attribute data before rasterization to ensure that the raster grids were aligned with the original forest attribute raster data.

Software

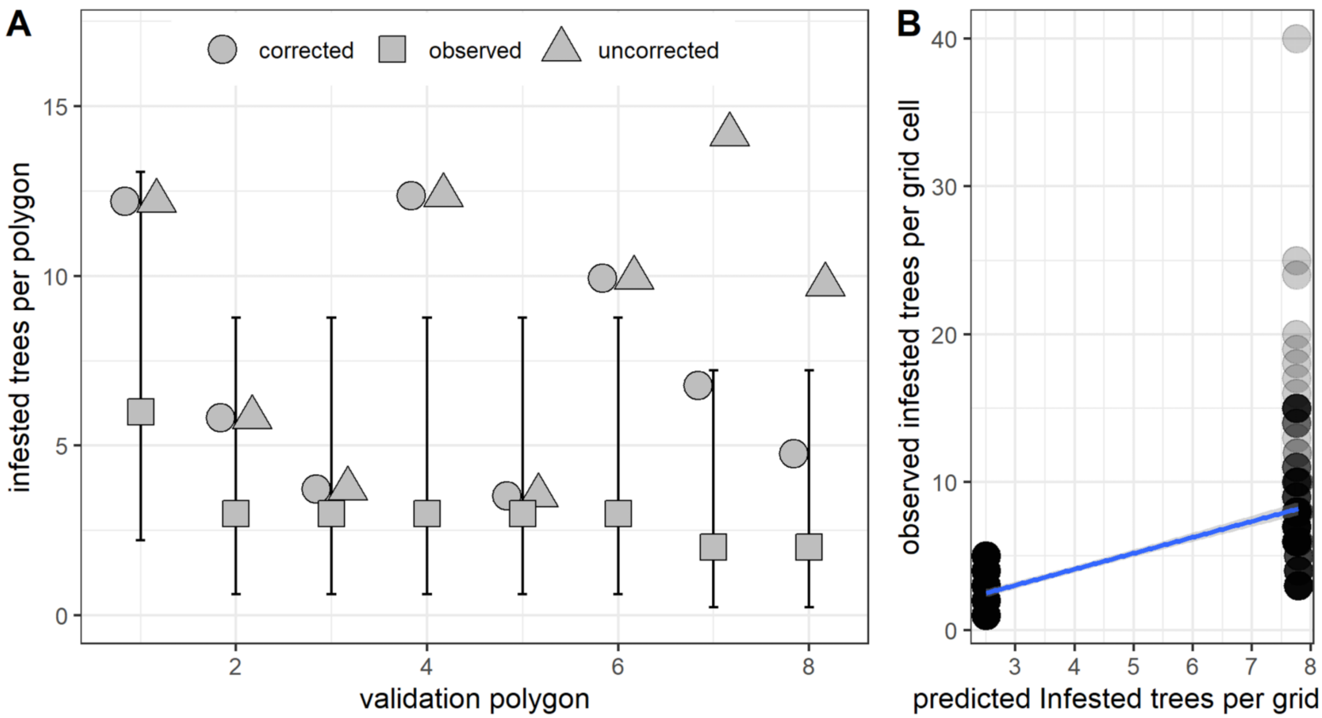
Data were processed using QGIS geographical information system software and in Python (Van Rossum and Drake Jr 1995) by calling GDAL software (GDAL/OGR contributors 2021). Some processing of spatial raster data was completed using R statistical software (R Core Team 2021) with the raster package (Hijmans 2020); figures were produced using ggplot2 (Wickham 2016), cowplot (Wilke 2020), sf (Pebesma 2018), stars (Pebesma 2021), and ggspatial (Dunnington 2021) packages in R.

Results

When proportions of infested land derived from polygons are summed in Jasper National Park from 2014 to 2020 using eq. 2, they often exceed one in the most severely impacted regions (Fig. 1A) and were as high as 1.51 indicating overestimation of tree mortality caused by mountain pine beetles. After applying our probabilistic correction based on eq. 7, overestimation was greatly reduced (Fig. 1B) and estimated cumulative proportions never exceeded one.

