Modelling Five-Needle Pine across the Crown of the Continent Ecosystem

Suggested Citation: Blackadder, S., Sexton, E., & McDermid, G. (2016). Modelling Five-Needle Pine across the Crown of the Continent Ecosystem. Report Prepared for Crown Managers Partnership.

Funding for this work was provided by the Great Northern Landscape Conservation Cooperative, in partnership the Crown Managers Partnership.

Contents

Summary

The Crown Managers Partnership, in collaboration with partners across the Crown of the Continent Ecosystem (CCE), developed a Crown-wide species distribution model of whitebark pine, which describes the relative probability of occupancy of this species across the landscape. The model was built with a Random Forest algorithm using 48 environmental variables and more than 20 datasets containing whitebark pine location records from collaborators across the CCE. Overall, the model performed well. Based on the data used to build the model, the model had an area under the Receiver Operating Characteristic (ROC) curve of 0.985, which means that 98.5% of the time the model can correctly differentiate between sites that are occupied by whitebark pine and those that are not, based on available data. It is important to note, however, that some portions of the CCE had no suitable data available, and data that was employed contains inconsistencies related to sampling strategy and effort, purpose of collection, data format, and time period inspected. This paper summarizes the modelling process and results, and discusses the challenges that needed to be overcome when working across multiple jurisdictions.

Background

The Crown of the Continent Ecosystem (CCE) is one of the most ecologically diverse and jurisdictionally fragmented landscapes in North America. Spanning 72,000km^2, the CCE covers land public and private lands in Alberta, British Columbia, and Montana. The Crown Managers Partnership (CMP) is a voluntary management partnership amongst federal, state, provincial, tribe and First Nations managers from the region. The CMP seeks to overcome transboundary ecological challenges through multi-jurisdictional partnerships. The Transboundary Conservation Initiative (TCI) is the CMP's flagship program and provides a framework for addressing shared conservation priorities among stakeholders in the Crown of the Continent.

Crown of the Continent Ecosystem

Whitebark pine was identified by Crown of the Continent stakeholders as a top conservation priority because of its ecological and cultural importance, conservation status across jurisdictions, and the serious threats it faces due to climate change. Whitebark pine is a keystone and foundation species. It facilitates tree island development by moderating harsh alpine conditions, stabilizing soil, and retaining snowpack, and whitebark pine seeds provide critical nutrition for many species including grizzly bears (Resler & Tomback, 2008; Smith et al., 2008a; D. F. Tomback & Achuff, 2010; Diana F Tomback, Chipman, Resler, Smith-McKenna, & Smith, 2014)

Unfortunately, whitebark pine face many threats which are causing critical levels of mortality across much of its range. The species was listed in 2012 by the Canadian Federal Government as endangered, and are under consideration by the United States Federal government for listing in that country. The threats to whitebark pine include: mountain pine beetle (*Dendroctonus ponderosae*, MPB); white pine blister rust (*Cronartium ribicola*; WPBR), an invasive fungal pathogen introduced from Europe; fire suppression leading to other more shade tolerant trees out competing WBP; and climate warming trends (Bockino, 2012; Chang, Hansen, & Piekielek, 2014; Keane, Morgan, & Menakis, 1994; Smith et al., 2008b; D. F. Tomback & Achuff, 2010; Diana F Tomback et al., 2014).

The epicentre of whitebark pine decline is the Crown of the Continent (personal communications Tomback 2016, CMP Annual Forum presentation), and stakeholders identified the following transboundary priorities: 1) promoting conservation efforts aimed at the coordination of information, 2) creation of restoration plans and results, and 3) compilation of data. These efforts are difficult to achieve across such a large multijurisdictional landscape. The Crown Manager Partnership applied for and received funding from the Great Northern Landscape Conservation Cooperative for the 2015-2016 year to pursue several projects centred on whitebark pine. One of these projects was to map the distribution of whitebark pine across the CCE using existing data from partners. This document summarizes those efforts.

