BioOne COMPLETE

Highlighting an Overlooked Hotspot for American Ginseng (Panax quinquefolius) in Virginia

Authors: Peters, Jacob D.J., and Griscom, Heather P.

Source: Natural Areas Journal, 42(2) : 109-119

Published By: Natural Areas Association

URL: https://doi.org/10.3375/21-25

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

Research Article

Highlighting an Overlooked Hotspot for American Ginseng (*Panax quinquefolius*) in Virginia

Jacob D.J. Peters[1,2] and Heather P. Griscom[1]

[1]James Madison University, 800 S. Main Street, Harrisonburg, VA 22807

[2]Corresponding author: peters.jdj@gmail.com; 276-219-1757; ORCID ID: 0000-0001-9387-3983

Associate Editor: Anatoliy Khapugin

ABSTRACT

American ginseng (*Panax quinquefolius*) is a globally desired medicinal plant that is becoming increasingly difficult to study due to harvest-induced rarity. Thus, this species' conservation could be greatly improved via species distribution models, making it a model organism for studying sampling bias. In an attempt to refine a state-derived distribution model for ginseng in Virginia, we conducted additional surveys in a biologically diverse yet under-sampled region of the state—the Cumberland Mountains—thereby increasing the number of documented ginseng occurrences in this region thirteen-fold ($N_1 = 16$, $N_2 = 214$). Our surveys resulted in the model predicting an increased probability of American ginseng occurrence not only statewide ($\mu_1 = 0.099$, $\mu_2 = 0.104$) but particularly so in the Cumberland Mountains ($\mu_1 = 0.170$, $\mu_2 = 0.278$), highlighting a consistently overlooked hotspot for biodiversity in Southern Appalachia. We suggest that more geographically balanced surveys and reduced overrepresentation of heavily protected and managed areas such as National Parks—in addition to heeding local knowledge—can be an effective method of mitigating geographic bias in predictions from species distribution models.

Index terms: Appalachia; biodiversity hotspots; non-timber forest products; sampling bias; sampling methods; species distribution models

INTRODUCTION

The herbaceous layer of a forest contributes significantly to the species diversity therein, has been shown to be linked positively with soil health, and provides food and shelter to many organisms (Gilliam and Roberts 2003; Gilliam 2007; Eisenhauer et al. 2011). As a result, the herbaceous layer may deserve more than being a dependent variable and perhaps is crucial to researchers and landowners attempting to understand and protect temperate forests (Gilliam and Roberts 2003). In North American deciduous forests, up to 80% of plant species can be found within the herbaceous layer, and the extinction rate of forest herbs is several times greater than that of canopy species, making them worthy of study in the realm of conservation (Levin and Wilson 1976; Gilliam 2007). In this study, we focus on the notable and highly desired non-timber forest product, American ginseng (*Panax quinquefolius* L.).

Wild ginseng populations tend to form clusters ranging from just a few to over 100 plants within several square meters, making this species particularly susceptible to overharvest (Cruse-Sanders and Hamrick 2004; Mooney and McGraw 2007). Additionally, ginseng is facing increased threat of extinction due to climate change, herbivory, and habitat loss (McGraw et al. 2003; Furedi and McGraw 2004; Case et al. 2007; Farrington et al. 2009; McGraw et al. 2013; Souther and McGraw 2014; McGraw and Chandler 2018). As a result, ginseng is a model organism for studying species distribution and population ecology, as well as impacts of landscape change, deer, climate change, and sampling bias.

When attempting to protect and conserve at-risk species such as ginseng, we must identify the habitat in which those species thrive and locate wild populations. This can be accomplished with species distribution models, which can facilitate more efficient field surveys and the location of wild plants in remote areas (Guisan and Thuiller 2005; Van Manen et al. 2005; Hackett et al. 2020). Predictive models implemented as conservation tools can simplify the work of locating, growing, or harvesting ginseng—in addition to aiding restoration and conservation of other taxa, as has been demonstrated by several studies (Evans et al. 2011; Oppel et al. 2012; Reynolds et al. 2016; Mohapatra et al. 2019; Hackett et al. 2020). However, models representing large geographies with diverse ecotypes may be of limited use in unique and under-sampled regions.

Recently, a species distribution model for American ginseng was constructed by the Virginia Department of Conservation and Recreation's Division of Natural Heritage (DCR-DNH), which provided ginseng presence data derived from vegetation surveys at 4727 long-term natural community study plots (typically 20×20 m in size) and a suite of 78 environmental variables used here (Virginia Natural Heritage Program 2017; Fleming and Patterson 2021; Virginia Department of Conservation and Recreation 2021). It should be noted that ginseng was not the specific target of these surveys, and DCR-DNH's species distribution models are made using data from these long-term plots that are placed in specific areas to capture different ecological communities (Fleming and Patterson 2021; Virginia Department of Conservation and Recreation 2021). Their model incorporated 1955 documented ginseng occurrences from study sites across the entire state of Virginia—spanning an area of around 100,000 km^2.

Virginia can be broken down into five physiographic provinces, each of which was sampled by DCR-DNH (Figure 1). Provinces are distinguished based on the underlying geology, which has a large impact on the soil and biological community

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

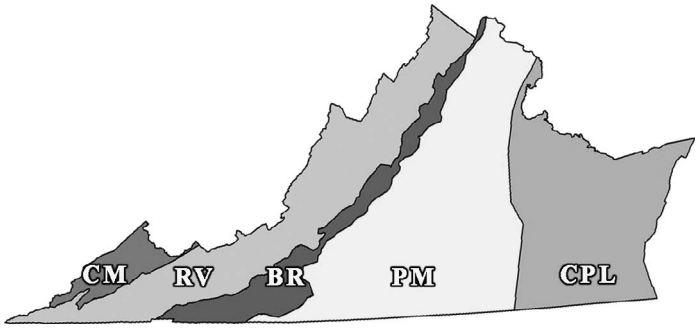

Figure 1.—Physiographic provinces of Virginia (simplified boundaries). CM = Cumberland Mountains, RV = Ridge and Valley, BR = Blue Ridge Mountains, PM = Piedmont, CPL = Coastal Plain.

composition (Wilson and Tuberville 2003). In this study, we focus on the two mountainous provinces of Virginia that are commonly thought to contain ginseng habitat: the Blue Ridge Mountains and the Cumberland Mountains.

The Blue Ridge province of Virginia covers ~9000 km^2 and consists of a narrow (5–50 mile wide) mountain range generally running north to south. According to Wilson and Tuberville (2003), this province is underlain by weather-resistant granites and metabasalts, with steep slopes, rugged peaks, and elevations ranging from 1000 feet (ft) to 5729 ft at the highest mountain peak in the state. Due to this variation in elevation, the Blue Ridge harbors various forest types including mixed mesophytic, xeric oak, heath and pine-oak, rich cove, hemlock, oak-hickory, and northern hardwood forests, among others—making this a noteworthy region to focus ecological studies (Wilson and Tuberville 2003). In DCR's initial species distribution model for ginseng, this province was represented by 744 documented ginseng occurrences derived from surveys at 949 long-term study plots. This is not surprising, as the Blue Ridge contains large amounts of protected land such as the ~800 km^2 Shenandoah National Park and ~1300 km^2 of national forest.