When we compared estimates of killed tree stem counts based on our probabilistic correction with validation data collected by ground survey crews, it was evident that the correction reduced overestimation relative to the uncorrected approach (Fig. 2A, polygons 6 and 7). p-values for the Poisson test of the null hypothesis that the true intensity (λ) of the Poisson distributed observed count of red-topped infested trees in polygons seven and eight was equal to the corrected estimate shown in figure (Fig. 2A) of 6.77 and 4.78 were $p = 0.08$, and $p = 0.35$ respectively. This is a positive validation result for the corrected estimates. Conversely, p-values for the Poisson test of the null hypothesis that the true intensity of the Poisson distributed observed count of red-topped infested trees in polygons seven and eight was equal to the uncorrected estimates of 14.16 and 9.75 were $p = 0.00016$ and $p = 0.006$, respectively. This is a negative validation result for the uncorrected estimates. In polygons which did not overlap areas of previous recorded infestation, corrected and uncorrected estimates were similar and p-values for the null hypothesis

Fig. 2. (A) Validation of tree mortality estimates made using the uncorrected approach and using our probabilistic correction with error bars on observed data that reflect our assumption that counts of infested trees are Poisson distributed random variables. Circular points represent predictions based on our corrected approach, square symbols represent observed data points, and triangular symbols represent estimates produced using the uncorrected approach. Only validation polygons 7 and 8 were infested multiple years in a row and were therefore appropriate tests of the probabilistic correction. (B) Validation plots for an approach for converting aerial surveyed point data that include only severity classifications of light, moderate, and severe to estimates of numbers of infested trees per grid cell, where each grid cell contains up to one point. The line in panel B represents a linear regression $y = -0.1971 + 1.0804x$, $p < 2.2\text{e}{-16}$ ($n = 560$) with adjusted $R^2 = 0.46$.



that the true intensities of the Poisson distributed observed count of red top trees were equal to the corrected and uncorrected estimates were only rejected at an alpha level of 0.05 for validation polygon 4 and polygon 6 (Fig. 2A). However, predictions based on both the corrected and uncorrected approach were consistently above the observed counts of infested trees in the validation polygons (Fig. 2A) indicating that even after correction, we still overestimated killed tree density although overestimation was greatly reduced by our correction.

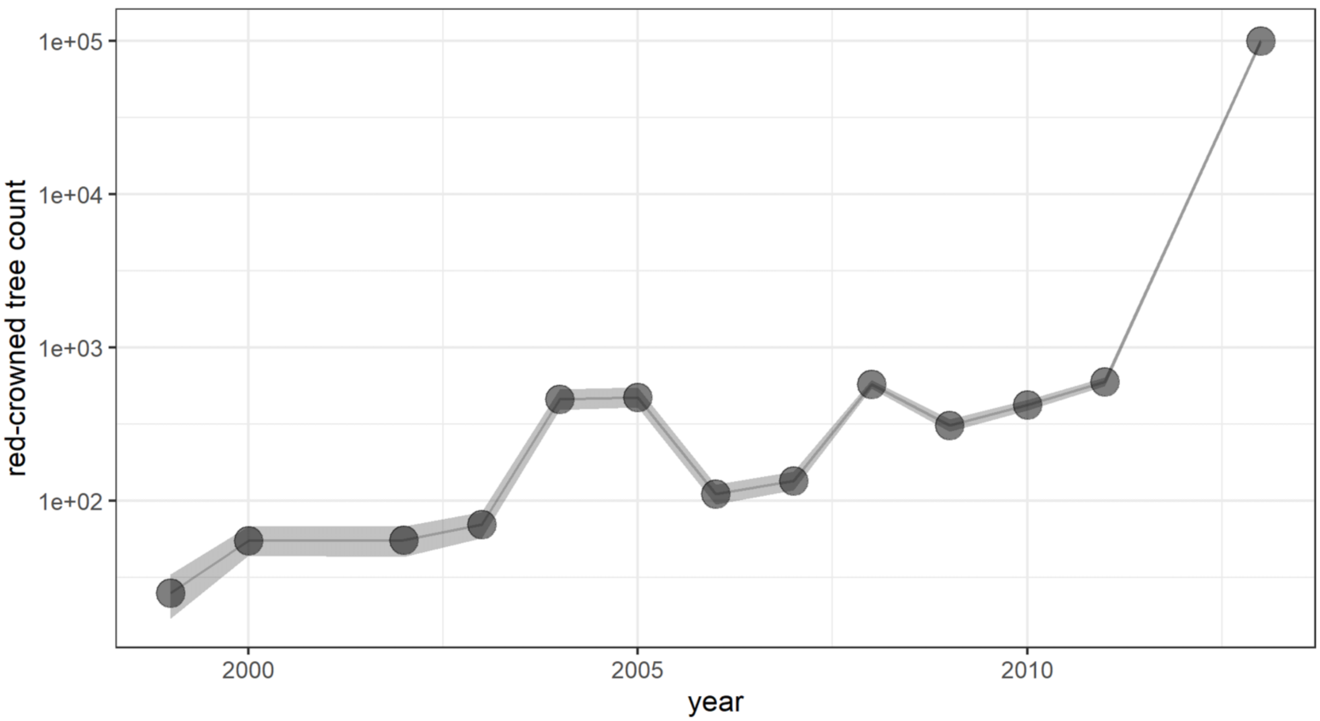
Internal validation of our approach for estimating densities of infestations from point data on mountain pine beetle infestations that lack estimates of numbers of trees per point and instead possess a severity classifier demonstrated that we expect to explain around 46% of the variability in tree counts per grid cell using our estimation approach (Fig. 2B).

