2013

MAPPING THE DISTRIBUTION AND ABUNDANCE OF WESTERN LARCH (LARIX OCCIDENTALIS NUTT.) WITH MULTI-TEMPORAL SATELLITE IMAGERY AND GRADIENT MODELING

Steven Joseph Touzel
The University of Montana

MAPPING THE DISTRIBUTION AND ABUNDANCE OF WESTERN LARCH (*LARIX OCCIDENTALIS* NUTT.) WITH MULTI-TEMPORAL SATELLITE IMAGERY AND GRADIENT MODELING

By

STEVEN JOSEPH TOUZEL

Bachelor of Arts in Geography, University of South Carolina, Columbia, SC, 2006

Thesis

Presented in partial fulfillment of the requirements for the degree of

Master of Science in Geography

The University of Montana

Missoula, MT

August 2013

Approved by:

Sandy Ross, Associate Dean of the Graduate School

Dr. David Shively, Committee co-Chair

Department of Geography

Dr. Zachary Holden, Committee co-Chair

Department of Geography

Dr. Andrew Larson

Forestry and Conservation

ABSTRACT

Touzel, Steven, M.S., Summer 2013 Geography

Mapping the Distribution and Abundance of Western Larch (*Larix Occidentalis* Nutt.) with Multi-Temporal Satellite Imagery and Gradient Modeling

Committee Chair: David Shively and Zachary Holden

Western larch (*Larix occidentalis* Nutt.) is one of three native North American larch species, it occupies the mountainous regions of northwestern North America, and it is a deciduous conifer. Western larch is among the most ecologically and economically important conifer tree species in the northern Rockies region. This study explores the viability of mapping western larch via the analysis of multi-temporal Landsat imagery and gradient modeling. Larch presence and abundance data from 300 field plots correlated with Normalized Difference Vegetation Index seasonal change ($NDVI_{sc}$) explains 46% of the variability in larch basal area. Multivariate models built with $NDVI_{sc}$ and additional climatic and topographic variables only slightly improved the models. These satellite imagery based models suggest that western larch tends to occur primarily on shaded, north-facing slopes within the study area. This analysis was contrasted with a gradient modeling approach using data from 4800 Forest Inventory and Analysis plots and a suite of fine scale (30-60 m) topographic and climatic data as predictors. These models correctly predicted larch presence with error rates of less than 20%. Presence or absence of western larch was found to be strongly dependent on minimum temperature and water balance variables (soil moisture deficit and actual evapotranspiration). Probability prediction rasters produced with these data also showed a noticeable northern aspect tendency. The accuracy of the remote sensing based models suggest that the method may be applied to other areas, and the output from both model types points to a strong relationship between larch presence and fine scale topographic and climatic factors, especially as they interact to affect soil moisture.

ACKNOWLEDGMENTS

I would like to thank Dr. Zachary Holden for all of his contributions to this project and his unceasing patience with me as I slowly, and sometimes painfully, worked to grasp theory and conduct this study.

I would also like to thank my co-chair, Dr. David Shively, for his guidance throughout this process which invariably kept me from straying too far from reaching my goals here at The University of Montana. I would like to thank my final committee member, Dr. Andrew Larson, for his unique point of view and the insight which he provided.

I would also like to thank all of the UM faculty and students, especially within the Geography Department, who helped to make my work possible. To all the Geography graduate students, thank you for your ability to tolerate and support me through this process.

Finally, I want to thank all my family and friends who in whatever large or small way helped to guide me through the past two years.

ABBREVIATIONS

AET	Actual Evapotranspiration
AIC	Akaike Information Criterion
ASP	Slope Aspect
ASR	Annual Solar Radiation
AUC	Area Under the Receiver Operating Characteristic Curve
BA	Basal Area
BAF	Basal Area Factor
CVER	Cross Validation Error Rate
CVK	Cross Validation Kappa Hat
CVRMSE	Cross Validation Root Mean Square Error
DBH	Diameter at Breast Height
DEF	Soil Moisture Deficit
DISS#	Dissection Index with Variable Window Sizes
Elev	Elevation
FIA	Forest Inventory and Analysis
GAM	Generalized Additive Model
GLOVIS	Global Visualization Viewer
GSP	Growing Season Precipitation
GSR	Growing Season Radiation
HLI	Heat Load Index
K	Kappa Hat
LAOC	*Larix occidentalis* Nutt.
LR	Multiple Linear Regression
MAT	Mean Annual Temperature
MODIS	Moderate Resolution Image Spectroradiometer
MRLC	Multi-Resolution Land Characteristics Consortium
MSE	Mean Squared Error

NDVI	Normalized Difference Vegetation Index
$NDVI_{sc}$	Normalized Difference Vegetation Index seasonal change
NED	National Elevation Dataset
NLCD	National Landcover Dataset
OOB	Out of Bag Error
PPT	Mean Annual Precipitation
RF	Random Forest
RMSE	Root Mean Squared Error
SDM	Species Distribution Model
TM	Thematic Mapper
TMIN	Minimum Temperature
TRASP	Transformed Aspect
USFS	United States Forest Service
USGS	United States Geological Survey
VI	Vegetation Index
WRS	Worldwide Reference System

CONTENTS

FIGURES

TABLES

1. Introduction

1.1. *Larix occidentalis* Nutt.

Western larch (*Larix occidentalis* Nutt.) is one of the most ecologically and economically important conifer tree species in western North America. It is a seral, shade intolerant, long lived, and deciduous species of pine with a life span of 400 to 900 years, and the largest of the *Larix* genus (Schmidt & Shearer, 1991). It can grow as fast as four feet in one year, and produces high density mature wood after only 15 years making it the most productive and economically important of the three native North American larch species (Parent, Mahoney, & Barkley, 2010). Western larch normally reaches 30 to 55m in height and a diameter at breast height (DBH) of 1 to 1.5 m, occasionally exceeding 60 m in height and two meters in diameter (Eckenwalder, 2009).

Western larch is the only native North American larch species to offer high timber values and desirable silvical attributes that make it an economically attractive and profitable species (Schmidt, Shearer, & Roe, 1976). It occupies an estimated 2.5 million acres in its U.S. range with an estimated volume of saw timber of 30.3 billion board feet; over half of the total acreage and board feet is found in western Montana. Economically, the wood products industry is heavily reliant on the species, and annual volume of larch harvest is second only to Douglas-fir in volume cut in the northern Rockies region (Schmidt et al., 1976). Breeding programs for larch have been in place since the 1980s (Rehfeldt & Jaquish, 2010) emphasizing the need for a predictive model for larch abundance as opposed to only larch presence, as well as an improved understanding of the interactions between western larch and the patterns of environmental conditions

suitable for growth. The ability to locate areas where larch is most abundant could improve breeding programs by informing managers of the environmental conditions that are most likely to increase seedling survival rates.

The range of western larch is limited to relatively moist and cool climate zones of the upper Columbia River basin of northwestern Montana, northern and west central Idaho, northeastern Washington, southeastern British Columbia, along the east slopes of the Cascade Mountains in Washington and northern Oregon, and in the Blue and Wallowa Mountains of southeastern Washington and northeastern Oregon (Figure 1) (Schmidt & Shearer, 1991).

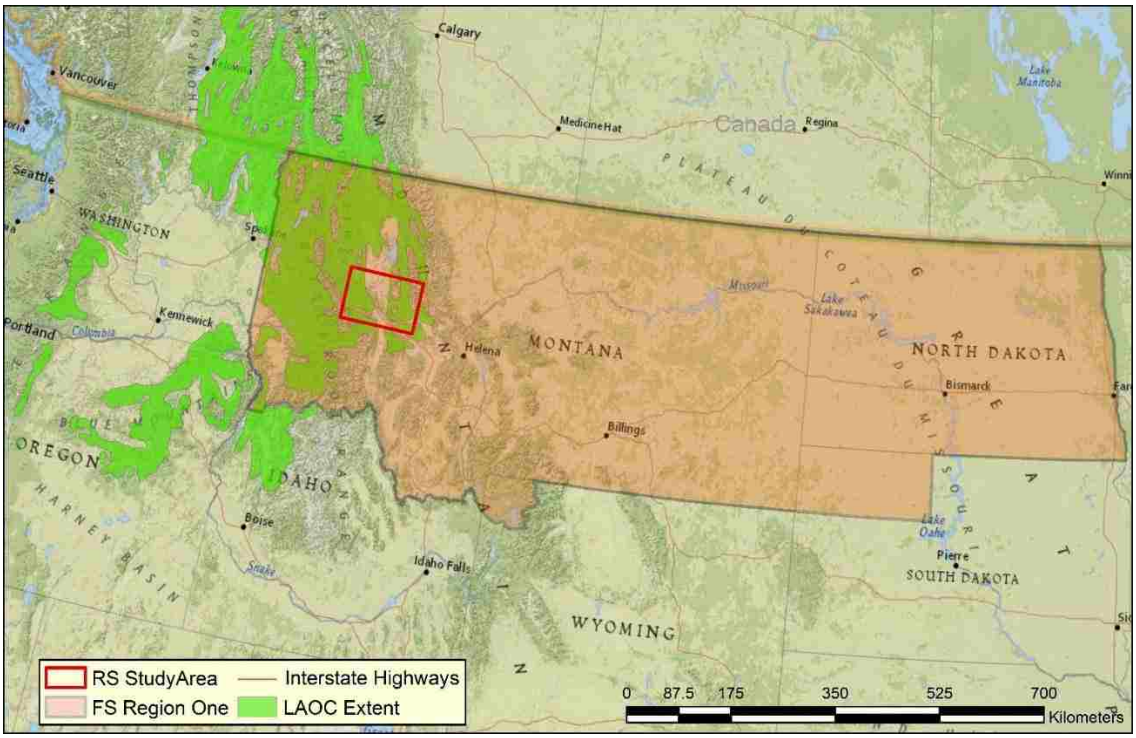


Figure 1. Western larch extent (Little, 1971), Forest Service Region 1 and Remote Sensing study area.

Growing best in the cool climates of the Northern Rocky Mountain slopes and valleys,

larch is often found on northeast facing slopes on deep and well drained Inceptisol and Alfisol soils (Schmidt et al., 1976, Larsen, 1916, 1925, Shearer, 1961). It grows in a wide range of elevations from 600 to 1600 meters (Eckenwalder, 2009), growing at higher elevations along its southern geographic extent, while low temperatures prevent its migration upward in areas farther north (Barbour, Burk, & Pitts, 1980). It most often forms an admixture with its common site associates, consisting of: ponderosa pine (*Pinus ponderosa*), Douglas-fir (*Pseudotsuga menziesii*), grand fir (*Abies grandis*), western white pine (*Pinus monticola*), western red cedar (*Thuja plicata*), western hemlock (*Tsuga heterophylla*), Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), and mountain hemlock (*Tsuga mertensiana*) (Schmidt & Shearer, 1991).

Western larch is dependent on frequent forest fire to reduce competition. As a result of thick bark and characteristic high branching, mature larch are among the most fire resistant trees in the northern Rockies (Schmidt & Shearer, 1991). Fire also thins stands, prepares seedbeds, and produces open canopy environments that allow seral species such as western larch to outperform competitor species, producing increasingly pure stands with increasing frequency of fires (Schmidt & Shearer, 1991). The speed at which larch recolonizes burned areas also helps to protect watersheds from prolonged erosion associated with such disturbances (Schmidt & Shearer, 1991). As a result of the standard practice of fire prevention and suppression in North American forests, pure stands of western larch are infrequent as less fire resistant species have expanded into these areas (Schmidt & Shearer, 1991, Powers, Adams, Joslin, & Fiske, 2005).

Western larch is a resilient species and generally insect and disease resistant.

Larch is resistant to windthrow due to its extensive root system and loses few branches resulting from the accumulation of snow during winter due to its deciduous nature (Schmidt & Shearer, 1991). It is generally unaffected by pine beetles. Dwarf mistletoe (*Arceuthobium laricis*), a parasitic plant, represents the greatest non-climatic disturbance to larch growth as it can reduce seed viability, limit growth, causes burls, brashness, weakens trees which enables further attacks by disease and insects, and in some cases causes mortality (Schmidt et al., 1976, Schmidt & Shearer, 1991).