All physiographic provinces were included in sampling and used for modeling ginseng distribution in Virginia. The Ridge and Valley, Piedmont, and Coastal Plain were represented by 586, 90, and 519 documented occurrences of ginseng, respectively. In contrast, only 77 of the 4727 long-term study plots are present in the Cumberland Mountains, resulting in only 16 documented occurrences of ginseng in this region (~4000 km^2) by DCR-DNH. According to Wilson and Tuberville (2003), the Cumberland Mountains, a subregion of the Appalachian Plateaus with underlying beds of mostly sandstone and shale, is a region of rugged mountains with branching valleys and elevations ranging from around 1000 to over 4000 ft. This province is mainly composed of mixed mesophytic forests in the form of rich cove, acidic cove, and oak-hickory forests, and contains less protected land in comparison to the Blue Ridge, with only ~650 km^2 of national forest (Wilson and Tuberville 2003; Virginia Natural Heritage Program 2017). From our initial predictive model using these data, the probability of ginseng presence across the Cumberland Mountains was only half that of the Blue Ridge—indicating a sampling deficiency in the former.

Uneven detection rates and geographical sampling bias are common issues across large areas and can have impacts on the predictive models constructed with those data (Stockwell and Peterson 2002; Phillips et al. 2009; Hui et al. 2011; Syfert et al. 2013). Contrary to initial model predictions, researchers involved in this study, state botanists, and locals suspect that the Cumberland Mountains harbor as much if not more American ginseng than any other region in the state, as productive mesophytic forests are present throughout. In many cases, having a wider distribution of survey sites across a study region could result in more accurate predictions for species distribution and suitable habitat. Identifying habitat characteristics in which ginseng best competes could enable more successful reintroductions and aid in its conservation.

We hypothesized that the presence of American ginseng is as probable in the Cumberland Mountains of southwestern Virginia as it is in the heavily protected lands of the Blue Ridge. This is not apparent in current models that are being used for conservation efforts, likely due to sampling deficiencies coupled with the unique physiography and microclimates in this region. The goal of these surveys was to improve upon the initial model's representation of the Cumberland Mountains by increasing field sampling in that region to highlight the simplest means of increasing accuracy when modeling appropriate ginseng habitat and distribution across Virginia, thus informing conservation and restoration undertakings.

METHODS

Sampling Ginseng Occurrence

We conducted surveys in random locations representing various habitat types throughout the Cumberland Mountains in an attempt to reduce biased field surveys of our own. We suspected that due to ginseng's apparent rarity, the probability of individuals being present on randomized transects would be low. To mitigate this, random points were generated within National Forest, State Parks, and select private properties (total area of ~800 km^2) to guide time-limited surveys in the summer of 2019. Any ginseng plants identified in the field while traveling to or from these random points were documented. We did not attempt to define or measure boundaries for ginseng populations. Each survey guided by these random points was limited to a single day of sampling (6 hr maximum spent surveying), to reduce the likelihood of one area being sampled much more heavily than another (totaling >25 surveys with 14 successful survey-days). The specific number of unsuccessful survey-days was not noted, although we suggest that doing so may be useful for better documenting detection rates and survey effort.

For each survey, we began at a road or trail and navigated to our randomly generated point. Any ginseng plants identified in the field while traveling to or from these random points were documented in an attempt to ensure wild plants were not ignored. We did not attempt to define populations or metapopulations in this study; rather, we focused simply on identifying individuals across the landscape. When ginseng individuals were located, we collected a GPS point using a Trimble GeoExplorer 6000 series (model 88950; Trimble Inc., California) unit and documented the height, leaf length and width, the number of leaves (prongs) present, and the reproductive capability of each plant. We measured plant height

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

Table 1.—Companion species abundance rankings. Describes system for ranking companion species abundance and proximity to ginseng. Rank is dependent on the number of individuals and their proximity to the nearest ginseng plant.

Rank	Description	Number of individuals (individuals/species)	Proximity to ginseng
1	Present	1–several	Same physiographic position, but not within 2 m
2	Infrequent	1–5	Within 2 m
3	Moderately abundant	6–9	Within 2 m
4	Abundant	10–19	0–5 m
5	Very abundant	20+	0–5 m

in centimeters (cm) from the ground or top of leaf litter to the top of the most prominent leaf. For plants with irregular growth (i.e., curved or damaged stem), we measured height in the same way (from the ground to the top of the leaf) rather than measuring the entire length of the stem. We considered individuals reproductive if they had evidence of flowering or fruiting.

Communicating with both private citizens and agency professionals aided detection of additional ginseng plants. Additionally, the United States Department of Agriculture Forest Service uses a suite of "companion" species to determine appropriate ginseng habitat (Kauffman 2006). Although efforts were made to survey in areas both with and without these species, their presence facilitated the detection of wild ginseng. Where ginseng was present, the prevalence of any nearby companions was ranked based on abundance and proximity (Table 1). Companion species were ranked using this method with the exception of bedstraw (*Gallium* spp.) due to their growth habit.

Preparing Model Predictions

We used 78 environmental variables created by Virginia's DCR-DNH, along with hundreds of presence points, acquired via a license for use of digital data in 2019. Although there is some correlation among variables, we did not attempt to remove them in order to keep our model identical to DNH's original species distribution model. This suite of raster data represented various climatic, topographical, geological, and land cover variables across the state of Virginia (see Appendix). We used background points in this study in lieu of absence points due to the uncertainty in concluding that ginseng was not present due to harvest. Background points used in this way serve to represent the average, or background, values of a variable across a study region, rather than classifying the environment in which a species does not or theoretically cannot exist (Phillips et al. 2009). A model can learn the general characteristics of a region from background points, which can then be compared to the characteristics of specific locations at which the species has been documented.

In this manuscript we focus on a relatively simple model with two iterations. The first iteration contained the original 1955 ginseng occurrences before our sampling was conducted, while the second iteration received our addition of 198 occurrences documented in the Cumberland Mountains (Table 2). To create the model, we utilized the *randomForest* package in R (Liaw and Wiener 2002). Random Forest is a machine learning algorithm that uses decision trees to classify data based on features, where decision trees are constructed using random subsamples of data and then are subsequently split based on randomly selected variables to learn the impact of each variable on the outcome (Breiman 2001; Liaw and Wiener 2002) (Figures 2a, b). Random Forest automatically calculates the importance of all variables used in the modeling process, defined as the increase in mean square error when a given variable is rearranged or removed within a decision tree (Liaw and Wiener 2002). A more important variable will have a greater impact on error when removed, while a less important variable will have a lower impact when removed (i.e., when an important variable is removed, the error of that model will increase).

