Geoderma 352 (2019) 330–341

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Mapping pedomemory of spodic morphology using a species distribution model

Adrienne C. Nottingham[a], James A. Thompson[a,*], Frederica Wood[b], Pamela J. Edwards[b], Michael P. Strager[c]

[a] West Virginia University, Division of Plant and Soil Sciences, PO Box 6108, Morgantown, WV 26506-6108, USA
[b] USDA Forest Service, Northern Research Station, PO Box 404, Parsons, WV 26287, USA
[c] West Virginia University, Division of Resource Management, P.O. Box 6108, Morgantown, WV 26506-6108, USA

ARTICLE INFO

Keywords:
Maximum Entropy
Digital soil mapping
Red spruce
Spodosol
Central Appalachians
Pedomemory

ABSTRACT

Red spruce (*Picea rubens*) ecosystems in the high elevations of the central Appalachians of the eastern United States are the focus of ongoing restoration efforts due to the valuable ecosystem services these forests provide. Recent research has shown that spodic materials still present in the soil represent the pedomemory of the historic extent of red spruce forests in the region. A dataset containing 221 points with varying spodic intensities and 29 environmental variables collected from the Monongahela National Forest in West Virginia, USA, was used to evaluate the utility of a species distribution model, Maximum Entropy (MaxEnt), for predicting the presence of spodic properties. MaxEnt was selected for evaluation because, as a presence-only model, it inherently omits absence locations and thereby reduces the risk of including false absences (i.e., herein, locations that have undergone some level of depodzolization) unlike other models previously used to predict pedomemory. Model outputs that employed three different spodic intensity class inputs—very weak to strong expression, weak to strong expression, and strong expression—resulted in similar spodic probability predictions, though there was less area mapped as transitional probabilities in the strong expression model than the two models that included weaker spodic intensity input data. Permutation importance indicated that no single or small subset of environmental variables controlled the three model outputs, perhaps because the environmental covariates may have been too coarse or not strongly enough associated with podzolization processes to be very important. When the output from the MaxEnt model using the full range of spodic intensities (very weak to strong) was compared to an output produced using a presence-absence model (random forests), there was approximately 62% agreement (where both models predicted presence or both predicted absence) for the cells in the top 40% of the predicted probabilities.

1. Introduction and background

Soils form as the result of five interacting environmental factors: climate, organisms, relief, parent material, and time (Dokuchaev, 1999; Jenny, 1941). The concept of using environmental factors to predict soil characteristics is accepted and provides the basis for traditional soil mapping and contemporary digital soil mapping (e.g., McBratney et al., 2003; Boettinger et al., 2010). Nauman et al. (2015a, 2015b) demonstrated that the inverse also may be possible; that is, current soil properties might be useful for predicting previous environmental conditions because certain soil characteristics can persist even after some environmental factors change—a concept termed pedomemory (Targulian and Goryachkin, 2010; Lin, 2011; Monger and Rachal, 2013; Nauman et al., 2015a, 2015b).

In the central Appalachians of the eastern United States, Spodosols are a soil order that is particularly well aligned with the concept of pedomemory. Spodosols form through the process of podzolization (Schaetzl and Isard, 1996; Lundström et al., 2000a; Lundström et al., 2000b; Sauer et al., 2007). In this region, podzolization is largely driven by three factors: organisms (conifer vegetation, particularly red spruce/hemlock ecosystems) climate (cool, moist) and parent material (base-poor geologies) (Oosting and Billings, 1951; Stanley and Ciolkosz, 1981), and the resulting spodic characteristics (typically thick organic horizons, albic horizons, and subsoil accumulations of organic matter and aluminum and iron sesquioxides) can persist for long periods after the vegetation from which they were formed is no longer present or is much less dominant in the over story stand (Soil Survey Staff, 2003). However, through time, spodic materials degrade if one or more of the

* Corresponding author.
E-mail address: james.thompson@mail.wvu.edu (J.A. Thompson).

https://doi.org/10.1016/j.geoderma.2017.10.044
Received 3 April 2017; Received in revised form 18 October 2017; Accepted 27 October 2017
Available online 10 November 2017
0016-7061/

conditions responsible for their formation are modified; this process is referred to as depodzolization (Barrett and Schaetzl, 1998). In the central Appalachians, depodzolization has resulted primarily from the loss of red spruce, which was much more dominant prior to 1880, but was largely lost from forest stands between 1880 and 1930 due to extensive logging and subsequent wildfires (Clarkson, 1964; Lewis, 1998).

The degree to which spodic characteristics are expressed (i.e., spodic intensity) at any point in time depends upon the state and pathways of progressive or regressive pedogenesis (Barrett and Schaetzl, 1998). Podzolization can take as long as hundreds to thousands of years (Lundström et al., 2000a, 2000b), whereas depodzolization can occur more rapidly, requiring as few as 30 yr or as many as 200 yr (Hole, 1975; Nornberg et al., 1993). The rates of both podsolization and depodzolization ultimately depend upon site-specific environmental characteristics (Barrett and Schaetzl, 1998).

Based on the knowledge that both podzolization and depodzolization are soil forming processes, some assumptions regarding spodic properties and historic red spruce cover in this region can be made. If spodic soil properties are observed today, one can reasonably assume that red spruce was present at that site in the past. However, if spodic soil properties are not visually evident in the observed soil morphology, one cannot assume that red spruce was not historically present, due to the potential for depodzolization.

This paper describes the results of a novel application of the species distribution model Maximum Entropy (MaxEnt) to predict soil pedomemory. MaxEnt is a presence-only model, which means that it uses only presence data as its input (in this case, locations where some level of spodic properties were observed). The first objective of this study was to determine how the mapped extent of podzolization employing three different spodic intensity ranges compare using MaxEnt. The second objective is based on the assumption that using a presence-only model will decrease the risk of including false absences; in other words, because it includes only presence data, it avoids the assumption that the lack of current spodic intensity is not interpreted as spodic characteristics were never present or the site is not conducive to spodic development. For the second objective, the model output from MaxEnt using the full range of spodic intensity (very weak to strong expression) was compared to the output generated by Nauman et al. (2015a) for the same area using the presence-absence model random forests. In the latter model, absence (i.e., no spodic morphology observed) also were included in the input data.

Studies have documented potential negative effects of including absence data in ecological modeling, mainly due to the possibility of including 'false absences' (in this case, locations where spodic morphology was not observed) (Svenning and Skov, 2004; Jimenez-Valverde et al., 2008). The inclusion of absence data may introduce confusion in the random forests model due to the potential for depodzolization (Nauman et al., 2015a, 2015b). Consequently, the resulting mapped extent of spodic properties using MaxEnt may be more representative of the 'fundamental niche' (Phillips et al., 2006) of spodic properties, which represents all places on a landscape which are conducive to Spodosol formation (and, therefore, likely within the historic extent of red spruce forests).

2. Materials and methods

2.1. Study area

This study involves approximately 124,687 ha of the Monongahela National Forest (MNF) in eastern West Virginia, USA (Fig. 1). It includes areas underlain by Chemung Group (Devonian-age acid siltstone and sandstone), Hampshire Formation (Devonian-age acidic shale and siltstone), and Pottsville Group (Pennsylvanian-age acid sandstone) geologies (West Virginia Geological and Economic Survey, 1968). The majority of the MNF has a moist climate (1184–1524 mm of precipitation per year) with cool mean annual temperatures (ranging from 6 to 8 °C) (NOAA NCDC, 2016). Much of the area considered in this study is among the wettest and coolest in the MNF and West Virginia. The elevation ranges from approximately 800 to 1300 m; associated soil temperature regimes span the bounds between mesic and frigid soil temperatures (Lietzke and McGuire, 1987; Stanley and Ciolkosz, 1981), with the colder soil temperatures generally present at higher elevations. On the MNF, these areas are typically transition zones between areas dominated by mixed hardwood species (*Acer rubrum, A. pennsylvanicum, Prunus serotina,* and *Fagus grandifolia*) and those dominated by red spruce and hemlock (*Tsuga canadensis*) (Shigo, 1972; Nauman et al., 2015a).

The presence-only data for this study are part of a dataset collected from 2010 to 2012 in an effort to understand the extent of podzolization across the MNF (Nauman, 2015). Point observations were collected from soil pits and soil transects across a variety of geology types and soil series (Dekalb, Berks, Mandy and Wildell), and under varying forest compositions (from no red spruce present to red spruce dominant). A total of 332 point observations obtained from Nauman (2015) were compiled (Fig. 2). Spodic intensities were assigned at the time of soil sampling to each of the 332 points using the spodic-intensity scale developed by the USDA NRCS, which has discrete values ranging from 0 to 2 based on soil morphological properties observed in the field (Table 1) (Nauman et al., 2015a). Spodic properties occur in varying intensities (Schaetzl and Isard, 1996; Lundström et al., 2000a, 2000b; Nauman et al., 2015a) depending on the degree of influence of each environmental factor at the specific location. Spodic properties include a lighter colored E horizon and subsoil accumulations of aluminum and iron sesquioxides that are darker and redder in color (Soil Survey Staff, 2003). A spodic intensity of 0 indicates that the soil has no visible spodic properties, while a spodic intensity of 2 indicates the soil has the strongest spodic expression.

