Modeling Contemporary Climate Profiles of Whitebark Pine (*Pinus albicaulis*) and Predicting Responses to Global Warming

Warwell, Marcus V.; Rehfeldt, Gerald E.; and Crookston, Nicholas L.

Rocky Mountain Research Station, USDA Forest Service, Forestry Sciences Laboratory, 1221 S. Main, Moscow, Idaho 83843, U.S.A.

The Random Forests multiple regression tree was used to develop an empirically-based bioclimate model for the distribution of *Pinus albicaulis* (whitebark pine) in western North America, latitudes 31° to 51° N and longitudes 102° to 125° W. Independent variables included 35 simple expressions of temperature and precipitation and their interactions. These climate variables were derived from a spline climate model that provides point estimates (latitude, longitude, and altitude). The analysis used presence-absence data from more than 119,000 plot locations largely from USDA Forest Inventory and Analysis. Of these plots, 2738 contained whitebark pine. Climate estimates were made for each plot location and assembled into 10 data sets containing all of the plots where whitebark pine was present at a proportion of 40%. The remaining 60% were randomly selected from the database.

The Random Forest regression tree analysis was then applied to each of the 10 data sets to develop a model for predicting the occurrence of whitebark pine from climate, using the general approach described in detail in Rehfeldt et al. (2006). The analyses consisted of constructing 10 forests each with 100 trees. The most parsimonious model was determined through an iterative process of removing variables and repeating the analysis until an increase in out-of-bag error was observed. Output from the analyses included statistics useful for evaluating the importance of input variables. We used importance scores from the Gini index to determine variable importance. To produce mapped projections of the biolclimate model, the estimated climate for 5.9 million pixels of 1 km resolution representing the terrestrial area of western USA were run down the regression tree. Each tree of the 10 forests provided a vote on whether or not the climate of the pixel was within the profile of the species.

The most parsimonious had 9 variables and had an overall error of classification of 2.47%. Among the 9 variables, the 3 most important were degree-days > 5 °C, mean temperature of the warmest month, and Julian date when degree-days > 5 °C reaches 100. When run independently, these 3 variables classified the species occurrence with an overall error of only 4.91%. The mapped prediction for the occurrence of *P. albicaulis* using its climate profile was more accurate than its published range map (Fig 1).