Each model contained 2000 decision trees (k) and each node of a tree contained a specific number of randomly selected variables (m) equal to the square root of the total number of variables present in the dataset ($m = 8$). The sample size used for stratification in the models was equal to 20% of the number of total presence points. Using this stratification method enables each tree to contain an equal number of presence and background points.

Our model was constructed using basic methods (i.e., we did not include any variations in data preparation or modeling parameters), with the first iteration excluding newly documented occurrences (hereafter referred to as the "initial model") from this study and the second including them (hereafter referred to as the "updated model"). This enables us to view the impact of increasing sampling in a known hotspot for biodiversity (Wilson and Tuberville 2003). The initial model was essentially identical to the original species distribution model published by Virginia's DCR-DNH in terms of data used and model parameters within R (Virginia Natural Heritage Program 2017).

Table 2.—Model descriptions and results. Displays modeling method, number of presence and background points, number of variables, out-of-bag (OOB) error, and area under the ROC curve (AUC) for both models. pres=presence, bg=background, prob=predicted probability, VA=Virginia, BR=Blue Ridge, CM= Cumberland Mountains.

Model	*n* pres	*n* bg	*n* variables	OOB error	AUC	Mean prob of pres (VA; BR; CM)
(initial) baseline	1955	5664	78	3.13%	0.9926	0.099; 0.288; 0.170
(updated) baseline with additional occurrences	2153	5664	78	3.33%	0.9927	0.104; 0.293; 0.278

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

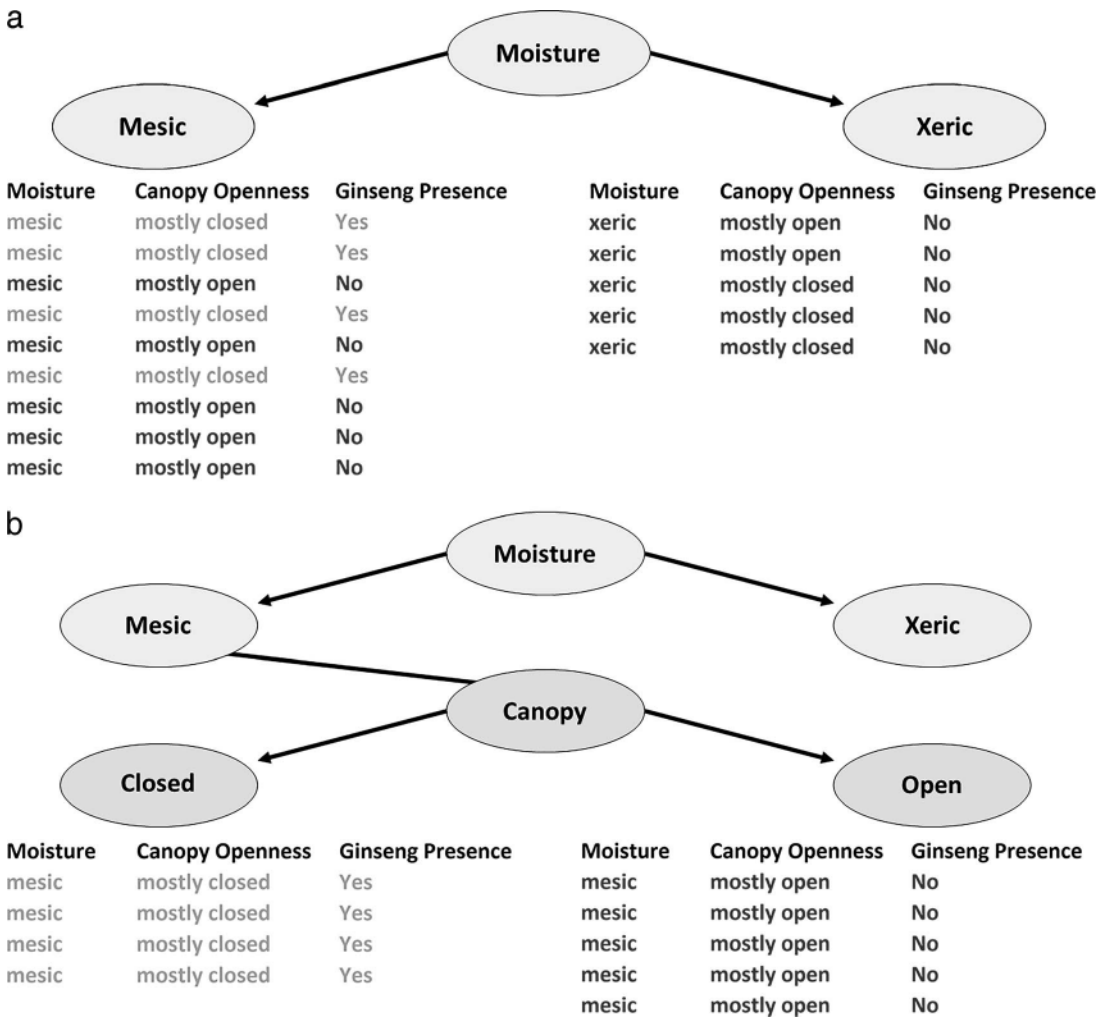

Figure 2.—(a) Initial branch of a hypothesized data set containing soil moisture, canopy openness, and ginseng presence. This decision tree first splits these data by soil moisture (mesic or xeric) and observes the resulting subsets of these data. (b) Second branch of a hypothesized data set containing soil moisture, canopy openness, and ginseng presence. This decision tree has further split the subset of data with mesic soil by the canopy openness and observed that canopy openness at mesic sites can be a useful predictor of ginseng presence.

Evaluating Model Performance

To ensure our models were performing appropriately, we conducted 1000 modeling runs and evaluated them using Random Forest's Out-Of-Bag error (OOB error) as well as the commonly used metric, Area Under the Receiver Operating characteristic Curve (AUROC or AUC). When training a Random Forest model, the OOB error is the median value from the error distribution calculated from all randomized decision trees ($k = 2000$) in each forest. Conducting 1000 modeling runs allowed us to have as many OOB error estimates, from which we could then calculate the median value from runs. Specifically, the OOB error is calculated by withholding one-third of the data from a sample and using those data to test the model's accuracy at predicting the outcome (i.e., we can train the model and then use that model on data for which we know the answer; Breiman 1996). Generally, an OOB error value of less than 10% indicates good performance in a model. The AUC is a common metric for evaluating predictive models and works similarly to the OOB error of Random Forests—where a certain percentage of data are withheld from training to be used later for testing the model's accuracy and a value greater than 0.9 generally indicates an excellent performance (Hanley and McNeil 1982; Fawcett 2006; Evans and Cushman 2009; Williams et al. 2009; Crawford and Hoagland 2010; Evans et al. 2011; Barbet-Massin et al. 2012; Narouei-Khandan et al. 2017; Hackett et al. 2020; Zhang et al. 2020). The average AUC was calculated from five-fold cross validation (i.e., we randomly withheld one-fifth of the data to be used for testing the model) repeated 200 times for a total of 1000 unique modeling runs. After evaluating each model iteration, they were used along with the set of environmental variables to create a prediction in the form of a new raster where each pixel (30×30 m) represents the probability of ginseng presence (Figure 3).

