# TREE COVER MAPPING FOR ASSESSING GREATER SAGE-GROUSE HABITAT IN EASTERN OREGON

Eric Nielsen[1] and Matt Noone[2]

[1]Institute for Natural Resources, Portland State University, Portland, OR
[2]Public Service Commission of Wisconsin, Madison, WI

February 2014

## 1. PROJECT OVERVIEW

We used a predictive model to map canopy cover of vegetation over seven feet in height ("tall woody vegetation") at 30-meter resolution over nearly 29 million acres within and adjacent to the range of the greater sage-grouse in Oregon (Figure 1). Texture measures computed at various resolutions from color-infrared aerial photography provided the main source of predictor data used to produce the map. Canopy cover was treated as a categorical variable using six cover classes: absent (cover class C0), present at less than 4% (C1), 4 – 10% (C2), 10 – 20% (C3), 20 – 50% (C4), and 50% and over (C5). The map is referenced to conditions in the years 2011 and 2012.

Although the specific target of the mapping was western juniper (*Juniperus occidentalis*), our reference data did not permit separating juniper from other tall woody vegetation during the predictive modeling process. The majority of the tall woody vegetation within the project area is western juniper. However, in high elevation regions, riparian, wetland, and residential areas, other vegetation is occasionally represented.

The methodology discussed here produces raw modeled data. It is recommended that prior to use in most applications this raw tree cover product be additionally filtered or masked to eliminate false detections which often occur adjacent to agricultural areas and roads. For species-specific applications, an additional modeling phase is necessary to either eliminate tree cover detections likely to be species other than the target, or to model species importance values associated with each tree occurrence.

## 2. METHODS

### 2.1. Predictor data

Uncompressed color-infrared aerial photography at 1-meter resolution provided by the National Agricultural Imagery Program (NAIP) and Landsat TM satellite imagery at 30-meter resolution were the main sources of predictor data. Although topographic metrics were generated at 10-meter resolution from the National Elevation Dataset (NED), these were not used for prediction because the reference data did not allow training across the full range of topographic conditions within the mapping area, which would result in spatial bias. In addition, excluding variables related to environmental setting allows the map to fully reflect existing rather than potential vegetation and forms a more suitable baseline for monitoring.

The red and near-infrared bands were extracted from the NAIP imagery, and the Normalized Difference Vegetation Index (NDVI, see Tucker 1979) was computed. The red band and NDVI were then used as complementary sources for creating a variety of texture metrics. Each was spatially degraded to several coarser resolutions (2-meter, 3-meter, 4-meter, 6-meter, 9-meter, and 12-meter) through aggregating by the median value of the corresponding 1-meter data. Two base texture metrics were then

created at each resolution: the 3x3-cell focal standard deviation and the absolute difference of each cell from the median of the values of its eight nearest neighbors.