As a deciduous conifer, larch has distinctive physiological and morphological characteristics that make it unique among other conifers that share its range. Western larch provides an aesthetically pleasing addition to the landscapes of the Northwest. The seasonal changes in its foliage present a major contrast in color even in stands in which larch is a small component. Newer, bright green needles begin to appear in May and June, becoming darker green during the summer, and turn bright yellow and senesce in September and October. Mature larch also contribute to wildlife diversity by providing habitat for pileated woodpeckers and other cavity nesters as well as food supply for squirrels, deer, elk, bears, and moose (Schmidt et al., 1976; McClelland & McClelland, 1999).

Western larch has distinctive physiological characteristics resulting primarily from its deciduous growth habit. It produces more photosynthetic leaf surface area each year than any other coniferous tree of a given size class (Eckenwalder, 2009), and leaves also have a relatively large specific leaf area (Gower & Richards, 1990). In terms of growth rates, this combination of factors has the useful consequence of allowing larch to assimilate more carbon via higher photosynthetic rates (Gower & Richards, 1990,

Eckenwalder, 2009). However, this also requires a relatively large amount of water and growth rates become limited when soil moisture is scarce (Higgins et al., 1987). While western larch has the ability to germinate under a wide variety of conditions, even at the dry end of its range (Oswald & Neuenschwander, 1993), higher surface temperatures on south and west facing exposures decrease moisture availability which reduces seedling survival. These physiological and growth characteristics of western larch, combined with its inability to tolerate shade, limits the distribution of larch to relatively mesic sites and establishment of stands to post disturbance, open canopy areas (Schmidt et al., 1976, Schmidt & Shearer, 1991).

There is growing concern among scientists and land managers about the persistence and spatial redistribution of vegetation given rapidly warming climate. The predictions of Rehfeldt and Jaquish (2010) suggest that the suitable habitat of western larch may face drastic changes in the coming century. Compounding these possible changes are results derived from a study concerning climate change and Siberian larch in which Shuman, Shugart, and O'Halloran (2011) establish that the genus may have important feedbacks to the earth's climate system. The study presented evidence that larch forests increase albedo over an area as the loss of needles during winter results in a reduced obscuring of snow build up beneath the canopy. This leads to an increase in albedo from 0.16 in non-deciduous dominated cover to 0.26 in areas of larch dominated cover. Due to its possible response to changes in climate as well as its ecological and economic value, western larch represents an important species for investigation.

1.2. Mapping and Modeling of Species Distribution and Abundance

In the past decade, classifying the presence or absence of tree species via the analysis of remotely sensed imagery has become increasingly prevalent (e.g., Pfeffer, Pebesma, & Burrough, 2003, Krishnaswamy, Kiran, & Ganeshaiah, 2004, Zimmermann, Edwards, Moisen, Frescino, & Blackard, 2007, Buermann et al., 2008, Maignan, Breon, Bacour, Demarty, & Poirson, 2008, Narayanaraj, Bolstad, Elliott, & Vose, 2010, Sanchez-Azofeifa et al., 2011). Increased quality of data and availability of data are the main drivers of this trend, and this has led to new and improved species distribution modeling predictor variable datasets in otherwise un-surveyed areas (Gould, 2000; Nagendra, 2001, Turner et al., 2003, Buermann et al., 2008). A parallel evolution has also occurred in modeling methodologies, increasing model prediction validity while using relatively small sample datasets through the implementation of such methods as non-parametric computer learning or neural network methods (Pandit, Hayward, Leeuw, & Kolasa, 2010, Cutler et al., 2007). As a result, species distribution models (SDM) are utilized more and more often as tools for assessing and motivating environmental conservation and management (Pandit et al., 2010).

Satellite imagery can provide high resolution temporal and reflectance information that can be analyzed for spectral signatures, phenological events (e.g., seasonal senescence, bud burst, and flowering), and canopy biomass which are often unique among tree species (Franklin & Miller, 2009). Within coniferous forests, species-level discrimination can be difficult as many species are not easily distinguishable via phenology (e.g., most do not senesce), spectral signature, or canopy biomass. Because western larch is a *deciduous* conifer it represents a unique exception to the common traits of coniferous trees making it a prime candidate for classification

via the analysis of remotely sensed imagery. The deciduous nature of western larch may be exploited in this case by applying seasonal change detection methods to vegetation metrics within coniferous forests. The change in larch needle presence causes a large shift in photosynthetic activity that is detectable with green biomass metrics such as the normalized difference vegetation index (NDVI) derived from remotely sensed imagery. The ability to detect changes in vegetation indices (VI) metrics over the yearly cycle of larch growing period to senescence period was pioneered in a study conducted by Busetto et al. (2010) concerning European larch (*Larix decidua*). The study utilized the time series of 250 meter resolution Moderate Resolution Image Spectroradiometer (MODIS) NDVI data to monitor larch phenology and found strong correlations between larch senescence and NDVI change. However, the study did not attempt to extrapolate these correlations to map larch abundance. Additionally, the relatively coarse spatial resolution of MODIS data could limit its utility for detecting larch in mixed species stands and in topographically complex and heterogeneous environments.

Previous studies' attempts to establish the spatial distribution of western larch in western North America have been based on bioclimatic characteristics such as temperature, precipitation, freezing dates, combinations of temperature and precipitation related metrics (Rehfeldt & Jaquish, 2010), and soil characteristics (New, 1999). These two studies, respectively, utilized relatively coarse resolution (1km resolution) datasets and study areas that neglected the majority of western larch habitat (only including south-eastern British Columbia).

It is important to note that the mountainous regions which western larch inhabits are subject to significantly complex spatial variation in climate as well as other

environmental factors. This can substantially alter the regularity of environmental conditions even at a relatively fine scale. Precipitation, soil characteristics, amount of incident solar radiation, and temperature ranges vary due to multiple factors associated with the topographically complex regions of north western North America. In these areas, vegetation is dependent on variables that operate on sub-kilometer scales, thus conclusions drawn from models based on predictor variables at kilometer scales may be erroneous or at least over-simplified (Pandit et al., 2010). Errors associated with modeled species distributions using coarse-scale datasets, when compared with more informed models, can lead to misinterpretation of species-environmental variable correlations and subsequent mismanagement of species (Pandit et al., 2010). Strahler, Logan, and Bryant (1978) cite indirect gradients such as fine scale topographic parameters as strongly influencing vegetation composition. Other studies have also shown that topographically complex regions create considerable uncertainty in the extent of vegetation change due to climate change (Bartlein, Whitlock, & Shafer, 1997), and that finer scale climatic factors (direct and resource gradients) are of importance to tree establishment (Baker, 1995). Topographic variation has a significant influence on direct and resource gradients through such interactions as elevation on temperature and precipitation, and slope angle and aspect with radiation, soil moisture, soil development, and wind exposure (Franklin, 1995, Baker, 1995). These factors combine to create considerable variation within kilometer-scale climate model grid cells, which become visible as distinct changes in forest vegetation at sub-kilometer spatial scales (Lutz, Wagtendonk, & Franklin, 2010).

Western larch seedlings are particularly dependent on topographic variation

through its effect on radiation and temperature levels as they influence available soil moisture (Oswald & Neuenschwander, 1993). Thus, due to the interactions between complex terrain and direct and resource gradients, it is beneficial to utilize finer resolution data. Previous studies have shown that macro-scale climatic predictors do not perform well at fine scales, and that SDMs which include fine scale predictors produce superior model performance (Franklin &Miller, 2009).

The goal of the present study is to assess the potential for mapping the distribution and abundance of western larch using multi-temporal Landsat TM 5 imagery, in conjunction with interpolated climatic and topographic variables. Seasonal needle loss should result in large decreases in greenness relative to other species, which should be detectable in the near infrared portion of the electromagnetic spectrum. The analysis of multi-temporal remotely sensed data to produce a western larch distribution map has not previously been attempted, and the moderately high (30m) resolution of Landsat imagery should provide further insights into the topographic influences on the distribution of the species. The results of remote sensing-based models are compared with a more traditional gradient modeling approach employed in this study using Forest Inventory and Analysis (FIA) plot data, developed without satellite-derived predictors.

This gradient modeling process takes a more direct route in attempting to improve upon the previous attempts by Rehfeldt and Jaquish (2010) by incorporating fine scale (30-60 meter) climatic, topographic, and water balance parameters to predict the probability of the presence of larch within the study area. This process essentially investigates the patterns that exist between western larch locations and generalized ecological conditions present over a large portion of the western larch range in the

contiguous United States through the use of FIA plots. The FIA plot data are a much larger dataset which covers the range of western larch habitat within United States Forest Service (USFS) Region 1 (Figure 1), and the systematic sampling methods which were utilized to produce them reduce bias. Though both methods seek to improve the knowledge of the spatial distribution of the species in a spatially heterogeneous environment, building a model based strictly on environmental gradients will allow for the prediction of the fundamental niche as determined by the range of environmental conditions within which larch currently exists (realized niche) (Franklin & Miller, 2009, Lomolino, 2006).

2. Materials and Methods

2.1. Study Site

The modeling of larch presence and abundance using multi-temporal satellite imagery was conducted on portions of the Lolo and Flathead National Forests in western Montana, contained within Landsat Thematic Mapper (TM) scene Worldwide Reference System (WRS) path: 41, row: 27 (Figure 2).

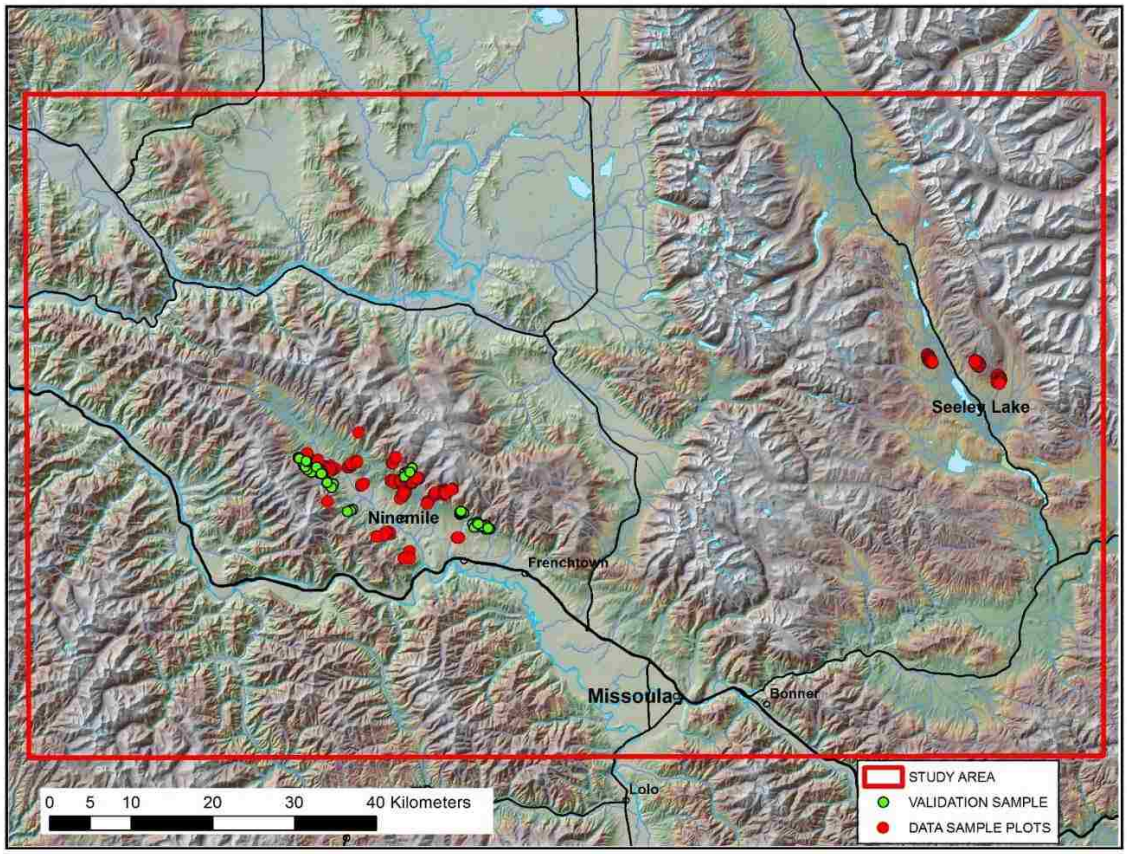


Figure 2. Remote Sensing Study area. Located in western Montana, with original model building samples and validation sample locations shown in the Nine Mile Valley area in the west and the Seeley-Swan Valley in the east.

