Estimating abiotic thresholds for sagebrush condition class in the western U.S.

Stephen P. Boyte [a*], Bruce K. Wylie [b], Yingxin Gu [c1], Donald J. Major [d]

[a] Senior Scientist, KBR, Contractor to the U.S. Geological Survey Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD, 57198, USA

[b] Research Physical Scientist, U.S. Geological Survey EROS Center, Sioux Falls, SD 57198, USA

[c] Scientist, InuTeq, Contractor to U.S. Geological Survey EROS Center, Sioux Falls, SD, 57198, USA

[d] Fire and Landscape Ecologist, Bureau of Land Management, Boise, ID 83709, USA

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

*Correspondence: Stephen P. Boyte, USGS, 47914 252nd St, Sioux Falls, SD 57198, USA Tel.: +1 605 594 6171. E-mail address: sboyte@contractor.usgs.gov (S.P. Boyte).

[1]Present address: I.M. Systems Group, Inc., at NOAA NESDIS Center for Satellite Applications and Research, College Park, MD 20740, USA

Abstract

Sagebrush ecosystems of the western U.S. can transition from extended periods of relatively stable conditions to rapid ecological change if acute disturbances occur. Areas dominated by native sagebrush can transition from species-rich native systems to altered states where non-native annual grasses dominate, if resistance to annual grasses is low. The non-native annual grasses provide relatively little value to wildlife, livestock, and humans and function as fuel that increases fire frequency. The more land area covered by annual grasses, the higher the potential for fire, thus reducing the potential for native vegetation to reestablish, even when applying restoration treatments. Mapping areas of stability and areas of change using machine-learning algorithms allows both the identification of dominant abiotic variables that drive ecosystem dynamics and the variables' important thresholds. We develop a decision-tree model with rulesets that estimate three classes of sagebrush condition [i.e. sagebrush recovery, tipping point (ecosystem degradation), and stable]. We find rulesets that primarily drive development of the sagebrush recovery class indicate areas of mid elevations (1 602 m), warm 30-yr July temperature maximums (tmax) (30.62 °C), and 30-yr March precipitation averages equal to 26.26 mm, about 10% of the 30-yr annual precipitation values. Tipping point and stable classes occur at elevations that are lower (1 505 m) and higher (1 939 m), respectively, more mesic during March and annually, and experience lower 30-yr July tmax averages. These defined variable averages can be used to understand current dynamics of sagebrush condition and to predict where future transitions may occur under novel conditions.

Keywords: climate, decision-tree model, machine learning, non-native annual grass, sagebrush, western U.S.

Introduction

Sagebrush (*Artemisia* spp.) ecosystems of the western U.S. are imperiled (Chambers and Wisdom 2009; U.S. Fish and Wildlife Service 2013). Threats to the ecosystems include wildfire, climate change, development, invasion of non-native annual grasses, and expansion of conifers (Chambers et al. 2017). The threats compromise the ecosystems' abilities to provide services like clean water and air, wildlife habitat, forage for grazing, recreational opportunities, and biodiversity (Rose et al. 2015). The amount of area sagebrush currently occupies is little more than half its historical range (Chambers et al. 2017; Davies and Bates 2019). Euro-American migration into the western U.S., and the accompanying increase in disturbances and invasion of non-native grasses, coincided with sagebrush range reduction (U.S. Fish and Wildlife Service 2013; Chambers et al. 2017). Disturbances (e.g. land-use change, fire, overgrazing), often multiple compounding disturbances, caused sagebrush ecosystems to transition from extended periods of relatively stable conditions where native shrub and perennial grass species dominated to ecologically degraded conditions where non-native annual grasses invaded and now dominate. To identify transitional locations, we defined criteria for three classes of sagebrush condition, developed a dataset that reflected the classes, integrated the dataset with relevant independent variables into a decision-tree model, and used the resulting model algorithms to develop spatially explicit maps of sagebrush condition class. For purposes of this study, we named the three classes sagebrush recovery, tipping point (representing ecological degradation), and stable.

Most restoration efforts in sagebrush ecosystems have been minimally effective (Blomberg et al. 2012; Svejcar et al. 2017). Therefore, recovery to a sagebrush-dominated system after a disturbance can be expensive and take many years, if recovery ever occurs (Svejcar et al. 2017), although recent studies have preliminarily shown enhanced success of sagebrush restoration (Davies et al. 2018; Germino et al. 2018; Davies and Bates 2019). Sagebrush ecosystems vary in their abilities to resist non-native annual grass invasion and recover from disturbance. Systems with low resistance and resilience were manifested in large geographical areas of non-native annual grass stands that have increased fire frequencies and threaten adjacent healthy rangeland systems. Chambers et al. (2007) found that specific factors influenced how vulnerable a sagebrush ecosystem was to cheatgrass (*Bromus tectorum* L) invasion, the most ubiquitous non-native annual grass in the study area. Climate, disturbance regime, the competitive abilities of the resident species, and traits of the invader were all influential factors of invasibility. Invasibility increased when resources were unused by native vegetation, such as after a fire (Rau et al. 2014; Roundy et al. 2018). Invasion also occurred when resource availability was inconsistent (Rau et al. 2014), which led to periods when resource supply exceeded the resident species' ability to utilize it while invasive species' propagule pressure existed (Davis et al. 2000). This phenomenon could have occurred in low to mid elevations of the sagebrush steppe (Chambers et al. 2014) where precipitation exhibited high temporal variability (Bradley and Mustard 2005). In areas of relatively high perennial vegetation productivity, greater resource utilization by perennial vegetation reduced ecosystems' invasibility (Chambers et al. 2014). How efficient an invading plant was at using resources when they become available could have determined a plant's invasion success (Bansal et al. 2014). Greater resilience levels have been positively associated with higher precipitation, greater soil resources, and higher plant productivity, and linked to higher levels of resistance (Chambers et al. 2014).

The goals of this study were to identify abiotic variables that most influenced the prediction of three classes of sagebrush condition and to establish the most common environmental thresholds that characterized each class. We clarified and defined some of the abiotic characteristics and associated thresholds that influenced sagebrush ecosystems' resilience to disturbance, invasibility to non-native annual grasses, and stability. We parameterized decision-tree software in two ways so that it generated 1) a predictive model based on a tree structure and 2) a descriptive model based on rulesets. These two models allowed us to achieve the following objectives:

1) Develop a spatially explicit predictive map of three classes of sagebrush condition.
2) Develop a spatially explicit ruleset map that shows where every ruleset occurred.
3) Identify abiotic variables that most strongly drive development of the sagebrush condition class model.
4) Establish thresholds of the most commonly used abiotic variables that delineate each sagebrush class.

Methods

Study area

We focused our study on arid and semiarid sagebrush ecosystems in the western U.S. where annual grass invasion was likely, sagebrush was native, and wildlife and livestock graze. The study excluded areas higher than 2 250 m elevation because cheatgrass was much less likely to invade at elevations above about 2 000 m in the northern Great Basin (Boyte et al. 2015) and because these areas were more resistant to cheatgrass invasion and more resilient to disturbances than areas at lower elevations (Chambers et al. 2014). Also excluded were areas where the 2011 National Land Cover Database (The National Land Cover Database 2011) classified a pixel as something other than shrub or herbaceous/grassland (Fig. 1). Excluded areas were masked and neither mapped nor included in processes and analyses.

The study area encompassed 286,574 km^2 including parts of 6 states. The 30-yr (1981 – 2010) climate averaged equal 337 mm of precipitation, a minimum temperature of 0° C, and a maximum temperature of 14° C (PRISM Climate Group). The elevation minimum equaled 568 m with a mean of 1 562 m when the 2 250 m limit was observed (North American Vertical Datum 88). The topography varied substantially with alternating mountains and valleys throughout much of the study area. The vegetation was composed of mostly shrubs, native grasses, and invasive grasses with the dominant shrubs being sagebrush (*Artemisia* spp.). Some of the common sagebrush species included Basin big sagebrush (*A. tridentata* Nutt. *ssp. tridentata*), Wyoming big sagebrush (*A.t.* Nutt. *ssp. wyomingensis* Beetle & Young), and low sagebrush (*A. arbuscula* Nutt.) Other common shrub species included shadscale saltbush (*Atriplex confertifolia* [Torr. & Frem.] S. Watson), rabbitbrush (*Chrysothamnus* spp.), greasewood (*Sarcobatus* spp. Nees), and some antelope bitterbrush (*Purshia tridentata* [Pursh] DC.). Grasses included Sandberg bluegrass (*Poa secunda*), basin wildrye (*Leymus cinereus* [Scribn. & Merr.] Ả. Löve), Idaho fescue (*Fescue idahoensis* Elmer), and the introduced crested wheatgrass (*Agropyron cristatum* L.). Several annual grasses have invaded the study area including cheatgrass, North African wire grass (*Ventenata dubia* [Leers] Coss.), medusahead (*Taniatherum* spp.), and, in the warmest and most arid places, red brome (*Bromus rubens* L.).