Quantifying Variable Importance

We quantified the importance of all environmental variables used in both iterations of our species distribution model (Figures 4, 5). From the updated model including newly documented occurrences, the importance of certain variables is demonstrated using Partial Dependence Plots (PDPs) (Figure 6). These plots illustrate the marginal effect that a certain variable

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

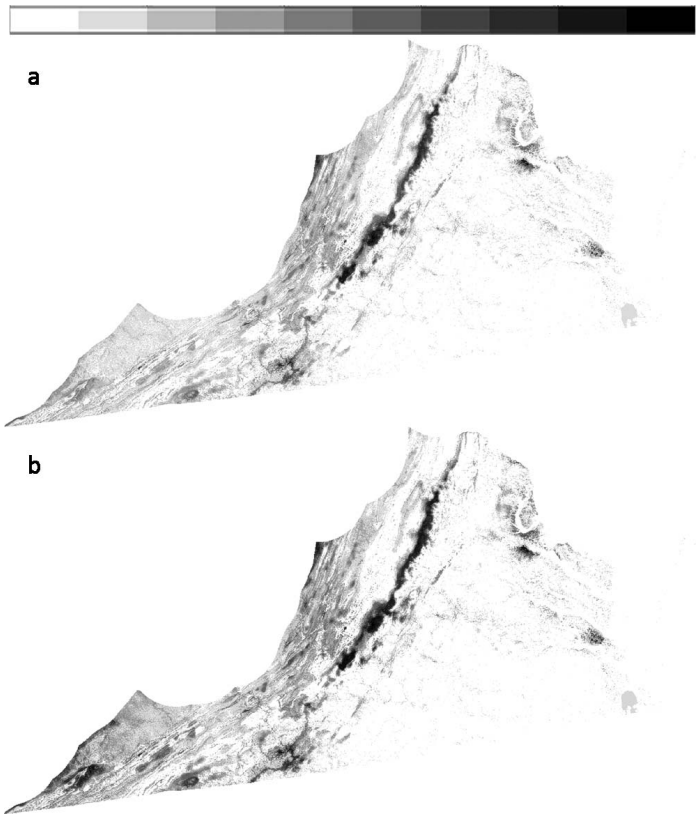

Figure 3.—Comparison of initial (a) and updated (b) distribution models for American ginseng. The Cumberland Mountains are in the southwestern portion of both models. The probability of ginseng presence ranging from 0 to 1 is shown above (light to dark).

has on the outcome of the model in question. For the PDPs displayed here, we selected variables that were calculated as highly important by Random Forest but were also deemed likely to be understood and useful in the field (i.e., intelligible variables).

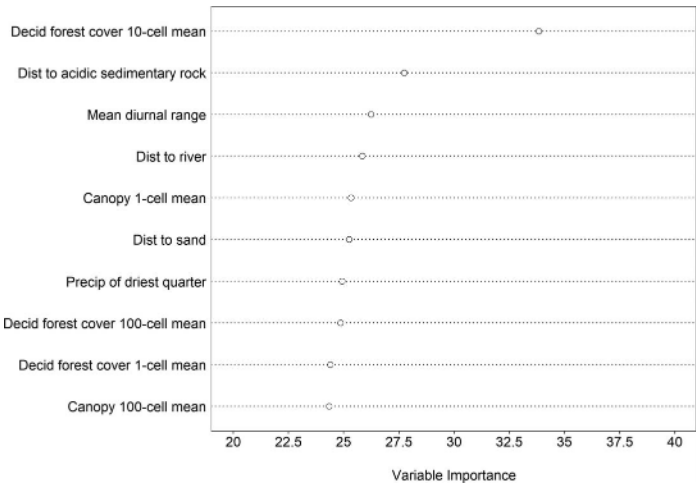

Figure 4.—Importance rankings of top 10 variables from the initial distribution model. Importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, precip-precipitation.

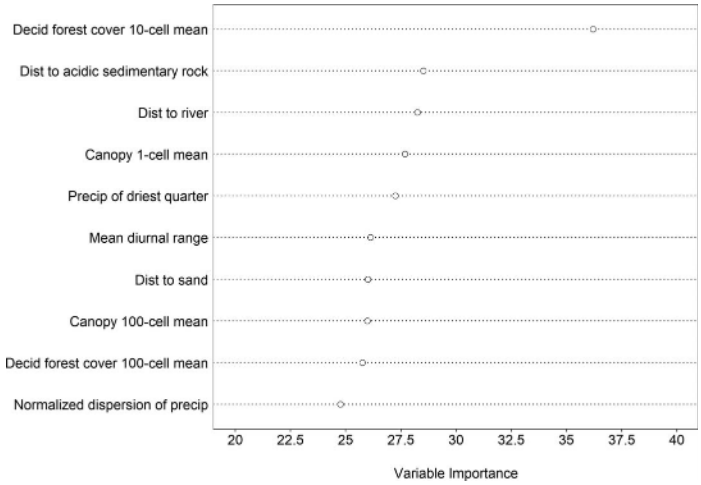

Figure 5.—Importance rankings of top 10 variables from the updated distribution model. Importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, precip-precipitation.

RESULTS

Field Surveys

We documented 198 ginseng plants across ~800 km^2 of the Cumberland Mountains, where there were previously only 16 documented plants. We estimate that these plants comprised around 25 populations—perhaps more accurately described as patches—with patch sizes ranging from a single plant to around 40 individuals. These individuals varied in size, leaf (prong) number, and reproductive maturity (Figure 7). The association between ginseng and companion species quickly became apparent and thereafter the presence of companion species was noted to be an efficient method for detecting wild ginseng during random surveys.

We found at least one herbaceous companion alongside 191 of the 198 documented ginseng plants, and a total of 703 occurrences of herbaceous companions were documented (Table 3). The most common companions were bedstraw (*Galium* spp.; $N = 165$) and black cohosh (*Actaea racemosa* L.; $N = 156$) (Table 3). Other frequent companion species were rattlesnake fern (*Botrypus virginianus* (L.) Michx.), blue cohosh (*Caulophyllum thalictroides* (L.) Michx.), and sweet cicely (*Osmorhiza claytonii* (Michx.) C.B. Clarke) (Table 3).