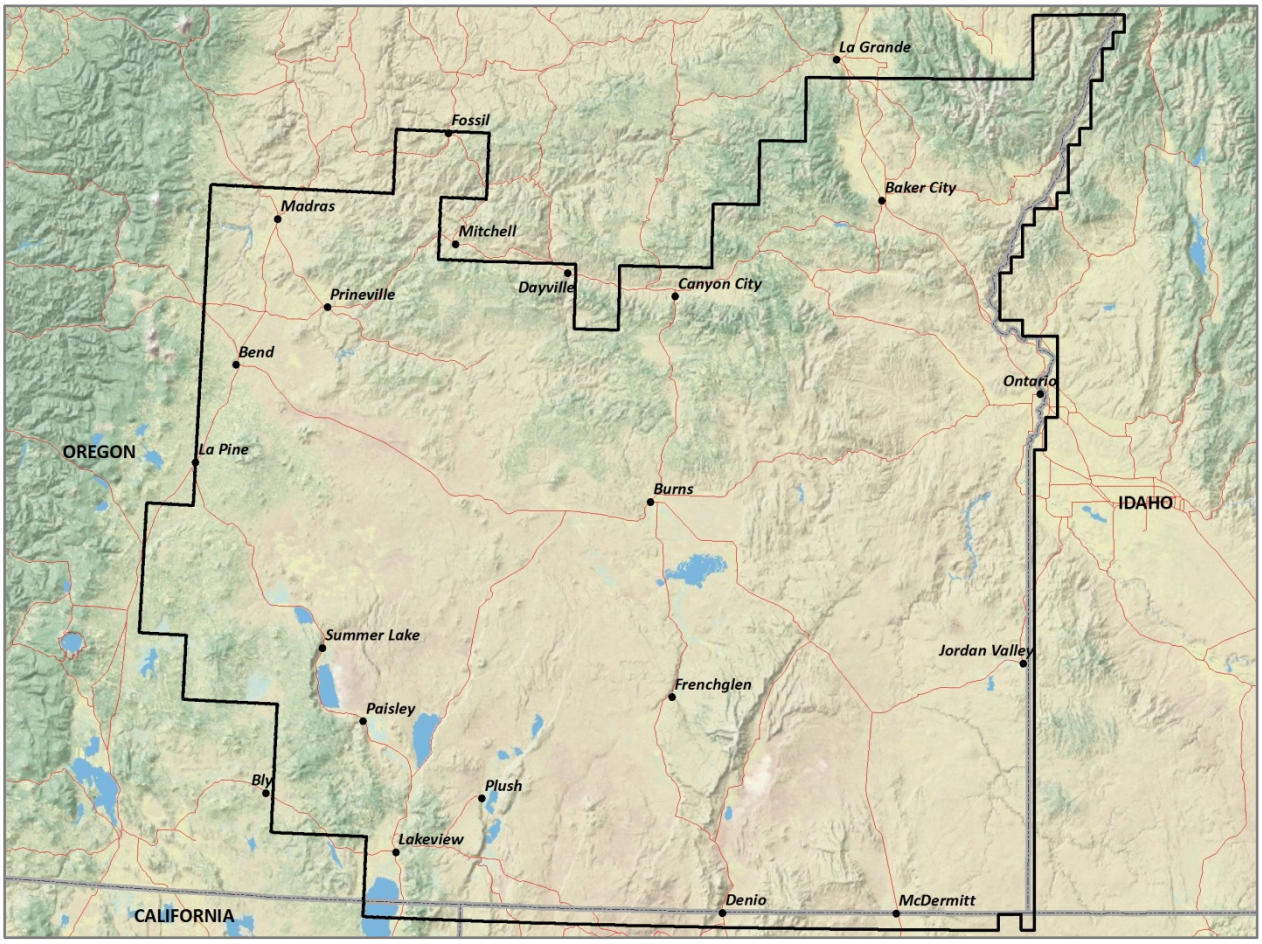


**Figure 1. Mapped area in eastern Oregon is within solid black line.**

Two combinations of the above metrics were created to attempt to compensate for contrast variability in the photography introduced by variations in sun-surface-sensor geometry, view angle, atmospheric conditions, and phenology. First, the Normalized Difference Texture Index (NDTI, introduced here) was created by pairwise combination of corresponding metrics at resolutions differing by a factor of three. The focal standard deviations at 1-meter and 3-meter resolution, 2- and 6-meter, 3- and 9-meter, and 4- and 12-meter resolution were combined analogously to the NDVI formula. Second, the focal standard deviations derived from the red band and the NDVI were combined at each resolution, in the same manner. Both of these combinations result in some cancellation of noise due to contrast variability.

All NAIP-based predictors were then degraded to 30-meter resolution, aggregating both by the median and the mean of the constituent finer resolution values. In addition, the maximum values of the near-infrared response and NDVI occurring in any 1-meter cell were extracted to the 30-meter grids. Finally, tiles from the western half of the project area were reprojected to UTM zone 11 (datum NAD83) via nearest-neighbor resampling, to match the projection of the tiles in the eastern half, and all tiles were mosaicked into a single image for each predictor.

To provide spectral information not available in the aerial photography, we downloaded cloud-free Landsat TM satellite data from the USGS EROS Data Center. The images, collected in late summer 2011, were reprojected via nearest-neighbor resampling to the common projection of UTM zone 11 (datum NAD83) and converted to exo-atmospheric reflectance using header file information. Adjacent paths were then radiometrically normalized via variance-preserving reduced major axis regression (see Cohen *et al.* 2003) before merging them across the project area. The three bands of the Tasseled Cap Transformation

(Crist and Cicone 1984) were calculated from the merged data using coefficients derived for exo-atmospheric reflectance (Huang *et al.* 2002). The resulting mosaicked images were inspected and bands 1 and 2 (representing reflectance in the blue and green wavelengths) were eliminated because of poor normalization across the project area. The TM data were kept at their native 30-meter resolution.