Datasets

We categorized the datasets used in this study into two types: 1) the sagebrush condition class reference dataset. This dataset was developed from four remotely sensed derived datasets, and parameters were defined by the sagebrush condition criteria. The four datasets were described in the next four subsections, and the sagebrush condition criteria were defined in the subsection “Sagebrush Condition Classes”; and 2) datasets used as drivers in the decision-tree model. The model drivers included an elevation dataset that was derived from The National Map (https://nationalmap.gov/elevation.html). The elevation dataset had a native 30-m spatial resolution, and we spatially averaged it using a 7x7 focal scan and resampled it to 250 m using nearest neighbor. Polaris soils data were 78-m native spatial resolution data. We spatially averaged those data using a 3x3 focal scan, and then resampled them to 250 m using nearest neighbor. The climate data, which included 30-yr annual, seasonal, and monthly precipitation and temperature minimums and maximums (PRISM Climate Group), were resampled from 800 m to 250 m using bilinear interpolation.

The four remotely sensed derived datasets described next used Moderate Resolution Imaging Spectroradiometer (MODIS) sensor’s Normalized Difference Vegetation Index (NDVI) data (Jenkerson et al. 2010) at 250-m spatial resolution. NDVI allows the monitoring of seasonal and interannual vegetation greenness and has been used to measure green biomass (Jensen 2005). We applied regression-tree optimization protocol (Gu et al. 2016; Wylie et al. 2018) to the development of all regression-tree models to minimize each models’ errors and overfitting tendencies.

Annual herbaceous percent cover

We developed and published a time series (2000 – 2016) of spatially explicit annual herbaceous percent cover estimates (Boyte and Wylie 2017) at 250-m spatial resolution throughout much of the western U.S. While the time series estimated annual herbaceous percent cover, the primary annual herbaceous type targeted was cheatgrass because of its pervasive dominance in sagebrush ecosystems. Cheatgrass was highly responsive to annual precipitation (Bradley and Mustard 2005; Pilliod et al. 2017), leading to temporal variation in annual herbaceous percent cover. Annual herbaceous percent cover demonstrated high spatial variation because of biophysical factors and a history of disturbances (Chambers et al. 2014). We developed the annual herbaceous estimates using a regression-tree model integrated with field-based, remotely sensed, climate, fire, and biophysical data (Boyte et al. 2019a). The model used more than 30,000 field-based training data points from 5 years (Boyte et al. 2019a) to develop spatially explicit estimates of annual herbaceous percent cover (see Table 1 for model accuracy assessments). We used 17 years of 7-day pixel composites of NDVI data to map the wide-ranging annual herbaceous percent cover dynamics inherent to the water-limited sagebrush ecosystem. Driving variables for the model included native 250-m remotely sensed derivatives – annual growing season NDVI, summer NDVI, and start of season time (a phenology measure). Additional variables consisted of a year-since-fire dataset, PRISM 30-yr precipitation (PRISM Climate Group), Polaris soils data – organic matter and available water capacity, and topographic data – elevation, a wetness index, and north- and south-facing slope indices. Aspects were defined by azimuths between 315° and 45° (north aspects) and 135° and 225° (south aspects) on slopes greater than 8.5% (where aspect angles were measured in a clockwise direction and north = 0°). Undefined aspects were assigned a null value.

Sagebrush percent cover

Aggregated 250-m enhanced MODIS NDVI from weeks 15 – 40 of each year of the study period (2001 – 2015) served as the growing season NDVI. Sagebrush percent cover was a derivative of growing season NDVI, calculated on a pixel-by-pixel basis using the following algorithm (Eq. 1) (Rigge et al. 2019).

$$[(0.4247 * Annual \text{ growing season NDVI}) - 43.839] \quad \textbf{[1]}$$

Sagebrush Ecosystem Performance Anomaly

In sagebrush ecosystems, annual grasses influenced the NDVI signal, which could have caused sagebrush percent cover to be overestimated, especially in recently disturbed areas. To mitigate this problem, we developed a regression-tree model that predicted sagebrush percent cover by incorporating seasonal and monthly weather data to separate effects of disturbances and land management from effects of weather (Wylie et al. 2012). The dataset was the sagebrush ecosystem performance anomaly, or sagebrush anomaly, and it introduced temporal variation throughout the study period.

We calculated the difference between the sagebrush percent cover and the predicted sagebrush percent cover datasets and used statistical confidence levels to define normal and abnormal ecosystem performance on an annual time step (Wylie et al. 2008; Gu and Wylie 2010; Wylie et al. 2012; Rigge et al. 2013). We labeled pixels of abnormal performance as overperformance and underperformance depending on whether a pixel fell above or below the 95% confidence level, respectively. Pixels with high normal or overperformance that also experienced high levels of annual grass percent cover ($\geq$ 10%) and low levels of tree cover ($\leq$ 15%) were pixels of likely sagebrush overestimation and not included as reference data.

Sagebrush site potential

Sagebrush site potential was developed as a measure of an ecosystem's inherent productivity (Wylie et al. 2008) and, in this study, represented the long-term average production of sagebrush biomass in a good sagebrush state, i.e. non-degraded or non-disturbed (Rigge et al. 2019). Site potential introduced spatial variability across the study area, whereas the sagebrush anomaly, sagebrush percent cover, and annual herbaceous percent cover datasets introduced temporal variability throughout the study period. We defined a pixel as in good sagebrush state if more than 40% of a pixel's vegetation cover was sagebrush, less than 10% of its absolute cover was annual herbaceous, and no fires burned in the pixel from 1993 – 2014. A final criterion required that more than 70% of the pixel be classified by the National Land Cover Database as shrub or herbaceous/grassland land cover.

Long-term (2000 – 2015), above-average growing season NDVI from MODIS at 250-m resolution served as a proxy for sagebrush site potential. The MODIS NDVI values ranged from -0.01 to 0.59 on a scale of -1 to 1. A random stratification of pixels that met the definition of a good sagebrush state were selected as training samples, with a low sagebrush productivity tier ranging from -0.01 to 0.14 (n = 3 543), a moderate tier ranging from 0.15 to 0.22 (n = 3 695), and a high tier ranging from 0.23 to 0.59 (n = 3 514). Site potential was modeled in regression-tree software, and model accuracy assessments are displayed in Table 1. The model was driven by Polaris soils data – organic matter and available water capacity – at 0 to 30 cm depth, a compound topographic index, steep north- and south-facing slopes, and Landsat-based NDVI. The Landsat-based NDVI was, on a pixel-by-pixel basis, the value equal to the NDVI's 90th percentile in the months of August and September from 1986 – 2013. Landsat NDVI values used in the study were unlikely to be affected by disturbances that occurred after 1986. Landsat NDVI data were native 30-m spatial resolution data that we spatially averaged using a 7x7 focal scan and then resampled to 250 m.

Sagebrush condition classes

We developed criteria that defined three classes of sagebrush condition – sagebrush recovery, tipping point, and stable. The criteria were limited by the study period (2000 – 2015) and determined using the four remotely sensed derived datasets just described. Any pixel that met the criteria for a specific class was a reference data candidate unless the pixel met the criteria for more than one class, then it was removed from the pool of reference data candidates. We classified a pixel as sagebrush recovery if it met two criteria: 1) for at least 3 years early (2001 – 2010) in the time series, that pixel had annual grass percent cover ≥ 10%, sagebrush site potential > 7%, sagebrush anomaly was high normal performance or above, and tree canopy ≤ 15%; and, 2) for at least 3 years late (2011 – 2015) in the time series, that same pixel had annual grass percent cover < 10%, sagebrush percent cover > 10%, and sagebrush anomaly was normal, i.e., it functioned as we expected sagebrush should. A total of 61 444 pixels, or 1.3% of unmasked pixels, met the sagebrush recovery criteria. Pixels that met the criteria for a specific class could be randomly selected as a dependent variable in the sagebrush condition class model.