2.2. MaxEnt modeling approach and settings

MaxEnt requires two sets of data: presence-only data (in this case, point locations where spodic soil properties were observed) and environmental variables believed to be important for the attribute of interest (Pearson, 2007, 2010). Environmental variables can be continuous (e.g., elevation) or categorical (e.g., geologic formations), although MaxEnt works best when the number of categorical variables is limited (Pearson, 2007, 2010). The 332 data points originally compiled included locations spanning the entire range of spodic intensities (Table 2). The MaxEnt model required locations that had evidence (i.e., presence) of podzolization, so only locations with a spodic intensity of at least 0.5 were utilized in this study. Consequently, 221 presence points were included. These data were further assembled into three presence-only data sets for use in MaxEnt based on three spodic intensity ranges: 0.5–2.0, 1.0–2.0, and 2.0 (Table 2).

Environmental variables believed to be important to the development of spodic soil properties and presence of red spruce were used to generate the background sample (a user-specified number of samples taken from all possible locations considered to be equally likely to be a presence locality) in the MaxEnt models (Merow et al., 2013b). To allow comparison to the result from Nauman et al. (2015a), we employed 29 of the same 32 environmental variables, derived from digital elevation models and Landsat Geocover data, that they used (Table 3). The three calculated ratios of Landsat Geocover data were excluded because their use resulted in too much missing data due to division by zero.

Coefficients of correlation (r) were calculated among all 29 variables and, as expected, many showed high correlation (which we defined as $r > 0.80$ or < -0.80). Undesired outcomes, such as increased model complexity and decreased confidence in interpretations, may result when correlated variables are used in MaxEnt (Phillips, 2017; Baldwin, 2009). Therefore, model runs using all 29 variables first were compared to runs without correlated variables to determine how model

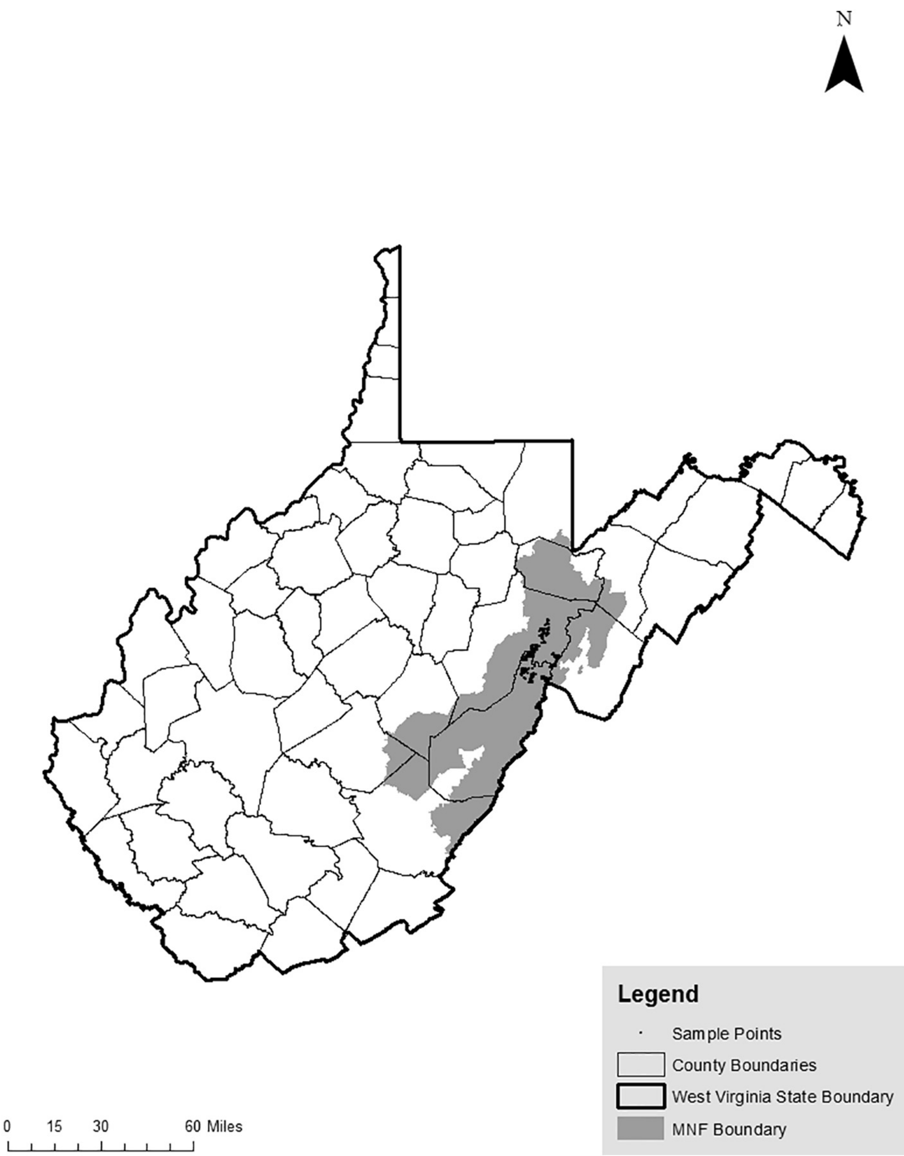

Fig. 1. Location of sample points across the Monongahela National Forest in eastern West Virginia.

gain was affected by including and excluding the correlated variables. Area under the curve (AUC) (i.e., under the receiver operating characteristic (ROC) curve) values (Elith et al., 2011; Beane et al., 2013; Merow et al., 2013a), described later, revealed the two types of runs resulted in only minor differences in AUC values (measures of model performance). Because the differences were minor and one of our objectives was to compare MaxEnt and the random forests results obtained by Nauman et al. (2015a) who employed the correlated variables, all 29 of the variables were utilized in the MaxEnt model runs described herein.

MaxEnt has numerous model settings that influence the outputs. For most settings, the model defaults were utilized in all runs (Pearson, 2007, 2010). Only those settings for which the defaults were not utilized are discussed further (Table 4).

MaxEnt has three methods of replication (used to quantify the variation in model results), but only bootstrapping was used in this analysis. Bootstrapping selects the user-specified number of points from the environmental variables across the study area to create the background sample. For each bootstrap analysis the background sample is randomly selected from the selected environmental variables, and then each point is returned back to the sample pool (Pearson, 2007, 2010; Beane et al., 2013) so the same record may be included in more than one replicate run (Pearson, 2007, 2010). We also used the random seed option in bootstrapping to ensure that each replicate data set was independent from all others (Pearson, 2007, 2010). Prior to running the replications, bootstrapping also randomly withholds a percentage of the available presence-only data to test model validation, and these random test data are not included in the replicate runs. We used 40% of the data as random test data because this is a percentage commonly used in MaxEnt modeling (Elith et al., 2006; Phillips and Dudik, 2008; Beane et al., 2013). Forty percent equates to 89, 80 and 35 withheld test sample sizes, respectively, for spodic intensity classes 0.5–2.0, 1.0–2.0 and 2. Ten replicates for each spodic intensity range were run.

MaxEnt has three output types: raw, and two transformed outputs derived from the raw data (Merow et al., 2013a; Pearson, 2007, 2010). For this analysis, logistic output, one of the transformed outputs, was used because it is useful for comparing model results (Merow et al., 2013a) and both study objectives involve comparisons of model runs. Some explanation of logistic outputs is warranted for understanding the results described later; a more detailed explanation of the logistic and other outputs is provided in Merow et al. (2013a).