Purpose

The purpose of this project was to develop a spatially explicit occupancy model for whitebark pine across the full extent of the CCE. The occupancy model communicates the relative probability of occurrence, and was created using seamless ecological variables covering the CCE, and known whitebark pine location data provided by partners in Alberta, British Columbia, and Montana. Our work provides a landscape-scale view of occupancy in the CCE and compliments other fine-scale modeling activities in the region.

Collaborators

This project was only possible because partners from across the CCE shared their tree location data. We'd like to sincerely thank everyone who took the time to gather and submit their data for this project. We recognize that the data shared with us is the result of significant time, effort, and financial resources. As such, we'd like to thank the following partners and agencies:

- US Forest Service
- National Park Service
- US Bureau of Land Management
- Whitebark Pine Ecosystem Foundation
- Alberta Environment and Parks
- BC Ministry of Forests
- Parks Canada

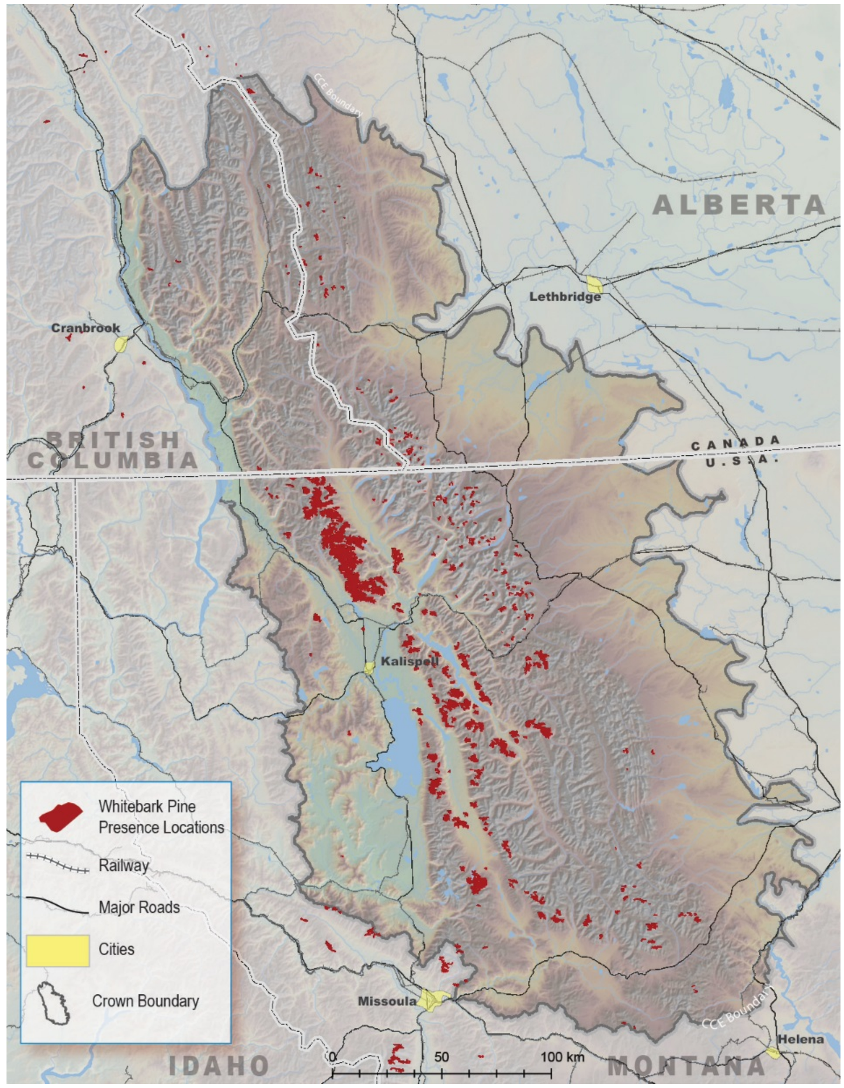
Funding was generously granted by the Great Northern Landscape Conservation Cooperative.