We classified a pixel as tipping point if it met two criteria: 1) for 2 years early (2001 – 2005) in the time series, the pixel had annual grass percent cover < 10%, sagebrush percent cover > 7%, and normal sagebrush anomaly; and, 2) for at least 2 years late (2011 – 2015) in the time series, that same pixel had annual grass percent cover ≥ 10%, sagebrush site potential was at least moderate (> 7%), and sagebrush anomaly was at least high normal – high normal to overperformance sagebrush anomaly can indicate cheatgrass presence. A total of 101 663 pixels, or 2.2% of unmasked pixels, met the tipping point criteria.

We classified a pixel as stable if it met two criteria: 1) for 13 or more years during the study period, the pixel had annual grass percent cover < 10%, normal sagebrush performance, and annual grass percent cover no more than one standard deviation from the study period mean for that pixel; and, 2) for 13 or more years during the study period, that same pixel had sagebrush percent cover > 7% and sagebrush percent cover no more than one standard deviation from the study period mean for that pixel. A total of 11 029 pixels, or 0.24% of unmasked pixels, met the stable criteria.

When the decision-tree software classified a pixel using the tree structure, it established a confidence level for that prediction using a value between 0 and 1 (Quinlan 2013). The class confidence levels for the sagebrush class condition model ranged from 0.33 to 1.00. We defined a class confidence level equal to or greater than 0.70 as high probability, a class confidence level equal to or greater than 0.50 and less than 0.70 as moderate probability, and a class confidence level equal to or greater than 0.33 and less than 0.50 as low probability.

Model development

The decision-tree model used, See5 (https://www.rulequest.com/), was designed to obtain information from databases of primary (dependent variable) and ancillary (independent variables) data and then construct diagnostic rules based on that information to predict discrete classes (Quinlan 2013). We developed a database with a randomly stratified sample of pixels (n = 16 550), of which 15% met the criteria for stable pixels, 32% for sagebrush recovery pixels, and 53% for tipping point pixels. We used a sample of possible pixels because pixels that met the criteria for the stable class equaled only about 6% of all possible reference data pixels. To avoid underrepresenting the stable class in our model, we set a minimum requirement of 2 500 pixels from each class. To avoid both spatial autocorrelation issues and overrepresenting the tipping point class, we limited the number of sample pixels from that class. We developed a test dataset of 2 500 pixels that was a random subset of all potential dependent variable pixels not used for model training. The test dataset provided an independent validation of model accuracy. We also calculated a ten-fold cross validation of the training data, which provided a second independent validation of model accuracy.

We developed two classification models (Fig. 2): one to predict classes and a second to describe rulesets that defined how independent variables were used. The predictive model used a tree structure and ran with no user-defined constraints to its number of nodes, allowing the modeling software to use an automated protocol to prune the tree. A mapping software, MapSee5, applied the See5 classifiers to the independent variables' imagery to generate a sagebrush condition class prediction map (Boyte et al 2019b) . The descriptive model utilized a user-defined number of rulesets. Limiting the number of rulesets greatly enhanced the user's ability to interpret the associated thresholds of independent variables but sacrificed some accuracy. A low number of rulesets facilitated the ecological interpretation of the rulesets, which increased the understanding of the descriptive model (Quinlan 2013). We included identical drivers for both models, and they were 30-yr climate, elevation, and soils data. The 30-yr climate data included average annual values, average values of selected months, and average seasonal values for spring (March – May) and summer (June – August). Precipitation (ppt), temperature maximums (tmax), and temperature minimums (tmin) were the 30-yr climate variables used. The soils data included Polaris soil organic matter and available water capacity (Table 2). Of the 30 variables integrated into the sagebrush condition models, 6 were identified as being most impactful to the development of both the predictive and the descriptive models. These six variables were analyzed and used to delineate thresholds that characterized sagebrush condition.

We also analyzed the output of the descriptive model that defined the rulesets, identified confidence levels for rulesets and their associated classes, and calculated the geographical space each ruleset covered. The rulesets were input into a conditional-statement model that mapped each ruleset in its spatial context based on stratification thresholds from the independent variables (Fig. 3). The descriptive model incorporated 21 rules using 15 independent variables. Nine rulesets defined the variable conditions that made up the tipping point class, and the ruleset from that class with the highest prediction confidence (0.839) is shown below. This ruleset constituted about 2.39% of all training points, 1 390 km^2 of the study area, and defined a tipping point as:

Elevation >1602 m AND ≤1798 m
April ppt > 19 mm
Summer ppt ≤ 26 mm
June tmin ≤ 26 °C
Summer tmin ≤35 °C.

Validation

We validated the sagebrush condition class map using two datasets and methods. First, we acquired from Bureau of Land Management (BLM) staff 2013 – 2016 BLM Assessment Inventory and Monitoring (AIM) plot-level data that were collected from the field, typically using the line-point intercept method at three transects of 30 m each within a spoke design (Herrick et al. 2017). Field data that can be used to validate remotely sensed data were difficult and expensive to gather and process and were therefore relatively scarce (Browning et al. 2015; Bradley et al. 2018). Field data that matched the spatial resolution of a 250-m eMODIS NDVI pixel were even more uncommon. Once obtained, field data can be hard to apply to remotely sensed data (Bradley et al. 2018), especially remotely sensed data with coarser spatial resolutions. The AIM data were ubiquitous for our study area and period, and for our purposes, provided an acceptable measure of validation.

We examined the AIM plot-level data for percent cover of sagebrush and non-native annual grasses and compared those data to corresponding pixels and their predicted classes. If the sagebrush and invasive annual grass percent cover values for an AIM plot matched a specific class' criteria, and that class corresponded to the class represented on the predictive map, then that data point was considered correct. If the non-native annual grass percent cover for any AIM field plot was equal to or greater than 15%, then the pixel had to be classified as tipping point to be correct, regardless of the percent cover of sagebrush because areas with this level of cheatgrass were more likely to have burned (Bradley et al. 2018) and therefore, for this study's purpose, were identified as a tipping point. Finally, if sagebrush and invasive annual grass percent cover values of a plot did not meet the criteria for any class, then the data point was removed from the analysis because the data point may have been associated with factors other than those that defined the sagebrush condition classes.

Second, we overlaid Monitoring Trends in Burn Severity (MTBS) polygons from 2001 – 2015 on the sagebrush condition class map (Monitoring Trends in Burn Severity 2018). MTBS polygons in the western United States represented wildland fires in almost all circumstances because of the relative infrequency of prescribed fires in the West and the minimum 1 000-acre fire-size threshold used to establish the polygon extents (Joshua Picotte, Fire Specialist, MTBS Science Support, written communication, 25 September 2019). We calculated the percent of total area burned by class for each year. While we display all study period years in Figure 4, we focus the validation on the last 5 years of the study period because of the time dimension used in our criteria to define classes. This time dimension focused the transition from a tipping point to a sagebrush recovery class and the transition from a stable sagebrush ecosystem state to a tipping point class on the last 5 years of the time series. Given this time dimension, the last 5 years of the time series were most critical in assessing the accuracy of these datasets. The MTBS polygons were used to help define sagebrush site potential, but the effect on the sagebrush condition classes would likely be relatively minimal as only two of its criteria used site potential in their definitions.

Results

Model development

The predictive and descriptive models used independent variables to drive model development, and although the same variables were available to both models, the predictive model used all available variables whereas the descriptive model used only some (Table 2). The two models also used variables at different frequencies. These phenomena primarily occurred because we developed the predictive model by applying few constraints to the decision trees that predicted classes, and we developed the descriptive model by severely limiting the number of rulesets that described how independent variables were associated with each class.

Model variable usage was presented as a percentage and calculated by the decision-tree software based on if the value of the variable was known and if the variable was used in the prediction of a class (Quinlan 2013). Multiple variables' model usage frequencies equal 100% in the predictive model, including elevation and 30-yr precipitation. These two variables were most frequently used in the descriptive model at 75% and 73%, respectively. The frequent use of these variables in both models indicated their heavy influence in classifying sagebrush condition. Other variables of importance for both the predictive and the descriptive models were available water capacity, 30-yr July temperature maximum, and 30-yr March temperature minimum. July weather variables, collectively, influenced the predictive model more than any other month's weather variables, although March weather variables were almost as influential. March and July weather variables were the weather variables used most frequently in the descriptive model, although their frequency of use was much less than in the predictive model and much less than elevation and 30-yr precipitation. May was the month with weather variables least used in the predictive model. Seasonal and other monthly weather averages were relatively influential in the predictive model, but many were used sparingly or not at all in the descriptive model. Summer temperature maximum was not used in the descriptive model and was used the least in the predictive model, indicating little influence on sagebrush condition class.