The predictions of western larch presence and abundance are calculated over an area covering approximately 10,500 km^2 encompassing a large portion of the western larch range of western Montana. The interaction of large-scale climate patterns, geologic events and subsequent landforms have produced a complex pattern of ecosystems in this region as climate, soils, and vegetation vary considerably even over small distances. The continental climate within this zone ranges from dry cool temperate, wet cool temperate, to subalpine boreal. The general climate pattern of this region is represented by cold to mild wet winters and warm dry summers. Forests are predominantly coniferous and species vary according to local environmental factors and disturbance history. Elevations range from 700 to 3,000 meters and this accounts for a majority of the climatic, soil, and vegetation differences. Annual precipitation in this region varies with elevation, ranging from 27 cm in lower elevation valleys to 97 cm at higher elevations. Ponderosa pine (*Pinus ponderosa*), Douglas fir (*Pseudotsuga menziesii*), and western larch are common at low-to mid-elevations, while lodgepole pine (*Pinus contorta*), Engelmann spruce (*Picea engelmannii*) and subalpine fir (*Abies lasiocarpa*) occur at higher elevations.

2.2. Field Data Collection

Field data of western larch presence and abundance were collected over a 3 month period from May to August, 2012. Sampling was conducted in two general areas, 220 of 300 samples were located in Nine Mile Valley, approximately 30 km west of Missoula. An additional 80 samples were located in the Seeley-Swan Valley (Figure 2), approximately 50 miles east of Missoula. Sample plots were selected purposively at sites that were relatively accessible by forest roads. To reduce edge effects, sample plots were located at least 50 meters from roads or trails. Plot sample sites were separated by at least 50 meters to prevent the overlap of data extracted from 30 meter resolution topographic and climatic layers, and Landsat TM imagery. Information gathered included: counts of the presence of all tree species; ocular estimates of the percent presence of deciduous plant material at levels of 0-1 meter, 1-5 meters, sub-canopy, and canopy; ocular estimates of percent canopy cover; latitude and longitude; and photographs of views from the center of the plot facing north and west. The abundance of trees in each plot was estimated using variable radius sampling, and presence of a tree within a plot was determined via the use of a Spiegel Relascope using a basal area factor (BAF) of 10 (Bitterlich, 1984). The 10 BAF, for the Spiegel Relascope, is utilized with the units of ft^2/acre. When converted to m^2/ha this becomes a BAF of 2.296. This BAF was chosen as it was large enough to preclude excessively large plots that often lead to an obscuring of the visibility of trees from the central survey position as a result of topography or tree density (Bitterlich, 1984). Deciduous undergrowth and deciduous tree information was recorded in order to document other plants that may affect NDVI (Krishnaswamy et al., 2004). The location of the center of each plot was recorded using

a Garmin GS60CSx global positioning system datalogger; at least 50 positions were logged and then averaged. Photographs were collected to document site locations for possible future site resampling and to help justify results or inspect unusual findings.

Independent samples for model validation were collected during April, 2013. All 50 samples were collected in the Nine Mile Valley area. Samples were randomly generated, and buffered to select those between 50m and 200m of forest roads. Data collected for each validation sample included counts of all western larch within a Relascope BAF of 10, and photographs of views from the center of the plot facing north and west. Sample points were navigated to via the use of a Garmin GS60CSx.

2.3. Topographic, Climatic, and Remotely Sensed Data

A suite of topographic, biophysical and climatic data were generated for the study area using 30m resolution digital elevation data from the USGS National Elevation Dataset (Gesch, 2007, Gesch et al., 2002). All variables are listed in Table 1 and are described in Holden, Morgan, and Evans (2009). These variables included growing season and annual cumulative incident clear sky radiation, using the model of Flint and Childs (1987), and heat load index (McCune, Grace, & Urban, 2002). Topographic predictors included slope, aspect, and metrics of topographic dissection (Pike & Wilson, 1971) calculated with multiple window sizes. Climatic data included 30 year average (1971-2000) mean annual precipitation and minimum temperature generated using the thin plate spline models implemented in Anusplin (Hutchinson, 1993). Additionally, a simple monthly climatic water balance model was generated for the study area following methods described by Dobrowski et al. (2013) and adapted to finer spatial resolutions by Holden, Landguth, and Purdy (in review) to produce soil moisture deficit (DEF), and actual evapotranspiration (AET) for the study area.

Table 1. Description of variables and associated abbreviations.

Variable Description	Abbreviation
Climatic Variables	
Minimum Temperature (°C)	TMIN
Mean Annual Precipitation (cm)	PPT
Growing Season Precipitation (cm)	GSP
Mean Annual Temperature (°C)	MAT
Topographic Variables	
Dissection Index with variable window sizes	DISS# (i.e. DISS27)
Elevation (m)	Elev
Aspect (degrees)	ASP
Transformed Aspect	TRASP
Annual Solar Radiation (w/m^2)	ASR
Growing Season Radiation (w/m^2)	GSR
Heat Load Index	HLI
Slope (%)	SLOPE
Water Balance Variables	
Soil Moisture Deficit (mm)	DEF
Actual Evapotranspiration (mm)	AET
Landsat Derived Variables	
Seasonally differenced Normalized Differenced Vegetation Index	$NDVI_{sc}$

Landsat Thematic Mapper (TM) 5 imagery for the study area was acquired through the Global Visualization Viewer administered by the US Geological Survey (USGS). For the analysis of seasonal vegetation index change, a summer image and a fall image were necessary. While there are many cloud free summer images available, the availability of corresponding relatively cloud free fall (> Julian Day 300) images for the study area are extremely limited. It is necessary to use a Julian Day 300 or soon thereafter image due to the likelihood of deciduous understory vegetation senescence

occurring prior to this date and the presence of snow cover obscuring image reflectance values soon after, both of which reduce the ability of NDVI to accurately detect vegetation (Busetto et al., 2010). Busetto et al. (2010) found that the timing of the full leaf off date or the end of the senescence period of the phenological cycle for European larch can be assumed to be relatively static as it is more dependent on photoperiod than factors such as temperature or drought (Migliavacca et al., 2008, Stöckli et al., 2008), and for this current study this characteristic is assumed to also apply to western larch. The use of early spring images was also considered, however no cloud-free early spring Landsat images were available for the study area. Additionally, snow cover at high elevations would have limited the extension of the model predictions to higher elevation areas. As a result of these limitations, the most recent clear images were utilized. Landsat TM5, WRS path: 41, row: 27 images representing summer from August 8th, 1998 (Julian day 220) and fall from October 27th, 1998 (Julian Day 300) were the most recent cloud free images. These images were imported into ERDAS Imagine 11 and inspected for irregularities (i.e. cloud cover, haze, and striping) and then processed for reflectance values. The processed imagery was then used to calculate NDVI values via the following equation:

$$NDVI = \frac{LandsatBand4 - LandsatBand3}{LandsatBand4 + LandsatBand3} \quad \text{Equation 1}$$

The two scenes were then differenced to create an NDVI seasonal change value ($NDVI_{sc}$).

$$NDVI_{sc} = Summer\ NDVI - Fall\ NDVI \quad \text{Equation 2}$$

The NDVI equation (Rouse, Hass, Schell, & Deering, 1974) produces values in the range of -1 to 1, with values of 0.3 to 1 representing vegetated surfaces with increasing values indicating increasing green vegetation and lower values indicating less vegetated surfaces such as defoliated trees, bare earth, or water (Jensen, 2000). Near infrared light is reflected and red light is absorbed at higher rates when they interact with healthy vegetation producing the higher values of NDVI. The opposite occurs when there is less green vegetation present. Due to a lack of sensitivity, there is also a possible source of error in these values caused by the presence of snow, cloud cover, or understory vegetation. NDVI values are highly correlated with canopy biomass and leaf area index, making it a suitable index for canopy vegetation change purposes. It is also relatively insensitive to atmospheric and topographic effects, enables scene-to-scene comparisons, and has been used frequently in SDMs (Jensen, 2000, Krishnaswamy et al., 2004, Buermann et al., 2008).

A 30m resolution, 2006 National Landcover Dataset (NLCD) produced by the Multi-Resolution Land Characteristics (MRLC) Consortium was used to discriminate between forested and non-forested areas. A mask was created using these two classification types combined with a cloud cover mask and applied to all predictions in order to remove the urban, agriculture, deciduous, and other land classifications that may create false positives as a result of the presence of non-larch deciduous material which exist outside of mixed and coniferous forests.

2.4. Data Analysis

A suite of statistical models including multiple linear regression (LR), generalized additive models (GAM) (Wood, 2008) and the machine learning algorithm Random Forest (RF) (Breiman, 2001) were used to quantify relationships between larch presence or abundance, and satellite and climatic, topographic, and water balance predictor variables. The Raster package (Hijmans & Etten, 2011) was used to import and manipulate all raster datasets. All statistical and modeling data analysis was performed in R (Development Core Team, 2011). Models were developed for predicting larch abundance as a continuous variable, and detecting presence or absence of larch at each plot. Presence/absence models were developed using various thresholds of larch abundance, with threshold values for presence set at a sequence of basal area (BA) values from 2.296 m^2/ha (10 ft^2/acre) to 34.44 m^2/ha (150 ft^2/acre).

2.4.1. Remote Sensing Based Models for Larch Abundance

Correlation analysis and simple linear regression were first used to examine relationships between field observations of larch abundance and $NDVI_{sc}$, as well as topographic, climatic and water balance variables. Predictor variables were screened prior to analysis, and variables with a correlation (Pearson's r) of 0.7 or higher were removed to avoid multi-colinearity. Correlations between larch abundance and predictor variables were also tested under a variety of predictor variable transformations where it was deemed appropriate.

Larch abundance was then modeled with GAM, LR, and RF models using all of the remaining variables and then pruned to produce more parsimonious models. For the

LR models, variables were selected which contributed significantly to the explanatory power of the model as determined through a stepwise bi-directional variable selection process using the Akaike Information Criterion (AIC) and using a threshold p value of less than or equal to 0.05 (Fisher, 1925). GAMs were built with Gaussian distribution and a log link in order to restrict model predictions to positive values. Smoothing terms were applied to predictor variables in GAM models and models with and without smoothers were compared via AIC and adjusted R^2 (R^2) values to determine their significance to model performance. Smoothers were applied with the mgcv optimized splining procedure which uses a cross validation approach with the data to choose an optimal value for the smoothing parameter (Wood, 2011). Candidate GAMs were further pruned to produce parsimonious models using the ≤ 0.05 p value, and AIC score optimization.

Random Forest models generate multiple classification trees (Breiman, 2001), each of which provide a vote for the final prediction of the given observation. Predictions of each tree differ due to a process of resampling the data and utilizing a randomly selected subset of the full list of the given predictor variables. Random Forest models were run with nodes split with 2 randomly selected variables and 2000 trees to insure convergence to the lowest mean squared error (MSE). Variables selected for RF models were determined by an iterative process whereby the inclusion of a given variable was dependent on whether it contributed to improvements in percent variation explained, and through variable importance data as derived from reductions in model error associated with each variable.

