

Regression Tree Modeling of Spatial Pattern and Process Interactions

Trisalyn A. Nelson, Wiebe Nijland, Mathieu L. Bourbonnais, and Michael A. Wulder

Abstract In forestry, many fundamental spatial processes cannot be measured directly and data on spatial patterns are used as a surrogate for studying processes. To characterize the outcomes of a dynamic process in terms of a spatial pattern, we often consider the probability of certain outcomes over a large area rather than on the scale of the particular process. In this chapter we demonstrate data mining approaches that leverage the growing availability of forestry-related spatial data sets for understanding spatial processes. We present classification and regression trees (CART) and associated methods, including boosted regression trees (BRT) and random forests (RT). We demonstrate how data mining or machine learning approaches are useful for relating spatial patterns and processes. Methods are applied to a wild-fire data and covariate data are used to contextualize the quantified patterns. Results indicate that fire patterns are mostly related to processes influenced by people. Given the growing number of multi-temporal and large area datasets on forests and ecology machine learning and data mining approaches should be leveraged to quantify dynamic space-time relationships.

Spatial Pattern and Processes

Many scientific disciplines are interested in quantifying the relationships between spatial patterns and spatial processes (Nelson 2012). Intuitively, we understand that geographic patterns present at a given date can tell a story about a prior sequence of

T.A. Nelson (✉)
School of Geographical Sciences and Urban Planning, Arizona State University,
Tempe, AZ 85287, USA
e-mail: trisalyn.nelson@asu.edu

W. Nijland • M.L. Bourbonnais
Spatial Pattern Analysis and Research (SPAR) Laboratory, Department of Geography,
University of Victoria, PO Box 3060, Victoria, BC, Canada, V8W 3R4

M.A. Wulder
Canadian Forest Service, Natural Resources Canada, Pacific Forestry Centre,
506 West Burnside Road, Victoria, BC, Canada, V8Z 1M5

events, or reveal information on the functioning of a system. Investigation of spatial patterns may lead to a better understanding of processes otherwise obscured from measurement. Popular culture is full of examples of the link between pattern and process. A familiar example of the link between mapped patterns and processes is the use of maps to plot and link events to solve a crime in TV dramas. Perhaps more relevant is the use of pattern-based jargon like hot spots, a spatial pattern of abundance, to indicate locations with important ecological processes (Nelson and Boots 2008).

Owing to the popularity of global positioning systems (GPS), geographic information systems (GIS), remote sensing, and increasingly user-friendly access to these data via Internet applications like Google Maps, there has been an explosion in the availability and interest of mappable data (Nelson 2012). Citizens, managers, and scientists have an increasing interest in extracting the information available from maps. As the amount of spatial data increases concurrently with improved access, the spatial co-location of data from different sources is becoming increasingly powerful for the analysis of spatial patterns and processes. Users from novices to experts, confronted with spatial data, are posing questions and requesting applications that continue to drive a need for new analytical methods. Rather than considering single points or events, increasingly knowledgeable users desire integration and refinement of the data available. We consider a spatial pattern as the expression of one or many spatial processes and a spatial process as a sequence of events carried out in some definite manner (Haining 1990). While spatial patterns are typically considered a snapshot in time, processes are temporally dynamic and expected to change (Getis and Boots 1978).

Describing Spatial Patterns

To describe the outcomes of a dynamic process in terms of a spatial pattern, we often consider the probability of certain outcomes over a large area rather than on the scale of the particular process. Forest fire occurrence is a good example of this. Fires are short-lived, localized events that are highly dependent on weather conditions and the presence of a source of ignition. The specific location of fires is therefore not readily predictable. However, over larger areas patterns in fire occurrence become clear with more frequent fires in areas with high fuel loads, dry climates, and more ignition sources (Parisien and Moritz 2009; Gralewicz et al. 2011; Gralewicz et al. 2012). Measurement of fuel or climate conditions on a landscape may therefore inform us about the probability of wildfire occurrence in a specific place (Chuvieco and Salas 1996; Parisien and Moritz 2009).

A similar ratio is used to map the ecological niche of a plant or tree species. We can create a model of the preferred climate and soil condition of a particular tree showing where it can grow and where it is more likely to thrive (Nijland et al. 2014; Waring et al. 2014), but within that range it still may or may not be present because of past events like the presence of a seed source and the availability of any free space.

Spatial patterns can either be directly related to the spatial processes by occurring in the same place or have a more complex relationship acting as a function of distance to another location. The nature of the relationship influences how we map the spatial patterns involved. The spatial pattern of forest logging operation provides an example of co-located and more complex spatial interactions: Forest logging is done in places with merchantable wood present, a simple co-location. Logging also occurs in places with access to a mill or other processing plant; the spatial relationship between logging and mills is more complex because they interact dependent on distance or connectivity to the mill. Even for simple co-location of spatial variables we still need to consider what “the same place” actually means.

When working with data from different sources, as is often the case, the data needs to be unified to a common spatial unit. The size of the principal spatial unit is dependent on the process we are studying (Nijland et al. 2009). In some cases a predefined or natural spatial unit is available, such as tenure areas in forests or watersheds in hydrology; in other cases, we may need to find our own solution by imposing a grid or other regular pattern, or else segment the area by other boundaries, perhaps roads or administrative areas (Dark and Bram 2007). Naturally, the detail of our environmental data needs to match the principal spatial unit and process. If we use climate patterns only to model the distribution of a plant or tree species, the results will be limited to regional patterns. If more detail is required, additional information with finer detail, for example, soil conditions or forest structure, may be included in the model (Nijland et al. 2014).

With complex spatial relationships we need to define the connection between our location and nearby features. In many cases our best approximation is by simple distance. With other cases the connection may be limited by physical boundaries or operate over a network. For instance, when transporting logs to a mill, the travel time or cost over the road network is more relevant than the simple Euclidean distance to the mill (Anderson et al. 2011).

Process Complexity

In forestry, many fundamental spatial processes cannot be measured directly and data on spatial patterns are used as a surrogate for studying processes (Levin 1992; Sokal et al. 1998; Jacquez 2000). As an example, consider landscape-scale forest insect infestations (e.g., Bone et al. 2013). The spatial processes of large-area insect infestations cannot be measured directly and the pattern of infested trees is the expression of the process of infestation (Robertson et al. 2008). By quantifying the spatial and temporal patterns of insect infestation, we generate new hypotheses or knowledge on the spatial processes of infestation. For instance, knowing the distance at which beetle infestation patterns are aggregated on the landscape provides information on the spatial scale of infestation processes (Powers et al. 1999) and identifying hot spots of infestation in space and time provides evidence of how forest susceptibility changes as an infestation progresses (Nelson et al. 2006).

Relating spatial pattern and process can be complex. Spatial patterns and processes are connected through a positive feedback; patterns are an expression of process, but processes are influenced by pattern (Fortin et al. 2003). For instance, in a forestry context the spatial pattern of the forest age is known to influence the risk of fire, while at the same time fire changes the age distribution of the forest (Gralewicz et al. 2012). The constant interplay between fire and forest age distribution is just one example of the feedback between pattern and process that can complicate interpretation. Interactions between pattern and process can be further concealed by the complicated one-to-many relationship between pattern and process. Many processes will express similar spatial pattern on the landscape and it can be near impossible to assign a pattern to a precise process (Fortin and Dale 2005 pp. 3–4; Langford et al. 2006). In reality most patterns are the result of many processes interacting together and through time.

Data Mining

Growing availability of forestry-related spatial data sets is creating an opportunity to use the spatial patterns in those data sets to explore and quantify spatial processes. Well-known spatial methods like kriging or k-nearest neighbor interpolation do use spatial patterns to generate predictions for unmeasured locations, but do not provide information on the process side (Hastie et al. 2009). Data mining methods are specifically suited to model patterns and processes in large volumes of information (Shekhar et al. 2003). Classification and regression trees (CART) (Breiman et al. 1984) and associated methods, such as boosted regression trees (BRT) (Elith et al. 2008; Hastie et al. 2009) and random forests (RT) (Breiman 2001), represent data mining or machine learning approaches useful for relating spatial patterns and processes. Exploratory in nature, CART and CART-based approaches are increasingly used in forestry and ecology research where there is an interest in quantifying and predicting spatial pattern and process dynamics and identifying influential drivers of these dynamics (De'ath and Fabricius 2000; Hawkins 2012). CART-based approaches have been used to understand natural and human drivers of spatial patterns of forest fire ignition (Gralewicz et al. 2012; Bourbonnais et al. 2013a, b), potential spatial variability in vegetation and forest composition under different climate change scenarios (Holmes et al. 2013), spatial patterns of species distributions based on environmental gradients (Elith et al. 2006; Leathwick et al. 2006), and wildlife health in the context of habitat conditions and human disturbance (Bourbonnais et al. 2013b).

CART-based methods are well suited to the study of spatial pattern and process as they can handle large datasets, complex nonlinear relationships, and missing data that are prevalent in spatial datasets, and can accommodate continuous and categorical variables, as well as variable interactions (De'ath and Fabricius 2000). The ability to handle mixed data types and the relatively straightforward interpretation of the resultant model structure put CART at an advantage over neural nets and support

vector machines which are other machine learning methods designed to handle large data volumes and complex relationships (Hastie et al. 2009). Unlike parametric regression methods, such as ordinary least-squares regression and generalized linear models, CART-based methods make no assumptions about the structure of the data and the underlying processes, which in forestry and ecology are complex and may be unknown, and as such represent a flexible data-driven nonparametric regression approach.