General characteristics of dominant variables

To establish the general characteristics of the dominant variables associated with each class – sagebrush recovery, tipping point, stable – we calculated a weighted average based on the percentage of study area covered by each ruleset and connected the averages to the most frequently used variables in the descriptive model (Table 3). We also analyzed the dominant variable thresholds defined by the rules in the descriptive model. Elevation and 30-yr average precipitation overwhelmingly drove model development, and distinct differences existed for elevation and 30-yr precipitation averages between sagebrush condition classes. The stable class was substantially more mesic and at higher elevations than either the sagebrush recovery or tipping point classes. Data also indicated that the stable class was cooler during March and July. The sagebrush recovery class occurred at elevations slightly higher than the tipping point class, but the tipping point class was more mesic, and, during mid-summer, cooler. The values for available water capacity varied little between classes, but soil organic matter increased from sagebrush recovery (38.85 kg • m^{-2}) to tipping point (79.92 kg • m^{-2}) to stable (108.10 kg • m^{-2}) classes. The tipping point class covered a substantial majority of the study area, so we expected, on average, this class to be the most variable.

Characteristic thresholds of dominant variables

Table 4 shows the variables that were the most impactful drivers for both the predictive and descriptive models and connects those drivers to the rulesets they influenced. Less commonly used drivers did influence the rulesets, but there were too many (30) to display concisely and cogently, and their impact on the development of the descriptive model was relatively minimal. Several elevation values served as thresholds to develop the descriptive model, but one value, 1 602 m, was present in 71% of all rulesets, including 6 of the 7 rulesets for the sagebrush recovery class. The tipping point class used the 1 602 m elevation threshold in 6 of its 9 rulesets, and for this class, at least 35% of the overall study area was defined by rulesets at or below 1 602 m with 22% at or below 1 339 m elevation. Less than 4% of the study area in this class was defined by rulesets higher than 1 602 m. All rulesets in the stable class were defined by elevations above 1 602 m with 3 of its 5 rulesets defined by > 1 602 m threshold. The weighted average elevation for the stable class was more than 338 m higher than either of the other two classes. The 30-yr precipitation variable threshold that predominated in the descriptive model's development was 338 mm, a number almost identical to the overall study area's 30-yr precipitation average (337 mm). This threshold value was present in 48% of all rulesets including all five of the stable class' rulesets, a class that was mostly present at higher elevations. Rulesets 14 (tipping point class) and 18 (stable class) showed 30-yr precipitation threshold values at > 338 mm, and these two rulesets' average elevations were 1760 m and 2129 m, respectively, substantially higher than their class means. Generally, rulesets that primarily drove development of the sagebrush recovery class indicated areas of mid elevations, warm July temperature maximums, and 30-yr March precipitation averages about 10% of the 30-yr average precipitation values. For the tipping point class, rulesets occurred at the lowest average elevations of the three classes, although the elevation range was widest. Annual precipitation totals were about halfway in between the other two classes, and March temperature minimums were lower than freezing (0 °C). Variables that drive development of the stable class indicated space that was at considerably higher elevations, received substantially more average annual precipitation, and had considerably lower average March temperature minimums than the other two classes.

Spatial representation of the descriptive model

Figure 3 displays the spatial arrangement of the rulesets defined by the descriptive model. Multiple rulesets can apply to an individual pixel, but we programmed the conditional-statement model so that a ruleset with the highest accuracy spatially defined a pixel, superseding rulesets with lower accuracies. The rules from the tipping point class (rulesets 8 – 16) cover 70.36% of the study area including highly populated areas in Idaho's Snake River Plain where elevations were lower than average, fires were relatively common, and ecosystem transformations were often driven by human activities. Northeastern and north-central Nevada were predominantly covered by the sagebrush recovery class (rulesets 1 – 7; 21.29% of study area), specifically rulesets 1 and 7. The stable class (rulesets 17 – 21; 8.35% of study area) mainly occupied areas of transitional and higher elevations. The average elevation of a stable pixel equaled 1939 m, substantially higher than either of the two other classes (Table 3). Ruleset 1 transitioned into and intermixed with ruleset 10 in north-central and northwest Nevada. Overall, ruleset 1 territory occupied a weighted-average elevation of 1442 m, whereas rule 10 occupied a weighted-average elevation of 1168 m and dominated the low elevation extents of the study area. In the geographic area of northwest Nevada where a basin and range topography of alternating elevations exists, rulesets transitioned relatively frequently with elevation changes. The sagebrush recovery and stable classes encompassed much of the northwesternmost corner of Nevada, coinciding with higher elevations. The stable class occupied the highest elevations in this geographic area and in northeast Nevada and in the northeast corner of the study area. With few exceptions, southeastern Oregon was covered by tipping points rules and mostly occupied low and relatively low elevations. The 30-yr annual precipitation averaged in this area experienced substantial variability with values ranging from 274 mm to 474 mm.

Spatial representation of the predictive model

The predictive model's spatial output (Fig. 5) displayed similar spatial patterns to the descriptive model's patterns of sagebrush recovery, tipping points, and stable classes. Differences exist between the two maps because, as discussed above, the two models were developed with different goals. One area where the two maps deviate was in Idaho on the eastern edge of the Snake River Plain and near the Oregon/Idaho border on the western edge of the Snake River Plain where the predictive model displays sagebrush recovery and the descriptive model displays tipping points. Sagebrush recovery pixels constituted 28.77% of all unmasked pixels, and they were located primarily in north-central and northeast Nevada with some scattered in and around the periphery of the Snake River Plain. Small pockets of recovery existed in topographically diverse areas just west of the Snake River Plain in southeast Oregon. A considerable section of the rest of southeast Oregon was modeled as sagebrush recovery. The majority (57.36%) of the predictive map was classified as tipping point, while much of the northern tier met the classification of high probability. A relatively sizeable section of the study area along the California / Nevada border was also classified as high probability of tipping point. The stable class covered 13.87% of the study area, mostly at higher elevations stretching along and across Nevada's northern border. In the northeast corner of the study area, the stable class frequently intermixed with both sagebrush recovery and tipping point classes.

Validation

Despite the spatial incongruence between the 250-m datasets used in this study and the AIM data (Boyte et al. 2019a), the overall agreement between the three classes and the field-based AIM data was relatively strong (64.09%) (Table 5). Each year experienced similar overall accuracy, ranging from 62.37% in 2013 to 68.54% in 2014. Accuracies deviated considerably when analyzed based on class. The tipping point class' overall accuracy equaled 73.16% with its highest accuracy during 2015 at 83.26%. This class covered most of the study area. These numbers were much higher than the sagebrush recovery class, which had an overall accuracy of 37.84% and a high in 2013 of 44.16%. The stable class performed much like the tipping point class with an overall accuracy of 65.12% and a high in 2013 of 73.91%.

When the MTBS polygons were overlaid on the predictive map and compared to the sagebrush condition class underneath, we expected that more of the tipping point class would be encompassed than the other two classes because, by definition, it had a higher percent of highly flammable annual grass cover. Figure 4 showed exactly that phenomenon throughout the time series with more of the burned area occurring in the tipping point class each year. At two points in the time series, 2005 and 2010, the area burned in the tipping point and the sage recovery classes were almost equal. After both of those years, the amount of area burned in the tipping point class increased when compared to the amount of area burned in the sagebrush recovery class. A consistent downward trend appeared from 2011 – 2014 when the sagebrush recovery class declined as a percentage of the total area burned and the tipping point class area encompassed almost 100% of the burned area. There was a slight upturn in burned area for the sagebrush recovery class in 2015. The inverse of that trend occurred in the tipping point class as it increased as a percentage of the total burned area until 2015 when it turned slightly downward.

The stable class experienced a consistent downward trend from 2007 until 2012 when a slight increase occurred. The trend for the stable class was downward during the last three years with less than 1.5% of the area burned in that class during both 2014 and 2015.

In the last 5 years of the study period, 71.08% of the area that burned was in the tipping point class, 22.50% was in the sage recovery class, and 6.42% in the stable class. In the last 3 years of the study period, 91% of burns occurred in the tipping point class. The 5-year burned-area percentages were similar to total-area percentages for the associated classes: tipping point class (70.36%); sage recovery class (21.31%); and stable class (8.35%).