2.4.2. Remote Sensing Based Binary models of Larch Presence/Absence

In order to establish thresholds of larch abundance at which $NDVI_{sc}$ can effectively predict larch presence in a cell, GAM and Random Forest modeling were applied to predict larch occurrence as a binary variable at various levels of larch abundance. Binary larch presence-absence response variables were created at increments of 2.296 m^2/ha (10 ft^2/acre), and models were generated to predict larch presence probabilities at each increment. GAMs were built using the binomial distribution with a logit link and variables were chosen in the same manner as described above.

RF models were optimized via the minimization of out of bag (OOB) error matrix percent error and kappa hat (K) statistics. Due to the way error matrix percent error is calculated (number of correct predictions / number of samples), increasing the frequency of absences when modeling for higher larch BA may result in high accuracy as there is a higher probability of guessing correctly. To account for this, it is important to adjust for the possibility of chance agreements in model output and this is accounted for in the K statistic (Cohen, 1960).

2.4.3. Gradient Modeling with Forest Inventory Plot Data of Larch Presence/Absence

Data from 4,824 FIA plots covering all of the USFS Region 1 (Figure 1) lands were used to develop gradient models of larch occurrence (presence-absence). Presence of larch at each plot was assigned based on a threshold of 10% abundance or greater. Predictor variables used in modeling were all related to water and energy either directly or indirectly and in some cases were strongly correlated with each other ($-0.7 > r > 0.7$)

due in part to the processes used in producing them. In cases such as interpolated temperature (MAT, TMIN), elevation data are directly used in the production process and output displays a resultant strong correlation between the two. Elevation was removed from models within which temperature was a significant predictor and TMIN was utilized in place of MAT as TMIN is known to limit the distribution of larch (Barbour et al., 1980). Larch presence of 10% or greater was then predicted with GAM and RF models built using the remaining variables. This modeling process was implemented in order to take a more direct route in investigating the relationship between larch presence and fine scale topographic, climatic, and water balance variables. A secondary objective here was to compare satellite-derived maps of larch occurrence with more traditional topographic and climate envelope based models. In this way, further inquiries can be made as to the effectiveness of such gradient models to accurately predict larch presence. However, this comparison is dependent on the accuracy of the remote sensing based model predictions. RF and GAM model predictor variables were selected and models were generated using the same criteria for binary models as described above.

2.4.4. Model Accuracy Assessment

Final models were chosen based on optimization of the aforementioned model performance metrics and then further tested through cross validation. In the cross validation procedure a partitioned dataset was used for model building containing a randomly selected 90% of the full dataset. This model was then used to predict the withheld 10% and the error rates between predicted and true values were used to calculate model accuracy. An average root mean squared error (RMSE) was calculated

for the continuous larch BA prediction models, and a percent classification error rate, area under the Receiver Operating Characteristic curve (AUC), and a K statistic were calculated for larch presence-absence prediction models. For GAMs, the predicted probability of larch presence was classified as present or absent using the conventional greater than 0.5 probability threshold. The cross validation process was repeated 100 times, each time with a new random draw of observations. The resulting 100 mean RMSEs, classification error rates, or K statistics were averaged to calculate a final mean cross validation root mean square error (CVRMSE), a mean cross validation percent error rate (CVER), or a mean cross validation K statistic (CVK) for each model and then compared.

The final best performing models for both continuous and binary larch predictions were then utilized to create prediction rasters over the entirety of the study area. Binary predictions were produced as probabilities of presence, and continuous predictions were produced as larch BA m^2/ha. The NLCD and cloud mask was then applied to each prediction to remove cloud cover and non-forested areas. Remote sensing based prediction rasters were validated by examining differences between model predictions and the 50 independent validation plot sample data as an RMSE statistic for continuous models or a percent error rate and K statistic for binary models. For the final, best performing, continuous larch BA model prediction raster, 1,000,000 (≈10% of the prediction raster) random sample points were generated and correlations between predicted larch BA and topographic, water balance, and climatic variables were examined. Polar plots were used to assess the predicted distribution of larch abundance relative to aspect.

Comparisons between binary remote sensing based models and FIA produced gradient model predictions were made in order to examine any correlations between the gradient model and remote sensing based models. Due to the differences in response variables between the FIA binary models and the remote sensing models, comparisons can only be made about the probability of larch presence. However, even these comparisons are tenuous due to the differences in methods of production of the plot data between the FIA dataset and the remote sensing model dataset. It is hypothesized that the predictions of larch presence generated via remote sensing methods will represent a subset of the gradient model predictions. These comparisons assume that, based on the theoretical application of topographic variables within a model utilizing a larger geographic range of larch plot data to produce presence or absence probabilities, certain aspects of the model may not be ideal for the study area in question. Rasters were imported into ArcGIS 10 where all map creation was performed.

3. Results

3.1. Remote Sensing Continuous Basal Area Model Performance

Larch abundance was moderately well correlated with $NDVI_{sc}$. Simple linear regression of $NDVI_{sc}$ and larch abundance produced an R^2 value of 0.47 (Figure 3). Results of continuous models of larch abundance are shown in Table 2.

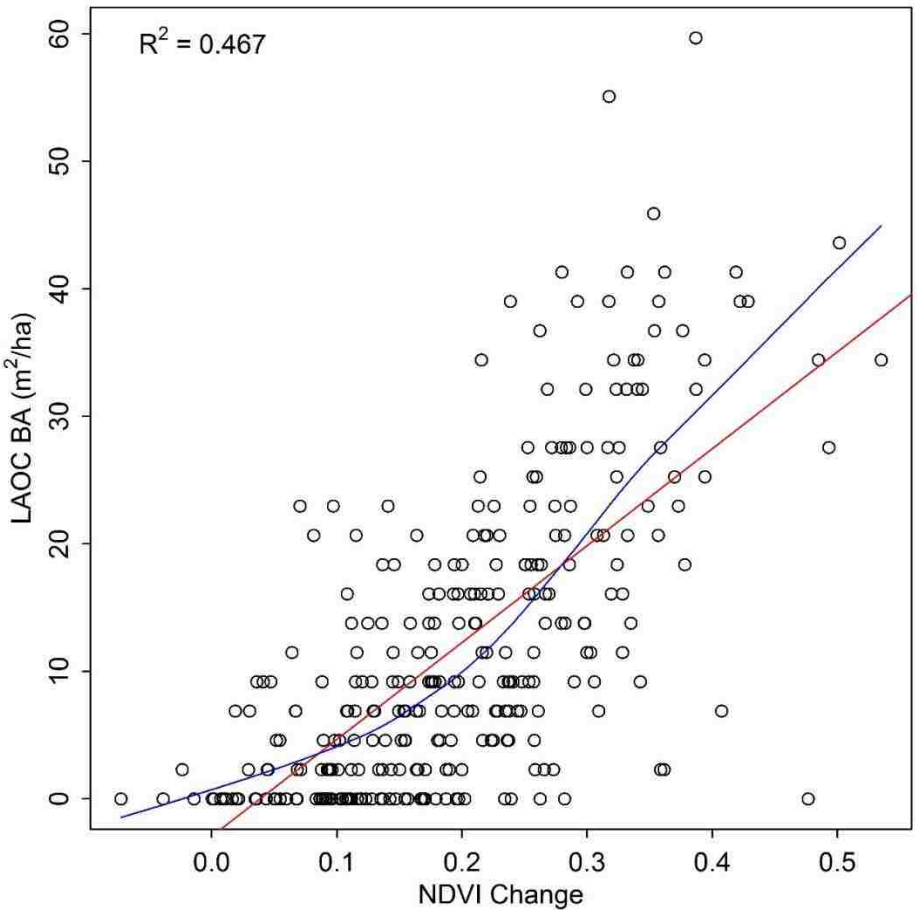


Figure 3. Simple linear regression (red line) of the correlation between larch abundance (m^2/ha) and $NDVI_{sc}$, with a locally weighted scatterplot smoother (blue line) also shown for reference.

Table 2. Diagnostic values from three candidate model types used to predict larch abundance.

Model Type	Predictors	R^2	Variation Explained (%)	Cross Val RMSE m^2/ha (ft^2/acre)	Independent Val RMSE m^2/ha	AIC
Linear Model	$NDVI_{sc}^3$, HLI, DEF, PPT	0.502		5.681 (24.91)	10.93	1290.81
Linear Model	$NDVI_{sc}$	0.467		5.836 (25.77)		+15.5
GAM	s($NDVI_{sc}$), HLI, DEF	0.519		5.610 (24.60)	11.73	2132.68
GAM	s($NDVI_{sc}$)	0.495		5.699 (25.00)		+11.2
Random Forest	$NDVI_{sc}^2$, ASR, TMIN, HLI		50	5.564 (24.57)	9.428	
Random Forest	$NDVI_{sc}$		29	6.512 (28.76)		

Parsimonious LR, GAM and RF models yielded similar accuracies. The strictly linear form of the $NDVI_{sc}$ variable was unable to accurately capture the confusion resulting from non-larch deciduous material as is illustrated by the locally weighted scatterplot smoother (Figure 3). To compensate for this trend, squared and cubed transformations were applied to this variable which reduced residuals at the lower end of the predicted western larch BA range. $NDVI_{sc}^2$ and $NDVI_{sc}^3$ were included as predictor variables and when either transformation was included in a final model all lower forms of the variable were also included. A final parsimonious LR model created with the inclusion of $NDVI_{sc}^3$, HLI, PPT, and DEF produced an R^2 of 0.50 but only reduced AIC scores from 1306.3 to 1290.81. When the correlation with $NDVI_{sc}$ was tested within a log link Gaussian distribution GAM with applied smoothers, the R^2 value increased to 0.495 (Table 2). Pruning a global GAM to produce a parsimonious final model generated predictor variables of $NDVI_{sc}$, HLI, and DEF and the R^2 value increased further to 0.52, representing the highest R^2 reached of any parsimonious model. AIC values for the final GAM with applied smoothing terms improved significantly when compared with the same model without smoothers and models without smoothers also performed poorly when compared via the R^2 model statistic. Based on these model diagnostics the smoothing terms appear to provide a stronger fit than the parametric transformations used in the LR model. When this final GAM is cross validated there is only a 0.09 m^2/ha improvement in CVRMSE from the simple $NDVI_{sc}$ GAM to the more complex model, indicating only marginal improvement when model complexity is increased. Residuals of this final GAM are normally distributed but heteroscedastic, exhibiting an increase in variance at higher basal area values. When the accuracies of the final LR and GAM models are tested on the independent validation dataset, LR

performs better, suggesting that the smoothing term within the GAM model may not be as effective as the parametric transformations utilized in the LR and RF models. Based on the model diagnostics and validation, there is some confusion as to the effectiveness of the smoothing term. It is possible that the smoothing term is inappropriate or that it may overfit the model to the original data and therefor does not predict well when extrapolated to new data.

Using Random Forest models to predict larch BA as a continuous variable showed similar explanatory power according to the model accuracy diagnostics with the original dataset. An RF model using only $NDVI_{sc}$ produced an explained variation of 28.94%. When a full parsimonious model was built with $NDVI_{sc}^2$, annual cumulative solar radiation (ASR), TMIN, and HLI, explained variation increased considerably to 50.13%. Inclusion of these predictor variables also showed corresponding improvements in CVRMSE from 6.51 m^2/ha to 5.56 m^2/ha. This final RF model produces the lowest CVRMSE of any model and when tested using the independent validation sample data the RF model performed significantly better than either the GAM or LR models. Based on the robust methodologies for model validation, it is apparent that Random Forest provides the best model for predicting larch basal area.

Final models for LR, GAM, and RF were utilized to predict larch abundance across the study area (Figure 4).