The mountain pine beetle outbreak in Jasper National Park increased at a low rate from 1999 to 2011. After 2011, a rapid increase in the count of red-crowned trees in the park was observed (Fig. 3). The difference between park level annual estimates of mountain pine beetle infested trees based on the uncorrected and corrected approaches was most evident from 2018 to 2019 (Fig. 4A). From 2018 to 2019 uncorrected estimates showed an increase in density of infested trees whereas the corrected estimates showed a flat curve (no increase) from 2018 to 2019 (Fig. 4A). Generally, cumulative densities of lodgepole pine trees per 625 m^2 varied between 0 and 50 stems across the national park (Fig. 4B). The outbreak was initially most severe towards the centre of the park and radiated outwards between 2015 and 2018 (Fig. 5). By 2019, the most severe infestations were at the extremities of infestations that occurred between 2015 and 2018 (Fig. 5). Note that infestations recorded as red-topped trees represented beetle activity from the year prior as trees require at least 1 year for foliage to become visibly red.

Discussion

Aerial overview surveys of insect damage are a valuable resource for documenting the impact of forest insects on landscapes that have been collected for nearly a century in North America by federal, provincial, and state agencies responsible for forest management (Wulder et al. 2004). Aerial overview survey data collected in Jasper National Park in Canada overestimated the annual proportion of the landscape that was impacted by mountain pine beetles because when estimated proportions were summed between 2014 and 2020, they often exceeded one. This phenomenon has been reported in several other studies that estimate propor-