Discussion

Mapping areas of stability and areas of change using machine-learning algorithms allowed both the identification of dominant abiotic variables that drive ecosystem dynamics and the variables' important thresholds. Both the predictive and descriptive models for this study were developed using multiple drivers that represented climatic, topographic, and soils data because these variables were considered among the most important drivers of invasibility of sagebrush ecosystems (Roundy et al. 2018). Resilience to disturbance and stress was associated with climatic and topographic gradients, resistance to cheatgrass invasion was driven by temperature and precipitation regimes (Chambers et al. 2014), and the driver of variable interannual cheatgrass growth was highly variable precipitation (Bradley and Mustard 2005). Periods of limited soil moisture reduced plant germination and establishment (Bishop et al. 2019), and management activities that sought to exclude non-native species introduced gaps in vegetation cover that caused increases in resource availability (Rau et al. 2014). Both phenomena created space for cheatgrass, an early-season and fast-growing plant, to have invaded by capitalizing on available moisture and nutrients when temperatures warmed, earlier than most native plants (Boyte et al. 2016).

Elevation was the most often-used variable in development of both the predictive and descriptive models, and the 1 602 m elevation threshold played an important role in ruleset development of the descriptive model. This elevation threshold was slightly above the mean elevation (1 562 m) for the overall study area, and substantially above the low end of the elevation range of 568 m. Elevation affected sagebrush invasibility indirectly through air and soil temperatures, plant communities and types, and plant productivity (Chambers et al. 2014). Historically, fire events were more common at higher elevations in the Great Basin than at lower elevations because higher plant productivity led to more continuous fuels that spread fire, and plants adapted to fire as a result (Chambers et al. 2014). The lower native plant productivity at lower, drier elevations left adequate space for annual grasses to invade and exploit unused resources.

The elevational influence was evident in the study as 50% of the area classified as tipping point was defined by elevation thresholds lower than 1 602 m with almost 33% being lower than 1 339 m. Less than 4% of the tipping point class' area had an elevation threshold above 1 602 m. Better soil productivity at higher elevations increased plant productivity, which increased resistance to annual grass invasion, and while available water capacity varied little among sagebrush condition classes, soil organic matter was positively correlated with elevation. While we did not show both resiliency and resistance in the same pixel, we can presume that the stable class was derived from a pixel's resistance to invasion, which was positively associated with resilience to disturbance (Chambers et al. 2014). The stable class was defined much differently than the other two classes when weighted averages of the most important variables were considered (Table 3). The stable class was substantially more mesic and much cooler, typical of higher elevation sites in the study area.

The 30-yr precipitation variable was used nearly as often as elevation in development of the predictive and descriptive models, demonstrating its strong influence on sagebrush condition class. The area mapped as tipping point class had higher average precipitation than the sagebrush recovery class, which seemed counter intuitive. However, given that the tipping point class covered more than 70% of the study area and with the wide diversity of sagebrush and annual grass productivity, these large-area precipitation (and elevation) means included a fair amount of site-specific variability. In addition, both elevation and precipitation played important roles in the separation of recovery versus tipping point areas, so the combined stratification of the classification-tree structure in precipitation and elevation may well account for much of the sub-regional deviations from these large and diverse area precipitation means.

In the northern Great Basin, March temperatures, much more than March precipitation, proved influential in a study that modeled and analyzed cheatgrass percent cover predictions using current and future climate data (Boyte et al. 2016). In the current study, March precipitation was one of the strongest drivers of sagebrush condition class while March temperature variables still exerted substantial influence on both the predictive and descriptive models. March's climate was important to cheatgrass's life cycle and its ability to compete with native plants in the semiarid Great Basin (Roundy et al. 2018), and with a weighted 30-yr average March temperature minimum below freezing for all classes (-5.38 °C for the stable class) (Table 3), cheatgrasses' ability to survive drops in temperatures below 0 °C (Bykova and Sage 2012) was observed. Annual average precipitation was identified by Bradley (2009) as a strong predictor of cheatgrass presence in the Great Basin, and this variable was strongly influential to the development of both the predictive and descriptive models. As a dominant driver of thresholds, 30-yr annual precipitation influenced 48% of the rulesets that covered 43% of the study area.

Implications

The spatially explicit maps revealed that most of the study area has been altered from its native state and has never recovered. These tipping points are likely to persist in the future as natural recovery in sagebrush ecosystems is a long-term process and restoration efforts can be marginally effective. Oftentimes subsequent disturbances interrupt natural recovery and restoration projects. Some areas have recovered or have been restored as evidenced by sagebrush recovery pixels. Stable pixels existed where sagebrush systems are most likely to persist in the future, although climate change can threaten the persistence of stable areas. The predictive model can be modified with future climate data to aid in identifying future achievable sagebrush recovery or stable persistence.

The study delineated the abiotic variables that most influence the development of the sagebrush condition models. The descriptive model is unique in that it provides relative values that define the most influential variable's thresholds. The values give clarity to some of the forces that drive sagebrush ecosystem stability and change. The study's findings have implications for land managers, ecologists, fire modelers, and policymakers as they identify potential areas of disturbance, recovery, and stability and determine the best way forward to preserve sagebrush ecosystems and the species that inhabit them. Given the scale of the study area and the adequate geographic distribution of sampling points, analyzing relationships among these forces and ecosystem change can elicit a better understanding of sagebrush ecosystems and enhance land management and policymaking efforts.

Acknowledgements

We thank Neal Pastick for his astute insights on the modeling process and his input to an earlier version of the manuscript. U.S. Geological Survey Land Change Science and National Land Imaging programs and the Bureau of Land Management funded the study. The funders had no role in the study or preparation of this manuscript.

References

[dataset] Boyte, S.P., Wylie, B., Gu, Y., 2019b. Estimating environmental thresholds for three classes of sagebrush condition in the western United States (2001 – 2015). U.S. Geological Survey data release. https://doi.org/10.5066/P98WBAL4.

[dataset] Boyte, S.P., Wylie, B.K., 2017. A time series of herbaceous annual cover in the sagebrush ecosystem. U.S. Geological Survey data release. https://doi.org/10.5066/F71J98QK.

Bansal, S., James, J.J., Sheley, R.L., 2014. The effects of precipitation and soil type on three invasive annual grasses in the western United States. J Arid Environ. 104, 38-42. http://dx.doi.org/10.1016/j.jaridenv.2014.01.010.

Bishop, T.B.B., Munson, S., Gill, R.A., Belnap, J., Petersen, S.L., St Clair, S.B., 2019. Spatiotemporal patterns of cheatgrass invasion in Colorado Plateau National Parks. Landsc Ecol. 34, 925-941. https://doi.org/10.1007/s10980-019-00817-8.

Blomberg, E.J., Sedinger, J.S., Atamian, M.T., Nonne, D.V., 2012. Characteristics of climate and landscape disturbance influence the dynamics of greater sage-grouse populations. Ecosphere. 3, 1-20. 10.1890/ES11-00304.1.

Boyte, S.P., Wylie, B.K., Major, D.J., 2016. Cheatgrass Percent Cover Change: Comparing Recent Estimates to Climate Change - Driven Predictions in the Northern Great Basin. Rangeland Ecol Manag. 69, 265-279. http://dx.doi.org/10.1016/j.rama.2016.03.002.

Boyte, S.P., Wylie, B.K., Major, D.J., 2019a. Validating a Time Series of Annual Grass Percent Cover in the Sagebrush Ecosystem. Rangeland Ecol Manag. 72, 347-359. https://doi.org/10.1016/j.rama.2018.09.004.

Boyte, S.P., Wylie, B.K., Major, D.J., Brown, J.F., 2015. The integration of geophysical and enhanced Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index data into a rule-based, piecewise regression-tree model to estimate cheatgrass beginning of spring growth. Int J Digit Earth. 8, 118-132. https://dx.doi.org/10.1080/17538947.2013.860196.

Bradley, B.A., 2009. Regional analysis of the impacts of climate change on cheatgrass invasion shows potential risk and opportunity. Glob Change Biol. 15, 196-208. http://dx.doi.org/10.1111/j.1365-2486.2008.01709.x.