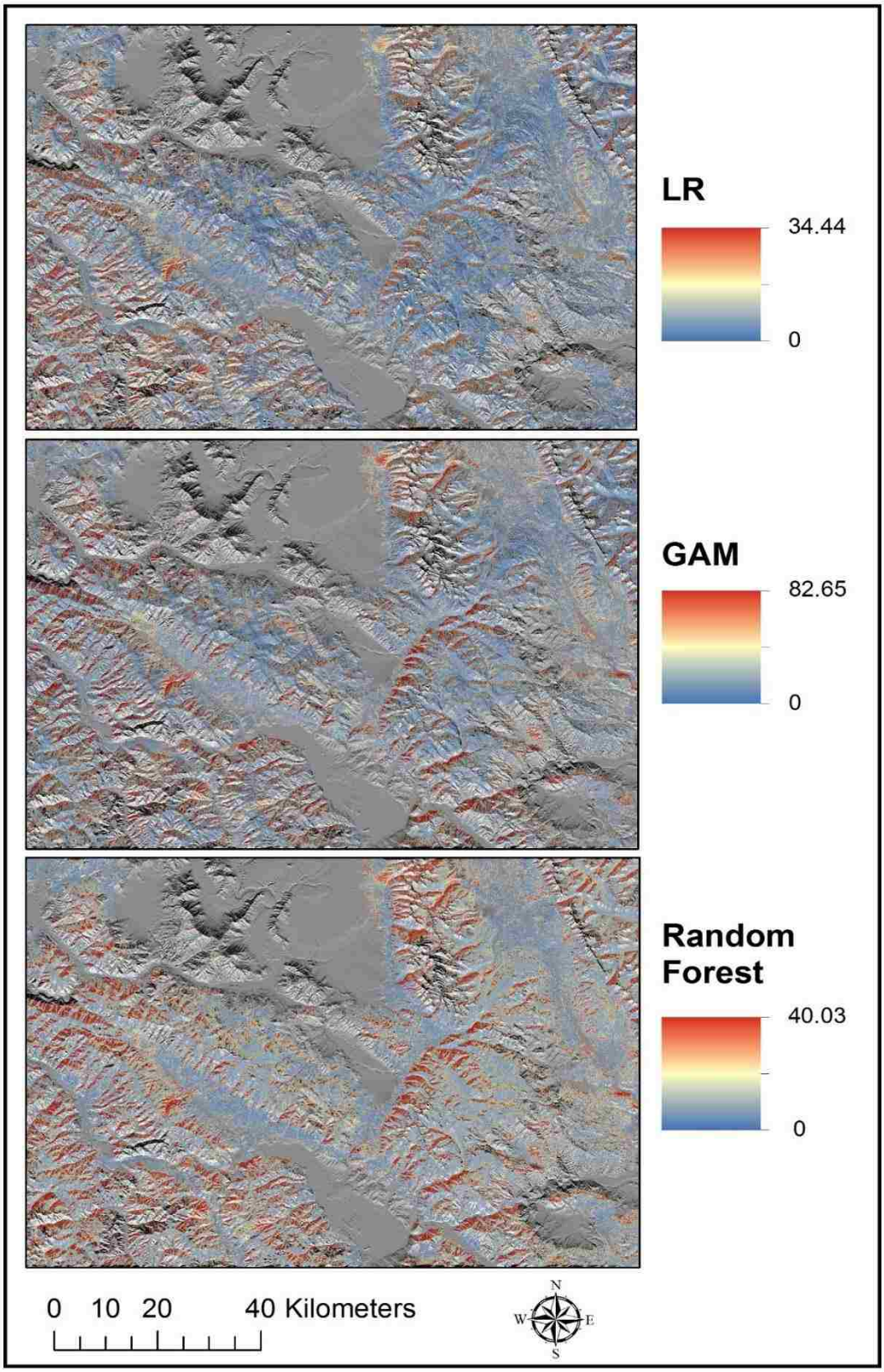


Figure 4. Optimized parsimonious model predictions of larch BA (m^2/ha) using LR, GAM, and Random Forest.

Output for each model displays a strong pattern of high larch abundance on northern facing slopes. This pattern is further corroborated via the averaged 1,000,000 randomly extracted samples from the RF model prediction raster. As shown in the polar plot (Figure 5A) it is clear that, based on the $NDVI_{sc}$ larch abundance predictions, there is a strong relationship between larch abundance and aspect, with higher levels of larch abundance on northern facing slopes within the study area. There is also a strong correlation with annual incident solar radiation, with larch abundance decreasing with increasing annual incident radiation (Figure 5B). Soil moisture deficit and AET display slightly weaker negative trends with larch abundance, and may be regarded as indicators of larch absence (Figure 5C, 5D). Variability in soil moisture deficit increases as larch abundance increases, which may indicate a lack of sensitivity to soil moisture deficit below a certain value.

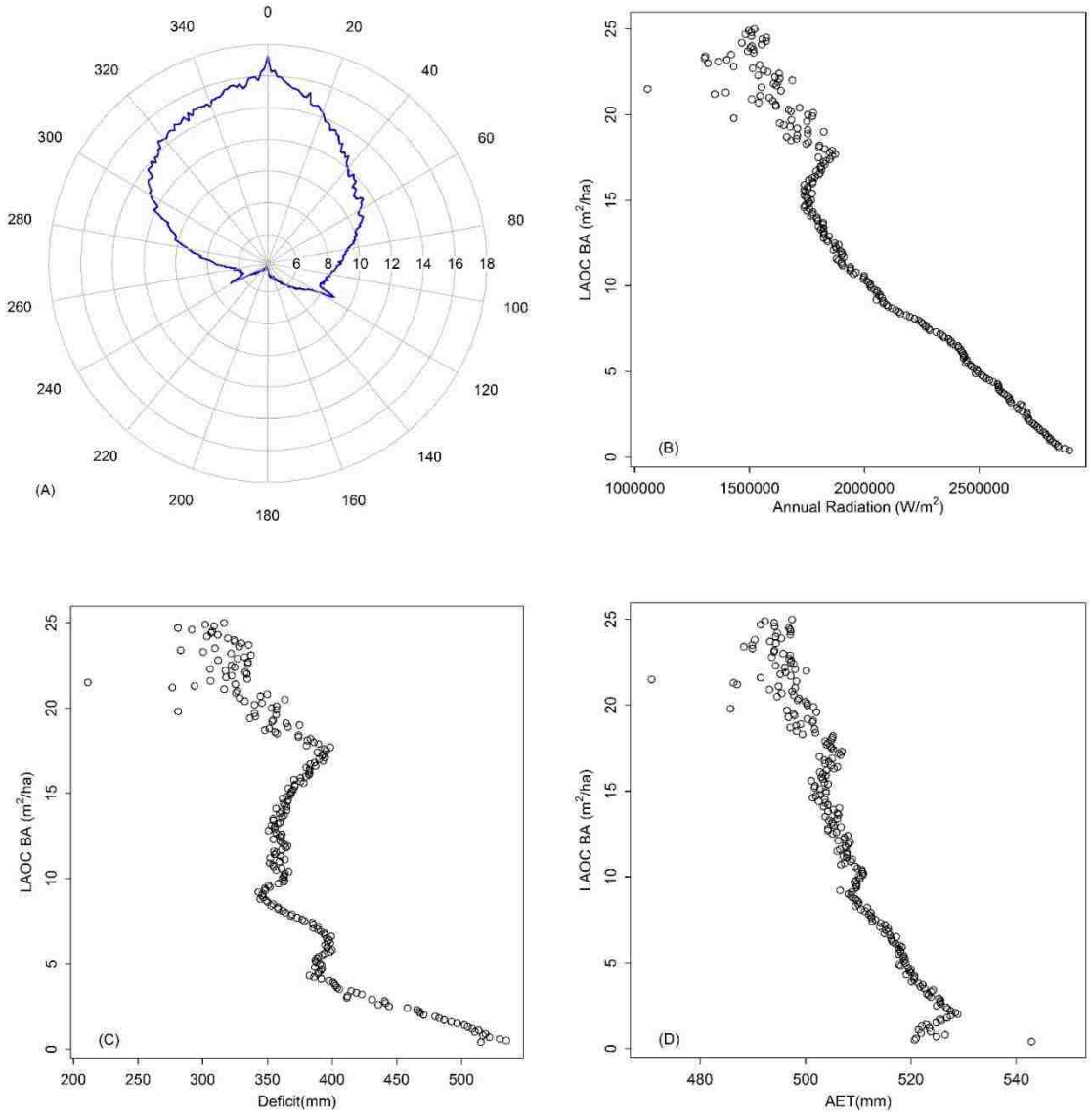


Figure 5. Polar plot and scatter plots illustrating the correlations of sampled predicted larch abundance and A) Aspect; B) Annual solar radiation, C) Soil moisture deficit, and D) Actual evapotranspiration.

3.2. Remote Sensing Presence-Absence Model Performance

Results of presence-absence models based on $NDVI_{sc}$ are shown in Table 3 and probability predictions generated for the study area with parsimonious models can be seen in Figure 6.

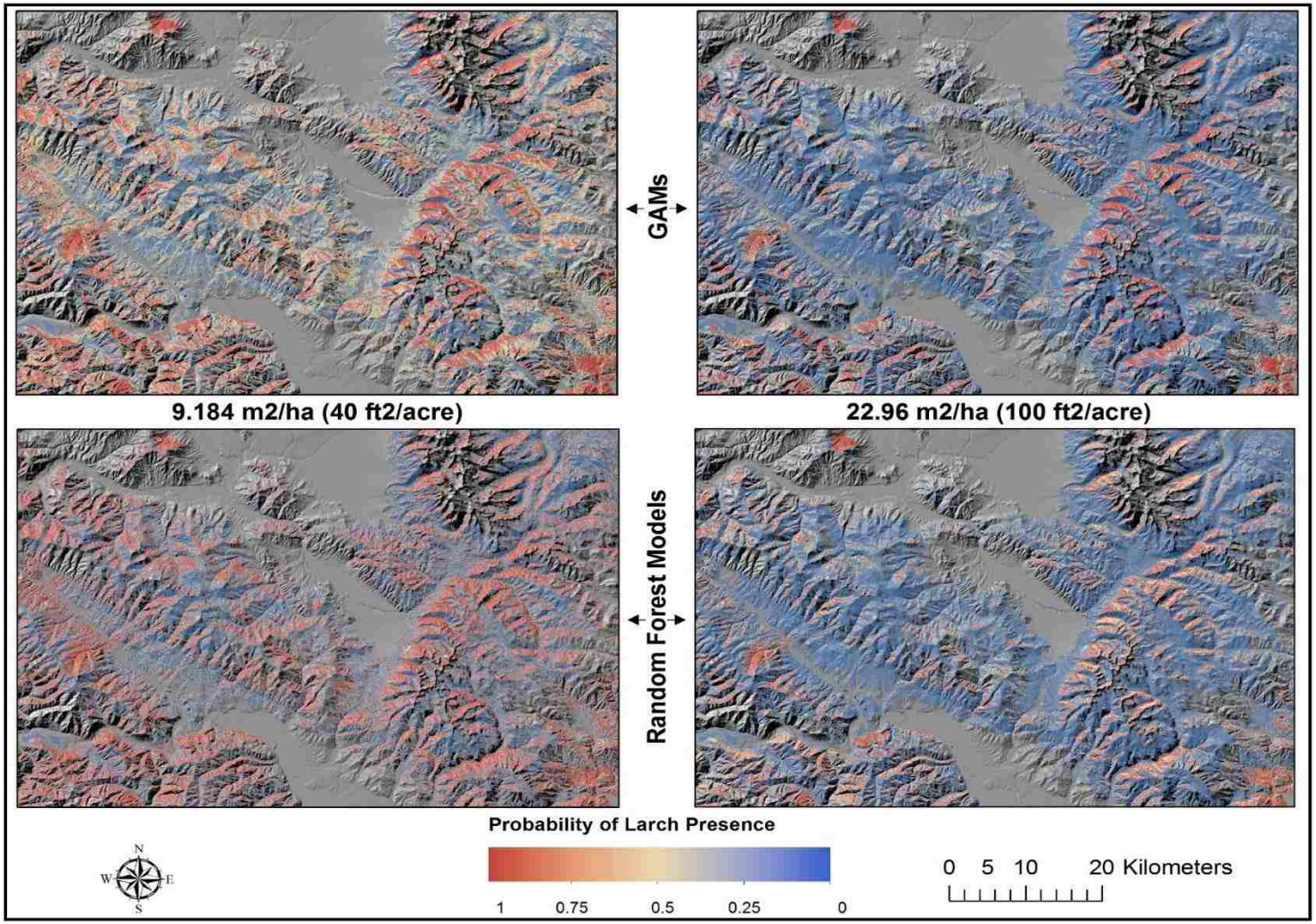


Figure 6. Remote Sensing based probability of Larch presence at 9.184 m^2/ha and 22.96 m^2/ ha thresholds.