Modeling Probability of Ginseng Presence

Both iterations of our species distribution model were calculated as accurate per the OOB error and AUC metrics, indicating that they were properly trained and are operating correctly—although they may be overfitting to some degree as these evaluation metrics (notably the AUC) were nearly perfect scores (Table 2). The predicted probability of ginseng presence across the state of Virginia in the initial and updated models averaged 0.099 and 0.104, respectively (Figure 3, Table 2). This was an expected result, as documenting additional points in one area should not drastically alter the statewide prediction. Specifically, within the Blue Ridge Mountains and Cumberland

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

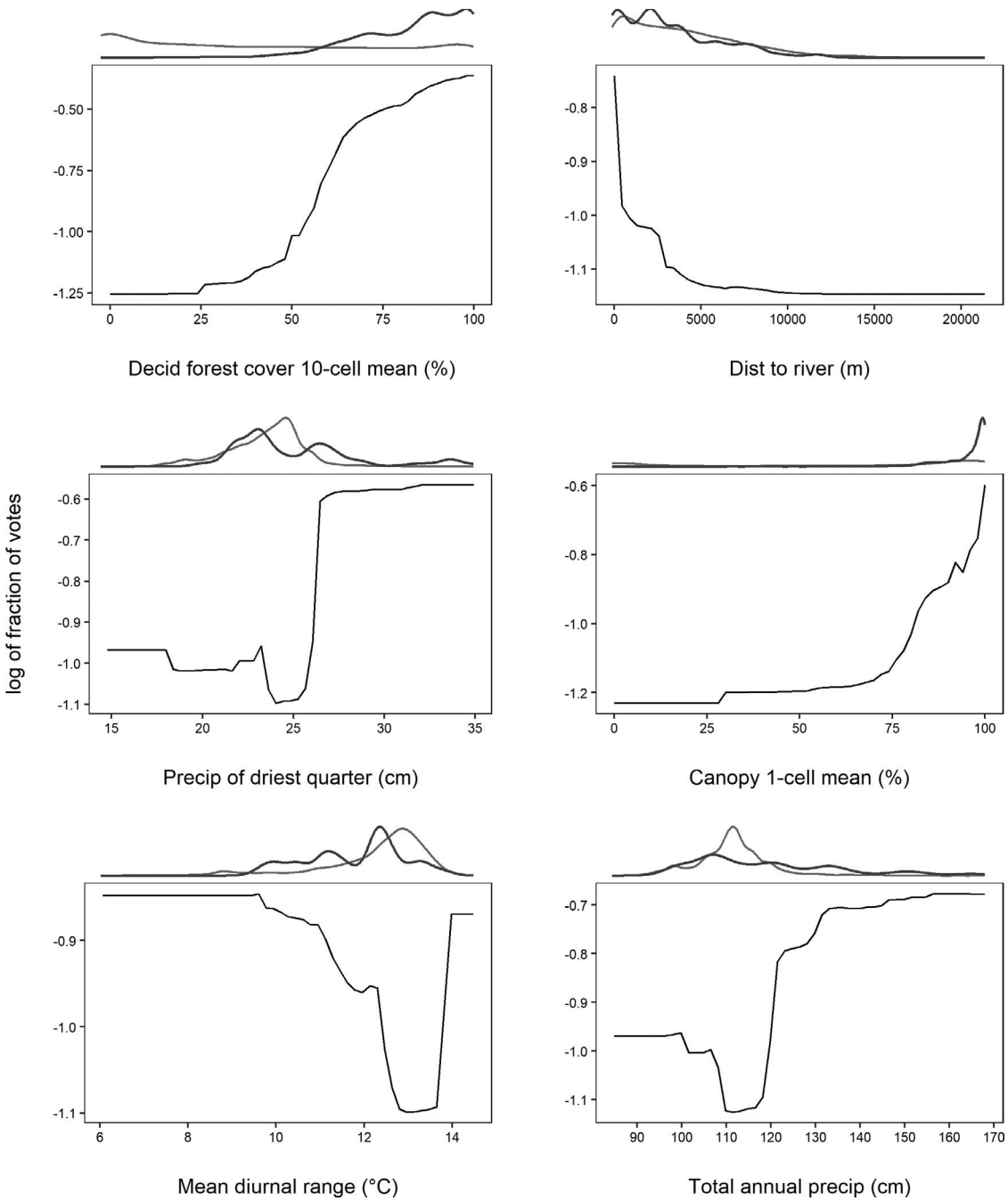

Figure 6.—Partial dependence plots of six selected variables from the updated distribution model. *Y*-axes demonstrate the marginal effect that the variable has on predicting the presence of American ginseng. Kernel density plots of presence and background data are displayed above plot margins via thick and thin lines, respectively. decid=deciduous, dist=distance, precip=precipitation.

Mountains the initial model predicted average probability of presence values of 0.288 and 0.170, respectively. With the addition of occurrence points from this study ($N = 198$), the mean predicted probability of presence of ginseng in both the Blue Ridge Mountains and Cumberland Mountains increased from 0.288 to 0.293 and from 0.170 to 0.278, respectively (Figures 3, 8; Table 2). Thus, increased sampling in the Cumberland Mountains led to a surprising increase of the predicted probability of ginseng presence in both of these regions.

Variable Ranking and Relationships to Ginseng Presence

Deciduous forest cover was calculated to be the most important variable for predicting ginseng in both models (Figures 4, 5). Perhaps more importantly from a field perspective, other variables frequently quantified as highly important were diurnal temperature range (the difference in monthly maximum and minimum temperature), canopy cover and openness, and distance to streams (Figures 4, 5). Per the updated model, deciduous forest cover and overall canopy cover showed a generally positive correlation with increased predictive

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

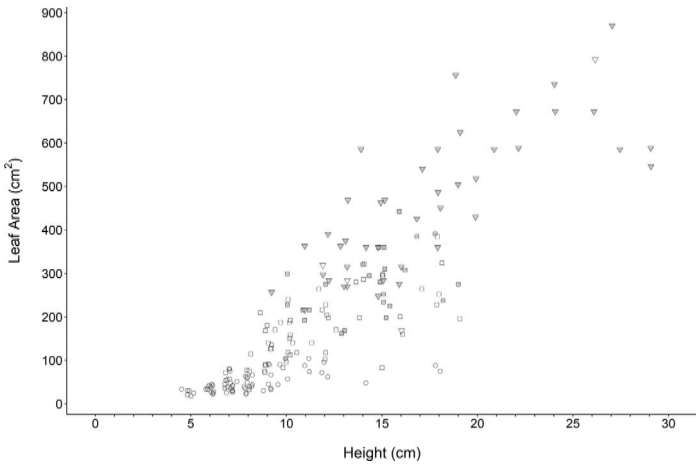

Figure 7.—Leaf area, height, leaf number (prongs), and reproductive maturity of ginseng individuals documented in this study. Circles, squares, and triangles denote leaf numbers of 1, 2, and 3, respectively. Filled and hollow shapes denote reproductive and nonreproductive individuals, respectively.

power of the model—while distance to streams was negatively correlated with predictive power (Figure 6). Other influential variables in the updated model, such as precipitation levels and mean diurnal range, had more complex relationships with predicting probability of presence, where specific ranges had the greatest influence on the model's predictive power (Figure 6). For example, our updated model experienced improvement in predictive power in areas where the diurnal range averages 10–12 °C, just as specific ranges of precipitation (neither too dry nor too wet) have a great impact on the accuracy of the model's predictive output (Figure 6).