Methods

Object-Oriented Modelling

The input data for the model (both independent and dependent variables) come from a wide range of sources, using a variety of methods, and in a number of resolutions. This poses challenges for modelling because we need to be able to identify the environmental conditions (independent variables) that exist at each location of interest on the landscape. In order to address this issue, we opted to use object-oriented analysis.

What object-oriented analysis does is group together similar explanatory variables into larger units of analysis called objects. Each object was then labelled “presence” or “available” based on the whitebark pine location data, and each explanatory variable is summarized over the entire object. This increases the likelihood that we accurately represent the environmental conditions at each location.



Whitebark pine presence locations contributed by collaborators across the Crown of the Continent. While some areas look like they have a lot of WBP it is important to note that this may be due to uneven sampling strategy and data type rather than the actual abundance of WBP

Data Inputs

This project used two types of data input: known whitebark pine locations, which are actual ground-truthed locations where whitebark trees have been confirmed; and environmental covariates derived from GIS and remote sensing.

Dependent Variables: The presence and absence of trees

Tree location data was graciously provided by a wide variety of partners across the CCE. We requested spatially explicit presence and absence data, however very little absence data was received so it was excluded. Data was received in a variety of formats: databases, GIS shapefiles, excel spreadsheets, and Google Earth KMZs. Furthermore, the type of data differed widely and included: point data, transect data, plot data, estimated tree locations, and data that had been fuzzed to prevent the actual location of the trees from being known.

In order to complete the modelling process, the data needed to cleaned and formatted for consistency across the Crown. This was a challenge in some cases as data was received without information on what projection the location data was collected in, what units data were collected in, what the plot size was, or how much tree locations were fuzzed. We used the following criteria to format the data and decide whether to keep or discard different datasets:

- Any data point where we were not reasonably sure of the location (unknown spatial references, estimated locations, fuzzed locations) was discarded
- All plot data was assigned "presence" over its entire area
- All transect data was assigned "presence" over its entire area

Once this was done, there were 2082 objects with whitebark pine presence in the CCE.

Species distribution models also require absence locations; in order for the algorithm to determine where whitebark pine is, it also has to know where whitebark pine isn't. True absences are rarely available for any species modelling and this project was no different. In order to fill this data gap, various methods have been developed to create pseudo-absences (also sometimes referred to as background or available data). Pseudo-absences are locations in the landscape which do not have a recorded presence, but we cannot be sure are true absences. There are many different ways to select pseudo-absences and all of them have strengths and weaknesses.   For our project, pseudo-absences were selected by randomly sampling 10% of objects greater than 500m from presence locations. This resulted in 23,743 pseudo-absences across the entire crown.

Independent Variables: Environmental Variables

All together 48 variables were used to model whitebark pine presence and absence. They were selected based on previous whitebark pine modelling work. The variables fall into 4 broad categories:

1. Climate Data:
    a. Bioclim 30 year normal from 1981 – 2010
    b. Resolution: $1000m^2$
2. Digital Elevation Model (DEM) derivatives:
    a. DEM developed by the CMP, based on SRTM and ASTER DEMs
    b. Resolution: $30m^2$
3. Spectral remote sensing variables:
    a. MODIS normalized difference vegetation index (NDVI) from the growing season of 2010
    b. Resolution: $250m^2$
4. Position variables
    a. Measures of latitude, longitude, eastness and northness

Model Development

The occupancy model was built using the Random Forest (RF) algorithm. RF is a computer learning algorithm that uses multiple independent classification trees. Classification trees are built on a random subset of the explanatory variables and a random sample of the dependent variable. Thousands of trees were built, and then aggregated through a majority voting process. RF is considered a very robust modelling technique for this type of application. The random subsetting and sampling that this method uses reduces overfitting and collinearity issues that other model types can suffer from (Chang et al., 2014).