Table 3. Model diagnostic values from FIA gradient models and $NDVI_{sc}$, topographic and climatic data model types used to predict larch presence.

Model Type	BA Threshold m^2/ha (ft^2/acre)	Predictors	R^2	Error Rate (%)	K	AUC	Cross Val Error Rate (%)	Cross Val K	Independent Val Error Rate (%)	Independent Val K
Remote Sensing Larch Presence Models										
GAM	9.18 (40)	$NDVI_{sc}$	0.329	24.6	0.506	0.831	25.3	0.475	32.6	0.372
GAM	22.957 (100)	$NDVI_{sc}$	0.382	13	0.53	0.890	13.0	0.527	20.4	0.385
Random Forest	9.18 (40)	$NDVI_{sc}^2$, AET		24	0.539	0.809	24.2	0.592	32.6	0.361
Random Forest	22.957 (100)	$NDVI_{sc}^2$, ASR, HLI		13	0.550	0.865	13.3	0.869	18.3	0.363
FIA Gradient Models										
GAM	10%	TMIN, AET, ASR, DEF, DISS27	0.26	19.5	0.335		20.2	0.279	0.856	
RF	10%	TMIN, AET, DEF, DISS27		19.7	0.364		20.1	0.303	0.840	

Model strength depended on the threshold BA (m^2/ha) value chosen for assigning larch presence-absence for each plot, and the strongest correlations were achieved when presence was assigned to plots with larch BA values of 9.183 m^2/ha (40 ft^2/acre) or greater. The RF model predictor variables at this threshold included $NDVI_{sc}^2$ and AET. This model returned OOB error rates of 24%, K values of 0.5397, and AUC values of 0.8095, suggesting a moderate to good performance when predicting larch presence at this density. Using a binomial GAM at this level incorporated a smoothed $NDVI_{sc}$ predictor variable only and generated slightly better results producing an R^2 of 0.329 and an AUC of 0.8313. To make Random Forest and GAM output more comparable, a binary classification was applied to the probability predictions of the GAM using the conventional greater than 0.5 as the reclassification rate for all classified presences. In this way, the agreement between the two model varieties is more apparent, as the error rates and K values are nearly identical. Utilizing the cross validation method produced similar error rates and K scores when compared with models using the entire dataset for model building and assessment. This was the case for all BA thresholds using the GAM and RF model types. Further analysis of these binary models using the independent validation points, showed similar model agreement. However, the GAM models produced lower commission error rates which led to slightly higher K values.

Predicting larch presence at a BA of 22.957 m^2/ha (100 ft^2/acre) yielded a more skewed split and was expected to decrease K values as a result. A binomial GAM including predictor variables of smoothed $NDVI_{sc}$, yielded increases in the GAM R^2 to 0.382 and AUC to 0.89. K statistics increased when predicting at the more skewed 22.957 m^2/ha threshold suggesting that model performance due to chance is limited.

However, in the more skewed model, the commission error rate for classified presences of 40.35% indicates that the model misclassifies presence at a noticeably higher rate. The corresponding commission error rate of 23.22% for the more balanced 9.183 m^2/ha threshold indicates an improvement in model performance in the most fundamentally important category for this study, predicting larch presence. Above the 9.183 m^2/ha larch BA threshold, larch abundance caused $NDVI_{sc}$ seems to be larger than most non larch source $NDVI_{sc}$, and therefore represents the lowest larch abundance at which accurate predictions can be extrapolated. A final binary GAM model was utilized to extract polygons of larch presence and a larch presence map was generated (Figure 7).

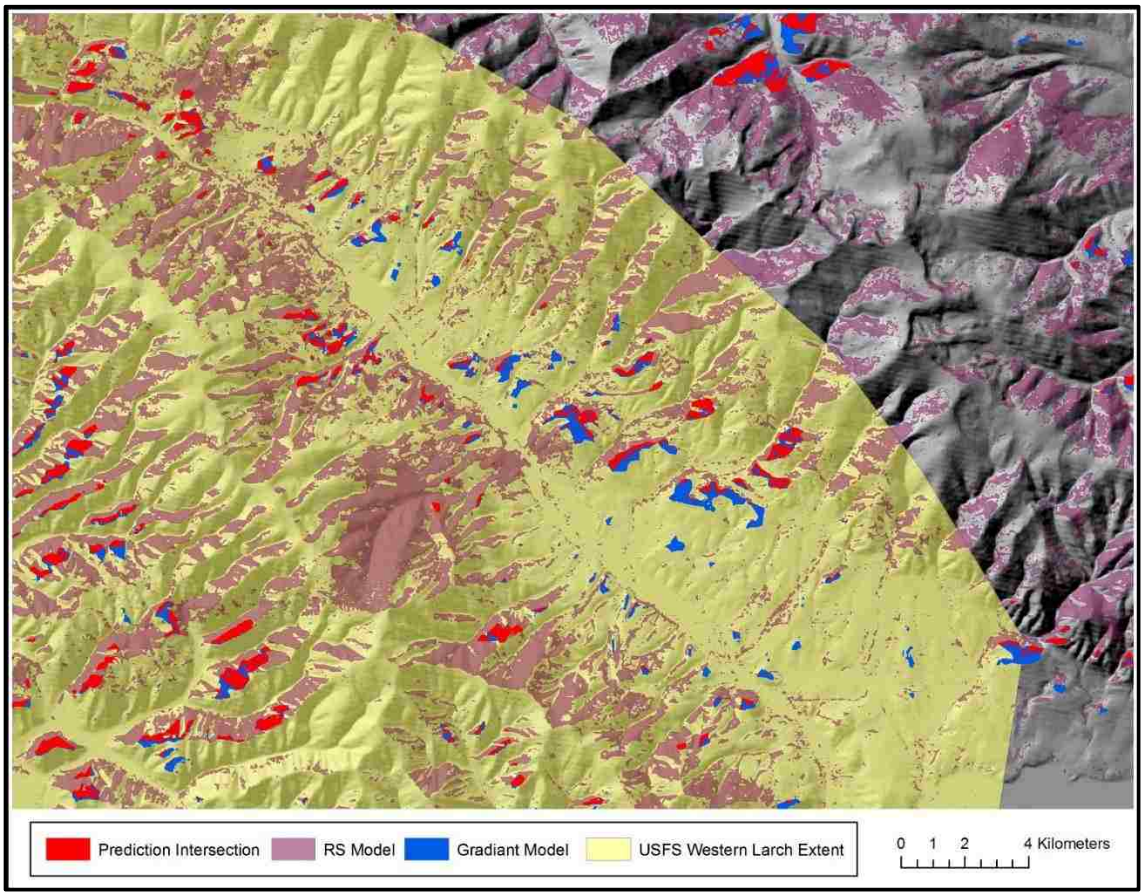


Figure 7. Western Larch presence predictions and intersection of remote sensing models and gradient models.

3.3. Gradient Modeling with FIA Data Performance

Simple gradient modeling with FIA plot data produced moderate model performance (Table 3). Probability predictions generated for the study area with final parsimonious models can be seen in Figure 8.

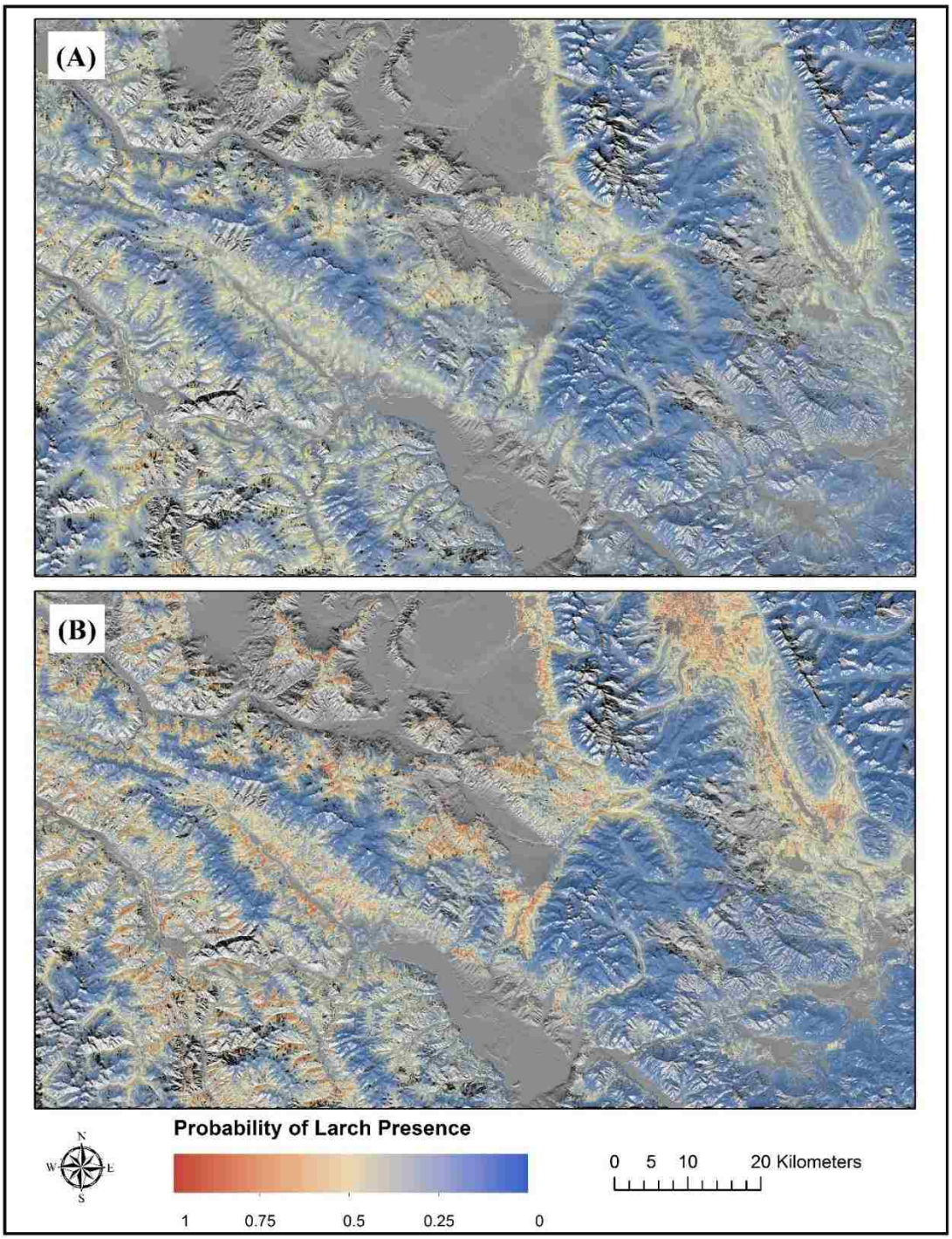


Figure 8. FIA data based Probability of Larch presence at 10% presence. A) Binomial GAM, B) Random Forest Model.