In this chapter we provide an overview of CART, and two CART-based methods, BRT and RF. We demonstrate each approach, illustrate how to interpret the results, and comment on the strengths of each method through a case study that aims to quantify how spatial patterns of mountain pine beetle infestation are changing fire processes in British Columbia, Canada. While we highlight the utility of these approaches, we refer the reader who requires additional theoretical background to comprehensive reviews provided by Berk (2008) and Hastie et al. (2009). We introduce the theory of each method and then demonstrate how it applies to the case study of the interaction of mountain pine infestations on large forest fires. We conclude with an interpretation of modeling results and highlight future directions.

Methods

CART Models

CART models, originally implemented by Breiman et al. (1984), recursively partition (i.e., split) the response, in our case the spatial pattern of interest, into increasingly homogeneous subsets based on information provided by the predictor variables considered (Berk 2008). At each binary split, a threshold value (for continuous variables) or group level (for categorical variables) that best reduces the error sum of squares in the case of a continuous response, or the Gini index in the case of a categorical response, is selected to partition the data into two subsets. Data partitioning continues in a stagewise manner, meaning earlier split values are not considered in subsequent partitions, until no further meaningful reductions in the error sum of squares or Gini index can be found based on the data. This exhaustive approach generally leads to a very large tree being grown, or many splits, which is then pruned to remove splits that over-fit the data identified through cross-validation (Hastie et al. 2009).

When displayed graphically, a CART model is an inverted tree with the root node representing the undivided data at the top and branches defined by partition values and leaves, or terminal nodes, representing the response values or groups beneath (De'ath and Fabricius 2000). In our case, the branches and partition values represent the spatial processes considered and the terminal nodes the spatial pattern of interest. The hierarchical structure of the CART model is interpreted based on the partition values and terminal node assignments. At each split, observations that satisfy the decision rule are assigned to the group to the left while those that do not are

assigned to the group to the right. The split values of the process variables, and associated terminal node value assignments representing the spatial pattern, allow us to infer the directionality of the spatial pattern-process dynamics. Additionally, the hierarchical structure of the CART model automatically incorporates interaction effects among process variables as terminal node assignments are dependent on all the preceding splits. For continuous response variables, CART model (i.e., regression tree) performance can be assessed based on the total sum of squares variance explained or the deviance explained (De'ath and Fabricius 2000). CART model performance for categorical response variables (i.e., classification tree) can be determined using a variety of classification accuracy assessments including misclassification error rates (De'ath and Fabricius 2000), confusion matrices (Berk 2008), and area under the receiver operating characteristic curve (Hastie et al. 2009). However, while CART models are easy to fit and interpret they do have a number of drawbacks. As mentioned, they are prone to over-fitting and defining stopping criteria and pruning large trees is not trivial (Murthy 1998; Berk 2008). CART models are also overly sensitive to changes in input data that can result in major changes in tree structure and split values (Hastie et al. 2009), making them a temporally static modeling approach. As a result, more robust methods that combine multiple stochastic trees have been developed.

BRT

BRTs use boosting algorithms to improve model accuracy by combining and averaging many CART models, rather than relying on a single tree to explain the association among spatial pattern and process(es) (De'ath 2007; Elith et al. 2008). Similar to CART, BRT is a stagewise procedure. However, unlike CART where binary splits are selected at each stage, BRT iteratively fits a completely new tree at each stage in order to minimize a loss function such as the deviance explained. Beginning with the first tree, a random subset of the data was selected and a tree is built that best minimized the loss in deviance explained. At each subsequent stage, a new tree using randomly selected data was built based on the residuals, or the unexplained variance in the response, from the combination of trees that already exist. Only the fitted values are reestimated at each iteration, while the existing trees and split values are unchanged. The final stochastic BRT model is a linear combination of hundreds or thousands of trees, rather than a single tree, resulting in a more robust model compared to the single tree produced by CART (Elith et al. 2008). However, unlike CART, which has few user-defined parameters, BRT models require the user to define the bag fraction, which specifies the proportion of the data randomly drawn at each iteration; the model learning rate, which determines the contribution of each tree to the model; and the tree complexity, which specifies the complexity of interaction effects included in the model (De'ath 2007; Elith et al. 2008). Combined, the learning rate and tree complexity determine the optimal number of trees required in the BRT model to minimize the loss in deviance explained

while avoiding over-fitting the data. We refer the reader to Elith et al. (2008) for an in-depth review of BRT parameters and model fitting procedures.

Similar to CART, the performance of BRT models can be assessed using the deviance explained or a classification accuracy assessment usually based on cross-validation using withheld data (De'ath 2007; Elith et al. 2008). While variable importance and directionality in CART models are easily interpreted using a tree diagram, no such output is produced by BRT as it combines numerous trees. Instead, the influence of each variable is determined based on the number of times each variable is chosen as a split, weighted by its improvement to the model at each split averaged over the total number of trees in the model (Friedman 2001; Friedman and Meulman 2003; Elith et al. 2008). Partial dependence plots, which average out the influence of all other variables besides the variable selected, are used to visualize the associations between influential process variables and the spatial pattern response (Friedman 2001; Friedman and Meulman 2003).

RF Models

RF models (for the statistical background see Breiman 2001) are another machine learning approach that combine and average many CART models. RFs use bootstrap samples of the data to fit numerous (generally 500–2000) individual regression or classification trees. Unlike BRT, a limited number of predictor variables are also drawn at random in each bootstrap sample and used for the recursive partitioning to fit each tree. The number of variables to be randomly selected in each bootstrap sample is the only user-defined parameter in a RF. Also, bootstrap sampling and recursive partitioning of individual trees are usually not done in a stagewise manner meaning influential predictors and thresholds may be selected more than once. However, as observations from all the trees are aggregated through averaging, RFs are quite robust to over-fitting. Observations in the data that do not occur in the bootstrap samples are referred to as the out-of-bag data (Cutler et al. 2007). Each tree is grown to its maximum size and used to predict the out-of-bag data, eliminating the need to retain data for cross-validation (Prasad et al. 2006; Cutler et al. 2007).

The comparison of predicted values or classes from the bootstrap aggregation of trees used to build the RF with those retained in the out-of-bag data provide the mean square error (regression) or misclassification error rate based on the Gini index (classification) of the RF model. Similar to BRT, partial dependence plots are used to visualize associations between process-based variables and the spatial pattern response (Cutler et al. 2007; Hastie et al. 2009), and variable importance is assessed based on the number of times a variable is included as a split in the model and how well it performs. In the case of RF, the accuracy of a variable is determined by randomly permuting values from the out-of-bag data and then comparing predictions made with these new data to those of the model. The difference, divided by the standard error, between the permuted and original out-of-bag data values or misclassification rate represents the importance of the variable (Cutler et al. 2007).

Case Study Context: Influence of Beetle Infestation Spatial Patterns on Fire Spatial Processes

Disturbance plays an important role in defining landscape pattern and can cause substantial change in ecosystem processes (Turner 1989). In Canadian forests, anthropogenic and natural disturbances, such as harvesting (Masek et al. 2011), forest fires (Stocks et al. 2002), and insect infestations and disease (Hall and Moody 1994; Volney and Fleming 2000), are the primary determinants of forest structure.

While mountain pine beetle (*Dendroctonus ponderosae* Hopkins [Coleoptera: Scolytidae]) infestations are endemic in North American lodgepole pine ecosystems (Amman 1977), the infestation that occurred in western Canada during the 1990s and 2000s was the largest on record and affected over 16 million ha (Walton 2010), leading to widespread mortality of adult lodgepole pine trees in the region. Across the large area affected the severity of infestation varied substantially (Robertson et al. 2009a; Wulder et al. 2010). However, the spatial pattern of forests affected by the infestation has been altered considerably, generally resulting in smaller, more complex, more numerous forest patches (Robertson et al. 2009b; Coops et al. 2010).

Impacts of mountain pine beetle infestation on forest fire regimes are of particular concern, as outbreak-related tree mortality is anticipated to increase the frequency and severity of forest fires (Shore et al. 2006; Negrón et al. 2008). The theory that mountain pine beetle-induced tree mortality results in more severe fires has only recently been tested empirically (Jenkins et al. 2012), and initial results of retrospective studies and empirical testing of mountain pine beetle-fire dynamics have been contradictory (e.g., Page and Jenkins 2007; Simard et al. 2011).

The complex spatial pattern-process interaction between mountain pine beetle infestations and fires seems dependent on the severity of mountain pine beetle attack (Hawkes et al. 2004). When trees are killed, foliar moisture content of both needles and fine fuels decreases (Reid 1961; Shore et al. 2006), causing severely affected mountain pine beetle stands to have increased flammability, higher capacity to support sustained crown fires, and high rates of spread (Turner et al. 1999, Page and Jenkins 2007; Jenkins et al. 2008, 2012; Jolly et al. 2012). However, a decreased amount and spatial continuity in crown fuel loading and contiguity, due to forests having a range of attack severity (i.e., light to severe), have also been found to lessen the probability of crown fire ignition (Klutsch et al. 2011; Simard et al. 2011).