A total of 90 predictor data layers were produced, all but seven of them derived from the NAIP data. To reduce data storage and processing requirements, all predictors were converted to unsigned 8-bit values by stretching each across three standard deviations from its mean value.

## 2.2. Training reference data

LiDAR data were obtained at 1-meter resolution from a repository at Oregon State University for eight imaged areas within the project area (Figure 2). Vegetation height was calculated by subtracting the bare earth elevation from the highest hit elevation and the fraction of each 30-meter pixel with height over seven feet was determined. 40,000 randomly located points were generated within the LiDAR areas.

The reference datasets were concentrated in the northern portion of the project area from areas not representative of the range of conditions to the south. In particular, a variety of geological types, burned areas, steep canyons, extensive sagebrush stands, and wetlands modeled poorly in early runs due to their lack of representation in the reference data. To alleviate this problem, 39 areas representative of these types were hand-digitized based on aerial photography (Figure 2). All appeared in photography to contain no tall woody vegetation. Additional random points were generated within these areas, equal in sum to the total number of absence points from the LiDAR-derived data. An equal number of points was chosen from each digitized polygon, totaling an additional approximately 6700 samples of absence data.

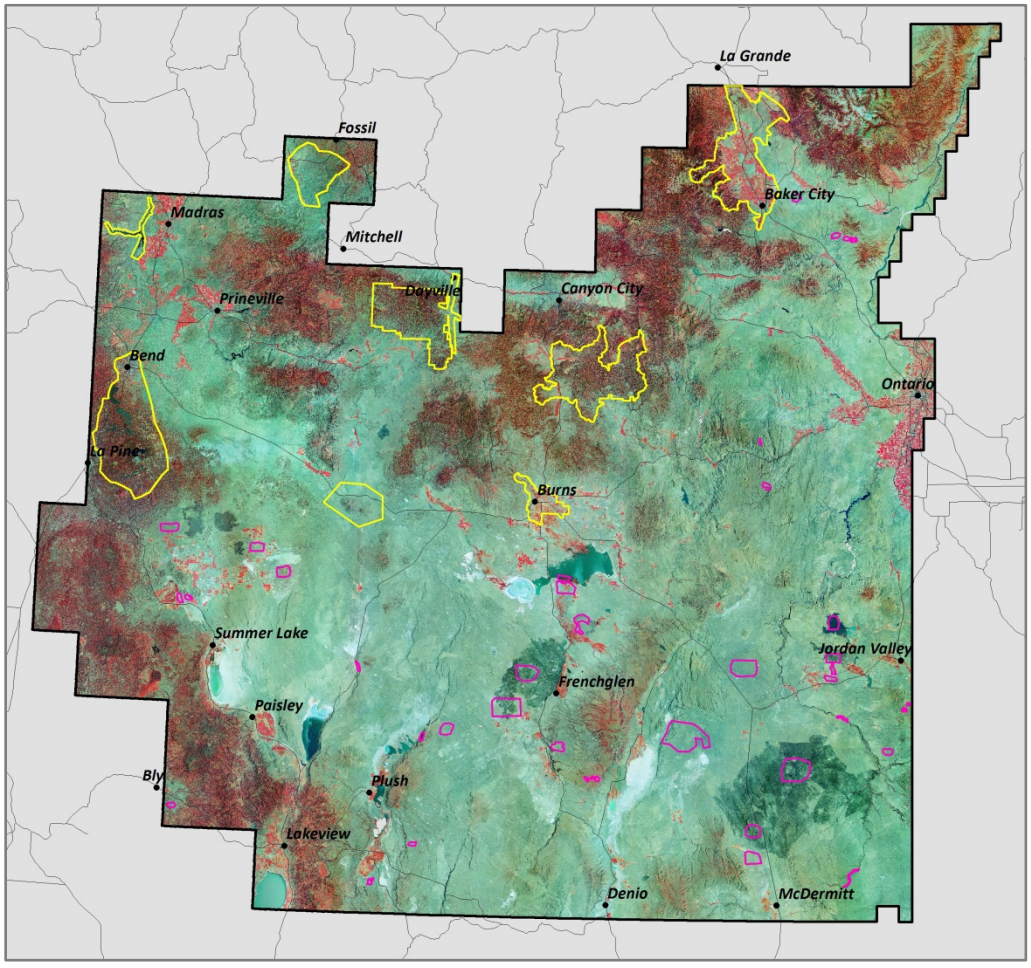


**Figure 2. Color-infrared 2012 NAIP mosaic of the project area, with LiDAR reference data areas outlined in yellow, and digitized absence polygons in pink.**