Error rates were low for both the binomial GAM and the Random Forest models (< 20%), however K statistics were also very low (< 0.37). The low K scores are as a

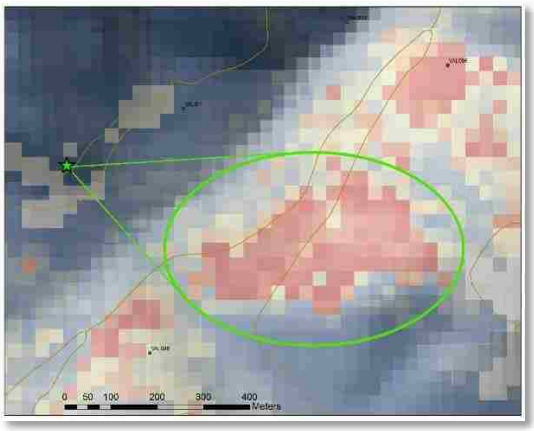
result of the skew in the number of absences in the FIA plot dataset, and the high error rates (57% for the Random Forest model) associated with predicted presence. The majority of errors within these models occurred as predictions of larch presence in plots where larch was absent. A pruned Random Forest gradient model included predictor variables of TMIN, AET, DEF, and DISS27. This model yielded K values of 0.36 and AUC values of 0.84. A final parsimonious model for the binomial GAM model type included predictor variables of TMIN, AET, ASR, DEF, and DISS27, all of which were included with applied smoothers. The binomial GAM produced a lower K value of 0.33, but a higher AUC value of 0.86. Minimum temperature represented the most significant predictor variable for both models, supporting previous work suggesting that minimum temperature is an important factor in limiting the distribution of western larch (Barbour et al., 1980, Schmidt & Shearer, 1991). The error rates for these models are deceiving and should not be regarded as indicative of a particularly good model fit. The K values provide more useful information in attempting to compare between the gradient model types and the remote sensing models. The gradient models show significantly lower K values, which suggests that the low error rates are not strongly related to the model fit. Comparing the K values between the gradient models and the remote sensing models suggests that seasonally differenced NDVI may significantly enhance the detection of western larch presence. Further analysis of these gradient models using the 10% withheld cross validation process, provided similar results. For the GAM, cross validation error rate and Kappa statistics rose and fell by 0.7% and 0.05% respectively as compared with the full dataset model rates. The RF model yielded a slightly lower increase in cross validation error rate (0.4%) and a slightly larger decrease in cross validation K (0.06%) than GAM models, suggesting that commission error rates were

more affected by the data withheld from model building.

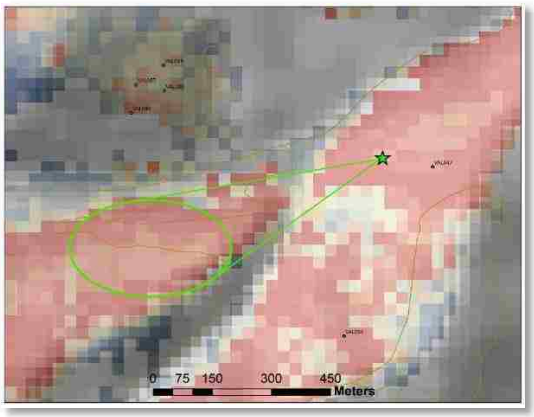
4. Discussion

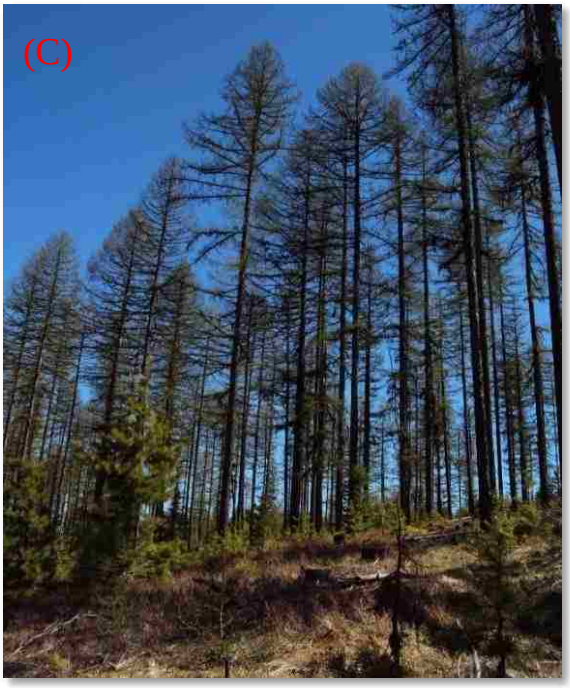
Remote sensing based models and climate envelope based models showed moderate ability at predicting the abundance and occurrence of western larch within the study area. When $NDVI_{sc}$ based Larch abundance models included topographic, climatic, and water balance predictors, model performance improved, but only slightly. This suggests that the majority of the variation is accounted for with $NDVI_{sc}$. Predictions from these models generally overpredict the density of larch within a cell but otherwise appear to provide a good indicator of the presence of high or low volumes of larch within the forest matrix. This is particularly visible in the photographs depicting larch abundance and the corresponding model predictions shown in Figure 9A-C. A number of other possible sample plot characteristics that may have affected the larch abundance-$NDVI_{sc}$ correlation were also examined, but yielded inconclusive results.


(A)


0 50 100 200 300 400
Meters


(B)


0 75 150 300 450
Meters


(C)

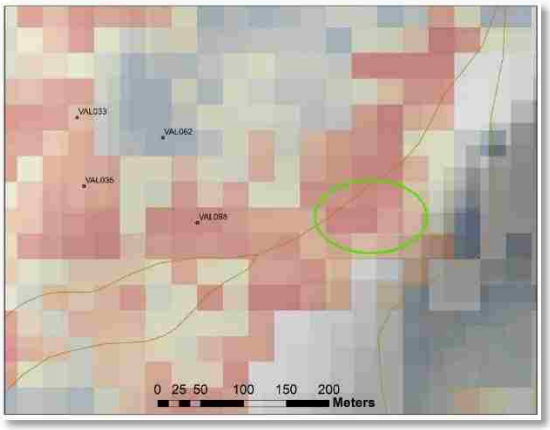

VAL033
VAL062
VAL035
VAL058
0 25 50 100 150 200
Meters

Figure 9. A-C: Model predictions of high larch density with corresponding photographs displaying corroborating larch abundance. D: Errors associated with predictions of high larch density within high elevation floodplains.

One possible source of error could arise from the presence of deciduous understory vegetation other than larch, to assess the potential influence of understory vegetation on larch detection, larch abundance and deciduous understory vegetation were combined into a single deciduous index. The resulting correlation between $NDVI_{sc}$ and the deciduous index was 0.49 which was not significantly different from the larch abundance model. This suggests that while other deciduous material in the area may affect predictions, the misclassifications or possible over predictions associated with this likely source of error are minimal. Percent canopy cover was also examined as a potential source of error, as it may affect the magnitude of $NDVI_{sc}$ by influencing the detectability of deciduous vegetation beneath the forest canopy. However, no relationships were found between plot canopy cover and the magnitude of NDVI change. A separation of the two sample sites (Nine Mile Valley and Seeley-Swan Valley) and examinations of the variation in $NDVI_{sc}$ response to larch abundance were conducted to test for possible inter-site variation. Model diagnostics suggest that there was no major

difference between sites. However, the sample size in the Seeley-Swan site was limited compared to the Nine Mile site, and a more even sampling may demonstrate variation between the sites. Visual inspection of the BA predictions combined with independent field verification data, suggests that models are particularly well able to predict the absence of larch within the coniferous forest matrix.

Presence-absence models utilizing $NDVI_{sc}$ were developed using varying thresholds of abundance to define the presence of larch at a plot. Model accuracy generally increased with increasing larch abundance, with the highest model accuracies achieved using a 40% threshold. These results are rather intuitive, but have important implications for the potential application of multi-temporal Landsat data for mapping the presence of western larch in new areas or at broader scales. These results suggest that accurate detection of larch presence as only a small fraction of total canopy cover may be limited. Seasonally differenced satellite indices show good potential for detection of moderately abundant or pure larch stands in this region.

Both the continuous larch abundance predictions and the presence-absence predictions showed western larch occurrence to be markedly limited on southern facing aspects. When topographic, climatic and water balance data were extracted from the remote sensing based larch abundance prediction rasters each model showed that larch abundance was negatively correlated with climatic water balance deficit (DEF) and ASR. This finding suggests that moisture limitations on exposed, dry, south-facing slopes may restrict the distribution of larch to more mesic aspect positions, supporting previous findings of increased regeneration success of larch in these areas (Schmidt & Shearer, 1991). The study region is relatively dry, receiving an average of 40 centimeters of

precipitation annually at lower elevations. Topography and principally solar insolation strongly mediates the moisture gradients in mountains of the region. The results of this analysis suggests that higher solar insolation on south-facing aspect positions result in soil moisture deficits that may limit larch seedling survival and therefor larch abundance.

Within this topographically complex study area, temperature, moisture and insolation included in the various models are highly variable even at sub-kilometer scales. Larch presence and abundance appears to be highly sensitive to aspect-scale moisture and temperature gradients. These results highlight the importance of including high spatial resolution, topographically sensitive bioclimatic predictors to adequately capture the distribution of western larch. The improved spatial resolution of the physically-based climatic water balance models used here may offer new insights into the mechanisms that govern the presence and abundance of western larch, as well as other species in this region. Past studies of the vulnerability of western larch to climate warming were performed using relatively coarse (1km) resolution climate data that ignored aspect influences on temperature, radiation and moisture and the potential for terrain to mediate larch distribution (Rehfeldt & Jaquish, 2010). The results of this study suggest that accounting for this fine-scale variation may be essential for understanding the potential response of western larch to climatic change.

This information should prove useful to management of the species especially as it pertains to the reforestation of burned areas or harvested areas utilizing larch's fast growth rates and improving BA yields of larch forests in areas that exhibit high growth capacity. The predictive models of larch abundance could also help to guide harvests or assist in the collection of larch occurrence data for further studies concerning phenology,

species health, and improving species distribution modeling.

Through the analysis of models via independent field validation samples, several key sources of error were established. Within the NLCD coniferous forest classification areas, there are frequent high elevation flood plains or bogs. In these flood plains, high volumes of grasses dominate with little or no canopy cover, which combine to produce an $NDVI_{sc}$ that is comparable to areas with high volumes of larch presence. One of the 50 randomly generated independent validation points was within an area which exhibited this land cover type and was therefore excluded from the model validation dataset (Figure 9D). Another source of model prediction error occurred in areas that displayed characteristic larch $NDVI_{sc}$, and were also within the optimal topographic, climatic, and water balance parameters but were unforested, containing mostly shrubs. It may be possible to remove these patches as potential larch growth areas with a more accurate coniferous forest landcover mask. Limited cloud cover within the fall Landsat image caused errors in predictions and these areas were removed. Where these clouds obscured ground cover, high values of $NDVI_{sc}$ were a result, and this led to areas of predicted high larch abundance. Upon further investigation, these areas do not contain high volumes of larch which supports the importance of obtaining a cloud free fall image.

While it is of interest to compare more commonly applied plot-based gradient models with models developed from remotely-sensed imagery, it is important to note that these analyses were conducted across very different geographic extents. The FIA sample data used to construct the gradient models represent a much larger portion of the distribution of western larch than the remote sensing study. This is significant because

distributions of larch outside of the remote sensing study area encompass a broader range of climatic characteristics, which could limit the distribution of western larch, or mediate the importance of aspect-scale, radiation driven influences on larch distribution and abundance. For example, slope and aspect may be more important within precipitation limited areas including northwestern Montana, but could be less important in much wetter areas of northern and central Idaho.