Given the extent of mountain pine beetle damage in British Columbia, Canada, and availability of spatial data, the interaction between infestation patterns and forest fire processes is an ideal case study for demonstrating how interactions between spatial pattern and process can be quantified using spatial data and regression trees.

Study Area

The study area includes the spatial extent of the mountain pine beetle infestation in British Columbia from 1999 to 2009, an area of ~16 million ha (Walton 2010). In order to represent the temporal variability observed in the spread of mountain pine

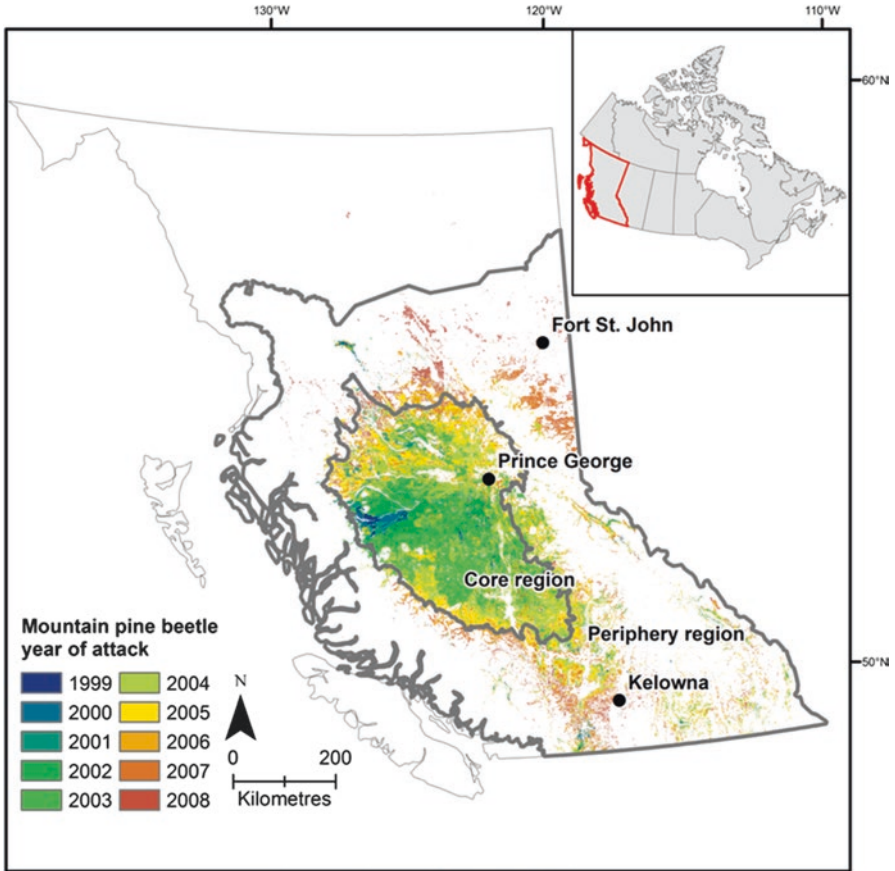


Fig. 1 The cumulative area impacted by mountain pine beetle from 1999 to 2009, divided into three regions (northern Periphery, Core, and southern Periphery) to account for spatial and temporal variability in the spread of the outbreak

beetle across the province, we divided the study area into Core and Periphery regions, based on ecoregions (Fig. 1). Ecoregions partition the province into regions of homogenous vegetation structure (Demarchi 2011). The Core region encompasses the epicenter of the mountain pine beetle epidemic, which transitioned from an incipient mountain pine beetle population in the mid-1990s to an epidemic population in 1999, and has experienced the most severe and widespread mortality (Aukema et al. 2006). While synchronous outbreaks were seen early in the outbreak in the south (Aukema et al. 2006), the epidemic predominantly spread from the study area center towards the south and north (Robertson et al. 2009b; Wulder et al. 2010). The Periphery region has a large abundance of host trees remaining, and mountain pine beetle-induced tree mortality is expected to continue (Walton 2010; Wulder et al. 2010).

Spatial Data

We use a spatial database of past fire activity as well as covariate data sets represented as anthropogenic influences, climate, terrain, elevation, and mountain pine beetle infestation to model the most influential predictors of the spatial pattern of large fires. An overview of data characteristics is provided in Table 1 and all data sets are visualized as maps in Fig. 2.

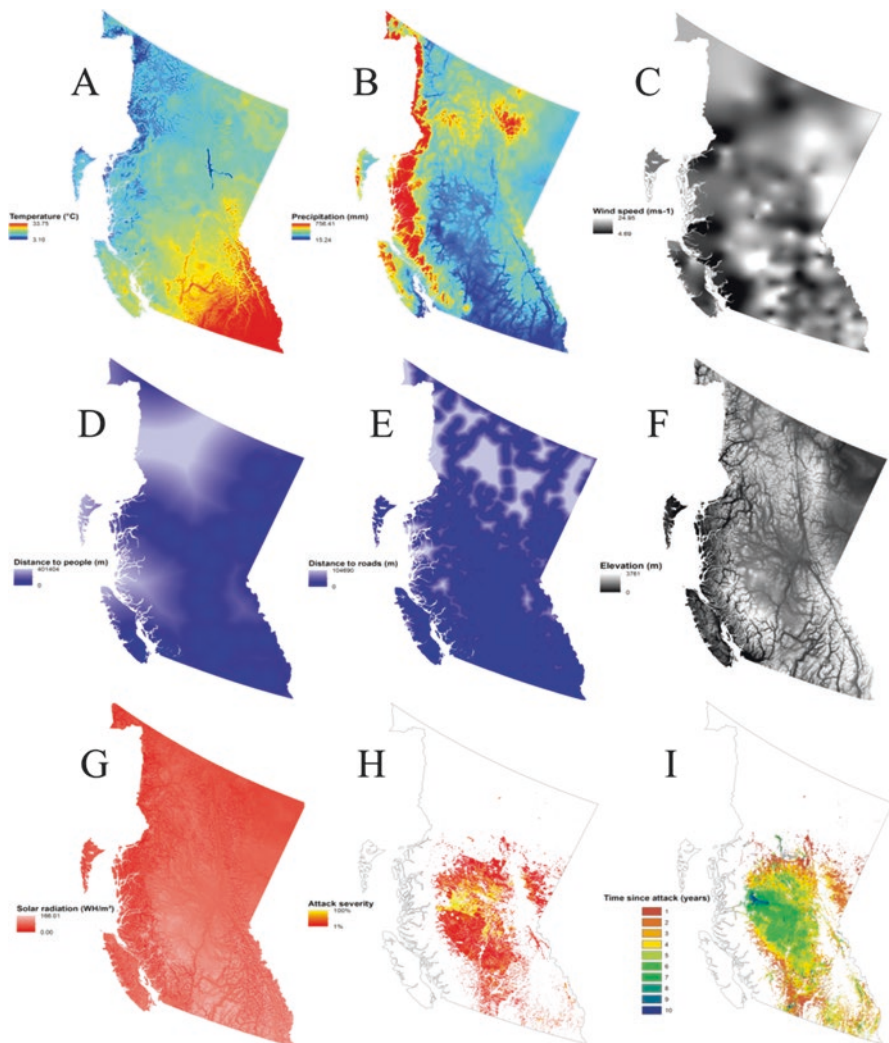


Fig. 2 Covariate layers used in classification tree modeling: **(a)** Temperature. **(b)** Precipitation. **(c)** Maximum average wind speed. **(d)** Distance to people. **(e)** Distance to roads. **(f)** Topography. **(g)** Solar radiation. **(h)** Time since mountain pine beetle attack. **(i)** Severity of mountain pine beetle attack

Wildfire Data

The Canadian National Fire Database (NFDB), a national repository of fire data from provincial, territorial, and Parks Canada fire agencies, provides spatial data for forest fires occurring in Canada (see Stocks et al. 2002). For this study, NFDB polygon data from 1999 to 2009 were used (Canadian Forest Service 2010). To promote reliable fire data only large fires were used. Large fires were defined as ≥ 32 ha. The 32 ha threshold removed the stochastic influence of small fires from the fire data distribution while providing an adequate number of mapped fires. Large fires were aggregated into the three study regions and stratified based on size into three classes: (a) 32 ha–200 ha; (b) 200 ha–1000 ha; and (c) > 1000 ha. In the Core region there were 82 large fires: 51 in Class A, 18 in Class B, and 13 in Class C. In the Periphery region there were 444 large fires: 256 in Class A, 119 in Class B, and 69 in Class C.

Climate Data

Climate is an important determinant of spatial patterns and processes of forest fire (Table 1) (e.g., van Wagner 1977; Flannigan and Harrington 1989; Flannigan et al. 2005; Parisien et al. 2006). In order to account for topographic variation in climate, ClimateWNA (Hamann and Wang 2005; Wang et al. 2011) was used to create weather variables for only the fire season (e.g., April 1–September 30). For the fire season, averages of maximum temperature and precipitation were calculated from 1999 to 2009 at a 1 ha spatial resolution. Hourly wind speed data were interpolated using spline interpolation from a provincial network of nearly 200 fire weather stations maintained by the BC Wildfire Management Branch. Data were stored in a 1 ha grid cell.