As noted, the logistic output is a transformation of the raw output. For the raw output, MaxEnt initially assumes that all cells in the study area are equally likely to contain the species or attribute of interest, so the initial relative occurrence rate value for each cell is calculated by the inverse of the total number of cells. For this analysis, there are

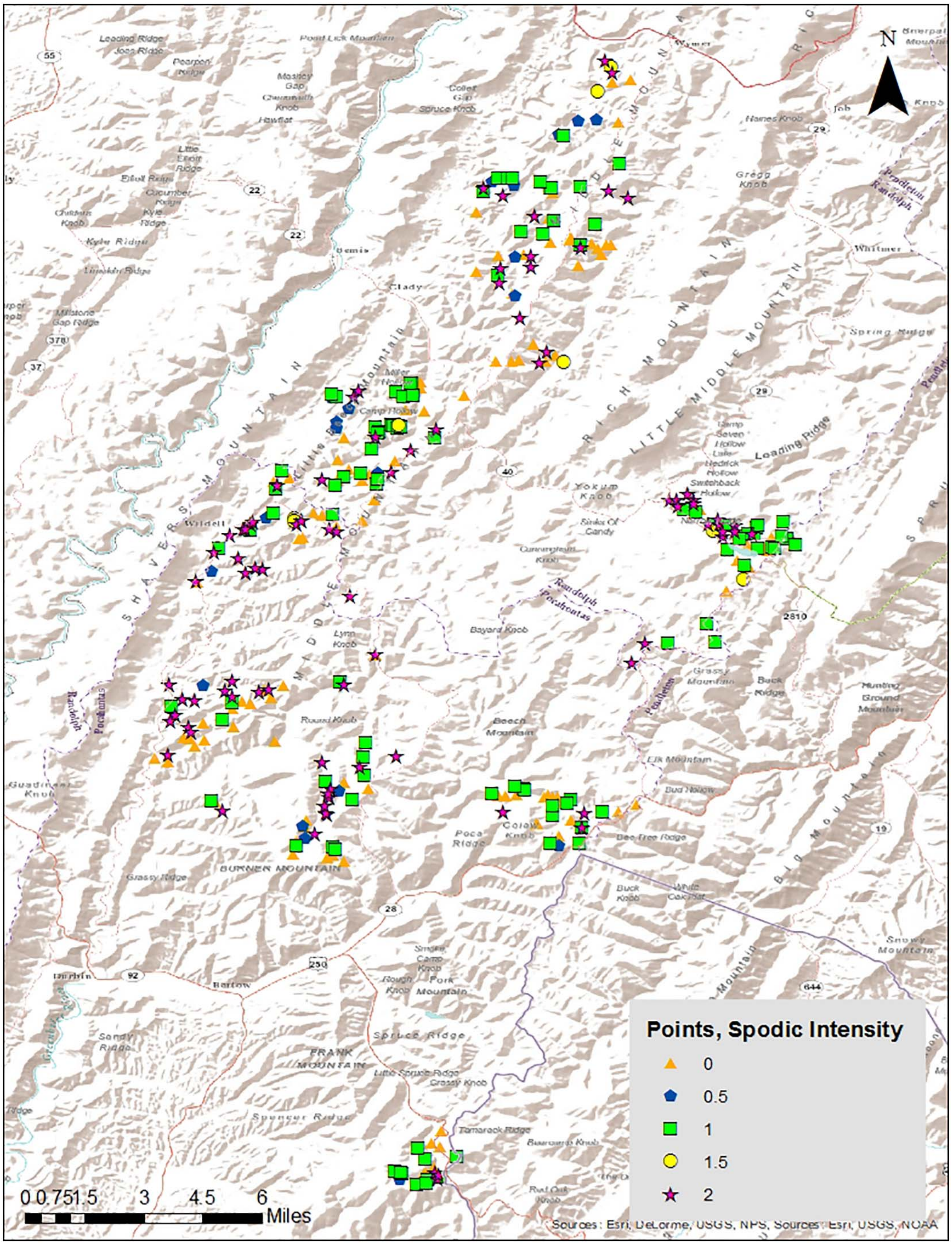

Fig. 2. Soil sampling and description points within the study area. Symbols denote the spodic intensity recorded at each point.

approximately 1.6 million cells, so the probability of spodic presence for each cell in the study area would have be approximately 1/1.6 million. Probabilities of each cell then are adjusted based upon environmental variable values at presence-only locations and the background sample such that the sum of the final probabilities for all cells in the raw output must sum to 1. Consequently, the cells in a raw output are not independent from one another. The transformation applied to the raw data for obtaining the logistic output amplifies higher probabilities while linearly scaling lower probability values. Therefore, the cell probability values for logistic output also are not independent, but they do not sum to 1 (Merow et al., 2013a).

2.3. Model comparison techniques

Comparison of maps and spatial modeling outputs often is necessary—and indeed was needed for this study—but no widely accepted protocols exist for such comparisons (Kuhnert et al., 2005; Visser and Nijs, 2005). Consequently, we used a combination of model-generated statistics, visual assessment, and model comparison tools for this paper (Kuhnert et al., 2005; Visser and Nijs, 2005).

To address the first study objective of comparing the MaxEnt model outputs of the three different spodic intensities, the three possible two-way comparisons were examined (i.e., 0.5–2.0 vs. 1.0–2.0, 1.0–2.0 vs. 2.0, and 0.5–2.0 vs. 2.0). However, because the probability data are continuous, and model agreement would result only for cells in which

Table 1
Spodic intensity classes and respective characteristics (adapted from Nauman et al., 2015a).

Rating	Level of podzolization	Soil properties associated with podzolization
0.0	No evidence	Not applicable
0.5	Very weak	Only slight physical evidence of podzolization; slightly redder hue and higher value is present at the top of the B horizon, but the hue is less than one Munsell hue redder than an underlying horizon; soil is non-smeary[a].
1.0	Weak, *spodic intergrade*[b]	Weak expression of podzolization; spodic materials are present, but do not meet the criteria for a spodic horizon; a weakly expressed Bs horizon is present, and is one Munsell hue redder than an underlying horizon. Bhs material is usually absent; no albic E horizon; spodic materials are sometimes weakly smeary
1.5	Moderate, *spodic intergrade*	Moderate expression of podzolization; spodic materials present as a spodic horizon; moderately expressed Bs horizon present, often with pockets of Bhs material; no albic E horizon; spodic materials are often weakly smeary
2.0	Strong, *Spodosol*	Strong expression of podzolization; spodic horizon is present usually underlying an albic E horizon; Bhs or Bh horizon is continuous across at least 85 percent of the pedon; spodic materials are often moderately smeary.

[a] See Schoeneberger et al., 2012 page 2–65 for a description of this metric.
[b] Spodic integrades are soils that may have some spodic properties or materials but do not fully meet the requirements of a Spodosol (Soil Survey Staff, 2003).

Table 2
Number of samples by spodic intensity (left), and number of samples by MaxEnt modeling classes. Points with a spodic intensity of 0.0 represent spodic absences, so they were not used for MaxEnt modeling because it employs presence-only data.

Spodic intensity	Number of observations	Class	Number in class
0.0	111		
0.5	22		
1.0	103	0.5–2.0	221
1.5	8	1.0–2.0	199
2.0	88	2.0	88

the probabilities were exactly equal (an almost impossible result), a more useful and interpretable procedure was used. Frequency distributions of the average probabilities were developed for each of the spodic intensity class model outputs. All three possible comparisons had similar distributions and there was a natural break in the distributions at the 0.6 probability level. Using that break, the MaxEnt modeled outputs were converted to binary outputs, where cells ≥ 0.6 and < 0.6 were assigned unique integer values and compared mathematically.

The AUC value was used to assess MaxEnt model performance. This value is the area underneath a ROC curve (Pearson, 2007, 2010; Elith et al., 2011; Beane et al., 2013; Merow et al., 2013a) (Fig. 3). A ROC curve is a plot of true positive vs. false positive rates (Pearson, 2007, 2010). The curve is a plot of *sensitivity* (number of presences correctly predicted) versus *1-specificity* (number of absences incorrectly predicted) (Pearson, 2007, 2010). An AUC value of 0.5 suggests that the model performed no better than a random model, while an AUC value of > 0.9 suggests the model excelled (Pearson, 2007, 2010; Young et al., 2011).

The importance of environmental variables was evaluated using permutation importance, normalized to percentages. The larger the permutation importance value, the more influence that variable has on the model outcome, particularly when it is followed by a marked decline in the permutation importance value for the next most important permutation value (in descending order) (Kalle et al., 2013). Permutation importance is not influenced by the paths that MaxEnt uses to generate the individual runs and final results (Phillips, 2006).

To evaluate the second objective of this paper, the MaxEnt model results were compared to the Nauman et al. (2015a) random forests model results. Because the random forests model employed presence and absence data (i.e., the entire 0–2.0 spodic intensity range), only the MaxEnt model using the most similar spodic intensity range (0.5–2.0) was employed in the comparison (i.e., 0 spodic intensity observations were not used because MaxEnt is a presence-only model). This comparison presents some challenges due to the inherent differences between the two models. Recall that cell values are not independent for MaxEnt, which often results in many cells with small predicted probability values (Merow et al., 2013b). By contrast, random forests output for each run is binary – after the model runs, each cell within the project area is classified as spodic or non-spodic. The cell probabilities reported by Nauman et al. (2015a), were based on the outcomes of 100 replicate runs: the value of each cell is equal to the number of times out of 100 that the cell was predicted to be spodic. For example, a cell with a value of 0.80, was classified as spodic (having some level of spodic expression) in 80 out of 100 model runs. Consequently, in random

Table 3
Digital elevation model-derived and Landsat Geocover environmental variables used to map spodic properties in MaxEnt (taken from Nauman et al., 2015b).