DISCUSSION

Species Distribution Models

By conducting additional field surveys with only 14 successful sampling days in a single summer, we expect that these predictive models are now more realistic in their predictions across the entire state. Surprisingly, increasing the documentation of ginseng from 16 occurrences in the Cumberland

Table 3.—Companion species presence, abundance, and relative frequency. Frequency was calculated as the proportion of companion occurrence (n) to ginseng occurrence ($n = 198$).

Companion	n presence	Median abundance rank	Relative frequency
Gallium spp.[a]	165	NA	0.83
Actaea racemosa	156	3	0.78
Botrypus virginianus	87	1	0.44
Caulophyllum thalictroides	85	2	0.43
Osomorhiza claytonii	70	2	0.35
Adiantum pedatum	66	1	0.33
Viola canadensis	43	1	0.22
Aristolochia macrophylla	18	1	0.09
Sanguinaria canadensis	13	1	0.07

[a] Abundance of *Galium* spp. was not quantified.

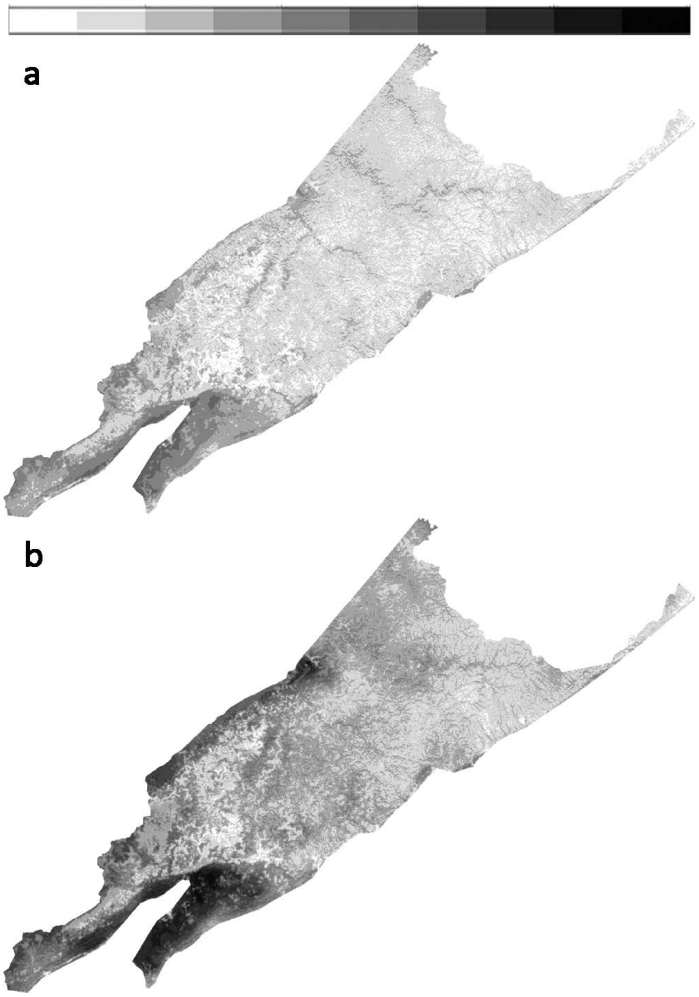

Figure 8.—Comparison of initial (a) and updated (b) distribution models cropped to the Cumberland Mountains in Southwestern Virginia. The probability of ginseng presence ranging from 0 to 1 is shown above (light to dark).

Mountains to over 200 did not only result in the model predicting a higher probability of ginseng presence there, but in the Blue Ridge Mountains as well (Figures 1, 3, 8; Table 2). This is somewhat counterintuitive, as we surveyed specifically in the Cumberland Mountains—which might suggest that the models would become biased in that direction. These final predictions along with similar variable importance rankings between the two models suggest that this was not the case; therefore, we are confident that our hypothesis was supported, and the updated model is creating less biased predictions.

Determining the modeling methods that most accurately predict habitat suitability rather than presence may be more complicated for ecologists than anticipated. Instinctively, researchers strive to create the most accurate model according to some evaluation metric—AUC and OOB error, in this case (Stockwell and Peterson 2002; Evans and Cushman 2009; Merow et al. 2014). However, the nearly perfect evaluation metrics of our model may indicate that they are overfitting (being overly conservative in their predictions) to some degree due to correlations among environmental variables and the ratio of

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

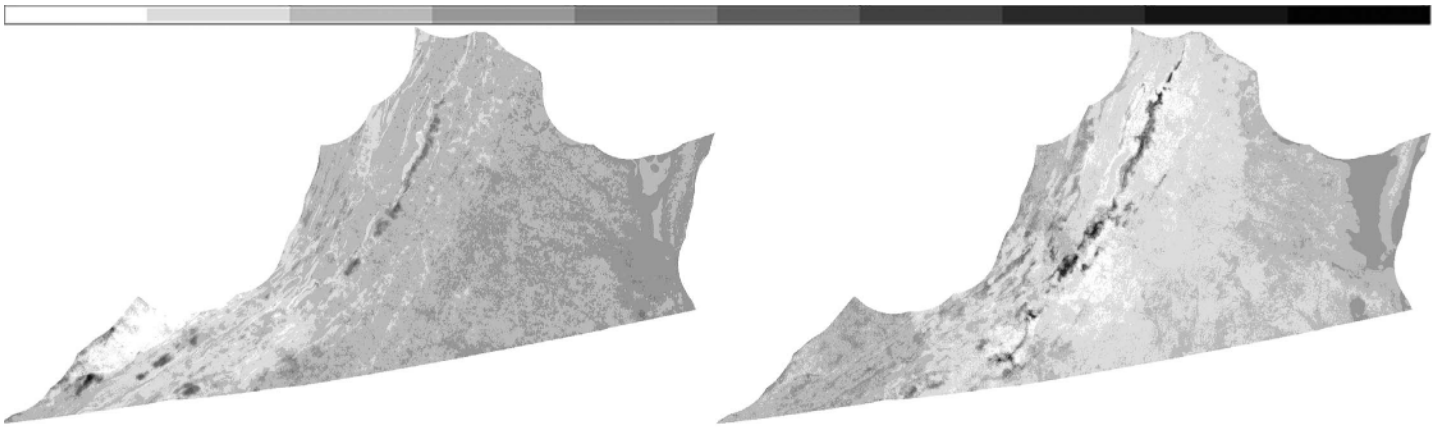

Figure 9.—Comparison of experimental model run. Left and right shows the predictive output of models trained exclusively on data from either the Cumberland Mountains or Blue Ridge Mountains, respectively. The probability of ginseng presence ranging from 0 to 1 is shown above (light to dark).

background to presence points (Table 2). It is a common issue for species distribution models to display overfitting, only predicting presence in small patches closely resembling locations where training data were located (Stockwell and Peterson 2002; Aguirre-Gutierrez et al. 2013). Therefore, they may not be practical for predicting habitat suitability, and we propose that these models (and models that perform similarly with some overfitting) may be more useful strictly as species distribution models for identifying wild ginseng occurrences for conservation efforts—rather than identifying generally suitable habitat.