Anthropogenic Covariates

Human proximity and access have been found to be drivers of elevated fire incidence in Canada (Gralewicz et al. 2012). However, areas with high densities of human settlement are also subject to extensive fire suppression efforts, which can be a limiting factor of fire size (Parisien et al. 2006). In order to assess anthropogenic influence on fire size, proximity to the nearest populated place was calculated for each 1 ha cell in British Columbia based on persistent nighttime light derived from the DMSP Operational Linescan System (see Wulder et al. 2011). Similarly, the Euclidean distance to the nearest road of any size was calculated for each 1 ha cell in British Columbia using the 2008 road network file from Statistics Canada (2008).

Table 1 Relevance of covariates as determinants of fire size

Covariate	Driver (units)	Abbreviation	Relevance to fire severity	Reference
Weather	1. Temperature (degrees Celsius) 2. Precipitation (mm) 3. Wind speed (m per second)	1. temp_avg 2. precip_avg 3. Wind	1. Contributes to drying of fuels and increased fire behavior. 2. Influences fuel moisture content. Acts as a moderator of fire severity. 3. Direct determinant of fire intensity, shape, and size.	Parisien et al. (2006) Flannigan et al. (2005) Flannigan and Harrington (1989) van Wagner (1977)
Mountain pine beetle	1. Time since attack (years) 2. Percent pine infested (%)	1. tsa 2. comp_mpb	1 and 2. Stand structure following mountain pine beetle infestation is changed significantly over a temporal scale. Changes in fuel loading, continuity, and moisture caused by mountain pine beetle mortality are believed to be key determinants of fire severity.	Simard et al. (2011) Klutsch et al. (2011) Axelson et al. (2010) Jenkins et al. (2008) Page and Jenkins (2007) Lynch et al. (2006) Shore et al. (2006) Bigler et al. (2005) Turner et al. (1999)
Topography	1. Elevation(m) 2. Solar radiation (WH/m ²)	1. Elev 2. Rad	1. Impacts temperature, precipitation, wind speed, and vegetation type. 2. Impacts air temperature and composition of vegetation.	Parisien et al. (2006) Díaz-Avalos et al. (2001) Miller and Urban (2000) Kumar et al. (1997) Franklin (1995)
Anthropogenic Land use	1. Proximity to roads (md) 2. Proximity to populated places (m)	1. dist2rd 2. dist2lt	1 and 2. Regions with greater anthropogenic influence can both contribute (i.e., ignition source) and restrict (i.e., fire suppression) fire size.	Gralewicz et al. (2012) Parisien et al. (2006)

Topographic Data

Elevation may influence temperature, precipitation, and wind speed, as well as vegetation type and contiguity, which have an impact on fire incidence (Díaz-Avalos et al. 2001; Gralewicz et al. 2012) and fire size (Miller and Urban 2000). We used a digital elevation model (DEM) obtained from the Government of Canada portal Geobase and resampled to 1 ha grid cells. Annual shortwave radiation (Watt hours per m²—WH/m²) (Wulder et al. 2010) as derived from the elevation data was also used because solar energy has been found to influence fire size (Kumar et al. 1997).

Mountain Pine Beetle Data

Previous studies of mountain pine beetle and fire dynamics used attack severity (e.g., Turner et al. 1999; Page and Jenkins 2007; Simard et al. 2011) and time since mortality (e.g., Bigler et al. 2005; Lynch et al. 2006) to predict interaction with fire. We employed spatial products generated by Robertson et al. (2009a) at a spatial resolution of 1 ha to represent the spatial pattern of infestation severity, as percent of pixel infested, and time (year) since infestation. Robertson et al. (2009a) integrated aerial overview surveys (AOS) and ground surveys of mountain pine beetle infestation with data on percent pine to map the annual percent pine infested, from 1999 to 2009 within a 1 ha pixel. Time since mortality was also calculated based on when the forest in a pixel reached 50% mortality.

Model Evaluation

In this section we evaluate the three regression tree methods, CART, RF, and BRT, and demonstrate how each can be applied to explore the impact of mountain pine beetle infestation on the spatial patterns of forest fires. Within our modeling framework we assume that covariate data are surrogates for spatial processes. Climate and topography are associated to fire by direct co-location, while anthropogenic influences are modeled as distance to roads and populated areas. By including covariate data sets that represent mountain pine beetle infestation conditions, as well as anthropogenic influences, climate and weather, and elevation, we can determine which processes are the most influential predictors of the spatial pattern of large fires.

We generated separate regression models for the Core and Periphery geographic regions in order to determine how and if various levels of infestation severity and/or duration influenced the spatial pattern of fire. To assess the accuracy of each method we used 70% of data for training and held back a random sample of 30% of data for testing each model. Confusion matrices for classes A, B, and C were provided for each of the models as well as overall accuracy as a percentage.

Cart

The CART model had 82.9% classification accuracy in the Core (Table 2) and 65.5% accuracy in the Periphery (Table 3). In the Core model, the largest fire class (>1000 ha) was most accurately predicted (92.3%). Some of the smallest fires (32–200 ha) were misclassified as midsize fires (11.8%) or as the largest fires (5.9%). The midsize fires (200–1000 ha) were classified 77.8% accurately with the remainder split between the smaller and larger fire classes, 16.7% and 5.6%, respectively. In the Periphery, the midsize fires (200–1000 ha) were most accurately predicted, with 87.5% correct. The smallest fires (32–200 ha) were classified with 66.1% accuracy and most of the misclassified fires were predicted to be midsize (26.5%). The largest fires were predicted with 60.0% accuracy and misclassified as both small (18.6%) and midsize (21.4%) fires.

In Fig. 3 we show CART results for both the Core and Periphery. In the model of the Core, the primary predictor of fire size class was percent pine infested. Values less than 35.0 percent pine infested (comp_mpb) were generally associated with the small and midsize fires (32–200 ha and 200–1000 ha), and the smallest fire class (32–200 ha) was predicted near roads (<3306 m) in locations with higher rainfall (precip_avg ≥ 52.4). Variable associations could indicate that wetter conditions and access that enables quick response to fires are helping to limit the size of fires in the Core, when percent pine infested was less than approximately one-third of a forest stand.

Table 2 For the Core, confusion matrices for classification and regression trees (CART), random forests (RF), and boosted regression trees (BRT)

	CART	CART	CART	BRT	BRT	BRT	RF	RF	RF
	A	B	C	A	B	C	A	B	C
A	82.4	11.8	5.9	72.1	19.8	8.1	63.0	23.5	13.6
B	16.7	77.8	5.6	16.1	79.0	4.8	30.0	20.0	50.0
C	0.0	7.7	92.3	12.2	8.2	79.6	25.0	56.3	18.8

Due to small sample sizes the confusion matrices are made from training and test data and are percentages. The overall accuracy of classification is calculated from only test data and for CART = 82.9%, RF = 49.6%, and BRT = 88.0%. A, B, and C are categories of smaller to larger fire size: 32–200 ha, 200–1000 ha, and >1000 ha

Table 3 For the Periphery, confusion matrices for classification and regression trees (CART), random forests (RF), and boosted regression trees (BRT)

	CART	CART	CART	BRT	BRT	BRT	RF	RF	RF
	A	B	C	A	B	C	A	B	C
A	66.1	26.5	7.4	85.7	10.0	4.3	64.7	25.5	9.8
B	12.5	87.5	0.0	3.4	86.2	10.3	39.6	52.1	8.3
C	18.6	21.4	60.0	0.0	0.0	100.0	15.4	7.7	76.9

Due to small sample sizes the confusion matrices are made from training and test data and are percentages. The overall accuracy of classification is calculated from only test data and for CART = 65.5%, RF = 64.4%, and BRT = 73.9%. A, B, and C are categories of smaller to larger fire size: 32–200 ha, 200–1000 ha, and >1000 ha

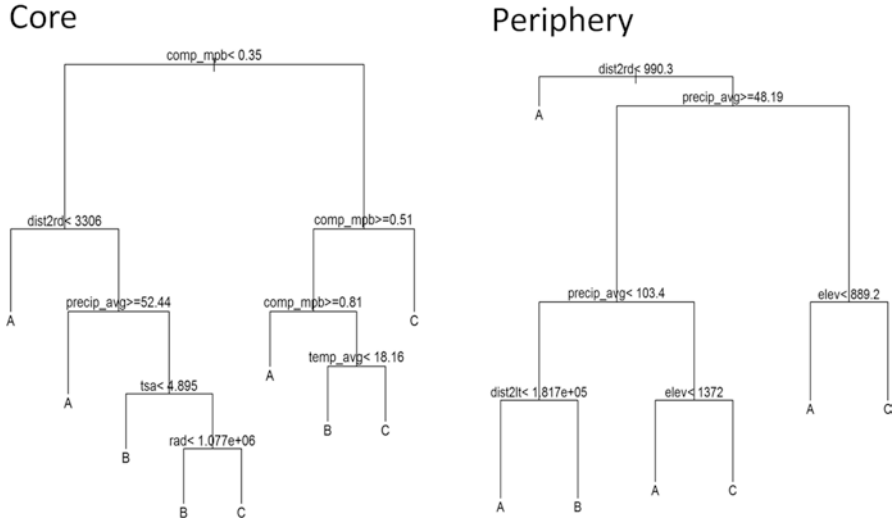


Fig. 3 CART results for Core and Periphery

A benefit of CART is shown by the nuanced predictions associated with the right branches. By allowing multiple splits on a single variable CART can represent complex relationships. When percent pine infested is $\geq 81.0\%$ the smallest fire class is predicted, which may be explained because very dead stands may limit the fuel load available for fire. However, the largest fires also occur when there is more fuel available, only 35.0% to 51.0% of the stand infested, or at high temperature (< 18.2 degrees).