Bradley, B.A., Curtis, C.A., Fusco, E.J., Abatzoglou, J.T., Balch, J.K., Dadashi, S., Tuanmu, M.N., 2018. Cheatgrass (Bromus tectorum) distribution in the intermountain Western United States and its

relationship to fire frequency, seasonality, and ignitions. Biological Invasions. 20, 1493-1506. https://doi.org/10.1007/s10530-017-1641-8.
Bradley, B.A., Mustard, J.F., 2005. Identifying land cover variability distinct from land cover change: Cheatgrass in the Great Basin. Remote Sens Environ. 94, 204-213. https://doi.org/10.1016/j.rse.2004.08.016.
Browning, D.M., Rango, A., Karl, J.W., Laney, C.M., Vivoni, E.R., Tweedie, C.E., 2015. Emerging technologies and cultural shifts advancing drylands research and management. Frontiers in Ecology and the Environment. 13, 52-60. https://doi.org/10.1890/140161.
Bykova, O., Sage, R.F., 2012. Winter cold tolerance and the geographic range separation of Bromus tectorum and Bromus rubens, two severe invasive species in North America. Glob Chang Biol. 18, 3654-3663. https://doi.org/10.1111/gcb.12003.
Chambers, J.C., Bradley, B.A., Brown, C.S., D'Antonio, C., Germino, M.J., Grace, J.B., Hardegree, S.P., Miller, R.F., Pyke, D.A., 2014. Resilience to stress and disturbance, and resistance to Bromus tectorum L. invasion in cold desert shrublands of western North America. Ecosystems. 17, 360-375. https://doi.org/10.1007/s10021-013-9725-5.
Chambers, J.C., Maestas, J.D., Pyke, D.A., Boyd, C.S., Pellant, M., Wuenschel, A., 2017. Using Resilience and Resistance Concepts to Manage Persistent Threats to Sagebrush Ecosystems and Greater Sage-grouse. Rangel Ecol Manag. 70, 149-164. https://doi.org/10.1016/j.rama.2016.08.005.
Chambers, J.C., Roundy, B.A., Blank, R.R., Meyer, S.E., Whittaker, A., 2007. What makes Great Basin sagebrush ecosystems invasible by Bromus tectorum? Ecological Monographs. 77, 117-145. https://doi.org/10.1890/05-1991.
Chambers, J.C., Wisdom, M.J., 2009. Priority research and management issues for the imperiled Great Basin of the western United States. Restor Ecol. 17, 707-714. https://doi.org/10.1111/j.1526-100X.2009.00588.x.
Davies, K.W., Bates, J.D., 2019. Longer-Term Evaluation of Sagebrush Restoration After Juniper Control and Herbaceous Vegetation Trade-offs. Rangeland Ecol Manag. 72, 260-265. 10.1016/j.rama.2018.10.006.
Davies, K.W., Bates, J.D., Boyd, C.S., 2018. Post-wildfire seeding to restore native vegetation and limit exotic annuals: an evaluation in juniper-dominated sagerush steppe Restoration Ecology. 27, 120-127. https://doi.org/10.1111/rec.12848.
Davis, M.A., Grime, P., Thompson, K., 2000. Fluctuating resources in plant communities: a general theory of invasibility. J Ecol. 88, 528-534. DOI 10.1046/j.1365-2745.2000.00473.x.
Germino, M.J., Barnard, D.M., Davidson, B.E., Arkle, R.S., Pilliod, D.S., Fisk, M.R., Applestein, C., 2018. Thresholds and hotspots for shrub restoration following a heterogeneous megafire. Landsc Ecol. 33, 1177-1194. https://doi.org/10.1007/s10980-018-0662-8.
Gu, Y., Wylie, B.K., 2010. Detecting ecosystem performance anomalies for land management in the Upper Colorado River Basin using satellite observations, climate data, and ecosystem models. Remote Sens. 2, 1880-1891. https://doi.org/10.3390/rs2081880.
Gu, Y.X., Wylie, B.K., Boyte, S.P., Picotte, J., Howard, D.M., Smith, K., Nelson, K.J., 2016. An Optimal Sample Data Usage Strategy to Minimize Overfitting and Underfitting Effects in Regression Tree Models Based on Remotely-Sensed Data. Remote Sens. 8, 943. https://doi.org/10.3390/rs8110943.
Herrick, J.E., Van Zee, J.W., McCord, S.E., Courtright, E.M., Karl, J.W., Burkett, L.M., 2017. Monitoring Manual for Grassland, Shrubland, and Savanna Ecosystems. USDA - ARS Jornada Experimental Range: Las Cruces, NM.
Jenkerson, C.B., Maiersperger, T.K., Schmidt, G.L., 2010. eMODIS—a user-friendly data source. U.S. Geological Survey Open-File Report 2010-1055: Reston, VA.
Jensen, J.R., 2005. Introductory digital image processing: A remote sensing perspective. Pearson Prentice Hall: Upper Saddle River, NJ. 526 p.

Monitoring Trends in Burn Severity. MTBS Data Access: Fire Level Geospatial Data. https://www.mtbs.gov/ accessed July 2018.

Pilliod, D.S., Welty, J.L., Arkle, R.S., 2017. Refining the cheatgrass-fire cycle in the Great Basin: Precipitation timing and fine fuel composition predict wildfire trends. Ecol Evol. 7, 8126-8151. https://doi.org/10.1002/ece3.3414.

PRISM Climate Group. Oregon State University. http://prism.oregonstate.edu/ accessed 27 August 2015.

Quinlan, J.R. An Overview of Cubist. RuleQuest. https://www.rulequest.com/ accessed 27 June 2019.

Quinlan, J.R. See5: An informal tutorial -- release 2.10. RuleQuest. https://www.rulequest.com/ accessed 27 June 2019.

Rau, B.M., Chambers, J.C., Pyke, D.A., Roundy, B.A., Schupp, E.W., Doescher, P., Caldwell, T.G., 2014. Soil Resources Influence Vegetation and Response to Fire and Fire-Surrogate Treatments in Sagebrush-Steppe Ecosystems. Rangeland Ecol Manag. 67, 506-521. https://doi.org/10.2111/REM-D-14-00027.1.

Rigge, M., Homer, C., Wylie, B., Gu, Y., Shi, H., Xian, G., Meyer, D.K., Bunde, B., 2019. Using remote sensing to quantify ecosystem site potential community structure and deviation in the Great Basin, United States. Ecological Indicators. 96, 516-531. https://doi.org/10.1016/j.ecolind.2018.09.037.

Rigge, M., Wylie, B.K., Gu, Y., Belnap, J., Phuyal, K., Tieszen, L.L., 2013. Monitoring the status of forests and rangelands in the western United States using ecosystem performance anomalies. Int J Remote Sens. 34, 4049-4068. http://dx.doi.org/10.1080/01431161.2013.772311.

Rose, R.A., Byler, D., Eastman, J.R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H., Hansen, M., Headley, R., Hewson, J., Horning, N., Kaplin, B.A., Laporte, N., Leidner, A., Leimgruber, P., Morisette, J., Musinsky, J., Pintea, L., Prados, A., Radeloff, V.C., Rowen, M., Saatchi, S., Schill, S., Tabor, K., Turner, W., Vodacek, A., Vogelmann, J., Wegmann, M., Wilkie, D., Wilson, C., 2015. Ten ways remote sensing can contribute to conservation. Conserv Biol. 29, 350-359. https://doi.org/10.1111/cobi.12397.

Roundy, B.A., Chambers, J.C., Pyke, D.A., Miller, R.F., Tausch, R.J., Schupp, E.W., Rau, B., Gruell, T., 2018. Resilience and resistance in sagebrush ecosystems are associated with seasonal soil temperature and water availability. Ecosphere. 9, e02417. https://doi.org/10.1002/ecs2.2417.

Svejcar, T., Boyd, C., Davies, K., Hamerlynck, E., Svejcar, L., 2017. Challenges and limitations to native species restoration in the Great Basin, USA. Plant Ecol. 218, 81-94. https://doi.org/10.1007/s11258-016-0648-z.

The National Land Cover Database. Department of Interior, U.S. Geological Survey. http://www.mrlc.gov/ accessed 28 June 2019.

U.S. Fish and Wildlife Service, 2013. Greater sage-grouse (Centrocercus urophasians) conservation objectives: Final report. U.S. Fish and Wildlife Service: Denver, CO.

Wylie, B.K., Boyte, S.P., Major, D.J., 2012. Ecosystem Performance Monitoring of Rangelands by Integrating Modeling and Remote Sensing. Rangeland Ecol Manag. 65, 241-252. https://dx.doi.org/10.2111/Rem-D-11-00058.1.

Wylie, B.K., Pastick, N.J., Picotte, J.J., Deering, C.A., 2018. Geospatial data mining for digitial raster mapping. GIScience & Remote Sensing. 56, 406-429. https://doi.org/10.1080/15481603.2018.1517445.

Wylie, B.K., Zhang, L., Bliss, N., Ji, L., Tieszen, L.L., Jolly, W.M., 2008. Integrating modelling and remote sensing to identify ecosystem performance anomalies in the boreal forest, Yukon River Basin, Alaska. Int J Digit Earth. 1, 196-220. http://doi.org/10.1080/17538940802038366.