5. Conclusion

This study examined the potential for multi-temporal Landsat derived vegetation indices and climatic envelope models to capture the current and potential distribution of western larch. Most importantly, change in NDVI between summer and fall images was moderately well correlated with the presence and abundance of western larch at field plot locations. Random Forest models generated using this correlation demonstrated the ability to relatively accurately predict larch abundance. This finding is of importance because of its possible application in monitoring larch presence and abundance without relying on costly and time consuming field data collection methods. Climatic envelope models based on 4800 FIA plots that incorporated fine-scale climate and biophysical predictors also predicted the presence of larch with moderate accuracy and captured similar features of larch distribution. Both methods produced larch presence maps that exhibited strong dependence on solar insolation and climatic water balance variables. These results suggest that within the study area, the occurrence of western larch is strongly dependent on topography, and in particular solar insolation. This conclusion is of importance due to the implications it may have in informing the management of the species. It can also help to update predictions of possible changes in the extent of suitable habitat due to climate change, considering the most recent predictions do not include topographic scale dependencies.

The moderately strong correlation between larch abundance and seasonally-differenced NDVI derived from Landsat imagery suggest that remotely sensed data have

some utility for improving efforts to map the current distribution of western larch in this region. However, there are important limitations to the analysis, and further study is warranted. One major limitation of using multi-temporal imagery as an indicator of larch presence is the general lack of cloud-free spring or fall scenes during snow-free periods, which is needed to capture seasonal differences of larch foliage. Future efforts to detect larch with multi-temporal imagery may benefit from the availability of more recent cloud free fall imagery. Furthermore, an investigation of a multiple year series utilizing this technique could also provide insight into the possible trends or migration of larch growth. In addition, alternative satellite sensors such as the MODIS with higher temporal revisit times, potentially fused with available cloud-free Landsat scenes should be evaluated as potential tools for mapping western larch.

6. References

Baker, W. L., & Weisberg, P. J. (1995). Landscape Analysis of the Forest-Tundra Ecotone in Rocky Mountain National Park, Colorado. *The Professional Geographer: The Journal of the Association of American Geographers, 47*, 361.

Barbour, M. G., Burk, J. H., & Pitts, W. D. (1980). *Terrestrial plant ecology*. Menlo Park, CA: Benjamin/Cummings Pub. Co

Bartlein, P. J., Whitlock, C., & Shafer, S. L. (1997). Future Climate in the Yellowstone National Park Region and Its Potential Impact on Vegetation. *Conservation Biology, 11*, 782-792.

Bitterlich, W. (1984). *The Relascope Idea: Relative Measurements in Forestry*. Slough: Commonwealth Agricultural Bureau.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32.

Buermann, W., Saatchi, S., Smith, T. B., Zutta, B. R., Chaves, J. A., Milá, B., & Graham, C. H. (2008). Predicting Species Distributions across the Amazonian and Andean Regions Using Remote Sensing Data. *Journal of Biogeography, 35*, 1160-1176.

Busetto, L., Colombo, R., Migliavacca, M., Cremonese, E., Meroni, M., Galvagno, M., Rossini, M., Siniscalco, C., Morra Di Cella, U., Pari, E. (2010). Remote Sensing of Larch Phenological Cycle and Analysis of Relationships with Climate in the Alpine Region. *Global Change Biology, 16*, 2504-2517.

Carlson, C. H., Dobrowski, S. Z., & Safford, H. D. (2012). Variation in Tree Mortality and Regeneration Affect Forest Carbon Recovery Following Fuel Treatments and Wildfire in the Lake Tahoe Basin, California, USA. *Carbon Balance and Management, 7*, 7.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*, 37-46.

Dobrowski, S. Z., Abatzoglou, J., Swanson, A. K., Greenberg, J. A., Mynsberge, A. R., Holden, Z. A., & Schwartz, M. K. (2013). The climate velocity of the contiguous United States during the 20th century. *Global Change Biology, 19*, 241-51.

Eckenwalder, J. E. (2009). *Conifers of the world: The complete reference*. Portland: Timber Press.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Flint, A. L., & Childs, S. W. (1987). Calculation of Solar Radiation in Mountainous Terrain. *Agricultural and Forest Meteorology, 40*, 233-249.

Franklin, J. (1995). Predictive Vegetation Mapping: Geographic Modelling of Biospatial Patterns in Relation to Environmental Gradients. *Progress in Physical Geography, 19*, 474-499.

Franklin, J., & Miller, J. A. (2009). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge: Cambridge University Press.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing, 68*, 5-32.

Gesch, D.B. (2007). The National Elevation Dataset. In David Maune, (Ed.), *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, 2nd Edition, (pp. 99-118). Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Gould, W. (2000). Remote Sensing of Vegetation, Plant Species Richness, and Regional Biodiversity Hotspots. *Ecological Applications, 10*, 1861–1870

Gower, S. T., & Richards, J. H. (1990). Larches: Deciduous Conifers in an Evergreen World. *Bioscience, 40*, 818-826.

Higgins, S. S., Black, R. A., Radamaker, G. K., & Bidlake, W. R. (1987). Gas Exchange Characteristics and Water Relations of Larix Occidentalis. *Canadian Journal of Forest Research, 17*, 1364-1370.

Holden, Z. A., Landguth, E., & Purdy, J. (in review). Modeling the spread of mountain pine beetle with time-space dependent environmental predictions

Holden, Z. A., Morgan, P., & Evans, J. S. (2009). A Predictive Model of Burn Severity Based on 20-year Satellite-Inferred Burn Severity Data in a Large Southwestern US Wilderness Area. *Forest Ecology and Management, 258*, 2399-2406.

Hutchinson, M. F. (1993). On Thin Plate Splines and Kriging. In M.E.Tarter and M.D.Lock (Eds.) *Computing Science and Statistics*, (Vol. 25), (pp. 55-62). Berkeley, CA: Interface Foundation of North America.

Jensen, J. R. (2000). *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, N.J: Prentice Hall.

Krishnaswamy, J., Kiran, M. C., & Ganeshaiah, K. N. (2004). Tree model based eco-climatic vegetation classification and fuzzy mapping in diverse tropical deciduous ecosystems using multi-season NDVI. *International Journal of Remote Sensing, 25*, 1185-1205.

Larsen, J. A. (1916). Silvical notes on western larch. *Proceedings of the Society of American Foresters. 11*, 434-440.

Larsen, J. A. (1925). Natural reproduction after forest fires in northern Idaho. *Journal of*

Agricultural Research, 12, 1177-1197.

Little, E. L. (1971). *Atlas of United States trees: Volume 1*. Washington, D.C: U.S. Dept. of Agriculture, Forest Service.

Lomolino, M. V., Riddle, B. R., & Brown, J. H. (2006). *Biogeography*. Sunderland, Mass: Sinauer Associates.

Lutz, J. A., Wagtendonk, J. W., & Franklin, J. F. (May 01, 2010). Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. *Journal of Biogeography, 37,* 936-950.

Maignan, F., Bréon, F.-M., Bacour, C., Demarty, J., & Poirson, A. (2008). Interannual Vegetation Phenology Estimates from Global AVHRR Measurements: Comparison with In-Situ Data and Applications. *Remote Sensing of Environment, 112,* 496-505.

McClelland, B. R., & McClelland, P. T. (1999). Pileated Woodpecker Nest and Roost Trees in Montana: Links with Old-Growth and Forest "Health". *Wildlife Society Bulletin,* 846-857.

McCune, B., Grace, J. B., & Urban, D. L. (2002). Analysis of ecological communities. Gleneden Beach, OR: MjM Software Design.

Migliavacca, M., Cremonese, E., Colombo, R., Busetto, L., Galvagno, M., Ganis, L., Meroni, M., Pari, E., Rossini, M., Siniscalco, C., Morra di Cella, U. (2008). European Larch Phenology in the Alps: Can we grasp the Role of Ecological Factors by Combining Field Observations and Inverse Modelling? *International Journal of Biometeorology, 52,* 587-605.

Nagendra, H. (2001) Using remote sensing to assess biodiversity. *International Journal of Remote Sensing, 22,* 2377–2400.

Narayanaraj, G., Bolstad, P. V., Elliott, K. J., & Vose, J. M. (2010). Terrain and Landform Influence on Tsuga Canadensis (L.) Carrière (Eastern Hemlock) Distribution in the Southern Appalachian Mountains. *Castanea, 75,* 1-18.

New, D. M. (1999). *Productivity of Western Larch in Relation to Categorical Measures of Climate, Soil Moisture, and Soil Nutrients* (Master's thesis). Vancouver: University of British Columbia.

Oswald, B. P., & Neuenschwander, L. F. (1993). Microsite Variability and Safe Site Description for Western Larch Germination and Establishment. *Bulletin of the Torrey Botanical Club, 120,* 148-156.

Pandit, S. N., Hayward, A., Leeuw, J., & Kolasa, J. (2010). Does plot size affect the performance

of GIS-based species distribution models? *Journal of Geographical Systems, 12*, 389-407.

Parent, D. R., Mahoney, R. L., & Barkley C. Y. (2010). *Western Larch: A Deciduous Conifer in an Evergreen World*. Moscow, Idaho: Idaho Forest, Wildlife and Range Experiment Station.

Pfeffer, K., Pebesma, E. J., & Burrough, P. A. (2003). Mapping Alpine Vegetation Using Vegetation Observations and Topographic Attributes. *Landscape Ecology, 18*, 759-776.

Pike, R. J., & Wilson, S. E., (1971). Elevation Relief Ratio, Hypsometric Integral and Geomorphic Area Altitude Analysis. *Bulletin of the Geologic Society of America, 82*, 1079-1084

Powers, R. F., Adams, M. B., Joslin, J. D., & Fiske, J. N. (2005). Non-Boreal Coniferous Forests of North America. In Folke Anderson (Ed.), *Ecosystems of the World* (Vol. 6), (pp. 221-292). Amsterdam: Elsevier.

R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rehfeldt, G. E., & Jaquish, B. C. (2010). Ecological Impacts and Management Strategies for Western Larch in the Face of Climate-Change. *Mitigation and Adaptation Strategies for Global Change, 15*, 283-306.

Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Montoring vegetation systems in the Great Plains with ERTS. 3rd ERTS symposium (pp. 309–317). Washington DC: U.S. Government Printing Office (ed. SP-351).

Sanchez-Azofeifa, A., Rivard, B., Feng, J.-L., Chong, M. M., Wright, J., Li, P., & Bohlman, S. A. (2011). Estimation of the Distribution of Tabebuia Guayacan (Bignoniaceae) Using High-Resolution Remote Sensing Imagery. *Sensors, 11*, 3831-3851.

Schmidt, W. C., Shearer, R. C., & Roe, A. L. (1976). *Ecology and Silviculture of Western Larch Forests*. Washington: U.S. Dept. of Agriculture, Forest Service

Schmidt, W.C. & Shearer, R.C. (1991). Larix Occidentalis Nutt. In R. M. Burns & B. H. Honkala (Ed.), *Silvics of North America*, (pp. 160-170). Washington: U.S. Dept. of Agriculture, Forest Service.

Shearer, Raymond C. (1961). First-Year Mortality of Coniferous Seedlings in the Western Larch-Douglas-Fir Type. *Montana Academy of Sciences Proceedings 20*, 18-19.

Shuman, J. K., Shugart, H. H., & O'Halloran, T. L. (2011). Sensitivity of Siberian Larch Forests to Climate Change. *Global Change Biology, 17*, 2370-2384.

Stöckli, R., Rutishauser, T., Dragoni, D., O'Keefe, J., Thornton, P. E., Jolly, M., Lu, L., & Denning, A. S. (2008). Remote Sensing Data Assimilation for a Prognostic Phenology Model. *Journal of Geophysical Research, 113*, 1-19.

Strahler, A. H., Logan, T. L., and Bryant, N. A. (1978), Improving Forest Cover Classification Accuracy from Landsat by Incorporating Topographic Information, *Proceedings of the Twelfth International Symposium on Remote Sensing of Environment*, ERIM, Ann Arbor, MI, 927-942

Wood, S. N. (2008). Fast Stable Direct Fitting and Smoothness Selection for Generalized Additive Models. *Journal of the Royal Statistical Society, 70*, 495-518.

Wood, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society, 73*, 3-36

Zimmermann, N. E., Edwards, T. C., Moisen, G. G., Frescino, T. S., & Blackard, J. A. (2007). Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, *44*, 1057-1067.