Variable name	Description
National elevation dataset (27.5 m resolution)	
nwness	Index from +1 to −1 of how northwest (+1) or southeast (−1) a site faces
eastness	Index from +1 to −1 of how east (+1) or west (−1) a site faces
southness	Index from +1 to −1 of how south (+1) or north (−1) a site faces
neness	Index from +1 to −1 of how northeast (+1) or southwest (−1) a site faces
dem	Elevation in meters
plan_curv	Curvature perpendicular to the slope direction
prof_curv	Curvature parallel to slope direction
ls_factor	Slope-length factor from USLE as calculated in SAGA GIS
convergence	Overall measure of concavity
slopepos	Index from 0 (valley floor) to 100 (ridgetop) of slope position (Hatfield, 1996)
slope	Slope gradient (rise/run) in fraction units
mrrtf	Multiple resolution ridgetop flatness index
mrvbf	Multiple resolution valley bottom flatness index
twi	Topographic wetness index
aacn2	Altitude above local stream channel
baselevel	Elevation of nearest channel point to each cell in its given watershed
contribarea	Upstream contributing area
relht1	Height of cell above the local minimum elevation in 1-cell radius
relht2	Height of cell above the local minimum elevation in 2-cell radius
relht3	Height of cell above the local minimum elevation in 3-cell radius
relh_5	Height of cell above the local minimum elevation in 5-cell radius
relht10	Height of cell above the local minimum elevation in 10-cell radius
relht20	Height of cell above the local minimum elevation in 20-cell radius
relht30	Height of cell above the local minimum elevation in 30-cell radius
relht50	Height of cell above the local minimum elevation in 50-cell radius
relht70	Height of cell above the local minimum elevation in 70-cell radius
Landsat Geocover 2000 (14.5-m resolution, resampled to 27.5 m)	
Nir	Near infrared band in 8-bit digital number units
Mir	Middle infrared band in 8-bit digital number units
Green	Green visible band in 8-bit digital number units

Table 4
MaxEnt settings used in this analysis and the justification for the setting choice. Model defaults were used for settings not listed in this table.

Setting	Setting utilized	Justification	References
Output format	Logistic	Recommended for comparing models	(Merow et al., 2013a)
Replication type	Bootstrapping	Bootstrapping samples with replacement; commonly used in the literature	(Elith et al., 2011; Merow et al., 2013a)
Random test percentage	40%	Commonly used in MaxEnt analyses	(Phillips, 2017; Phillips and Dudik, 2008)
Replicates	10	Chosen due to computing restraints	(Merow et al., 2013a; Pearson, 2010)

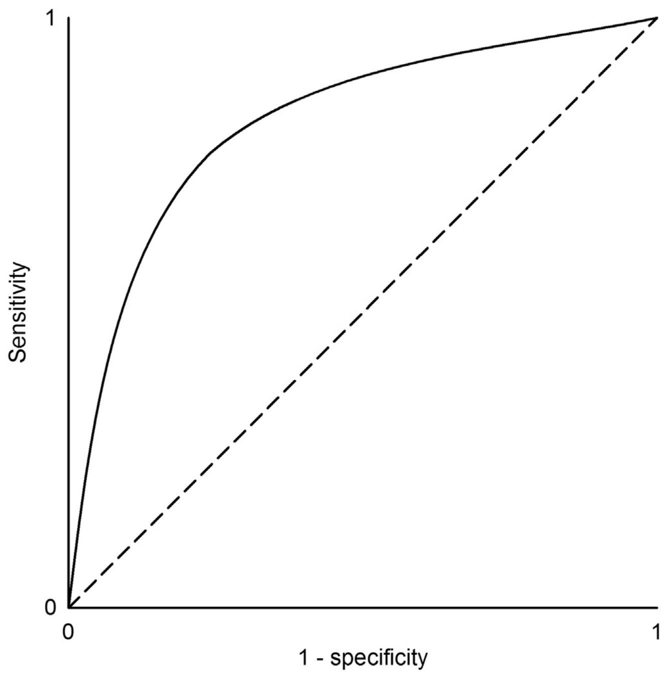

Fig. 3. A generalized example of an area under the receiver operating characteristic (ROC) curve value. The dashed line is included in all generated AUC values and is not reflective of data used in the model; it indicates an AUC value of 0.5, and represents a model that performs no better than one with random output. A model that performs perfectly would have an AUC value of 1. The solid line represents an actual model run (after Pearson, 2010).

forests output, each cell is independent of all other cells (Nauman et al., 2015a).

Due to the differences between MaxEnt and random forests, we focused on the general rank of cells for comparing the two model outputs. An agreement/disagreement analysis, similar to that for the three MaxEnt comparisons described previously, was performed. In this case the upper 40% of cells (in terms of predicted probability) were considered to be high probability (of spodic presence), and lower 60% of cells were considered low probability, but it should be noted that the upper 40% of cells had very different ranges of probability of presence (MaxEnt = 0.092589–0.937165; random forests = 0.610–0.994).

We also were interested in examining the environmental conditions present at locations where both models agreed (e.g., both predicted high probability of spodic presence) and disagreement (e.g., one predicted a low probability of spodic presence and the other a high probability) to get some idea of model drivers, as well as the degree of overlap in the environmental covariates where the models did not agree. Because the data were not normally distributed, the nonparametric Wilcoxon rank-sum test was used for the latter analysis to determine if environmental conditions were significantly different between the two models where their outcomes disagreed.

3. Results

3.1. MaxEnt comparisons

The modeled outputs for each of the three spodic intensity classes are shown in Fig. 4. Red and orange cells in Fig. 4A indicate a high probability that some degree of spodic expression are present because the 0.5–2.0 range includes the entire range of spodic intensity expression from very weak to strong (Table 1). By comparison, red and orange cells in Fig. 4B indicate a high probability of the presence of weak to strong spodic intensities since the 1.0–2.0 class data are employed, and in Fig. 4C those colors indicate high probability of only the strong spodic expression (2.0 class).

The distributions of all the MaxEnt probability data are heavily tailed, and follow a second order decay function. This is due to the fact that in a MaxEnt output cells are not independent (Merow et al., 2013b). Approximately 80% of the cells have probability of presence values ≤ 0.2 for all three models (Fig. 5). Because no cells are modeled as 0 probability in MaxEnt, probability values between the smallest value and 0.01 were used to approximate 0 probability. Using that approach, 25, 24, and 30% of the entire area had approximately 0 probability of presence for the 0.5–2.0, 1.0–2.0, and 2.0 spodic intensity class models respectively. Those areas are concentrated primarily in the four corners of the area, particularly in the northwest and southeast corners (Fig. 4).

Mean AUC values for the outputs of three spodic intensity classes (0.5–2.0, 1.0–2.0, 2.0) were identical at 0.96 (96%), and the standard deviations were all quite small, respectively, 0.003, 0.003, and 0.007. That the largest standard deviation was associated with the spodic intensity 2.0 class is not surprising given that it contained less than half the number of presence points than the other two classes (Table 2). Overall, the high AUC values and low standard deviations indicate excellent model performance (Pearson, 2007, 2010).

The results of the binary agreement/disagreement analysis of the MaxEnt model outputs indicate very high agreement (98–99%) among the three MaxEnt models (Table 5). The greatest amount of disagreement (cells predicted by one model as high probability, and predicted by another model as low probability) occurred for the spodic intensity class comparison of 0.5–2.0 vs. 2.0, which was due at least in part to those two models, respectively, having the greatest and fewest numbers of initial presence points (Table 2). However, it should be noted that only a small number of the cells had probabilities ≥ 0.6 (Table 6), so the agreement/disagreement results are reflective of only a small percentage of the entire area that was originally modeled in Fig. 4.

Interestingly, there are fewer cells with ≥ 0.6 probability in the 1.0–2.0 spodic intensity class than in the 2.0 spodic intensity class (or the 0.5 to 2.0 class) (Table 6), even though the 2.0 class had less than half the number of original presence sites than the 1.0–2.0 class (Table 2). This result suggests that for the original data there was a much narrower range of conditions for the environmental covariates for the 1.0 and 1.5 spodic intensity classes than for the 2.0 class, or that the covariates were most strongly associated with the strongest spodic characteristics (i.e., the 2.0 intensity class model). There were marked drops in permutation importance values in the 1.0–2.0 and 2.0 models, such that there were three variables for each of those two models that were more important than the other environmental covariates