Table 1. Model accuracy assessments for the three datasets developed with regression-tree models. The relative error magnitude is the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value (Quinlan 2008).

	Training			Test		
	Correlation coefficient (*r*)	Mean absolute error	Relative error	Correlation coefficient (*r*)	Mean absolute error	Relative error
Annual herbaceous	0.92	4.4	0.33	0.91	4.6	0.34
Sagebrush expected performance	0.97	1.7	0.20	0.97	1.7	0.20
Sagebrush site potential	0.98	1.1	0.16	0.98	1.1	0.17

Table 2. Driving variables for the sagebrush condition class models. Frequency (%) of usage is shown for each variable for both models. Dashes indicate that a variable was not used. A 10-boost option was applied to the predictive model to improve model accuracy.

	% Variable usage	
Variable	Predictive	Descriptive
Elevation	100	75
30-yr precipitation (ppt)	100	73
July temperature maximum (tmax)	100	14
Soil organic matter	100	4
March ppt	99	7
June temperature minimum (tmin)	95	6
July ppt	95	3
Available water capacity	94	18
March tmin	94	11
June ppt	93	4
July tmin	92	--
April ppt	90	6
March tmax	89	2
30-yr tmin	89	6
May ppt	88	--
August ppt	87	3
Spring ppt	87	--
April tmin	87	--
30-yr tmax	87	--
Summer ppt	85	--
April tmax	84	--
May tmax	79	--
August tmax	79	--
August tmin	77	--
Spring tmin	70	--
May tmin	69	--
June tmax	65	6
Spring tmax	57	--
Summer tmin	57	--
Summer tmax	27	--

The predictive model's training and test accuracies (*r*) equal 87.30% and 74.60%, respectively.
The 10-fold cross validation accuracy (*r*) equals 73.70%.
The descriptive model's training and test accuracies (*r*) equal 67.70% and 69.40%, respectively.

Table 3. Weighted average by class and variable. These averages generally define the thresholds of the most influential variables within each class of the descriptive model.

	Variable						
	Elevation (m)	30-y ppt (mm)	AWC	July tmax °C	March tmin °C	March ppt (mm)	Total area (%)
Class							
Recovery	1601.87	267.14	44.12	30.62	-2.74	26.26	21.29
Tipping point	1504.72	349.57	44.35	29.18	-2.8	28.49	70.36
Stable	1939.30	416.77	45.07	26.65	-5.38	40.49	8.35

Table 4. Dominant variable thresholds by class gleaned from the descriptive model's rulesets. Threshold range is given, along with the ruleset(s) to which the threshold range applies. The weighted mean value for each dominant variable is reported for each class.

Driver	Model threshold	Ruleset(s)	Ruleset(s) as % of total land area	Weighted class mean (μ)	Class range
Recovery class dominant thresholds Rulesets 1 - 7					
Elevation (m)	>1451 to ≤1602	2,3,4,6	4.53	1602	1442 - 1760
	>1339 to ≤1602	1	7.16		
	>1602	7	9.11		
30-yr July tmax (°C)	>31	1,5	7.66	30.62	29.21 - 32.57
	≤30	2	<1		
	>30	4	1.46		
	>28 to ≤31	6	<1		
30-yr March ppt (mm)	≤36	2,4,6	2.94	26.26	18.93 - 29.70
30-yr annual ppt (mm)	≤244	3	1.59	267.14	221.21 - 293.38
	≤338	7	9.11		
Tipping point class dominant thresholds Rulesets 8 - 16					
Elevation (m)	>1602	9,12,15	2.51	1504.72	1169 - 1762
	≤1602	11,16	13.14		
	>1602 to ≤1798	8	<1		
	≤1339	10	22.23		
30-yr annual ppt (mm)	≤338	9,12,13,	11.21	349.57	252.08 - 474.01
	>338	14	22.29		
March tmin (°C)	≤-4	8	<1	-2.80	-4.27 - -1.09
	>-3	15	1.00		
Stable class dominant thresholds Rulesets 17 - 21					

Elevation (m)	>1602	19,20,21	2.84	1939.30	1666 - 2129
	>2002	18	3.99		
	>1798	17	1.51		
30-yr annual ppt (mm)	≤338	17,19,20,21	4.35	416.77	264.50 - 562.94
	>338	18	3.99		
March tmin (°C)	≤-3	19,20,21	2.84	-5.38	-6.84 - -1.63

Table 5. The Assessment Inventory and Monitoring (AIM) accuracy assessment by class and year for the predictive sagebrush condition classes.

	Year	Class	Sample size	Number accurate	Percent accurate
	2013	Recovery	77	34	44.16
		Tipping point	234	152	64.96
		Stable	69	51	73.91
Sub total			380	237	62.37
	2014	Recovery	76	27	35.53
		Tipping point	198	159	80.30
		Stable	47	34	72.34
Sub total			321	220	68.54
	2015	Recovery	101	31	30.69
		Tipping point	227	189	83.26
		Stable	59	32	54.24
Sub total			387	252	65.12
	2016	Recovery	153	62	40.52
		Tipping point	485	337	69.48
		Stable	126	79	62.70
Sub total			764	478	62.57
Total			1852	1187	64.09

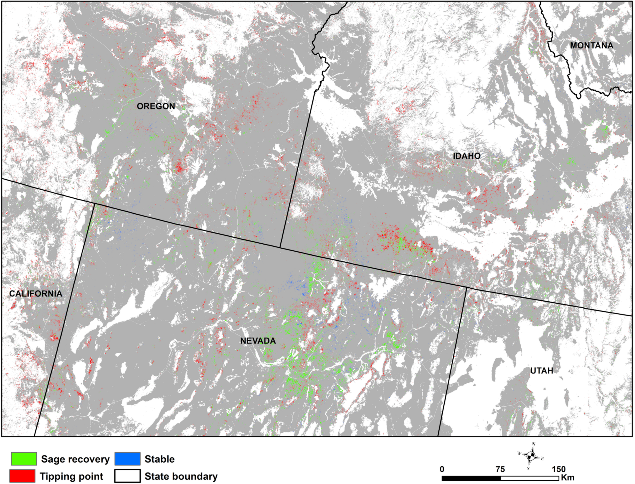
Figure 1. Potential reference data points. Pixels that met the criteria from 4 datasets for one of 3 classes of sagebrush condition are displayed. The mask (white inside map's borders) covers areas not classified as herbaceous/grassland or shrubs by the 2011 National Land Cover Database (NLCD) and/or areas that are higher than 2 250 m elevation.

Figure 2. Data and processes. The flowchart shows the data and outlines the processes used to develop the 2 sagebrush condition class maps. The data and processes are identical until the decision-tree modeling step where we developed 1 model that used a tree-structure for predictive purposes and a second model that used rulesets for descriptive purposes and the interpretation of abiotic thresholds of sagebrush condition class.