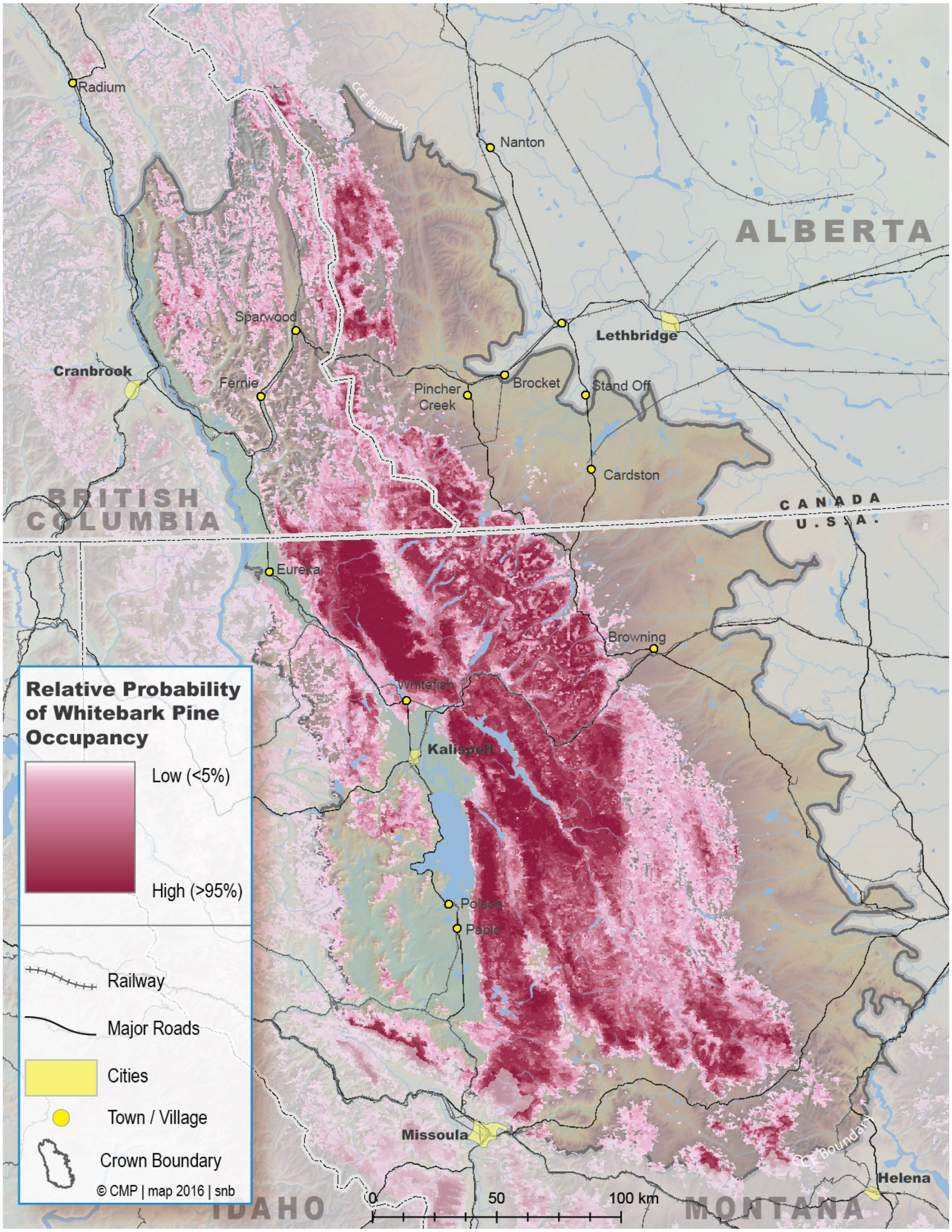
Model Testing and Results

Overall, the model performed well. The random forests algorithm uses "Out of Bag" testing to test the model: individual trees are grown using a random subset of the input data (both the explanatory and response variables) – this is called the "in bag" data, the trees are then tested on the remaining "out of bag" data to see how well they perform. In the case of the Crown-wide whitebark pine model, the area under the Receiver Operating Characteristic (ROC) curve is 0.985; this means that 98.5% of the time the model can correctly differentiate between sites that are occupied by whitebark pine and those that are not (Pearce & Ferrier, 2000).

The top 10 most important predictor variables in the model were:

- Mean annual precipitation
- Precipitation in the winter (dec – feb)
- Precipitation as snow
- Difference between mean temperature of coldest month and warmest month (a measure of continentality)
- Annual heat moisture index
- Summer heat moisture index
- Hargreave's climatic moisture index
- Longitude
- Mean summer precipitation
- Degree-days above 5°C

Although these variables are the most important in the model, this does not mean that they are the most important biologically to whitebark pine. This model is suitable for prediction of whitebark pine rather than inference about their biological needs and requirements.



Relative probability of occupancy of whitebark pine in the Crown of the Continent Ecosystem

Sources of Error

Input whitebark pine location data

Any species distribution model is only as good as the input data it is built from. This model was built on more than 20 different data sets. The datasets lacked consistency in:

- sampling strategy and effort
- purpose of collection
- data format
- time period

A significant amount of data manipulation was needed to get the data into a similar format. Any step of the data preparation could lead to error that persists throughout the modelling process.

Data availability

Some areas of the CCE had no presence locations from which to build the model. While it is possible for the model to interpolate in the areas without information, the true impact of the missing data on the model output is unknown and should be tested.

Very little absence data was received. Species distribution models require both presence and absence locations to build models. Without knowing true absence locations, pseudo-absences were used.