Several other data layers were needed for the model-building process. The mean of the 1-meter pixel height values over seven feet was determined for each 30-meter pixel, giving an indication of the dominant canopy height of the woody vegetation present at any training location. This information was used to reduce training on taller conifer forest types and focus model-building on stands of shorter trees such as juniper. Topographic slope derived from the NED was used to restrict analysis to areas of less extreme slopes; at higher slopes, artifacts in the LiDAR data become significant, making vegetation height data less reliable, especially for small trees and shorter vegetation. The most recent National Land Cover Dataset (NLCD) was also used to focus training on areas potentially containing western juniper.

At each sample point, the LiDAR-derived canopy cover and the value of each of the predictor grids were extracted. In addition, the mean tall woody vegetation height, the NED slope in degrees, and the NCLD land cover code were extracted for each point. Points with mean vegetation height over 50 feet, slope over 40 degrees, or NLCD codes corresponding to agriculture, wetlands, or developed areas were eliminated from the training data pool prior to modeling.

## 2.3. Modeling

A machine learning algorithm, Random Forests, was used for predictive modeling, running in the R statistical computing environment. The modeling process was broken into three stages in order to better represent distinctions among low canopy cover stands. In the first model stage, a binary split was made between samples with 4% or greater cover of tall woody vegetation (cover classes C2 – C5), and those with less than that amount (cover classes C0 and C1). This split was done first because 4% has been found to be a critical threshold for sage grouse reproduction. Samples with less than 4% cover were subjected to a second binary modeling stage which split samples where taller vegetation was totally absent (cover class C0) from samples with low cover (C1). A third modeling stage was performed on stands with 4% or greater cover, to divide them into four cover classes: 4 – 10% (C2), 10 – 20% (C3), 20 – 50% (C4), and 50% and over (C5). Random Forests regression was explored as an alternative to classification for prediction of a continuous cover amount, but model accuracies after binning to the above cover classes were significantly inferior.

In the first model stage, the training data in the four cover classes over 4% canopy cover were balanced between classes before modeling, by randomly downsampling until the classes were approximately equal. This was done in order to prevent the 4% and over class from being dominated by very high canopy cover stands which are often characterized by tree species other than juniper and for which detection in any event doesn't present a major challenge. In all model runs, each Random Forests tree was generated using equal sample sizes among the classes being modeled, in order to achieve a reasonable balance between omission and commission errors. The optimal set of predictor layers was determined for each model via a variable selection process that gradually reduced the number of predictor layers used to only those with the highest determined model importance values (see Evans and Cushman 2009). Overall error rates and maximum class error rates were compared for each reduced set of predictor layers. For the binary split model stages, the model with the lowest maximum class error rate was selected. For the 4-class model stage, the model with the lowest overall error rate was selected. The model generated during each stage consisted of 1000 trees.

An optimal probability threshold was determined for each of the binary split models, in order to equalize the relative frequency of false negative and false positive predictions in the training data. Using the optimal predictor layer set for each model stage, 5% of the training data was withheld and a model was generated from the remaining data. This model was then used to predict the class outcome probabilities for the withheld data. A confusion matrix was generated for each possible probability threshold value ranging from 0.01 to 0.99 in intervals of 0.01. This procedure was repeated twenty times and the cumulative confusion matrix was tallied for each threshold value. The interpolated threshold value resulting in an equal rate of false negative and false positive predictions was selected as the optimal threshold and was used during the subsequent prediction process, in addition to a simple probability threshold of 0.5.

### 2.4. Prediction

The models were applied across the project area, resulting in 30-meter resolution maps for each of the three model stages. Class probabilities were written out for the binary split stages, while only the most likely class was written out for the 4-class model. For a given pixel, the first model stage prediction determined whether the cover class prediction would be drawn from the second or third model stage. The cover class output contained six cover classes corresponding to the canopy cover of tall woody vegetation. Two versions of the map were made, one based on using a simple probability threshold of 0.5 for each of the binary split models, and one using the optimal probability thresholds determined above. Simple presence/absence maps were created from the cover class maps by combining all cover classes other than the absence class (cover classes C1 – C5) into a single class representing presence.