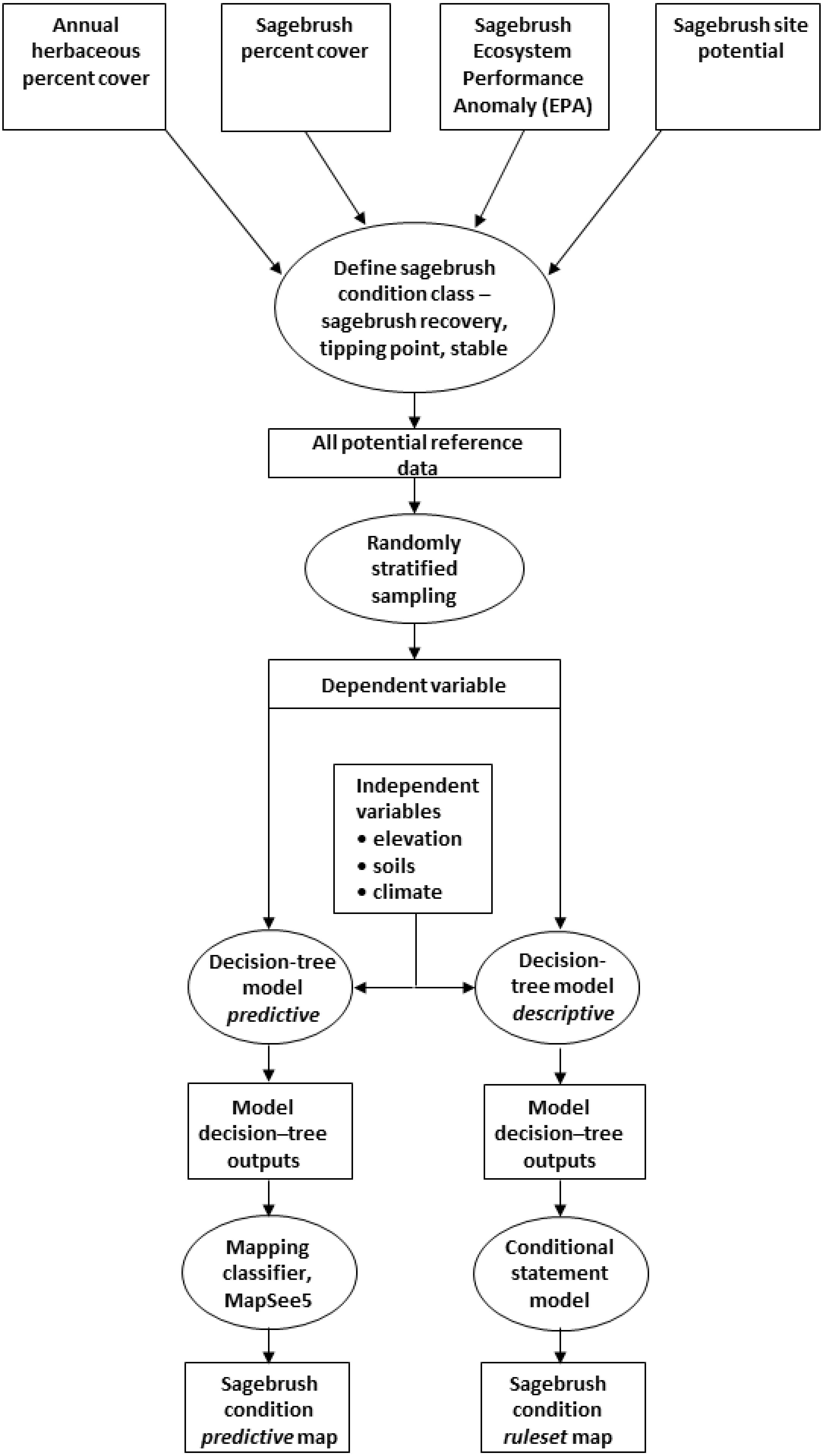
Figure 3. A spatially explicit ruleset map. Each ruleset, or classifier, is defined by the variable(s) and its value(s) used to establish if-then rules in the descriptive model. Rules 1 – 7 represent sagebrush recovery areas, 8 – 16 represent tipping point areas, and 17 – 21 represent stable areas. The rulesets are prioritized so that those with higher accuracy spatially supersede those with lower accuracy when a pixel is defined by more than 1 ruleset. The mask (white) covers areas not classified as herbaceous/grassland or shrubs by the 2011 National Land Cover Database (NLCD) and/or areas that are higher than 2 250 m elevation.

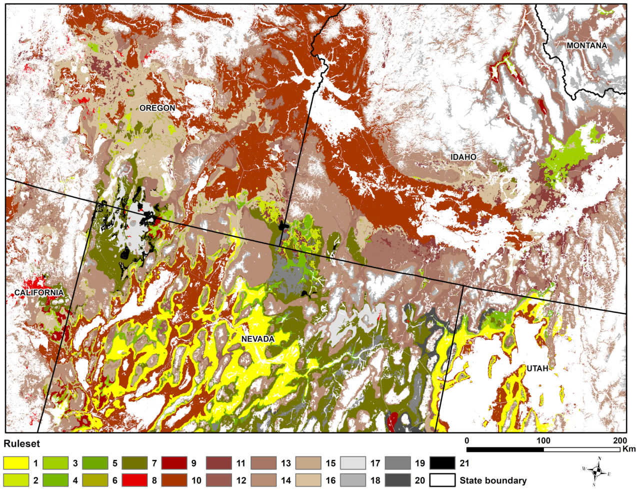
Figure 4. A chart that displays each sagebrush condition class – sagebrush recovery, tipping point, or stable – as a percent of the total area burned in the study area by year. We used a time dimension in our criteria to define each class. This time dimension focused the transition from a degraded state to a sagebrush recovery class and the transition from a stable sagebrush ecosystem state to a degraded state, or tipping point class, on the last 5 years of the time series. Given this time dimension, the last 5 years of the time series are most critical in assessing the accuracy.

Figure 5. A spatially explicit predictive map. Each pixel is classified as one of 3 sagebrush condition classes and delineated by confidence levels that ranged from 0.33 to 1.00 determined by a decision-tree model. A class confidence level equal to or greater than 0.70 was labeled high probability; a class confidence level equal to or greater than 0.50 and less than 0.70 was labeled moderate probability; and a class confidence level equal to or greater than 0.33 and less than 0.50 was labeled low probability. The mask (white) covers areas not classified as herbaceous/grassland or shrubs by the 2011 National Land Cover Database (NLCD) and/or areas that are higher than 2 250 m elevation.

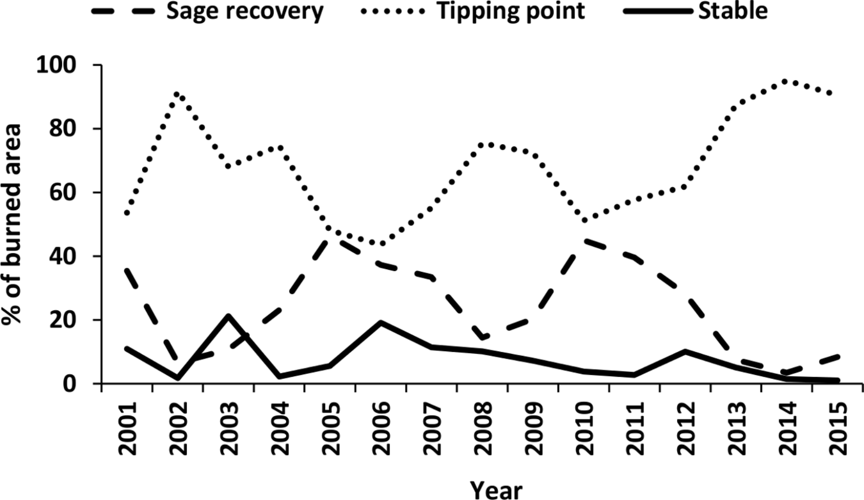


MONTANA
OREGON
IDAHO
CALIFORNIA
NEVADA
UTAH
Sage recovery
Stable
Tipping point
State boundary
N
S
E
W
0
75
150
Km

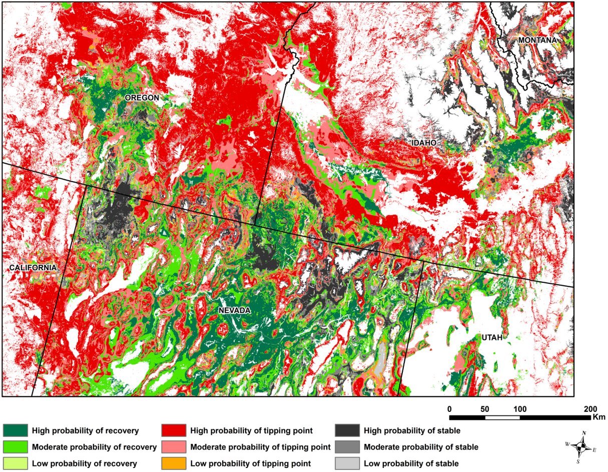

Annual herbaceous percent cover
Sagebrush percent cover
Sagebrush Ecosystem Performance Anomaly (EPA)
Sagebrush site potential
Define sagebrush condition class – sagebrush recovery, tipping point, stable
All potential reference data
Randomly stratified sampling
Dependent variable
Independent variables
• elevation
• soils
• climate
Decision-tree model *predictive*
Decision-tree model *descriptive*
Model decision–tree outputs
Model decision–tree outputs
Mapping classifier, MapSee5
Conditional statement model
Sagebrush condition *predictive* map
Sagebrush condition *ruleset* map



MONTANA
OREGON
IDAHO
CALIFORNIA
NEVADA
UTAH
Ruleset
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
State boundary
0
100
200
Km
N
S
E
W



Sage recovery
Tipping point
Stable
% of burned area
100
80
60
40
20
0
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year



MONTANA
OREGON
IDAHO
CALIFORNIA
NEVADA
UTAH
0
50
100
200
Km
N
W
E
S
High probability of recovery
Moderate probability of recovery
Low probability of recovery
High probability of tipping point
Moderate probability of tipping point
Low probability of tipping point
High probability of stable
Moderate probability of stable
Low probability of stable