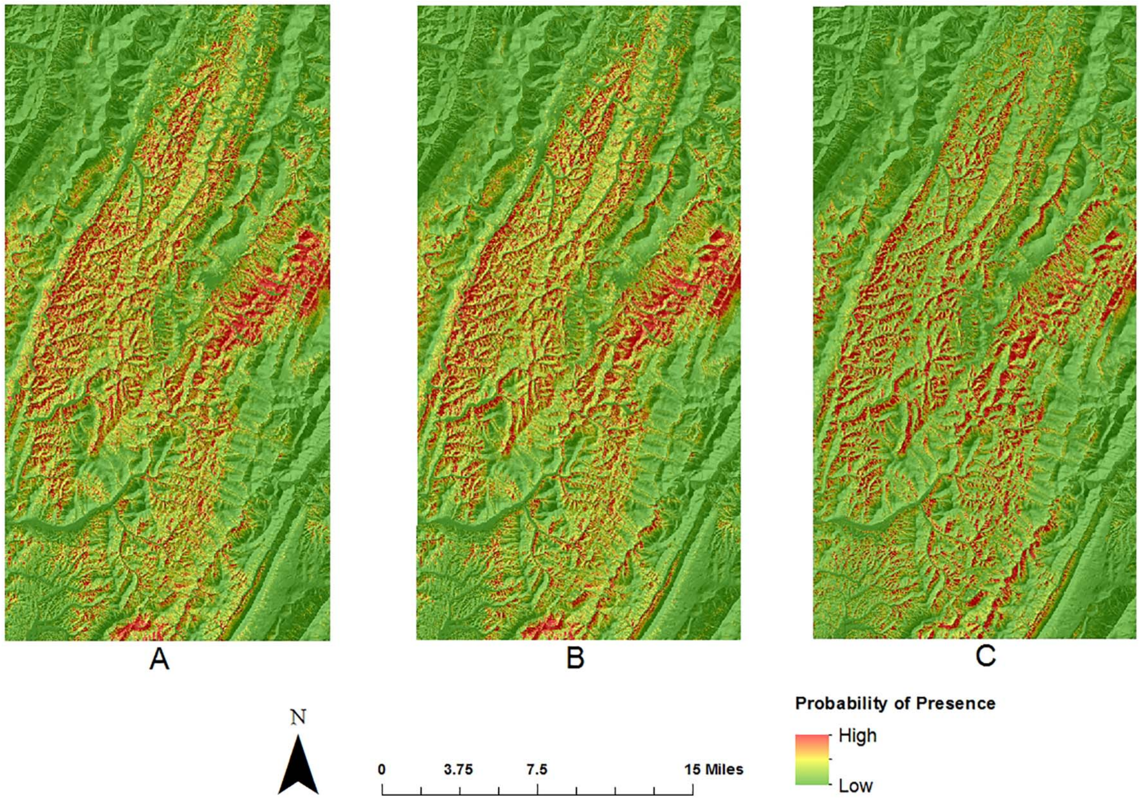

Fig. 4. The three MaxEnt spodic intensity model outputs: (A) spodic intensity class = 0.5–2.0, (B) spodic intensity class = 1.0–2.0, and (C) spodic intensity class = 2.0.

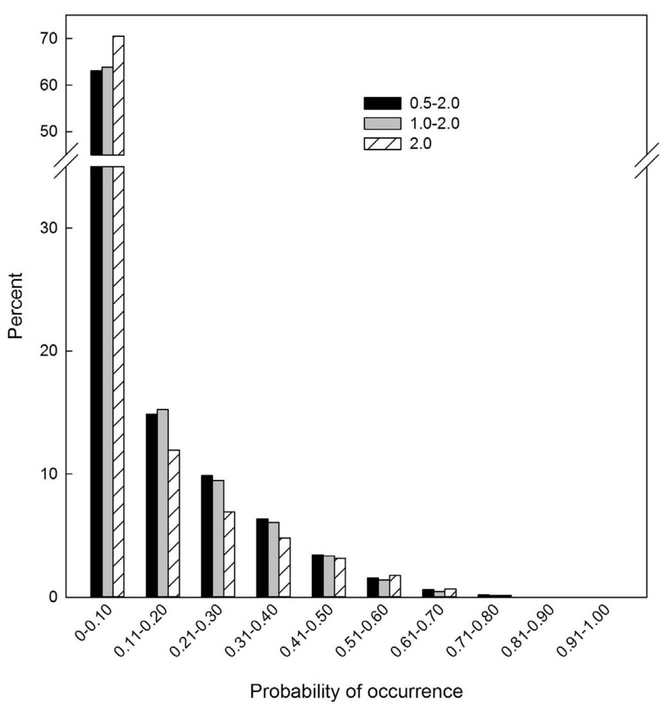

Fig. 5. Frequency distributions of MaxEnt probability of presence of spodic expression for the 0.5–2.0, 1.0–2.0, and 2.0 spodic intensity classes:

(Table 7). For the 0.5–2.0 class model, all the permutation importance values were relatively small and there were no distinct breaks across the 29 environmental variables (Table 7). Rather, according to the permutation importance metric, all covariates had similarly important influence. In all three models, an elevation-associated covariate was most important; *dem* for 0.5–2.0 and 1.0–2.0, and *baselevel* for 2.0. This is not surprising given that spodic characteristics are primarily associated with red spruce in this region, and red spruce is known to compete best at high elevations (Adams et al., 2010; Rentch et al., 2007).

Table 5
Percentages of agreement and disagreement probability of presence for the pairwise comparisons produced by the three spodic intensity classes (0.5–2.0, 1.0–2.0 and 2.0) using MaxEnt.

	Pairwise comparisons		
Modeled probability of presence (agreement or disagreement)	0.5–2.0 vs. 1.0–2.0	1.0–2.0 vs 2.0	0.5–2.0 vs 2.0
	% agreement or disagreement		
Both models predict low probability or both models predict high probability (agreement)	99.35	98.96	98.76
One model predicts high probability, one model predicts low probability (disagreement)	0.65	1.04	1.24

Table 6
Number of cells and percent of cells (or area) in the high probability (≥ 0.60) and low probability (< 0.6) classes for the MaxEnt models for each spodic intensity class.

Probability range	Spodic intensity range		
	0.5–2.0	1.0–2.0	2.0
		Number of cells	
Less than 0.6	1,601,664	1,605,569	1,601,847
Greater than or equal to 0.6	13,623	9718	13,440
		Percent of cells	
Less than 0.6	99.16	99.40	99.17
Greater than or equal to 0.6	0.84	0.60	0.83

Table 7
Permutation importance values for each of the 29 environmental covariates for the three MaxEnt spodic intensity class outputs. Variable definitions are provided in Table 3.

Spodic intensity range					
0.5–2.0		1.0–2.0		2.0	
Variable	Permutation importance	Variable	Permutation importance	Variable	Permutation importance
dem	9.4	dem	15.6	baselevel	13.2
relht70	7.7	mrvbf	8.8	relht50	12.9
slpos	6.6	relht70	8.3	mir	11.2
baselevel	5.5	relht20	4.5	eastness	5.4
mir	5.2	baselevel	4.4	southness	5.1
relht20	5.1	green	3.8	nwness	4.9
relht1	4.6	relht5	3.8	relht70	4.9
lsfactor	4.5	eastness	3.6	mrvbf	3.3
relht50	4.4	nwness	3.5	green	3.2
eastness	4.3	mir	3.5	slope2	3.0
nir	3.9	nir	3.2	relht5	3.0
green	3.4	slpos	3.2	contribare	2.8
relht5	3.2	relht10	2.8	neness	2.7
nwness	3.1	mrrtf	2.8	relht3	2.6
mrvbf	3.1	relht50	2.7	relht1	2.5
relht10	2.5	relht3	2.6	aacn2	2.5
slope2	2.5	aacn2	2.6	relht20	2.4
aacn2	2.4	relht30	2.2	relht2	2.2
plan_curv	2.3	southness	2.2	relht30	2.2
southness	2.3	convergenc	2.2	slpos	2.1
convergenc	2.1	relht1	2.0	relht10	2.1
relht3	1.8	plan_curv	1.9	dem	1.5
relht30	1.7	prof_curv	1.8	mrrtf	1.1
mrrtf	1.7	relht2	1.8	plan_curv	0.9
contribare	1.6	slope2	1.6	prof_curv	0.7
twi	1.5	contribare	1.4	convergenc	0.5
prof_curv	1.3	lsfactor	1.4	nir	0.4
neness	1.1	neness	1.0	lsfactor	0.4
relht2	1.0	twi	0.7	twi	0.3

3.2. MaxEnt and random forests output comparisons

The MaxEnt and random forests outputs are shown in Fig. 6. The highest probabilities in random forests output are distributed relatively evenly over the entire study area, compared to MaxEnt, which as noted previously, are primarily concentrated in the area where most of the original soil samples were collected. The agreement/disagreement comparison of 40% highest probability cells for the MaxEnt and random forests output showed approximately 62% agreement and 38% disagreement between the models (Table 8). Much of the concentration of cells where the models disagreed were again in the corners of the study area (Fig. 7).

The mean, minimum, and maximum values for each of the environmental covariates for the top 40% of highest probabilities of spodic occurrence in which there was agreement between the MaxEnt and random forests model are shown in Table 9. The same statistics are presented for each model for the top 40% of cells in which they did not agree in Table 10. The Wilcoxon rank sum test indicate that there was a significant difference ($p < 0.0001$) between the two models for every covariate.