Pseudo-Absence Selection:

There are many factors to consider when selecting pseudo-absences, namely how many to select, where to select them from. We opted to use a random selection of points more than 500m from recorded presence points. We selected 10% of the objects to ensure we had sufficient samples across the ecologically diverse crown. It is highly likely that some objects marked as absence actually have whitebark pine in them. A growing body of literature is examining the impact of pseudo-absence selection methods and number of pseudo-absences, however no consensus has been reached. The impact of using pseudo-absences as we did is unknown, and further testing is underway to look at the impact of pseudo-absence collection

Object delineation:

This modelling project was done using object-oriented analysis. The objects delineated by the computer algorithm are meant to assist in overcoming problems associated with the wide variety of input data types, as described above. However, it is possible that the object boundaries are not representative of the species-specific requirements of whitebark pine. Species modelling should always be done at a resolution that is meaningful to the species, however no testing has been done to determine whether the defined objects and the scale of the objects are the most appropriate for whitebark pine.

Limitations of the Model

This model is a landscape-scale look at the probability of occupancy in the CCE, it is meant to compliment finer-level analysis going on throughout the region. This model is best suited for looking at those broad-scale patterns and to direct more fine-scale investigations. For example, a model at this scale would not be suitable to planning on-the-ground restoration activities, but could help to define areas of interest where more detailed work needs to be done.

Work in Progress and Future Work

Current work is being conducted to understand the impact of both pseudo-absence selection strategy as well as the impact of data type and sampling strategy on the model output. Understanding these factors will assist in building models that are better at predicting high elevation 5NP and will also help to inform future sampling design and data collection methods.

References:

Bockino, N. K. (2012). Interactions of White Pine Blister Rust, Host Species, and Mountain Pine Beetle in Whitebark Pine Ecosystems in the Greater Yellowstone. *Natural Areas Journal*, *32*(1), 31–40.