### 2.5. Focal and topographic filtering

Focal filtered versions of the maps were created in order to reduce the degree of pixelation and remove some false detections due to linear features. Two successive 3x3-cell majority filters were applied to the presence/absence map for pixels modeled with tree presence. Pixels modeled as absence were not filtered because false positives appeared to be much more prevalent than false negatives. For pixels predicted as presence in the filtered map, a filtered cover class was determined by recoding each of the nine nearest neighbors to the midpoint of their predicted cover class, averaging them, and then reclassifying the averaged value using the original cover classes. Therefore, a total of four alternate map outputs were created, a smoothed and an unsmoothed version of each of the probability threshold alternatives. Finally, outputs in areas exceeding 40 degrees slope were masked since predictive models could not be parameterized under these conditions due to artifacts in the LiDAR reference data.

### 2.6. Map accuracy assessment

A map accuracy assessment was performed on the unfiltered, non-equalized error rate map, using independent reference data obtained for several parts of the project area. The accuracy assessment is documented separately in Nielsen *et al.* (2014).

## 3. RESULTS

### 3.1. Tree Cover

The smoothed map outputs for presence/absence and cover class derived from the non-equalized error rate model are shown in Figures 3 and 4.

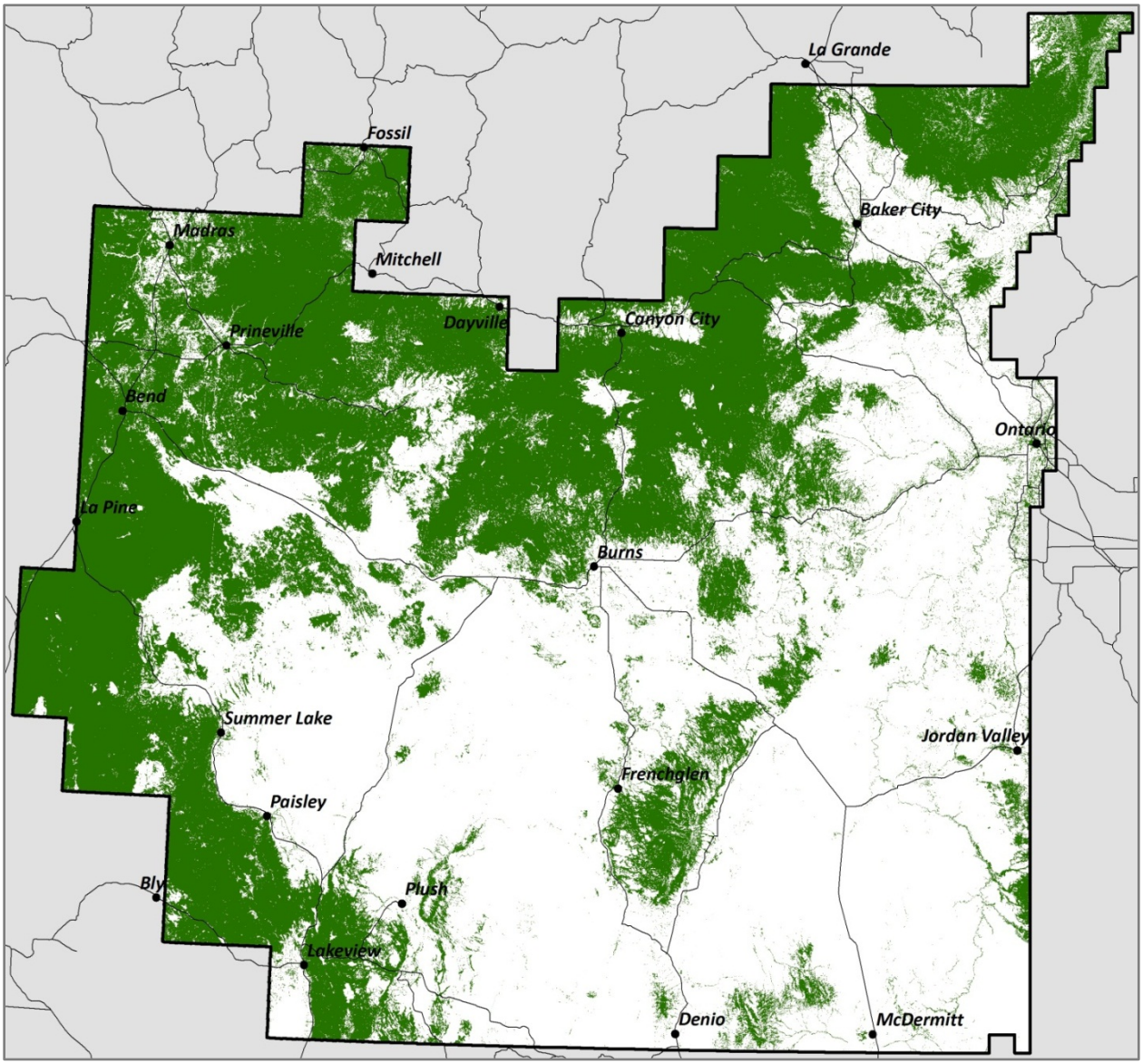


**Figure 3. Presence (green) and absence (white) of tall woody vegetation.**

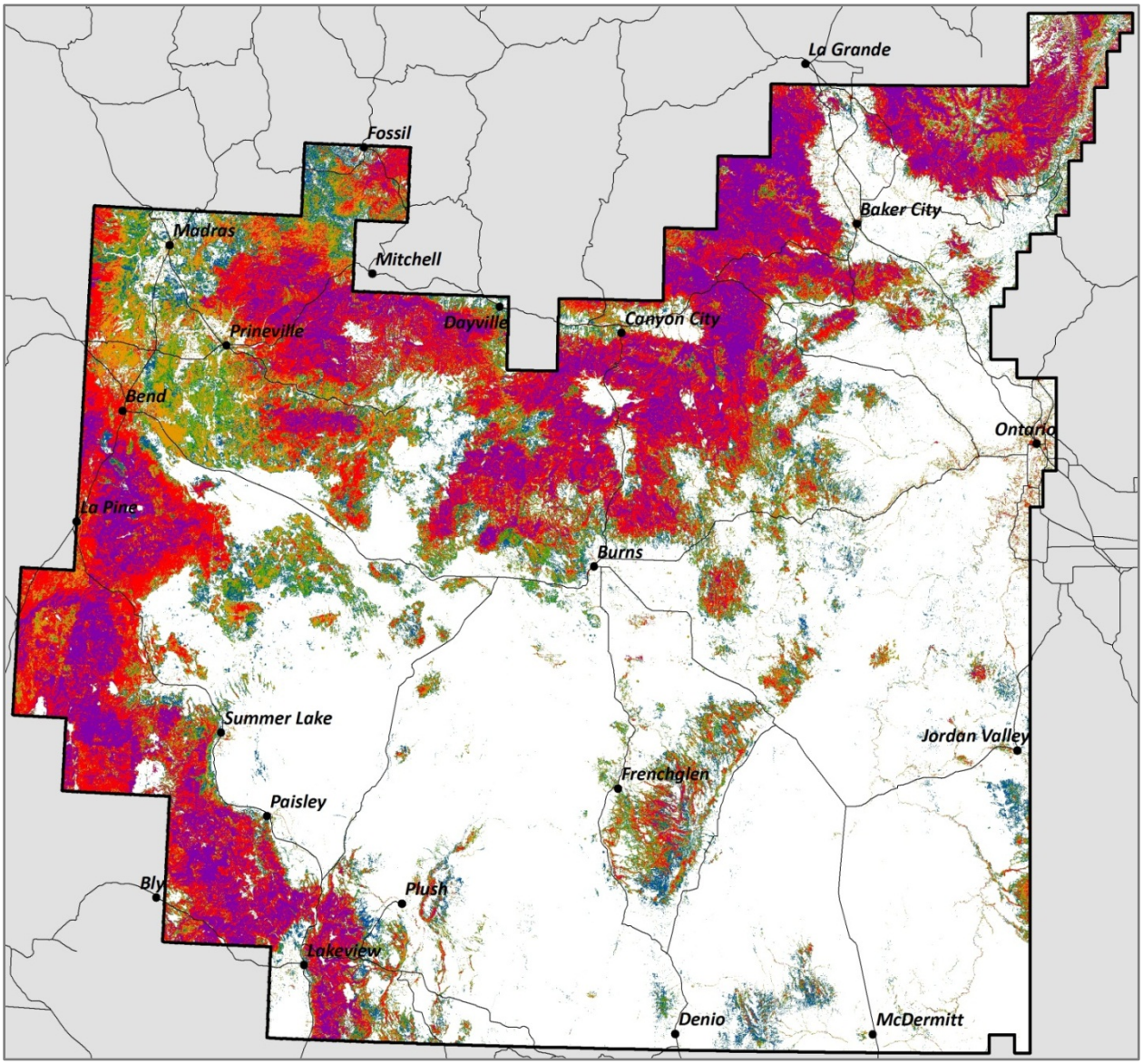