4. Discussion

All three of the MaxEnt models yielded very similar results spatially and with respect to their binary predictions of spodic presence. The model was not overly sensitive to the degree of spodic expression for any of the three data sets, whether including the entire range of spodic intensities or only the highest degree of spodic intensity. This similarity suggests that even though MaxEnt was developed as a species distribution model, it has application in physical analyses and is capable of identifying spodic potential.

The environmental covariates used in this examination had only minor influence on contributing to the spodic intensity predictions in MaxEnt. The highest permutation importance value was 15.6% for the spodic intensity 1.0–2.0 class model. This value is fairly low compared to permutation importance values typically obtained for many of the more traditional species distribution model applications described in the literature (e.g., Adhikari et al., 2012; Brambilla and Ficetola, 2012; Smart et al., 2012). The decreases in permutation importance values for the 2.0 intensity class (i.e., that included only strong spodic expression) showed three variables (*baselevel*, *relht50*, and *mir*) that were useful for model prediction (Table 7), but even these were not particularly strong relationships based on the permutation values and degree of drop over those three values. It is possible that the environmental covariates used are at scales that are too coarse or they have little influence on processes required for podzolization. The environmental covariates were least sensitive when presence data included the widest possible spodic intensity range. For any covariate to be identified as important in this situation, it would have to be uniquely associated with Spodosol formation or presence, including weak expression, while simultaneously not associated with locations where spodic expression is absent.

The highest probabilities for all three of the MaxEnt outputs (Fig. 4) tend to be concentrated near the area of the original model input presence points (Fig. 2). Farther from those points, such as toward the northwest and southeastern corners of the areas, the probabilities were predominantly approximately 0. This is likely due to the fact that these areas were not sampled as intensively as the central portion of the study area. MaxEnt is known to be constrained by sampling bias, so it performs poorly when predicting outside of the range of conditions from which the original presence data were collected (Elith et al., 2011; Hernandez et al., 2006; Kramer-Schadt et al., 2013). Consequently, it follows that even if environmental conditions in those locations were conducive to spodic presence, no observations were made there, so the full range of conditions that promote spodic presence were not sampled, limiting the modeled probability of presence in those areas. This is an important consideration for future MaxEnt modeling efforts in soil

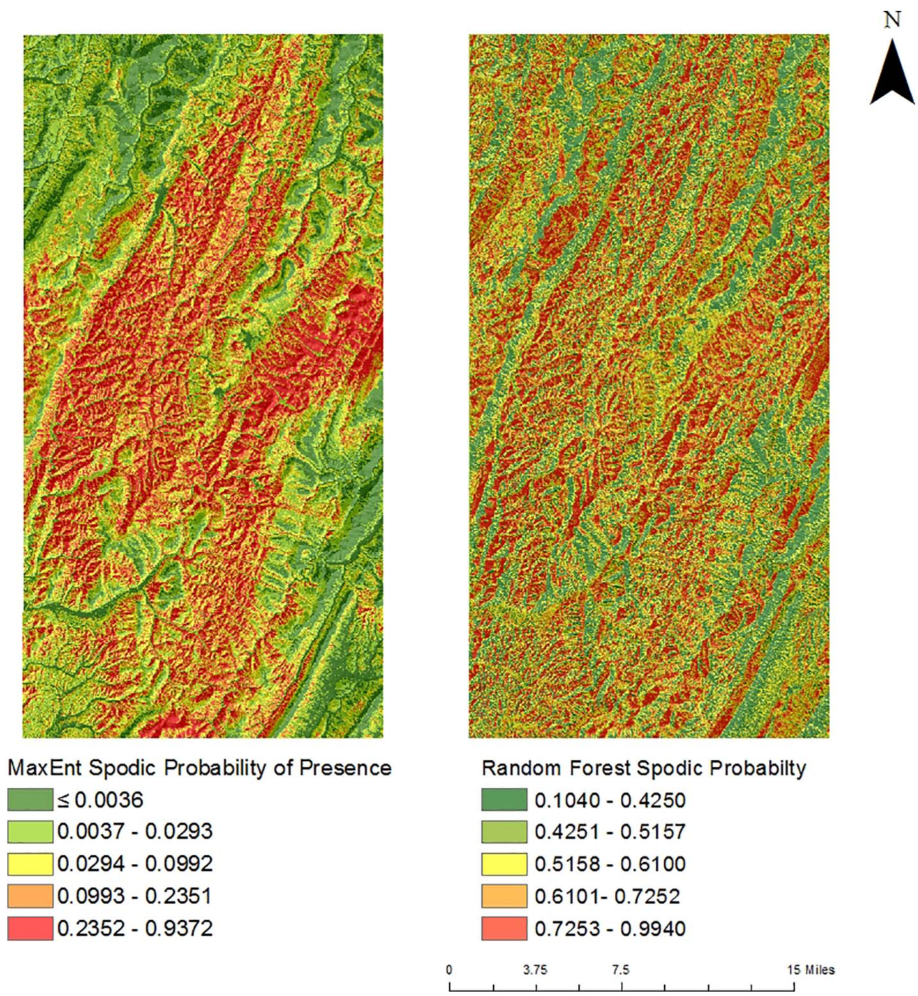

Fig. 6. Comparison of MaxEnt (A) and random forests (B) outputs. The MaxEnt model employed presence data using the range of 0.5–2.0 spodic intensity, while random forests used 0–2.0 spodic intensity data, since random forests includes absence data. Cells were separated into five equal quantiles to compare these outputs by rank. Colors represent probability that spodic properties exist, not intensity of spodic expression.

Table 8
Percentages of agreement and disagreement probability of presence for the MaxEnt-random forests comparison.

	Agreement	Disagreement
% agreement	61.85	38.15
# of cells	403,901	249,152

applications. Sampling for the dependent variable should extend outside the area of interest in all directions to adequately capture predictions over the area of interest.

One visually apparent difference in the MaxEnt modeled outputs between the spodic intensity class that contained only a single rating (2.0) and the other two classes that included ranges of spodic intensities is the amount of area mapped as transitional areas (i.e., yellow shades) (Fig. 4). The output for the single spodic rating 2.0 (Fig. 4C) has very little transitional area; instead red shades tend to transition directly to green shades. By comparison, the 0.5–2.0 and 1.0–2.0 modeled outputs have substantially more area in those transitional yellow shades (Fig. 4a and b). The frequency distributions of the three models (Fig. 5) show this response more quantitatively—the number of cells in the 2.0 class for the low probabilities (≤ 0.1) are greater than the other two classes, and there are more cells for the mid-range values (i.e., yellows) of the 2.0 intensity class than for the two other classes. It is not known which of these representations (more or less transitional presence) is more accurate, but the presence of transitional areas where spodic expression is not at its maximum seems intuitive and therefore reasonable. Regardless, these differences among the MaxEnt results indicate that a range of values for the variable utilized as presence data (e.g., spodic intensity classes, species abundance, density, etc.) can influence MaxEnt outputs differently than presence variables that do not include a range of values (the variable is either present or not, such as the 2.0 intensity class alone), at least when covariates do not exert strong influence on the model outputs.

Even though metrics for MaxEnt output generated here and random forests output generated by Nauman et al. (2015a) indicated good to excellent model performance for predicting spodic expression, there was still substantial amount of area in disagreement between the two models. Where the high probability cells for the two models were in disagreement, all of the values of the environmental covariates were significantly different between models. In terms of magnitude, the contributing area (*contribare*) was the environmental variable that was most different for cells that were not in agreement between the two models. The mean and maximum contributing area (m^2) for MaxEnt cells were both an order of magnitude smaller than the mean and maximum contributing areas for random forests (Table 10). MaxEnt tends to predict the higher probabilities for spodic characteristics higher up in watersheds. Another striking difference between the two models for high probability (i.e., top 40%) predictions pertains to aspect. The mean values for *eastness*, *nwness*, *southness*, and *neness* from MaxEnt were associated with the opposing aspect than that for random forests, as denoted by the opposite signs of each between models (Table 10).