Chang, T., Hansen, A. J., & Piekielek, N. (2014). Patterns and Variability of Projected Bioclimatic Habitat for Pinus albicaulis in the Greater Yellowstone Area. *PLoS ONE*, *9*(11), e111669. http://doi.org/10.1371/journal.pone.0111669

Keane, R. E., Morgan, P., & Menakis, J. P. (1994). Landscape Assessment of the Decline of Whitebark pine (Pinus albicaulis) in the Bob Marshall Wilderness Complex, Montana, USA. *Northwest Science*, *68*(3), 213–229.

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, *133*(0), 225–245. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304380000003227

Resler, L. M., & Tomback, D. F. (2008). Blister Rust Prevalence in Krummholz Whitebark Pine: Implications for Treeline Dynamics, Northern Rocky Mountains, Montana, U.S.A. *Arctic, Antarctic, and Alpine Research*, *40*(1), 161–170. http://doi.org/10.1657/1523-0430(06-116)

Smith, C. M., Wilson, B. C., Rasheed, S., Walker, R. C., Carolin, T. T., & Shepherd, B. (2008a). Whitebark pine and white pine blister rust in the Rocky Mountains of Canada and northern Montana. *Canadian Journal of Forest Research*, *38*(5), 982–995. http://doi.org/10.1139/X07-182

Smith, C. M., Wilson, B., Rasheed, S., Walker, R. C., Carolin, T., & Shepherd, B. (2008b). Whitebark pine and white pine blister rust in the Rocky Mountains of Canada and northern Montana. *Canadian Journal of Forest Research*, *38*(5), 982–995. http://doi.org/10.1139/X07-182

Tomback, D. F., & Achuff, P. (2010). Blister rust and western forest biodiversity: Ecology, values and outlook for white pines. *Forest Pathology*, *40*(3–4), 186–225. http://doi.org/10.1111/j.1439-0329.2010.00655.x