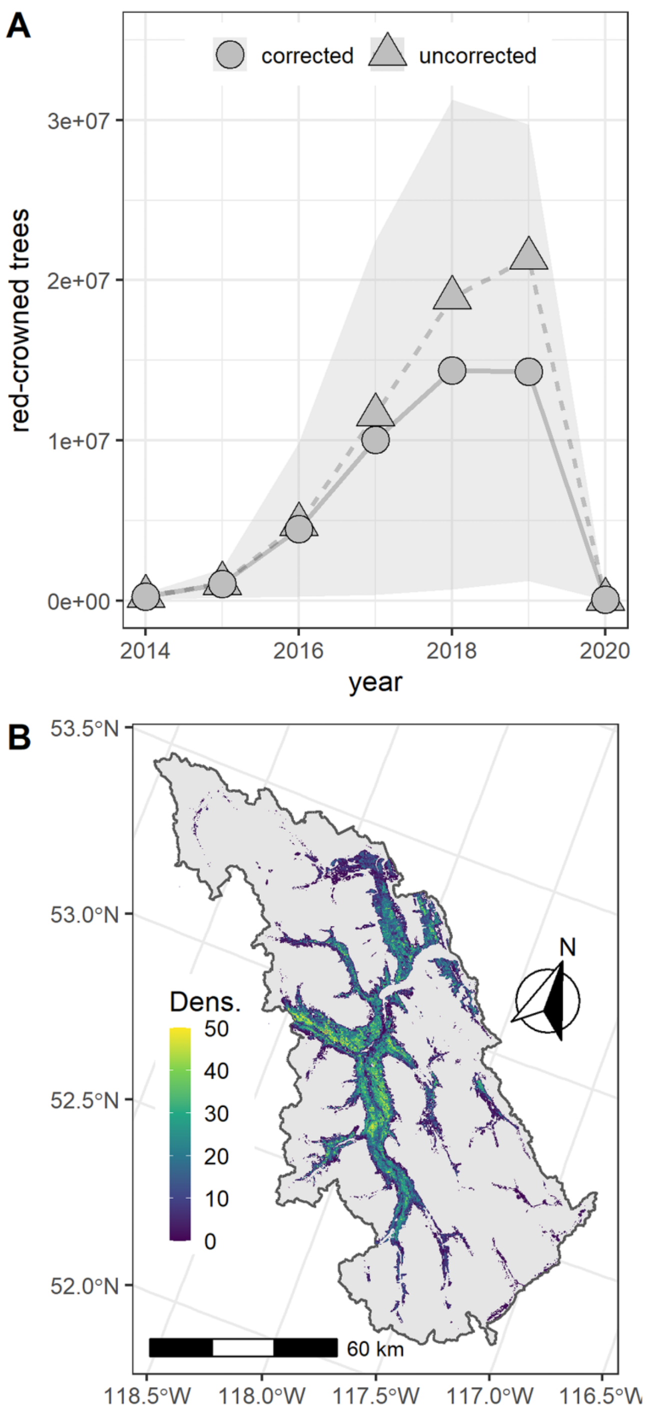
Fig. 3. Estimates of total counts of mountain pine beetle infested trees with red crowns in Jasper National Park from 1999 to 2013 with estimates derived from data recorded by aerial surveyors as points. Each point represented a tree or a cluster of several trees. Note the vertical axis is on a logarithmic scale and data were not available in 2012. The upper and lower boundaries of the shaded region represent the upper and lower 95% confidence intervals on the annual counts estimated using a bootstrap approach.



tions or percentages of pines killed by mountain pine beetles using aerial survey data (Wulder et al. 2009; Egan et al. 2019, 2020). We hypothesized that part of this overestimation was due to nonindependence in estimates of tree mortality in areas where infestations overlapped from one year to the next. Our analysis provides evidence that supports this hypothesis because estimates that did not account for serial nonindependence in areas infested multiple years in a row overestimated numbers of trees killed by mountain pine beetles in our validation data set, whereas estimates that were corrected using our probabilistic method did not. Conversely, estimates in regions that were not serially infested did not overestimate infested tree densities regardless of whether they were corrected or not. Moreover, once estimates of proportions of pine stands that were infested by mountain pine beetles were corrected to account for serial nonindependence, their sum over the outbreak never exceeded one in any location in our study area. We expect that part of the nonindependence in serially infested regions where data are recorded using sketch-mapped polygons arises when a proportion of red-crowned trees are double counted during aerial surveys because they retained red foliage for longer than 1 year — a problem that is likely ubiquitous across all aerial overview survey data of mountain pine beetle damage.

Aerial overview survey data were not originally conceived to give accurate estimates of densities of killed trees but were rather used to assess relative abundance of damage over time and space. However, several studies have proposed methods of translating aerial survey data into mortality estimates (Coops et al. 2006; Robertson et al. 2009; Meddens et al. 2012; Meigs et al. 2015). Most approaches for translating aerial survey data to mortality estimates rely on remotely sensed data and forest inventory data. As methods of monitoring forest insects including the mountain pine beetle are modernized, they will rely more heavily on remotely sensed imagery (Coops et al. 2006) and less on aerial survey data. In this century, the authors anticipate that most landscape scale monitoring of forest insects will be done using aerial, satellite, and drone imagery in combination with machine learning algorithms. In this context, historical aerial overview survey data may be useful for training algorithms to perceive insect damage the way that aerial surveyors do. The same hurdle of double counting red tree crowns and serial nonindependence may inhibit accurate estimation of tree mortality caused by infestations of mountain pine beetle and other forest insects using remotely sensed imagery. If the remotely sensed images that are used to assess forest insect damage have a resolution equivalent to the area of individual tree crowns, it is possible to avoid the problem of double counting affected tree crowns by omitting crowns that were affected in previous years. Until imagery at this high resolution is readily available, the probabilistic approach we used in this study will be useful for correcting inaccuracies in estimates based on remotely sensed imagery as well as estimates using aerial overview survey data.