**Figure 4. Modeled cover class of tall woody vegetation. Classes shown are absent (C0: white), present at less than 4% (C1: blue), 4 – 10% (C2: green), 10 – 20% (C3: orange), 20 – 50% (C4: red), and 50% and over (C5: magenta).**

## 3.2. Estimated model accuracy of products

The model-based accuracies of each of the three modeling stages were determined separately. The two binary split models were assessed by both Random Forests out-of-bag error estimates and bootstrapped estimates (20 repetitions with 5% of the training data withheld each time) during the threshold optimization procedure. Only out-of-bag error estimates were used to assess model accuracy of the four-class model. The results are shown in Tables 1 – 3. Without threshold optimization to equalize false positive and false negative error rates, false negatives were significantly more common in both binary model stages. However, since these error rates were assessed against the training data, which is not representative of the project area as a whole, this does not necessarily imply that the same pattern would hold true against independent test data. The preferred version of the map was instead determined through inspection of the alternate versions.

Random Forests Out-of-Bag Class Error Estimates

| | PREDICTED CLASS | | |
|---|---|---|---|
| REFERENCE CLASS | 0-4% | 4%+ | Error |
| 0-4% | 14556 | 983 | 6.33% |
| 4%+ | 683 | 6455 | 9.57% |

Bootstrapped Class Error Estimates for Equal Error Rates

| | PREDICTED CLASS | | |
|---|---|---|---|
| REFERENCE CLASS | 0-4% | 4%+ | Error |
| 0-4% | 14160 | 1271 | 8.24% |
| 4%+ | 610 | 6792 | 8.24% |

**Table 1. Out-of-bag and bootstrapped class error estimates for first binary split model (stage 1).**

Random Forests Out-of-Bag Class Error Estimates

| | PREDICTED CLASS | | |
|---|---|---|---|
| REFERENCE CLASS | abs | pres < 4% | Error |
| abs | 11493 | 1520 | 11.68% |
| pres < 4% | 589 | 1937 | 23.32% |

Bootstrapped Class Error Estimates for Equal Error Rates