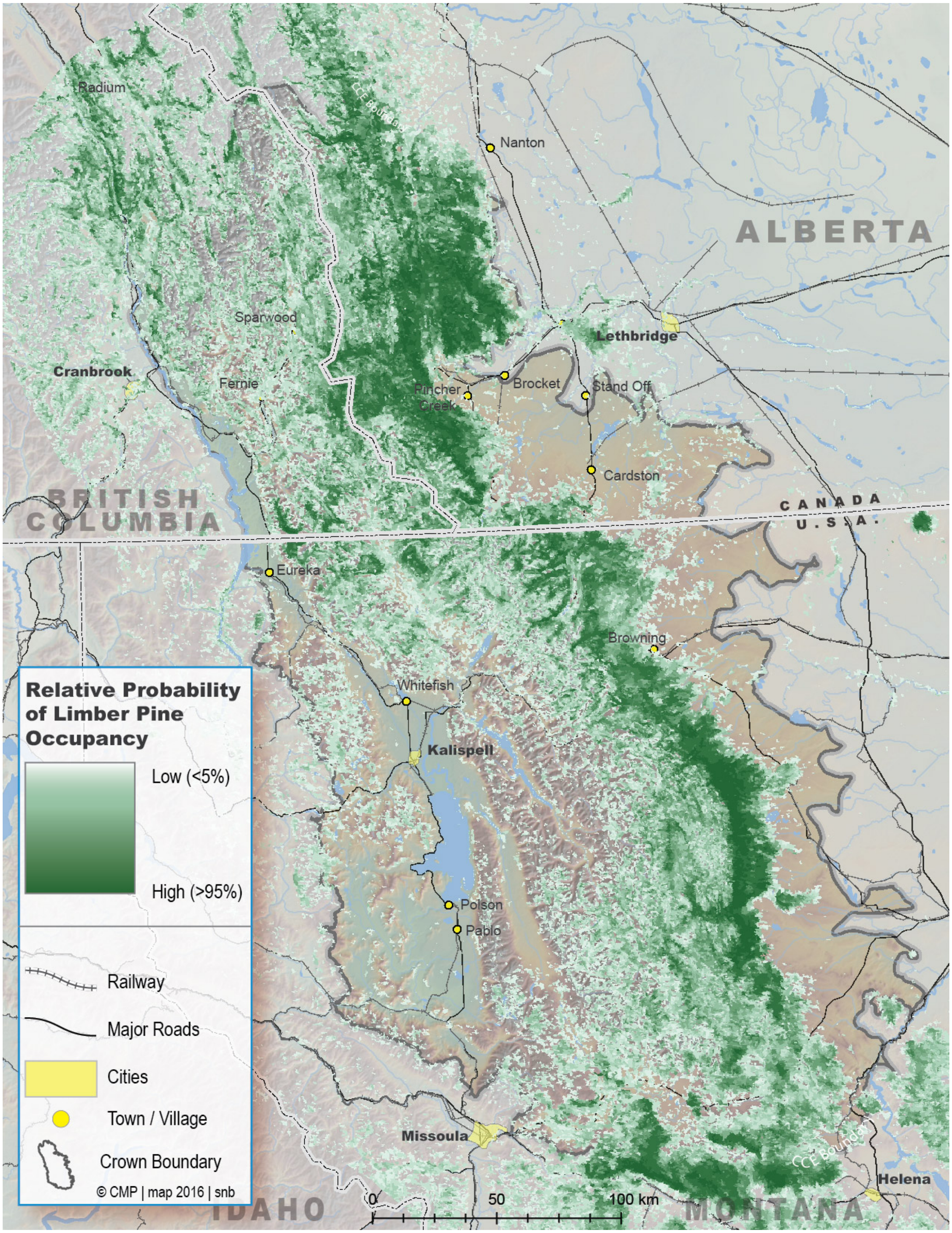
Tomback, D. F., Chipman, K. G., Resler, L. M., Smith-McKenna, E. K., & Smith, C. M. (2014). Relative Abundance and Functional Role of Whitebark Pine at Treeline in the Northern Rocky Mountains. *Arctic, Antarctic, and Alpine Research*, *46*(2), 407–418. http://doi.org/10.1657/1938-4246-46.2.407

Appendix A: Limber Pine Model

Limber pine are another high elevation five-needle pine in the CCE. They are categorized as Species at Risk in Alberta, and are a species that other jurisdictions in the CCE are carefully monitoring. Limber pine face many of the same threats as whitebark pine, and provide many of the same ecosystem benefits, but not as much research has been done on limber pine as compared to whitebark pine.

Given the importance of limber pine on the landscape, in addition to the whitebark Pine model, a limber pine model was also built. The same modelling approach was used to build a limber pine model as the whitebark pine model. The major differences in the limber pine model when compared to the whitebark pine model was the relative paucity of limber pine locations. The limber pine model only had 303 objects with presence as opposed to whitebark pine which had 2082 objects with presence. In addition to the lack of data, the limber pine data, like the whitebark pine data, suffered from similar problems of inconsistent sampling intensity, sampling design mismatches, and data collected for different purposes in different formats over a variety of timeframes. These data challenges are common when working over such a broad landscape that covers so many different jurisdictions, but the impacts of these challenges on the resulting model are unknown.

Using the data provided, the limber pine model we built has an the area under the Receiver Operating Characteristic (ROC) curve is 0.957; this means that 95.7% of the time the model can correctly differentiate between sites that are occupied by limber pine and those that are not (Pearce & Ferrier, 2000). These results are quite good, but it is important to note that a model is only as good as its input data. The limber pine data assembled likely does not represent the full environment of conditions in which limber pine would be present, this means that the model likely underestimates the probability of occupancy of limber pine on the landscape. However, without additional data it is not possible to make this assessment with any level of certainty. Additional limber pine location data could be used to improve and validate the model.



Relative probability of occupancy of limber pine in the Crown of the Continent Ecosystem