Using Variable Importance for Location and Reintroduction of Species

Although we do not recommend these prediction maps be used to identify suitable habitat for restoration work due to overfitting present in the model, there is still use in identifying variables important for ginseng success (either for future harvest or restoration). For example, deciduous forests with increased canopy closure (at least 75%) are correlated with correctly predicting ginseng presence (Figures 4, 5, 6). This result should not come as a surprise—stating that ginseng grows in the woods is not groundbreaking—but nevertheless variables such as deciduous forest must be included to distinguish this important landscape feature from others such as bodies of water, impervious surfaces, evergreen forests, pastures, and grasslands (Figures 4, 5, 6).

This result aligns with findings from Wagner and McGraw's 2013 study on sunflecks and their importance for ginseng, where they propose that ginseng is likely adapted to heterogeneous light regimes of old-growth forests, and with other studies modeling ginseng habitat and distribution, which noted that forest canopy cover is an important habitat requirement (McGraw et al. 2003; Van Manen et al. 2005; Hackett et al. 2020). This idea is both contrasted and upheld by the widely varied recommendations for and experimental results testing specific shade levels for growing *Panax* sp. in wild-simulated and greenhouse settings, which range from 70% to 95% (Lee et al. 1980; Proctor 1988; Proctor et al. 1990; Parmenter and Littlejohn 2000; Hsu 2002; Pritts 2010; Proctor and Palmer 2017). These results appear to support the idea that ginseng grown under artificial shade cloth (i.e., with a fairly constant light level) tends to be smaller, less valuable, less healthy, and has a shorter lifespan than ginseng grown under a true forest canopy (Pritts 2010; Suh et al. 2011). Information on optimal light levels for wild American ginseng in North America is limited, but these studies as well as our results suggest that perhaps there is not a magic number for canopy cover, but rather a heterogeneous light scheme may be more important.

Unfortunately, initial interpretations of some variables may not be immediately intuitive. For example, roughness (calculated as the 11th and 13th most important variable from the initial and updated model, respectively) may be interpreted by some as more "rugged" terrain (see Appendix). Upon closer inspection, however, this variable is highlighting cliffs and other very steep areas. Regardless, rugged terrain tends to contain more protected hollows and generally inaccessible areas that hinder foot traffic and could therefore be more likely to harbor healthy populations of ginseng. Lastly, some variables such as annual precipitation may be less informative because the range (100–130 cm/year) is typical for Virginia's overall climate (Figure 6). Therefore, we suggest that using statistically important yet intuitive variables in addition to considering companion species' presence may be the best method for locating American ginseng for harvest or conservation.

Sampling Bias

Although the initial sample size in this study ($N = 1955$) was adequate for the broad area of Virginia, the initial sample size in the Cumberland Mountains ($N = 16$) was insufficient for modeling species distribution or habitat suitability in that region. As a result, the initial model predicted that ginseng presence was twice as likely in the Blue Ridge (Table 2, Figure 3). In a tangential modeling run to investigate geographical bias, we trained models exclusively on presence data from either the Cumberland Mountains or the Blue Ridge Mountains and made subsequent predictions for the entire state (Figure 9). Notably, when we used data exclusively from the Blue Ridge, the Cumberland Mountains were not highlighted in the subsequent statewide prediction. However, when we trained the model

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

exclusively on data from the Cumberland Mountains, the Blue Ridge still appeared somewhat highlighted (Figure 9). Although these are purely visual interpretations of tangential models, they nevertheless support our hypothesis that the initial model overlooked the Cumberland Mountains when predicting ginseng distribution across Virginia.

The presence of unique microclimates and cold air drainage events common in the Cumberland Mountains has led to unexpected patches of northern hardwood forests and some of the coldest temperatures recorded in Virginia occurring here (Braun 1950; unpub. data). If these phenomena do not occur frequently in areas such as Shenandoah National Park in the Blue Ridge, it is likely that models biased toward this region will predict that ginseng is a specialist species that prefers very specific climates and habitats. However, when including ginseng presence data from the Cumberland Mountains across these climatic ranges, the resulting models may not be as likely to make this prediction—indicating that American ginseng may be more of a generalist than is currently assumed.

The Blue Ridge Mountains of central Virginia (more specifically, the area including and surrounding Shenandoah National Park) was indicated to be a ginseng hotspot by our models as well as those constructed by others (Van Manen et al. 2005; Virginia Botanical Associates 2020). We propose that our initial model, although biased toward the Blue Ridge, was correct in the prediction that the region is a hotspot for American ginseng. However, other regions with appropriate habitat (e.g., the Cumberland Mountains) may be overlooked in the predictive output due to oversampling in the Blue Ridge. The Cumberland Mountains was not predicted to be another ginseng hotspot until we conducted additional sampling and incorporated the resulting data into the model. Our findings align with those from other studies—although there are statistical methods that may help alleviate impacts from unequal sampling, there is likely no substitute for increased sampling effort (Phillips et al. 2009; Hui et al. 2011; Syfert et al. 2013; Fithian et al. 2015).

Sampling should occur in as many ecoregions as possible but also be distributed geographically to avoid overlooking potential hotspots of the study taxon. We suggest that areas such as National Parks may not be representative of the broader regions in which they are located due to the protections they receive, their geographical isolation, and the resulting uniqueness of their biota. As such, researchers should exercise caution when extrapolating from these heavily protected and managed areas to the broader landscape.

Suggestions for Modeling Species Distribution and Suitable Habitat

Species distribution models such as these can improve the effectiveness of conservation planning. Identifying where threatened species like ginseng currently reside—in addition to where they could be successfully reintroduced—is invaluable to conservation and restoration efforts and as such is a widely used approach for conservation work (McGraw et al. 2003; Guisan and Thuiller 2005; Van Manen et al. 2005; Williams et al. 2009; Hackett et al. 2020). Increased sampling is likely to improve the reliability of species distribution and habitat suitability models alike. By increasing our sample size in an under-sampled region, the resulting species distribution model now highlights an overlooked hotspot for American ginseng in south central Appalachia.

Furthermore, identifying companion species may be an invaluable tool for the conservation of plant species. Many ginseng harvesters recognize companion species (including those used here among others), and it has been demonstrated that companion plants can be used to enhance growth of wild simulated ginseng (Pritts 2010; Tringovska et al. 2015). It is also possible that companion species may impact soil chemistry or decrease competition with ruderal species in field cultivation (Liphadzi and Reinhardt 2006; Tringovska et al. 2015). This may explain some of the challenges in cultivating ginseng in a commercial setting, as companion species are absent, which warrants further investigation—but may also explain the connection between species of interest, such as ginseng, and their companions.

We found that detecting companion species was a reliable means of locating wild populations of American ginseng. This finding supports the continued use of companions as a means of selecting appropriate habitat for locating, harvesting, and reintroducing ginseng by agencies and citizens alike (Kauffman 2006; Pritts 2010; Cech 2013; Davis and Persons 2014; Woods 2015; Bonnabeaux 2016; Hackett et al. 2020; Hall 2021). Notably, black cohosh (*Actaea racemosa*), blue cohosh (*Caulophyllum thalictroides*), and bloodroot (*Sanguinaria canadensis* L.) appeared to be some of the most reliable for identifying areas in which ginseng may be present or more likely to grow well.