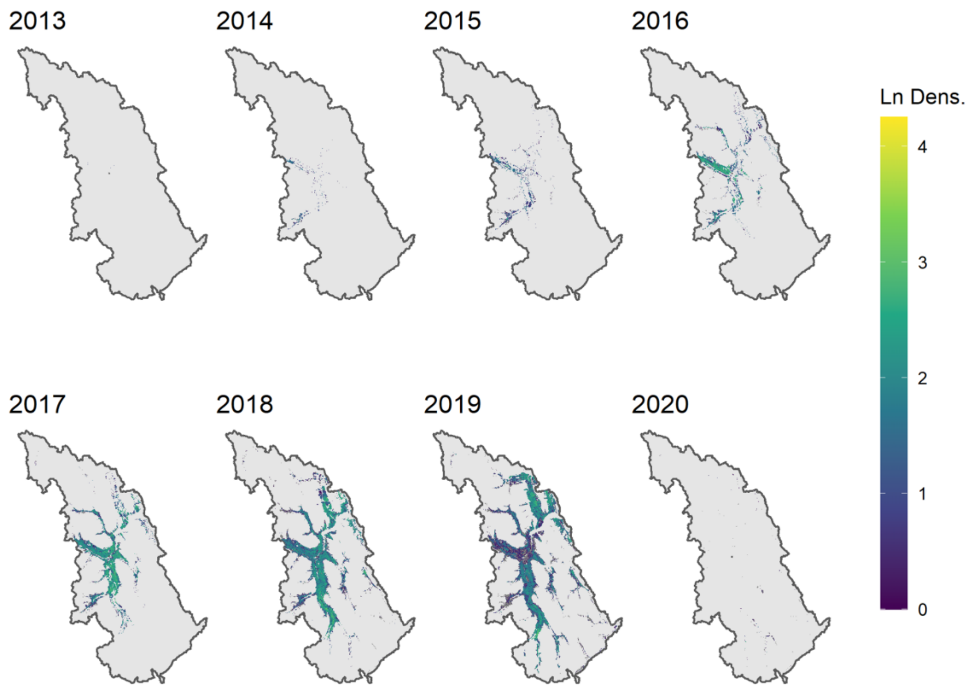
Fig. 4. (A) The trajectory of the mountain pine beetle outbreak in Jasper National Park starting in 2014, when infestations were recorded by aerial surveyors using polygons. The circular symbols represent the estimates derived from the corrected approach (eq. 6), whereas the triangle symbols represent estimates based on the uncorrected approach (eq. 1). The upper and lower boundaries of the shaded region represent upper and lower 95% CI for estimates of the corrected annual numbers of infested trees. (B) The cumulative density of killed trees per 625 m^2 as estimated using the probabilistic correction (eq. 6). The map was produced using the ggplot2 and cowplot packages in R with aerial overview survey data collected by the Canadian Forest Service and National Parks Canada that were rasterized using GDAL software in Python.



Limitations on positional accuracy of aerial survey data (Wulder et al. 2004) as well as uncertainties that arise when raster data are refactored to a different resolution using the nearest neighbour approach may constrain the predictive accuracy of our data sets at the 25 m × 25 m resolution. Therefore, comparisons of our predicted tree mortality estimates to observed data would likely be more appropriate at the 250 m × 250 m (6.25) ha resolution. However, we chose to leave decisions on the desired resolution of data in the hands of potential users of our data sets rather than decrease the resolution of our predictions ourselves. Because estimates of susceptible host tree densities across our study area were unavailable, we used remotely sensed data in tandem with forest inventory data to estimate host pine density. Estimating initial host tree density in this way undoubtedly adds uncertainty to our predictions of densities of killed trees in our raster data sets. Moreover, because aerial surveyors view deceased trees from above, they are more able to detect dominant and codominant killed trees. Thus, assessing mortality from above means that aerial surveyors are more likely to miss suppressed and intermediate trees that have been killed by mountain pine beetles. Although mountain pine beetles in their epidemic phase tend to attack dominant and codominant pines, they will gradually be forced into smaller trees as they deplete their preferred host size classes (Safranyik and Carroll 2006). For this reason, estimates based on aerial surveys of heavily attacked regions may sometimes underestimate the density of trees that have been killed. Another limitation of the current study is the paucity of validation data available to test our tree mortality predictions. Our validation polygon data set for Jasper National Park only contained eight polygons, in which ground crews counted infested trees. Of these eight polygons, only two delineated regions with records of ongoing serial infestation. Therefore, our validation data set for assessing our probabilistic correction was limited to these two polygon plots. Based on our comparison of predictions to infested tree counts in the validation polygons, it is likely that even with the probabilistic correction we have applied, our predictions of densities of lodgepole pine trees killed by mountain pine beetles in Jasper still overestimate tree mortality. Jasper National Park is currently commissioning a park-wide vegetation inventory. Once completed, this inventory will allow comparison of predictions of cumulative pine mortality from this study to inventory-based assessments of tree mortality. We look forward to this opportunity to assess the accuracy of our predictions and correct for discrepancies that may be revealed.