The two models clearly resulted in different extent and location of spodic presence; however, a comparison between these two models was not a primary focus of this study. The model which most accurately

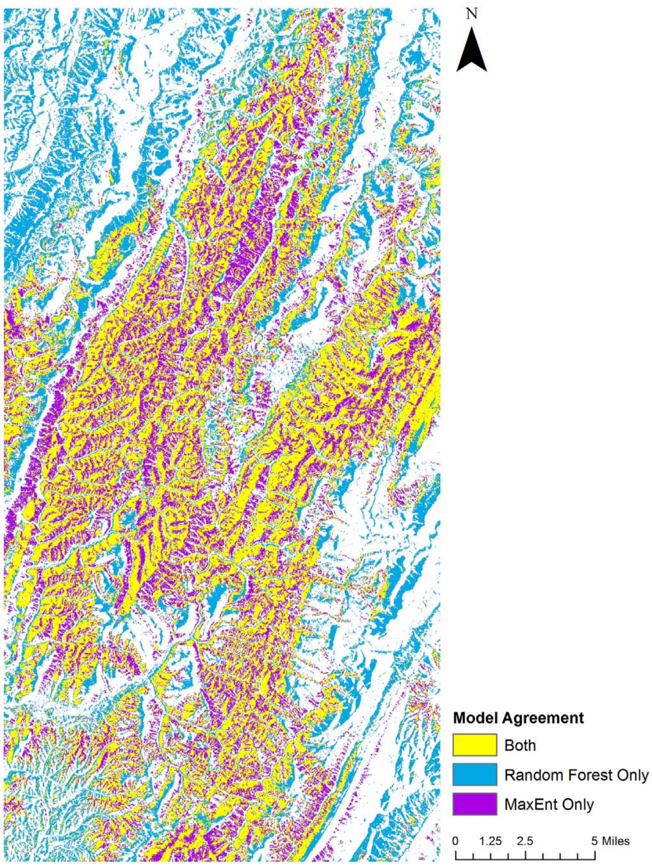

Fig. 7. Spatial comparison of the MaxEnt and random forests agreement and disagreement, using the top 40% of probabilities for each model. White areas are cells in the lower 60% of probabilities for both the MaxEnt and random forests outputs, so they were not included in this comparison.

Table 9
Mean, minimum, and maximum environmental variable values for the top 40% of highest probability cells where MaxEnt and random forests models were in agreement.

Variable	Mean	Minimum	Maximum
aacn2	64.56	0	463.18
baselevel	998.22	633.88	1263.65
contribare	5515	772	29,312,400
convergenc	3.87	− 61.21	93.73
dem	1062.78	645.22	1461.23
eastness	− 0.49	− 1.00	1.00
green	43.44	0	247.30
lsfactor	3.74	0	17.81
mir	56.20	0	255.00
mrrtf	0.11	0	3.90
mrvbf	0.05	0	3.96
neness	− 0.21	− 1.00	1.00
nir	134.61	0	255.00
nwness	0.48	− 1.00	1.00
plan_curv	0.00055	− 0.0055	0.0082
prof_curv	0.00007	− 0.0088	0.0085
relht1	6.75	0	24.56
relht10	53.68	0	165.69
relht2	13.33	0	45.86
relht20	77.38	0	275.95
relht3	20.51	0	70.22
relht30	93.12	0.17	360.40
relht5	32.84	0	94.59
relht50	116.26	0.96	491.42
relht70	135.74	0.96	576.65
slope2	0.248	0.000	0.759
slpos	50.04	0	102.00
southness	− 0.20	− 1.00	1.00
twi	5.77	3.18	21.82

describes current presence cannot be determined without some field verification. Yet even field verification of spodic intensity presents a conundrum for validation because a lack of spodic intensity may be due to complete depodzolization or that podzolization never occurred in that location. Spodic characteristics can be determined chemically after visual indicators of spodic morphology have degraded (Soil Survey Staff, 2003), but it is not known how long these chemical signatures remain after spodic expression is gone, and these techniques are cost prohibitive to undertake on a landscape scale.

A pressing research need is to understand where podzolization could possibly occur to inform red spruce restoration—specifically, to identify landscape locations that have the greatest potential for successful expansion of red spruce. There are many reason for encouraging its expansion; in addition to providing greater forest diversity and highly-valued recreational opportunities, this forest type is intimately tied to and essential for the survival of threatened and endangered wildlife species (Menzel and Ford, 2004; Dillard et al., 2008; Pauley, 2008).

The potential for depodzolization creates more difficulty for modeling spodic potential than many typical modeling applications in which presence of a specific modeled feature or attribute can be verified by field validation. As such, success in modeling spodic potential, and thus, conditions conducive to red spruce survival, may require the use of environmental variables that include fine-scale microclimate metrics as these may be more discerning for identifying conditions associated with spodic formation, and for differentiating between weak spodic presence and spodic absence. Inclusion of microclimate variables may yield models that are more robust and effective at predicting the occurrence of Spodosols. In addition, rather than relying solely on physical attributes, other local biological or chemical soil characteristics may be more robust and effective for informing MaxEnt about spodic prediction.

5. Conclusion

The probability of occurrence of spodic expression was modeled in a portion of the Monongahela National Forest in eastern West Virginia using the presence-only species distribution model MaxEnt. Three different spodic intensity classes were modeled, but the results were relatively similar, suggesting that MaxEnt modeling may be a useful tool for predicting spodic expression—the model was not overly sensitive to the degree of spodic intensity in its prediction, and the use of presence-only data may reduce the risks associated with including false absences that can occur in presence-absence models. The primary difference among the MaxEnt outputs was the model employing only the strongest spodic intensity values tended to have more abrupt transitions between cells with high probability of occurrence and those with low probability compared to the models that employed wider ranges of spodic intensity.

Individually, the 29 environmental covariates did not contribute substantially to any of the MaxEnt the modeled outputs, though the two models that excluded the very weak spodic intensity class found a few variables with some influence. Most of these were related in one way or another to elevation, which was expected because podzolization in this region is associated historically with red spruce presence, and red spruce is primarily found at high elevations.

The MaxEnt output determined from the widest spodic intensity class was compared to results from random forests, which employed presence and absence data from the same data set. Outputs using the 40% of cells with the highest probability of occurrence from both models were compared. Approximately 60% of those cells were in agreement between the two models. Areas of disagreement were primarily concentrated in areas that were located far from the original soil sampling points.

Covariates may not have been unique enough to the conditions that control spodic formation, which may have caused the prediction of low probabilities with distance from the original soil sampling locations. It is possible that more important environmental variables might include

Table 10
Mean, minimum, and maximum environmental variable values for the top 40% of highest probability cells where MaxEnt and random forests models did not agree. All means were significantly different, with Wilcoxon rank-sum test probability values < 0.0001.

Variable	MaxEnt			Random forests		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
aacn2	65.33	0	409.02	100.23	0	463.89
baselevel	993.85	633.45	1265.19	903.76	616.07	1224.17
contribare	14,380	772	65,630,600	130,037	772	177,626,000
convergenc	2.02	− 99.41	97.37	2.07	− 90.71	96.04
dem	1059.18	646.94	1465.21	1003.99	616.07	1458.14
eastness	0.28	− 1.00	1.00	− 0.59	− 1.00	1.00
green	55.41	0	255.00	59.34	0	255.00
lsfactor	4.19	0	18.59	3.71	0	23.09
mir	79.00	0	255.00	71.17	0	255.00
mrrtf	0.12	0	4.60	0.12	0	4.16
mrvbf	0.07	0	3.98	0.18	0	4.98
neness	0.11	− 1.00	1.00	− 0.25	− 1.00	1.00
nir	153.86	0	255.00	141.65	0	255.00
nwness	− 0.29	− 1.00	1.00	0.59	− 1.00	1.00
plan_curv	0.00	− 0.01	0.01	0.00040	− 0.01	0.01
prof_curv	0.00	− 0.01	0.01	− 0.00001	− 0.01	0.01
relht1	7.01	0	28.86	6.27	0	28.13
relht10	51.59	0	160.91	62.47	0	198.02
relht2	13.53	0	50.05	12.47	0	47.15
relht20	75.35	0	263.75	106.41	0	335.43
relht3	20.50	0	66.52	19.66	0	66.28
relht30	91.62	0	358.32	135.89	0	431.26
relht5	32.04	0	101.19	33.29	0	98.31
relht50	114.58	0	492.49	173.77	0	502.49
relht70	132.80	0	578.24	196.40	0	574.64
slope2	0.264	0.000	0.751	0.234	0.000	0.815
slpos	49.72	0	103.00	51.04	− 1.00	102.00
southness	0.13	− 1.00	1.00	− 0.24	− 1.00	1.00
twi	5.93	3.31	21.33	6.26	0	22.78

site-specific biological and chemical soil characteristics, rather than just coarse-resolution environmental characteristics alone. The overall interest in predicting spodic potential is to apply this information to identify areas for red spruce restoration. Because this forest type is highly valued for many reasons, including providing habitat for several threatened and endangered species, further investigation into model improvement will likely continue.

References

Adams, H.S., Stephenson, S.L., Rollins, A.W., Adams, M.B., 2010. The isolated red spruce communities of Virginia and West Virginia. In: Rentch, J.S., Schuler, T.M. (Eds.), Proceedings from the Conference on the Ecology and Management of High-Elevation Forests in the Central and Southern Appalachian Mountains. GTR-NRS-P-64. USDA Forest Service, Northern Research Station, Newtown Square, PA, pp. 1–12.

Adhikari, D., Barik, S.K., Upadhaya, K., 2012. Habitat distribution modeling for reintroduction of *Ilex khasiana* Purk., a critically endangered tree species for northeastern India. Ecol. Eng. 40, 37–43.

Baldwin, R.A., 2009. Use of maximum entropy modeling in wildlife research. Entropy 11 (4), 854–866.

Barrett, L.R., Schaetzl, R.J., 1998. Regressive pedogenesis following a century of deforestation: evidence for depodzolization. Soil Sci. 163 (6), 482–497.