| | PREDICTED CLASS | | |
|---|---|---|---|
| REFERENCE CLASS | abs | pres < 4% | Error |
| abs | 10819 | 2278 | 17.39% |
| pres < 4% | 436 | 2070 | 17.40% |

**Table 2. Out-of-bag and bootstrapped class error estimates for second binary split model (stage 2).**

| | PREDICTED CLASS | | | | | |
|---|---|---|---|---|---|---|
| REFERENCE CLASS | C2 | C3 | C4 | C5 | Error | Fuzzy Error |
| C2 (4-10%) | 1261 | 452 | 81 | 21 | 30.52% | 5.62% |
| C3 (10-20%) | 470 | 949 | 316 | 49 | 46.80% | 2.75% |
| C4 (20-50%) | 36 | 306 | 1092 | 398 | 40.39% | 1.97% |
| C5 (50%+) | 6 | 26 | 343 | 1377 | 21.40% | 1.83% |

**Table 3. Out-of-bag class error estimates for four-class cover model (stage 3). Fuzzy error estimates give the rate of errors of more than one class from the correct answer.**

### 3.3. Relative importance of predictors

Predictor importance in each model was assessed using the mean decrease in accuracy statistic reported by Random Forests. This measures the decrease in model accuracy observed during Random Forests trees which exclude a certain predictor.

The most important predictors for the first model stage which separated cover less than 4% from cover greater than that were textures at various resolutions computed using the NDVI based on NAIP imagery, and the maximum value of the 1-meter NDVI within each 30-meter cell. Textures based on the red band were only slightly less significant than those based on the NDVI. Textures at 3- and 4-meter resolution were the most important for both the NDVI and the red band. In general, the mean aggregation method was more effective than the median. The combination metrics were useful, particularly the NDTI at the finest resolution. Landsat TM predictors were somewhat less helpful; the most important was the Tasseled Cap brightness.