Compared to the CORE, different variables in the Periphery were found to be important predictors of fire size, indicating that the fire processes in the Core and Periphery likely vary. Most notably, no mountain pine beetle infestation variables were important predictors of fire size in the Periphery. The lack of importance of beetle infestation indicates that in the Periphery from 1999 to 2009 the mountain pine beetle infestation processes were not sufficiently severe to be a dominant driver of fire process. Distance to road, average precipitation, and elevation were the only variables used to predict fire size. The smallest fires (32–200 ha), which are more plentiful, occurred near roads. The largest fires occurred far from roads (≥ 990.3 m) with average precipitation < 48.2 mm, and at elevations > 889.2 m, or else far from roads (≥ 990.3 m), with average precipitation between 48.2 and 103.4 mm and elevation > 1372 m.

BRTs

The BRT model had the best overall classification accuracy with 88.0% of fires in Core being correctly classified (Table 2) and 73.9% in the Periphery (Table 3). In the Core, the smallest (32–200 ha), midsize (200–1000 ha), and largest (> 1000 ha)

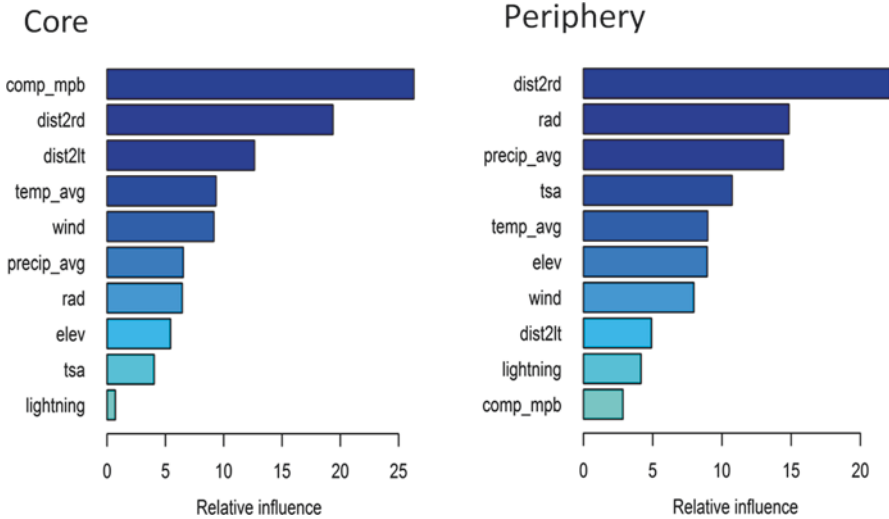


Fig. 4 Boosted regression tree results for Core and Periphery

fire classes had accurate predictions for 72.1%, 79.0%, and 79.6% of fires, respectively. In the Periphery the accuracies were even higher with 85.7%, 86.2%, and 100.0% of the smallest, midsize, and largest fires accurately predicted.

Examples of BRT outputs are shown in Fig. 3. The variable importance plot (Fig. 4) indicates the importance, in terms of rank and strength, of each variable for prediction. In the case of the Core, the percent pine infested was the most important predictor of fire size. Distance to road and populated place were the next strongest predictors, followed by average temperature and wind. In the Periphery, distance to road was the most important predictor. Weather variables were the next most important (solar radiation and average precipitation). Time since attack was the fourth most important variable, while percent pine infested was the least important predictor.

Directionality of associations between fire size and each predictor variable can be explored in the partial dependence plots which are shown for the Core in Fig. 5. Each class has a unique line and color: green, red, and black are the smallest (32–200 ha), midsize (200–1000 ha), and largest (>1000 ha) fire classes, respectively. Partial dependence plots can provide very useful information. For instance, where average temperatures were approximately 19 °C and higher, there was an increased probability of the largest fires. The highest probabilities of midsize fires occurred when temperatures ranged from 16 to 18 °C. Average temperature did not have a large influence on prediction of the smallest fire class. As indicated by the improved accuracy of prediction, there are statistical benefits to using BRT over CART. However, the CART regression trees allowed for intuitive exploration of variable relationships. Though similar information is available from partial dependence plots, the visualization was not as easy to interpret.

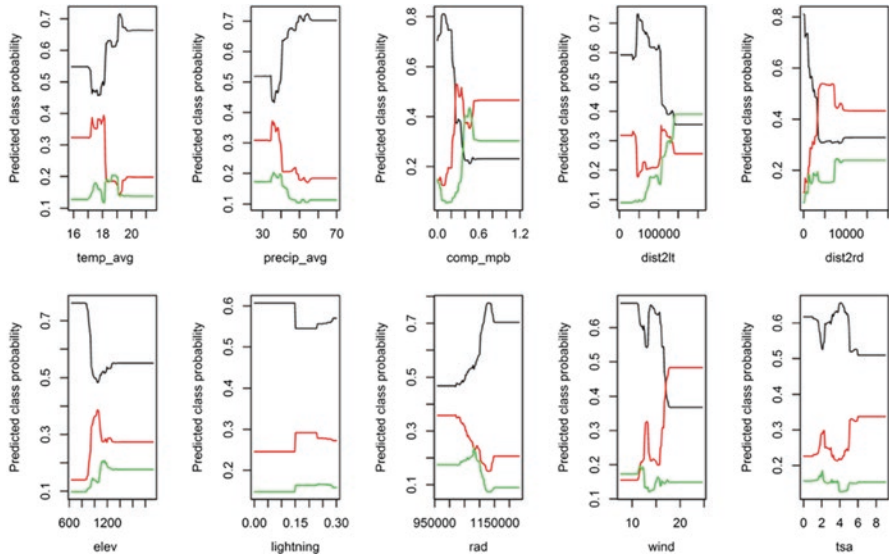


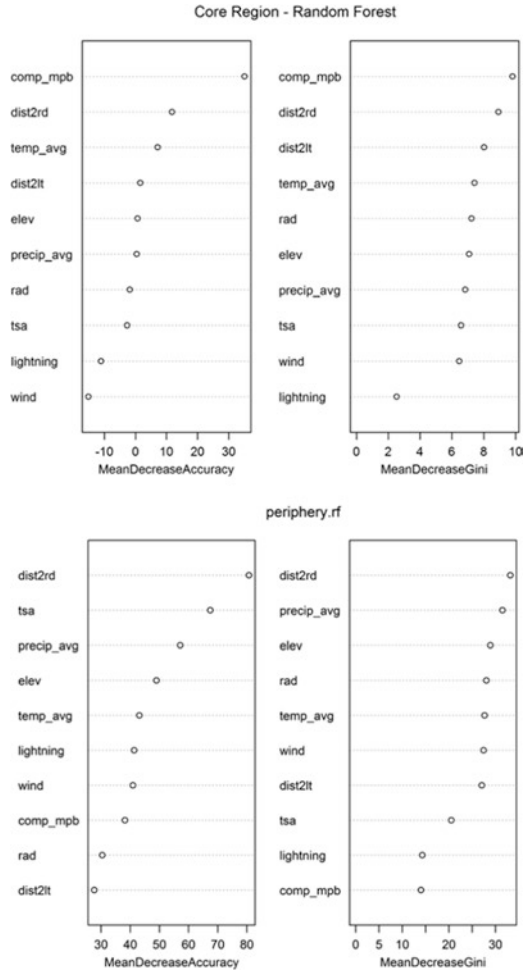
Fig. 5 Boosted regression tree partial dependence plots for Core. The *green*, *red*, and *black* lines are the smallest (32–200 ha), midsize (200–1000 ha), and largest (>1000 ha) fire classes, respectively

RF Models

The RF model had 49.6% classification accuracy in the Core (Table 2) and 64.4% accuracy in the Periphery (Table 3). Compared to both the CART and BRT models, the RF model performed poorly, particularly in the Core. The smallest fires (32–200 ha) were most accurately predicted, 63.0%, and the largest fires (>1000 ha) had the lowest number of correct classifications (18.8%), which may reflect sensitivity to sample size. In the Periphery, where the sample size was much larger, the largest fires (>1000 ha) were most accurately predicted (76.9%).