Beane, N.R., Rentch, J.S., Schuler, T.M., 2013. Using Maximum Entropy Modeling to Identify and Prioritize Red Spruce Forest Habitat in West Virginia. RP-NRS-23. USDA Forest Service, Northern Research Station, Newtown Square, PA.

Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S., 2010. Current state of digital soil mapping and what is next. In: Hartemink, A.E., McBratney, A.B. (Eds.), Progress in Soil Science, Digital Soil Mapping: Bridging Research, Environmental Application and Operation. Springer, New York, pp. 4–10.

Brambilla, M., Ficetola, G.F., 2012. Species distribution models as a tool to estimate reproductive parameters: a case study with a passerine bird species. J. Anim. Ecol. 81, 781–787.

Clarkson, R.B., 1964. Tumult on the Mountains: Lumbering in West Virginia. McClain Printing Co., Parsons, WV, pp. 1770–1920.

Dillard, L.O., Russell, K.R., Ford, W.M., 2008. Macrohabitat models of occurrence for the threatened cheat mountain salamander, *Plethodon nettingi*. Appl. Herpetol. 5 (3), 201–224.

Dokuchaev, V.V., 1999. On the concept of natural zones — St. Petersburg, 1899. Eurasian Soil Sci. 32, 726–727.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17 (1), 43–57.

Hatfield, D.C., 1996. TopoTools - a Collection of Topographic Modeling Tools for ArcINFO: Slope Position. US Forest Service, ESRI.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29 (5), 773–785.

Hole, F.D., 1975. Some relationships between forest vegetation and podzol B horizons in soils of Menominee tribal lands, Wisconsin, USA. Sov. Soil Sci. 7, 714–723.

Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York.

Jimenez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modeling. Divers. Distrib. 14, 1372–1381.

Kalle, R., Ramesh, T., Qureshi, Q., Sankar, K., 2013. Predicting the distribution pattern of small carnivores in response to environmental factors in the western Ghats. PLoS One 8 (11), e79295. http://dx.doi.org/10.1371/journal.pone.0079295.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schroder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, Sm, Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Biodivers. Res. 19, 1366–1379.

Kuhnert, M., Voinov, A., Seppelt, R., 2005. Comparing raster map comparison algorithms for spatial modeling and analysis. Photogramm. Eng. Remote. Sens. 71, 975–984.

Lewis, R.L., 1998. Transforming the Appalachian Countryside: Railroads, Deforestation, and Social Change in West Virginia, 1880–1920. University of North Carolina, Chapel Hill, NC.

Lietzke, D.A., McGuire, G.A., 1987. Characterization and classification of soils with spodic morphology in the southern Appalachians. Soil Sci. Soc. Am. J. 51, 165–170.

Lin, H., 2011. Three principles of soil change and pedogenesis in time and space. Soil Sci. Soc. Am. J. 75 (6), 2049–2070.

Lundström, U.S., van Breemen, N., Bain, D., 2000a. The podzolization process. A review. Geoderma 94, 91–107.

Lundström, U.S., van Breemen, N., Bain, D.C., van Hees, P.A.W., Giesler, R., Gustafsson, J.P., Ilvesniemi, H., Karltun, E., Melkerud, P.A., Olsson, M., Riise, G., Wahlberg, O., Bergelin, A., Bishop, K., Finlay, R., Jongmans, A.G., Magnusson, T., Mannerkoski, H., Nordgren, A., Nyberg, L., Starr, M., Tau Strand, L., 2000b. Advances in understanding the podsolization process resulting from a multidisciplinary study of three coniferous forest soils in the Nordic countries. Geoderma 94, 335–353.

McBratney, A.B., Santos Mendoca, M.L., Minasny, B., 2003. On digital soil mapping.

Geoderma 117, 3–52.
Menzel, J.M., Ford, W.M., 2004. Nest tree use by the endangered Virginia northern flying squirrel in the central Appalachian Mountains. Am. Midl. Nat. 151 (2), 355–368.
Merow, C., Smith, M.J., Silander, J.A., 2013a. A practical guide to MaxEnt for modeling species' distributions: what it does and why inputs and settings matter. Ecography 36 (10), 1058–1069.
Merow, C., Smith, M.J., Silander, J.A., 2013b. A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does and Why Inputs and Settings Matter. http://ww.oikosoffice.lu.se/appendix ((accessed 17.04.03) Appendix ECOG-07872, Appendix 1–6).
Monger, C., Rachal, D.M., 2013. Soil and landscape memory of climate change - how sensitive, how connected? Soc. Sediment. Geol. 104, 63–70.
Nauman, T.W., 2015. Spatializing the Soil-ecological Factorial: Data Driven Integrated Land Management Tools. PhD Diss.. West Virginia University, Morgantown, WV.
Nauman, T.W., Thompson, J.A., Teets, S.J., Dilliplane, T.A., Bell, J.W., Connolly, S.J., Liebermann, H.J., Yoast, K.M., 2015a. Ghosts of the forest: mapping pedomemory to guide restoration. Geoderma 247, 51–64.
Nauman, T.W., Thompson, J.A., Teets, J., Dilliplane, T., Bell, J.W., Connolly, S.J., Liebermann, H.J., Yoast, K., 2015b. Pedoecological modeling to guide forest restoration using ecological site descriptions. Soil Sci. Soc. Am. Jo. 79 (5), 1406–1419.
NOAA NCDC, 2016. Northeast Regional Climate Center. https://www.ncdc.noaa.gov/climate-information/regional (accessed 16.12.05).
Nornberg, P., Sloth, L., Neilson, K.E., 1993. Rapid changes of sandy soils caused by vegetation changes. Can. J. Soil Sci. 73 (4), 459–468.
Oosting, H.J., Billings, W.D., 1951. A comparison of virgin spruce-fir forest in the northern and southern Appalachian system. Ecology 36 (1), 82–103.
Pauley, T.K., 2008. The Appalachian inferno: historical causes for the disjunct distribution of *plethodon nettingi* (Cheat Mountain Salamander). Northeast. Nat. 15 (4), 595–606.
Pearson, R.G., 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. American Museum of Natural History, New York.
Pearson, R.G., 2010. Species' distribution modeling for conservation educators and practitioners. Lessons Conserv. 3, 54–89.
Phillips, S.J., 2017. A Brief Tutorial on Maxent. Available from url:. http://biodiversityinformatics.amnh.org/open_source/maxent/, Accessed date: 17 November 2002.
Phillips, S.J., Dudik, M., 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. Ecography 31 (2), 161–175.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.
Rentch, J.S., Schuler, T.M., Ford, M.W., Nowacki, G.J., 2007. Red spruce stand dynamics, simulations and restoration opportunities in the Central Appalachians. Restor. Ecol. 15 (3), 440–452.
Sauer, D., Sponagel, H., Sommer, M., Giani, L., Jahn, R., Strahr, K., 2007. Podzol: soil of the year 2007. A review on its genesis, occurrence and functions. J. Plant Nutr. Soil Sci. 170, 581–597.
Schaetzl, R.J., Isard, S.A., 1996. Regional-scale relationships between climate and strength of podzolization in the Great Lakes region, North America. Catena 28, 47–69.
Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., Soil Survey Staff, 2012. Field Book for Describing and Sampling Soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
Shigo, A.L., 1972. The beech bark disease today in the northeastern US. J. For. 70, 286–289.
Smart, L.S., Swenson, J.J., Christensen, N.L., Sexton, J.O., 2012. Three-dimensional characterization of pine forest type and red-cockaded woodpecker habitat by small-footprint, discrete-return lidar. For. Ecol. Manag. 281, 100–110.
Soil Survey Staff, 2003. Soil Taxonomy - A Basis System of Soil Classification for Making and Interpreting Soil Surveys. US Government Printing Office, Washington, DC.
Stanley, S.R., Ciolkosz, E.J., 1981. Classification and genesis of Spodosols in the Central Appalachians. Soil Sci. Soc. Am. J. 45, 912–917.
Svenning, J.C., Skov, F., 2004. Limited filling of the potential range in European tree species. Ecol. Lett. 7, 565–573.
Targulian, V.O., Goryachkin, S.V., 2010. Soil memory and environmental reconstructions. Eurasian Soil Sci. 44 (4), 464–465.
Visser, H., de Nijs, T., 2005. The map comparison kit. Environ. Model. Softw. 21 (3), 346–358.
West Virginia Geological and Economic Survey, 1968. Surface Geology - Rock Units. Digitized Map. http://wvgis.wvu.edu/ (accessed 17.04.03).
Young, N., Carter, L., Evangelista, P., 2011. A MaxEnt Model v3.3.3e Tutorial (ArcGIS v10). Natural Resource Ecology Laboratory, Colorado State University. http://ibis.colostate.edu/webcontent/ws/coloradoview/tutorialsdownloads/a_maxent_model_v7.pdf (accessed 17.04.03).