The maximum value of the 1-meter NDVI was by far the most important predictor in the second model stage. The red/NDVI texture difference combination metrics at 3-meter and 1-meter resolutions were the next most important predictors. In general, NDVI-based textures were very important in this model, but at finer resolutions than during the first stage model. Again, red-based textures were slightly less significant. Landsat TM predictors were more important in this model, especially the Tasseled Cap greenness and near-infrared reflectance.

The third model stage, the 4-class cover model, was dominated by very different predictors. The most important texture predictor was the 1-meter red band texture, and the Landsat TM Tasseled Cap brightness and red band reflectance were among the most important predictors. NDVI-based texture predictors at intermediate resolutions remained important, and the TM mid-infrared bands were also effective. All the combination texture metrics performed fairly poorly here.

### 3.4. Map accuracy assessment

Results of the independent map accuracy assessment are documented in Nielsen *et al.* (2014).

# 4. POTENTIAL FOR MAP IMPROVEMENT

The best way to improve the current map would be incorporating reference data from a more representative portion of the project area. In addition to refining the model's ability to distinguish between juniper and other land cover types with similar appearance, sufficient training data collected across the project area would allow model optimization based on realistic estimates of map class proportions within the project area. They would also allow use of topographic metrics as predictors, if desired, because positive and negative occurrences could be sampled across the range of predictor values.

Some of this reference data would require field collection, because while junipers can often be identified with near certainty through photointerpretation, it can be difficult to establish whether trees visible in aerial photography meet the height criteria defining the map target. If a goal is to map early stages of juniper invasion, with trees less than seven feet tall, field-derived data is absolutely essential since it is often impossible to distinguish smaller juniper trees from other woody vegetation. It would be helpful if sampling locations were selected with reference to the current map, in order to provide training data in borderline situations.

Achievement of significantly higher accuracy levels at the lowest cover classes could likely be realized by implementing an individual tree detection approach developed at INR. The method is a variant of spatial wavelet analysis (Falkowski *et al.* 2006) but with exponentially decreased processing time because key calculations are performed at reduced resolutions.

Modest improvements in map accuracy in the lowest cover classes might be realized by switching to a finer mapping resolution of 10 meters. Since individual trees would occupy a larger portion of these smaller pixels, accuracies at low cover amounts might be improved. However, increased edge effects due to image misregistration errors between the training and predictor datasets, and shading and radial distortion artifacts from adjacent pixels might cause a practical reduction in accuracy. At any rate, the prediction process would be much slower at 10-meter resolution, and could be more complicated to implement, as memory limitations might be reached without subdividing the mapping area.

Incremental improvements also might be accomplished by incorporating additional texture metrics, such as those based on the Gray Level Co-occurrence Matrix (Haralick *et al.* 1973) or other metrics with strong response to anisotropy. Incorporating satellite data derived from superior Landsat OLI imagery also might allow slight improvements. However, the benefits of these changes would likely not be realized in more than marginal amounts without first obtaining more representative training data.

## ACKNOWLEDGMENTS

Funding for this work was provided by The Nature Conservancy.

## REFERENCES

Cohen, W.B., Maiersperger, T.K., Gower, S.T., and Turner, D.P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment* 84: 561-571.

Crist, E.P., and Cicone, R.C. (1984). A physically-based transformation of Thematic Mapper data: the TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing* 22: 256-263.

Evans, J.S., and Cushman, S.A. (2009). Gradient modeling of conifer species using random forests. *Landscape Ecology* 24: 673-683.

Falkowski, M.J., Smith, A., Hudak, A.T., Gessler, P.E., Vierling, L.A., & Crookston, N.L. (2006). Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of LiDAR data. *Canadian Journal of Remote Sensing* 32: 153-161.

Haralick, R.M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics.* SMC-3: 610-621.

Huang, C., Wylie, B., Yang, L., Homer, C., and Zylstra, G. (2002). Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. *Report* (10 pp.). Sioux Falls, SD: USGS EROS Data Center.

Nielsen, E.M., Poznanovic, A.J., and Popper, K. (2014). Accuracy comparison of tree mapping methods in eastern Oregon. *Report* (12 pp.). Portland, OR: Portland State University and The Nature Conservancy. February 2014.

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8: 127-150.