The RF results are shown in Fig. 6 and included both the mean square error or accuracy plot and the plot of misclassification error rate-based change in the Gini index. We have not included partial dependence plots, though they were available in similar format to the plots shown for the BRT. The accuracy plot was similar to the BRT variable importance plot. In the Core, the percent pine infested is the most important variable for accurately predicting fire size, followed by distance to road and average temperature. Typically, the misclassification plot will rank variables similarly to the variable importance plots. The key difference is how much the prediction is influenced by the removal of a variable. As with all the models, the results for the Periphery are quite different from the Core and indicate different spatial processes operating in each region. In terms of variable importance, distance to road was the most important variable. The second most important variable was time since attack. However, time since attack has a much lower impact on the Gini index, suggesting that it does not impact the misclassification rate. The next highest ranked variables in the accuracy plot were all weather related.

Fig. 6 Random forest results for Core and Peripheral regions showing differences in the predictive ability of each variable



Comparing Modeling Approaches

CART, BRT, and RF each have unique strengths and weaknesses. CART results are easy to interpret. Relationships between variables are intuitively observed making it possible to develop hypotheses about spatial pattern and process relationships. Information on the directionality of variable relationships is available from partial dependence plots, but they are not as intuitive and require a careful eye to examine and summarize. The difficulty with CART is that a different tree may be produced each time the model is run and as such the results may not be robust. The BRT and

RF have statistical benefits. While they are similar in that they fit many forests, the algorithms are sufficiently unique that the accuracy of each was quite different in our case study. The key difference is that BRT fits each tree to a different subsample of unique data. Given that each sample is unique, each new fit is to the residuals of the previous model. In contrast, RFs create a cross validation from the bootstrapped samples and therefore do not require separate testing data to determine how well the model fits. As a general guide, we recommend using CART for exploratory analysis and BRT for prediction. RF may be advantageous when samples are small.

Beyond the statistical fits, it is important to consider the different results obtained from each model that will impact our interpretation of pattern and process. However, it is encouraging that there was consistency in the results. In the Core, consistently, percent pine infested and distance to road were important predictors. Weather variables also emerged consistently, though depending on the model average temperature or average precipitation they may be more important. The variables that were lowest ranked on the variable importance plots were also not included in the pruned trees. In the Periphery, the distance to road variable was consistently the most important predictor. It is interesting that in both the CART and BRT, which performed better than the RF, weather variables were the next most important predictors. The RF includes time since attack as the second most important predictor, which was not included in the pruned CART though it is the fourth most important variable in the BRT. Given the complexities of relating spatial patterns and processes, scientists have advocated for confirmatory analysis whereby several analyses are carried out and the consistent trends become the strongest signal of pattern and process interactions. Ensemble modeling, for example, is a similar idea and allows the strengths of many models to be leveraged together (Grenouillet et al. 2011). In many instances, we strongly advocate for using CART in conjunction with either BRT or RF to provide a more complete understanding of interactions.

Interpreting Regression Tree Results within the Context of Spatial Pattern and Process

In pattern and process studies, often the most difficult part of the research is the interpretation. By including spatial data sets representing different spatial processes in a model, we invite the spatial patterns to indicate which processes are most important. While CART and related methods are very flexible in taking on large volumes of data, it remains of crucial importance to have a theoretical basis for each of the covariate data sets or in the presented study for stratifying the area in a Core and Periphery. One of the most notable results of our models was the difference in the variables that were important for predicting fire in the Core and Periphery. The Core has experienced extensive mountain pine beetle attack (Wulder et al. 2010) and the level of severity impacts the spatial pattern of large fires. From CART, we see evidence that the relationship is not linear. Rather, moderate levels of beetle

infestation in locations with high temperature were predicted to support the largest fires. As well, large fires occurred where beetle infestations were relatively low, but forests were far from roads and had dry conditions, and trees had been attacked by beetles about 5 years previously.

In the Periphery, our results indicated that the mountain pine beetle infestation did not impact the spatial pattern of fire size. However, distance to road and weather were both consistently important predictors of fire size.

The relationship between roads and fire patterns we found in our spatial data is consistent with existing knowledge. Gralewicz et al. (2011, 2012) documented that fire patterns in Canada were more related to people than any other variable. Forests closer to roads have more opportunities for a human-induced fire to start. On the other hand, fires may burn longer and larger in areas with little access for fire suppression and a lower economic stake in the standing wood. Given the extensive logging road networks in British Columbia, access to forested areas is greatly influenced by roads. Fuel moisture content (Hayes 1942) and temperature have also been well documented to directly impact surface fire intensity and crown fire initiation (van Wagner 1977; Turner and Romme 1994; Bessie and Johnson 1995; Turner et al. 1999; Hély et al. 2000; Simard et al. 2011). In general, our model supports weather as an important driver of spatial patterns of large fires (Parisien et al. 2006; Parisien and Moritz 2009).

We included topographical variables as generic spatial covariates in the spatial model, but hypothesized that their relation to fire severity is indirect through weather conditions or vegetation composition (Table 1). In the model results they have low relative importance compared to more the more direct measures included. The inclusion of generic variables provides the model with a way to account for spatial patterns unresolved by the selected theory-loaded data sets. Elevation is a commonly used generic variable, but even latitude and longitude or projected coordinates can function as generic covariates (Guisan and Zimmermann 2000; Michaud et al. 2014; Nijland et al. 2014). A high importance of generic spatial patterns in selected models is an indication that spatial processes are present that are not well represented by more specific covariates (Cressie and Chan 1989). In such cases model inputs and underlying hypotheses should be reevaluated. The low importance of generic descriptors in our models strengthens our confidence that all relevant processes are included in the spatial covariates.

In our own work we have found CART, BRT, and RF as useful tools for making linkages between spatial patterns and processes. Data mining methods are usually bound by the assumption of spatial stationarity. Spatial stationarity occurs when the mean of a spatial pattern, the expression of a process, is similar in all parts of the study area (Bailey and Gatrell 1995, pp. 33–35). When data sets are small, spatial stationarity is possible, but as study extents increase this assumption often becomes invalid. In our case study we accounted for spatial non-stationarity by using two study regions, each with unique mountain pine beetle infestation processes and patterns. However, other methods such as geographically weighted regression are gaining momentum due to inherent ability to deal with variation in process interactions that are expected at landscape and regional scales (Brunsdon et al. 1996; Wang et al. 2005).

Future Direction

Geographers have developed a host of statistical techniques for quantifying spatial pattern (Nelson and Boots 2005; Robertson et al. 2007) and ecologists are adept at developing and applying methods that explore how covariate data predict patterns. There are many existing useful statistical methods for analyzing interaction between pattern and process including the tree-based models we focus on in this chapter. Inherent to spatial pattern analysis is the assumption that patterns observed in data represent a process. Historically, we have been limited to a few snapshots of spatial patterns in time. With the growing availability of satellite remotely sensed data, the temporal resolution and extent of data have changed. Archives of Landsat data are freely available from the mid-1980s at a temporal resolution of 16 days and a spatial resolution of 30 m, and back to 1972 with a slightly more coarse spatial resolution (resampled into products at 60 m). The United States Geological Survey Landsat archive has over 500,000 images of Canada (White and Wulder 2014). Landsat data are of special interest due to the capture of relatively large areas over a single imaging footprint at a level of detail that is informative of anthropogenic activities (Wulder et al. 2012). Moderate-resolution imaging spectroradiometer (MODIS) is also freely available but collects data with a more coarse spatial resolution, and images the entire earth every 1–2 days. Higher spatial resolution images are available from a number of commercial vendors, including RapidEye that can be captured daily. Changing temporal resolution of imagery requires that we consider the temporal resolution maps of a spatial pattern in the measurement of a process. Remote sensing science is moving away from temporally static representations of space to more dynamic representations of pattern (Verbesselt et al. 2010; Gómez et al. 2011). Multi-temporal spatial pattern data sets are better representing spatial processes, and in some cases, the temporal resolution is allowing broad-scale measurement of dynamic spatial process. As has been a constant issue in spatial sciences, development of methods to harness the content of new data sets has not kept pace with data acquisition technology. As data sets continue to grow, it is our view that data mining approaches, such as regression trees and related methods, will become more heavily utilized.

In addition to data mining approaches, it is common to model spatial processes and compare the patterns generated by models to observed data (Nelson and Boots 2005). Research is required to explore the benefits of integrating data-driven and modeling-based approaches for studying pattern and process interactions. Bayesian statistics are gaining momentum and offer a unique mechanism for linking spatial data and process modeling perspectives (Ghazoul and McAllister 2003; van Oijen et al. 2005). Bayesian statistics represent variables using distributions, allowing patterns to be represented by a range of values (Gelman et al. 2009). Given that patterns are only one possible realization of a process, representing patterns as distributions of values is more realistic. From a practical standpoint, Bayesian methods deal well with uncertainty as the distribution can be thought of as a mechanism for providing a confidence interval around observed values (van Oijen and Thomson 2010). Given

uncertainty in both spatial data and our understanding of forestry spatial processes, Bayesian approaches offer support for multiple issues. Growing spatial data archives provide data for building informative priors. It is common in Bayesian statistics to use uninformed priors based on uniform distributions, but archives, such as that available with Landsat (Wulder et al. 2012), are a mechanism for informing priors and analysis with multi-temporal representations of spatial patterns.