In the coming decades, damage caused by forest insects will be quantified using high-resolution multispectral imagery that will allow researchers and their algorithms to distinguish individual tree crowns, to classify them as killed or not killed, and possibly to identify the disturbance agent responsible for damage. However, many important historical outbreaks were quantified by aerial surveyors and approaches such as this one that allow more accurate assessment of historical damage will remain important. In the immediate future, we anticipate that the approach and the raster data associated with this study will be valuable to forest fire researchers quantifying the fire fuels that the mountain pine

Fig. 5. The progression of the mountain pine beetle outbreak from 2013 to 2020 with estimates natural logarithm transformed densities of red-crowned trees per 625 m^2 grid cell based on the corrected approach (eq. 6). Maps were produced using the ggplot2 and cowplot packages in R.



beetle outbreak generated in Jasper National Park, to conservationists seeking to understand the impact of mountain pine beetle outbreaks on endangered species, and to population ecologists seeking to analyse outbreak dynamics.

Acknowledgements

We thank two anonymous reviewers for insightful comments that improved this manuscript. We are also grateful for Roger Brett's (Canadian Forest Service – Parks Canada) contribution to this work, primarily through many hours of aerial surveillance, data cleaning, analysis and interpretation of Canadian forest insect and forest health data, but also because of numerous discussions that clarified the uses, strengths, and weaknesses of aerial survey data of forest insects. We are appreciative of edits to an earlier version of this manuscript by Landon Shepherd and Chris Watson of Jasper National Park (Parks Canada), which improved its readability and logical flow. This work was funded by the Pest Risk Management and Cumulative Effects programs of Natural Resources Canada and was produced while both authors were employed at the Canadian Forest Service.

Article information

History dates

Received: 11 July 2021
Accepted: 11 February 2022
Accepted manuscript online: 27 April 2022
Version of record online: 27 July 2022

Data availability statement

The data generated by this research will be submitted to the Dryad data repository to facilitate public access.

Author information

Author ORCIDs

Devin W. Goodsman https://orcid.org/0000-0003-1935-5779

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjfr-2021-0200.

References

Alberta Forest, Lands and Wildlife. 1985. Alberta phase 3 forest inventory. Technical report, Edmonton. Available from https://open.alberta.ca/dataset/fdee43d7-d296-41fa-9934-8fe78239ecf7/resource/4b0b3f11-510e-461a-9887-5cba15f2d786/download/afred-alberta-phase-3-forest-inventory-an-overview-1985.pdf

Beaudoin, A., Bernier, P.Y., Guindon, L., Villemaire, P., Guo, X.J., Stinson, G. et al. 2014. Mapping attributes of Canada's forests at moderate resolution through *k*NN and MODIS imagery. Can. J. For. Res. **44**(5): 521–532. doi:10.1139/cjfr-2013-0401.

Coops, N.C., Wulder, M.A., and White, J.C. 2006. Integrating remotely sensed and ancillary data sources to characterize a mountain pine beetle infestation. Remote Sens. Environ. **105**(2): 83–97. doi:10.1016/j.rse.2006.06.007.

Dunnington, D. 2021. ggspatial: spatial data framework for ggplot2. Available from https://cran.r-project.org/package=ggspatial.

Efron, B., and Tibshirani, R.J. 1994. An introduction to the bootstrap. 1st ed. Chapman & Hall/CRC, Boca Raton, FL. Available from https://www.routledge.com/An-Introduction-to-the-Bootstrap/Efron-Tibshirani/p/book/9780412042317 [accessed 1 May 2021].