References

- Amman G (1977) Role of the mountain pine beetle in lodgepole pine ecosystems: impact on succession. In: Mattson WJ (ed) *The role of arthropods in forest ecosystems*. New York, Springer-Verlag, pp 3–18
- Anderson NM, Germain H, Bevilacqua E (2011) Geographic information system-based spatial analysis of sawmill wood procurement. *J For* 109:34–42
- Aukema BH, Carroll AL, Zhu J, Raffa KF, Sickley TA, Taylor SW (2006) Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak. *Ecography* 29:427–441
- Axelson JN, Alfaro RI, Hawkes BC (2010) Changes in stand structure in uneven-aged lodgepole pine stands impacted by mountain pine beetle epidemics and fires in central British Columbia. *For Chron* 86:87–99
- Bailey TC, Gatrell AC (1995) *Interactive spatial data analysis*, vol 413. Longman Scientific & Technical, Essex
- Berk RA (2008) *Statistical learning from a regression perspective*, vol 14. Springer, New York
- Bessie A, Johnson E (1995) The relative importance of fuels and weather on fire behaviour in subalpine forests. *Ecology* 76:747–762
- Bigler C, Kulawkowski D, Veblen TT (2005) Multiple disturbance interactions and drought influence fire severity in rocky mountain subalpine forests. *Ecology* 86:3018–3029
- Bone C, Wulder MA, White J, Robertson C, Nelson TA (2013) A GIS-based risk rating of forest insect outbreaks using aerial overview surveys and the local Moran's I statistic. *Appl Geogr* 40:161–170
- Bourbonnais ML, Nelson TA, Cattet MRL, Darimont CT, Stenhouse GB (2013b) Spatial analysis of factors influencing long-term stress in the grizzly bear (*Ursus arctos*) population of Alberta, Canada. *PLoS One* 8(12):e83768. doi:10.1371/journal.pone.0083768
- Bourbonnais ML, Nelson TA, Wulder MA (2013a) Geographic analysis of the impacts of mountain pine beetle infestation on forest fire ignition. *Can Geogr/Le Géographe Canadien* 58(2):188–202. doi:10.1111/j.1541-0064.2013.12057.x
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. Retrieved from <http://www.springerlink.com/index/UOP06167N6173512.pdf>
- Breiman L, Friedman J, Olsen R, Stone C (1984) *Classification and regression trees*. Chapman and Hall/CRC, Boca Raton, p 368
- Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal* 28(4):281–298
- Canadian Forest Service (2010) National Fire Database – agency fire data. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre (Edmonton, AB). <http://cwfis.cfs.nrcan.gc.ca/ha/nfdb>. Accessed 28 Apr 2014
- Chuvieco E, Salas J (1996) Mapping the spatial distribution of forest fire danger using GIS. *Int J Geogr Inf Syst* 10:333–345. doi:10.1080/02693799608902082
- Coops NC, Gillanders SN, Wulder MA, Gergel SE, Nelson T, Goodwin NR (2010) Assessing changes in forest fragmentation following infestation using time series Landsat imagery. *For Ecol Manag* 259:2355–2365

- Cressie NA, Chan NH (1989) Spatial modeling of regional variables. *J Am Stat Assoc* 84:393–401. doi:[10.2307/2289922](https://doi.org/10.2307/2289922)
- Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88(11):2783–2792
- Dark SJ, Bram D (2007) The modifiable areal unit problem (MAUP) in physical geography. *Prog Phys Geogr* 31:471–479. doi:[10.1177/0309133307083294](https://doi.org/10.1177/0309133307083294)
- De'ath G (2007) Boosted trees for ecological modeling and prediction. *Ecology* 88(1):243–251. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17489472>
- De'ath G, Fabricius K (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81(11):3178–3192. Retrieved from [http://www.esajournals.org/doi/pdf/10.1890/0012-9658\(2000\)081\[3178:CARTAP\]2.0.CO;2](http://www.esajournals.org/doi/pdf/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2)
- Demarchi, D. A. (2011) The British Columbia Ecoregion classification, third edition. Ecosystem information section, Ministry of Environment. (Victoria, BC)
- Díaz-Avalos C, Peterson DL, Alvarado E, Ferguson SA, Besag JE (2001) Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Can J For Res* 31:1579–1593
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A et al (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129–151
- Elith J, Leathwick J, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813. doi:[10.1111/j.1365-2656.2008.01390.x](https://doi.org/10.1111/j.1365-2656.2008.01390.x)
- Flannigan MD, Harrington JB (1989) A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953–80). *J Appl Meteorol* 27:441–452
- Flannigan MD, Logan KA, Amiro BD, Skinner WR, Stocks BJ (2005) Future area burned in Canada. *Clim Chang* 72:1–16
- Fortin M-J, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge
- Fortin MJ, Boots B, Csillag F, Rempel TK (2003) On the role of spatial stochastic models in understanding landscape indices in ecology. *Oikos* 102(1):203–212
- Franklin J (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog Phys Geogr* 19:474–499
- Friedman J (2001) Greedy function approximation: a gradient boosting machine, 1999. *Ann Statist* 29(5):1189–1232. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Greedy+Function+Approximation:+A+Gradient+Boosting+Machine#3>
- Friedman JH, Meulman JJ (2003) Multiple additive regression trees with application in epidemiology. *Stat Med* 22(9):1365–1381. doi:[10.1002/sim.1501](https://doi.org/10.1002/sim.1501)
- Gelman A, Carlin JB, Stern HS, Rubin DB (2009) Bayesian data analysis, 2nd edn. Chapman and Hall/CRC, Boca Raton, FL
- Getis A, Boots BN (1978) Models of spatial processes: an approach to the study of point, line, and area patterns. Cambridge University Press, Cambridge
- Ghazoul J, McAllister M (2003) Communicating complexity and uncertainty in decision making contexts: Bayesian approaches to forest research. *Int For Rev* 5(1):9–19
- Gómez C, White JC, Wulder MA (2011) Characterizing the state and processes of change in a dynamic forest environment using hierarchical spatio-temporal segmentation. *Remote Sens Environ* 115(7):1665–1679
- Gralewicz NJ, Nelson TA, Wulder MA (2011) Spatial and temporal patterns of wildfire ignitions in Canada from 1980 to 2006. *Int J Wildland Fire* 21:230–242
- Gralewicz NJ, Nelson TA, Wulder MA (2012) Factors influencing national scale wildfire susceptibility in Canada. *For Ecol Manag* 265:20–29
- Grenouillet G, Buisson L, Casajus N, Lek S (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34(1):9–17
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135:147–186. doi:[10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Haining R (1990) Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge

- Hall JP, Moody B (1994) Forest depletions caused by insects and diseases in Canada 1982–1987. Natural Resources Canada, Canadian Forest Service, Ottawa, ON
- Hamann A, Wang TL (2005) Models of climatic normals for genecology and climate change studies in British Columbia. *Agric For Meteorol* 128:211–221
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, p 745
- Hawkes B, Taylor SW, Stockdale C, Shore TL, Alfaro RI, Campbell R, Vera P (2004) Impact of mountain pine beetle on stand dynamics in BC. In: Shore TL, Brooks JE, Stone JE (eds) Mountain pine beetle symposium: challenges and solutions. October 30–31, Kelowna, British Columbia. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-399, Victoria, BC
- Hawkins BA (2012) Eight (and a half) deadly sins of spatial analysis. *J Biogeogr* 39(1):1–9. doi:[10.1111/j.1365-2699.2011.02637.x](https://doi.org/10.1111/j.1365-2699.2011.02637.x)
- Hayes GL (1942) Differences in fire danger with altitude, aspect, and time of day. *J For* 40:318–323
- Hély C, Flannigan M, Bergeron Y, McRae D (2000) Role of vegetation and weather on fire behavior in the Canadian mixedwood boreal forest using two fire behavior prediction systems. *Can J For Res* 31:430–441
- Holmes K, Nelson T, Coops N, Wulder M (2013) Biodiversity indicators show climate change will alter vegetation in parks and protected areas. *Diversity* 5(2):352–373. doi:[10.3390/d5020352](https://doi.org/10.3390/d5020352)
- Jacquez GM (2000) Spatial analysis in epidemiology: nascent science or a failure of GIS? *J Geogr Syst* 2:91–97
- Jenkins MJ, Hebertson E, Page W, Jorgensen A (2008) Bark beetles, fuels, fires and implications for forest management in the intermountain west. *For Ecol Manag* 254:16–34
- Jenkins MJ, Page WG, Hebertson EG, Alexander ME (2012) Fuels and fire behavior dynamics in bark beetle-attacked forests in western North America and implications for fire management. *For Ecol Manag* 275:23–34
- Jolly WM, Parsons RA, Hadlow AM, Cohn GM, McAllister SS, Popp JB, Hubbard RM, Negrón JF (2012) Relationships between moisture, chemistry, and ignition of *Pinus contorta* needles during the early stages of mountain pine beetle attack. *For Ecol Manag* 269:52–59
- Klutsch JG, Battaglia MA, West DR, Costello SL, Negrón JF (2011) Evaluating potential fire behavior in lodgepole pine-dominated forests after a mountain pine beetle epidemic in north-central Colorado. *West J Appl For* 26:101–109
- Kumar L, Skidmore AK, Knowles E (1997) Modelling topographic variation in solar radiation in a GIS environment. *Int J Geogr Inf Sci* 11:475–497
- Langford WT, Gergel SE, Dietterich TG, Cohen W (2006) Map misclassification can cause large errors in landscape pattern indices: examples from habitat fragmentation. *Ecosystems* 9(3):474–488
- Leathwick JR, Elith J, Francis MP, Hastie T, Taylor P (2006) Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Ecol Prog Ser* 321:267–281.
- Levin SA (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* 73(6):1943–1967
- Lynch HJ, Renkin RA, Crabtree RL, Moorcroft PR (2006) The influence of previous mountain pine beetle (*Dendroctonus ponderosae*) activity on the 1988 Yellowstone fires. *Ecosystems* 9:1318–1327
- Masek, J. G., Cohen, W. B., Leckie, D., Wulder, M. A., Vargas, R., de Jong, B., Healy, S., et al. (2011) Recent rates of forest harvest and conversion in North America. *J Geophys Res* 116, G00K03, doi: [10.1029/2010JG001471](https://doi.org/10.1029/2010JG001471)
- Michaud JS, Coops NC, Andrew ME, Wulder MA, Brown GS, Rickbeil GJM (2014) Estimating moose (*Alces alces*) occurrence and abundance from remotely derived environmental indicators. *Remote Sens Environ* 152:190–201. doi:[10.1016/j.rse.2014.06.005](https://doi.org/10.1016/j.rse.2014.06.005)
- Miller C, Urban DL (2000) Connectivity of forest fuels and surface fire regimes. *Landsc Ecol* 15:145–154

- Murthy S (1998) Automatic construction of decision trees from data: a multi-disciplinary survey. *Data Min Knowl Disc* 2:345–389. Retrieved from <http://link.springer.com/article/10.1023/A:1009744630224>
- Nelson TA (2012) Trends in spatial analysis. *Prof Geogr* 64(1):1–12
- Nelson TA, Boots B (2008) Detecting spatially explicit hot spots in landscape-scale ecology. *Ecography* 31(5):556–566
- Nelson T, Boots B (2005) Identifying insect infestation hot spots: an approach using conditional spatial randomization. *J Geogr Syst* 7(3–4):291–311
- Nelson T, Boots B, Wulder MA, Shore T, Safranyik L, Ebata T (2006) Rating the susceptibility of forests to mountain pine beetle infestations: the impact of data. *Can J Forest Res* 36(11):2815–2825.
- Negrón JF, Bentz BJ, Fettig CJ, Gillette N, Hansen EM, Hayes JL, Kelsey RG et al (2008) US Forest Service bark beetle research in the western United States: looking toward the future. *J For* 106:325–331
- Nijland W, Addink EA, De Jong SM, Van der Meer FD (2009) Optimizing spatial image support for quantitative mapping of natural vegetation. *Remote Sens Environ* 113:771–780. doi:10.1016/j.rse.2008.12.002
- Nijland W, Nielsen SE, Coops NC, Wulder M a, Stenhouse GB (2014) Fine-spatial scale predictions of understory species using climate- and LiDAR-derived terrain and canopy metrics. *J Appl Remote Sens* 8:083572. doi:10.1117/1.JRS.8.083572
- Page W, Jenkins MJ (2007) Predicted fire behaviour in selected mountain pine beetle-infested lodgepole pine. *For Sci* 53:662–674
- Parisien M-A, Moritz MA (2009) Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecol Monogr* 79:127–154
- Parisien M-A, Peters VS, Wang Y, Little JM, Bosch EM, Stocks BJ (2006) Spatial patterns of forest fires in Canada, 1980–1999. *Int J Wildland Fire* 15:361–374
- Powers JS, Sollins P, Harmon ME, Jones JA (1999) Plant-pest interactions in time and space: Douglas-fir bark beetle outbreak as a case study. *Landsc Ecol* 14:105–120
- Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9(2):181–199. doi:10.1007/s10021-005-0054-1
- Reid RW (1961) Moisture changes in lodgepole pine before and after attack by the mountain pine beetle. *For Chron* 37:368–375
- Robertson C, Nelson TA, Boots B (2007) Mountain pine beetle dispersal: the spatial-temporal interactions of infestation. *For Sci* 53(3):395–405
- Robertson C, Wulder MA, Nelson TA, White JC (2008) Risk rating for mountain pine beetle infestation of lodgepole pine forests over large areas with ordinal regression modelling. *For Ecol Manag* 256:900–912
- Robertson C, Farmer CJQ, Nelson TA, Mackenzie IK, Wulder MA, White JC (2009a) Determination of the compositional change (1999–2006) in the pine forests of British Columbia due to mountain pine beetle infestation. *Environ Monit Assess* 158:593–608
- Robertson C, Nelson TA, Jelinski DE, Wulder MA, Boots B (2009b) Spatial-temporal analysis of species range expansion: the case of the mountain pine beetle, *Dendroctonus ponderosae*. *J Biogeogr* 36:1446–1458
- Shekhar S, Zhang P, Huang Y, Vatsavai RR (2003) Trends in spatial data mining. In: Kargupta H, Joshi A, Sivakumar K, Yesha Y (eds) *Data mining: next generation challenges and future directions*. AAAI/MIT Press, Cambridge, MA, pp 357–380
- Shore TL, Safranyik L, Hawkes BC, Taylor SW (2006) Effects of the mountain pine beetle on lodgepole pine stand structure and dynamics. In: Safranyik L, Wilson B (eds) *The Mountain pine beetle: a synthesis of biology, management, and impacts on lodgepole pine*. Natural Resources Canada, Canadian Forest Service, Pacific forestry Centre, Victoria, BC, pp 95–114
- Simard M, Romme WH, Griffin JM, Turner MG (2011) Do mountain pine beetle outbreaks change the probability of active crown fire in lodgepole pine forests? *Ecol Monogr* 81:3–24

- Sokal RR, Oden NL, Thomson BA (1998) Local spatial autocorrelation in biological variables. *Biol J Linn Soc* 65:41–62
- Statistics Canada (2008) .Road Network File, Reference Guide 92-500-GWE Ottawa, On. Available at: www.statcan.ca/bsolc/english/bsolc?catno=92-500-x
- Stocks BJ, Mason JA, Todd JB, Bosch EM, Wotton BM, Amiro BD, Flannigan MD, Hirsch KG, Logan KA, Martell DL, Skinner WR (2002) Large forest fires in Canada, 1959–1997. *J Geophys Res Atmos* 108:5.1–5.12
- Turner MG (1989) Landscape ecology: the effect of pattern on process. *Annu Rev Ecol Syst* 20:171–197
- Turner MG, Romme WH (1994) Landscape dynamics in crown fire ecosystems. *Landsc Ecol* 9:59–77
- Turner MG, Romme WH, Gardner RH (1999) Prefire heterogeneity, fire severity, and early post-fire plant reestablishment in subalpine forests of Yellowstone National Park, Wyoming. *Int J Wildland Fire* 9:21–36
- van Oijen M, Thomson A (2010) Toward Bayesian uncertainty quantification for forestry models used in the United Kingdom greenhouse gas inventory for land use, land use change, and forestry. *Clim Chang* 103(1-2):55–67
- van Oijen M, Rougier J, Smith R (2005) Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol* 25(7):915–927
- van Wagner CE (1977) Conditions for the start and spread of crown fire. *Can J For Res* 7:23–33
- Verbesselt J, Hyndman R, Newnham G, Culvenor D (2010) Detecting trend and seasonal changes in satellite image time series. *Remote Sens Environ* 114(1):106–115
- Volney WJA, Fleming RA (2000) Climate change and impacts of boreal forest insects. *Agric Ecosyst Environ* 82:283–294
- Walton, A. (2010). Provincial-level projection of the current mountain pine beetle outbreak: update of the infestation projection based on the 2009 provincial aerial overview of forest health and the BCMPB model (year 7). Research branch, BC Forest Service. (Victoria, BC)
- Wang T, Hamann A, Spittlehouse DL, Murdock TQ (2011) ClimateWNA – high-resolution spatial climate data for western North America. *Am Meteorol Soc* 51:16–29. doi:[10.1175/JAMC-D-11-043.1](https://doi.org/10.1175/JAMC-D-11-043.1)
- Wang Q, Ni J, Tenhunen J (2005) Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Glob Ecol Biogeogr* 14(4):379–393
- Waring RH, Coops NC, Mathys A, Hilker T, Latta G (2014) Process-based modeling to assess the effects of recent climatic variation on site productivity and forest function across western North America. *Forests* 5:518–534. doi:[10.3390/f5030518](https://doi.org/10.3390/f5030518)
- White JC, Wulder MA (2014) The Landsat observation record of Canada: 1972–2012. *Can J Remote Sens* 39(06):1–13
- Wulder MA, Ortlepp SM, White JC, Nelson TA, Coops NC (2010) A provincial and regional assessment of the mountain pine beetle epidemic in British Columbia: 1999–2008. *J Environ Inform* 15:1–13
- Wulder MA, White JC, Coops NC (2011) Fragmentation regimes of Canada’s forests. *Can Geogr* 55:288–230
- Wulder MA, Masek JG, Cohen WB, Loveland TR, Woodcock CE (2012) Opening the archive: how free data has enabled the science and monitoring promise of Landsat. *Remote Sens Environ* 122:2–10