Egan, J.M., Kaiden, J., Lestina, J., Stasey, A., and Jenne, J.L. 2019. Techniques to enhance assessment and reporting of pest damage estimated with aerial detection surveys. Report No. R1-19-09. US Forest Service, Missoula, Montana, USA. pp. 1–33. doi:10.13140/RG.2.2.34512.94729.

Egan, J.M., Kaiden, J., and Lestina, J. 2020. Multi-scale validation of aerial detection survey accuracy and post-processing techniques to depict mountain pine beetle outbreak severity and extent in the northern region.

Garwood, F. 1936. Fiducial limits for the poisson distribution. Biometrika, **28**(3/4): 437–442. doi:10.2307/2333958.

GDAL/OGR contributors. 2021. {GDAL/OGR} Geospatial data abstraction software library. Available from https://gdal.org.

Goodsman, D.W., Koch, D., Whitehouse, C., Evenden, M.L., Cooke, B.J., and Lewis, M.A. 2016. Aggregation and a strong Allee effect in a cooperative outbreak insect. Ecol. Appl. **26**(8): 2623–2636. doi:10.1002/eap.1404. PMID: 27862568.

Hijmans, R.J. 2020. raster: geographic data analysis and modeling. Available from https://cran.r-project.org/package=raster.

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

Meddens, A.J.H., Hicke, J.A., and Ferguson, C.A. 2012. Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States. Ecol. Appl. **22**(7): 1876–1891. doi:10.1890/11-1785.1. PMID: 23210306.

Meigs, G.W., Kennedy, R.E., Gray, A.N., and Gregory, M.J. 2015. Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the Pacific Northwest Region, USA. For. Ecol. Manage. **339**(1): 71–86. 10.1016/j.foreco.2014.11.030.

Monserud, R.A., Huang, S., and Yang, Y. 2006. Predicting lodgepole pine site index from climatic parameters in Alberta. For. Chron. **82**: 562–571.10.5558/tfc82562-4

Pebesma, E. 2018. Simple features for R: standardized support for spatial vector data. R J. **10**(1): 439–446. doi:10.32614/RJ-2018-009.

Pebesma, E. 2021. stars: spatiotemporal arrays, raster and vector data cubes. Available from https://cran.r-project.org/package=stars.

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.r-project.org/.

Robertson, C., Farmer, C.J. Nelson, T.A., Mackenzie, I.K., Wulder, M.A., and White, J.C. 2009. Determination of the compositional change (1999–2006) in the pine forests of British Columbia due to mountain pine beetle infestation. Environ. Monit. Assess. **158**(1–4): 593–608. doi:10.1007/s10661-008-0607-9. PMID: 19020984.

Safranyik, L., and Carroll, A. 2006. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. *In* The mountain pine beetle: a synthesis of its biology, management and impacts on lodgepole pine. *Edited by* L. Safranyik, and W. R. Wilson. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia. pp. 3–66. doi:10.1016/j.giec.2010.09.011.

Van Rossum, G., and Drake, F.L., Jr. 1995. Python tutorial. Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag: New York. Available from https://ggplot2.tidyverse.org.

Wilke, C.O. 2020. cowplot: streamlined plot theme and plot annotations for "ggplot2." Available from https://cran.r-project.org/package=cowplot.

Wulder, M., Dymond, C., and Erickson, B. 2004. Detection and monitoring of the mountain pine beetle. Available from www.pfc.cfs.nrcan.gc.ca [accessed 1 May 2021].

Wulder, M.A., Dymond, C.C., White, J.C., Leckie, D.G., and Carroll, A.L. 2006. Surveying mountain pine beetle damage of forests: a review of remote sensing opportunities. For. Ecol. Manage. **221**(1–3): 27–41. doi:10.1016/j.foreco.2005.09.021.

Wulder, M.A., White, J.C., Grills, D., Nelson, T., Coops, N.C., and Ebata, T. 2009. Aerial overview survey of the mountain pine beetle epidemic in British Columbia: communication of impacts. BC J. Ecosyst. Manage. **10**(1): 45–58. Available from www.forrex.org/publications/jem/ISS50/vol10_no1_art5.pdf [accessed 1 May 2021].