Finally, we recommend additional research into companion plants for species such as ginseng to aid detection and conservation. Although we did not conduct exhaustive vegetation surveys in order to maximize the efficiency of surveying for ginseng and refining the species distribution models, we anecdotally noted that some species appear to be frequent in locations where ginseng was found but are not considered companions by the USFS. These species include Jack-in-the-pulpit (*Arisaema triphyllum* L.), wild yam (*Dioscorea villosa* L.), smooth Solomon's-seal (*Polygonatum biflorum* (Walter) Elliott), hairy Solomon's-seal (*Polygonatum pubescens* (Willd.) Pursh), red trillium (*Trillium erectum* L.), and white trillium (*Trillium grandiflorum* (Michx.) Salisb.). This is not uncommon, and it appears that the species considered companions varies by region and may differ between researchers as they develop their own gestalt image for what they feel is adequate ginseng habitat (Pritts 2010; Cech 2013; Hackett et al. 2020; Hall 2021). In Michigan, Hackett and colleagues (2020) noted some of the same companions used by the USFS in southern Appalachia (such as *Osmorhiza* sp. and *Gallium* sp.), some that we have discussed herein as being potential companion species (such as *Polygonatum pubescens* and *Arisaema triphyllum*), in addition to other species not discussed here such as wild sarsaparilla (*Aralia nudicaulis* L.), sedges (*Carex* spp.), and wild leek (*Allium tricoccum* Ait.) (Kauffman 2006; Hackett et al. 2020). We suggest that additional research be conducted to identify possible companion species for ginseng as well as other plant species that are threatened or endangered, as the presence of companion species should be considered when attempting to classify suitable habitat for restoration and conservation efforts of this species.

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

CONCLUSIONS

The Cumberland Mountains and Blue Ridge Mountains are both unique hotspots for American ginseng in Virginia. However, the former did not gain equal standing in predictive models until the addition of nearly 200 occurrence points in that region. Although it is likely that the accuracy of predictions in the Cumberland Mountains could be further improved with continued sampling and the establishment of additional survey plots by DCR-DNH, our surveys markedly improved ginseng distribution models for this region as well as for Virginia as a whole. This work has built upon tools that will aid the conservation of American ginseng in the state of Virginia—but perhaps more importantly, we have shed light on a consistently overlooked area with a high level of biodiversity in Southern Appalachia.

ACKNOWLEDGMENTS

Special thanks to David Bucklin, Anne Chazal, and Karen Patterson of Virginia's Natural Heritage Program for providing guidance with modeling methodology, permitting the use of Natural Heritage's suite of environmental variables, and providing us with information regarding survey methods; to Michelle Davalos and Lois Boggs of the USDA Forest Service for assisting with field work planning and permitting us to conduct this research; to Julia Portmann, Wayne Browning, and Dr. Ryan Huish for assisting with field data collection; and to Gary Kauffman of the USDA Forest Service; and to Chris Ludwig of Virginia's Natural Heritage Program for providing early guidance for the project.

Jacob Peters completed his M.S. in Biology at James Madison University. He currently is a contracted researcher for Conservation International, teaches Biology at Blue Ridge Community College, and serves as a GIS intern with the Smithsonian Conservation Biology Institute.

Heather Griscom is a professor of Biology at James Madison University. Her research is focused on forest restoration in the Shenandoah Valley and Central America.

LITERATURE CITED

Aguirre-Gutierrez, J., L.G. Carvalheiro, C. Polce, E.E. Van Loon, N. Raes, M. Reemer, and J.C. Biesmeijer. 2013. Fit-for-purpose: Species distribution model performance depends on evaluation criteria — Dutch hoverflies as a case study. PLoS ONE 8(5):e63708.

Barbet-Massin, M., F. Jiguet, C.H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: How, where and how many? Methods in Ecology and Evolution 3:327–338.

Bonnabeaux, M. 2016. Ginseng: The black market herb of the Appalachian. Accessed 2 December 2020 from <http://www.technicianonline.com/arts_entertainment/article_771f9b02-eb2b-11e5-936c-132b98d780d8.html>.

Braun, E.L. 1950. Deciduous Forests of Eastern North America. Blackburn Press, Philadelphia, PA.

Breiman, L. 1996. Out-of-Bag estimation. <https://www.stat.berkeley.edu/pub/users/breiman/OOBestimation.pdf>

Breiman, L. 2001. Random Forests. Machine Learning 45:5–32.

Case, M.A., K.M. Flinn, J. Jancaitis, A. Alley, and A. Paxton. 2007. Declining abundance of American ginseng (*Panax quinquefolius* L.) documented by herbarium specimens. Biological Conservation 134:22–30.

Cech, R. 2013. American ginseng. United Plant Savers. <https://unitedplantsavers.org/187-american-ginseng/>

Crawford, P.H.C., B.W. Hoagland. 2010. Using species distribution models to guide conservation at the state level: The endangered American burying beetle (*Nicrophorus americanus*) in Oklahoma. Journal of Insect Conservation 14:511–521.

Cruse-Sanders, J.M., and J.L. Hamrick. 2004. Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L. (Araliaceae). American Journal of Botany 91:540–548.

Davis, J., and S.W. Persons. 2014. Growing and Marketing Ginseng, Goldenseal, and other Woodland Medicinals. New Society Publishers, Gabriola Island, BC, Canada.

Eisenhauer, N., K. Yee, E.A. Johnson, M. Maraun, D. Parkinson, D. Straube, and S. Scheu. 2011. Positive relationship between herbaceous layer diversity and the performance of soil biota in a temperate forest. Soil Biology and Biochemistry 43:462–465.

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

Evans, J.S., M.A. Murphy, Z.A. Holden, and S.A. Cushman. 2011. Modeling species distribution and change using Random Forest. Pp. 139–159 *in* C. Drew, Y. Wiersma, and F. Huettmann, eds., Predictive Species and Habitat Modeling in Landscape Ecology. Springer, New York, NY.

Farrington, S.J., R.M. Muzika, D. Drees, and T.M. Knight. 2009. Interactive effects of harvest and deer herbivory on the population dynamics of American ginseng. Conservation Biology 23:719–728.

Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27:861–874.

Fithian, W., J. Elith, T. Hastie, and D.A. Keith. 2015. Bias correction in species distribution models: Pooling survey and collection data for multiple species. Methods in Ecology and Evolution 6:424–438.

Fleming, G.P., and K.D. Patterson. 2021. Natural Communities of Virginia: Ecological Groups and Community Types. Natural Heritage Technical Report 21-15. Virginia Department of Conservation and Recreation, Division of Natural Heritage, Richmond, VA.

Furedi, M.A., and J.B. McGraw. 2004. White-tailed deer: Dispersers or predators of American ginseng seeds? American Midland Naturalist 152:268–276.

Gilliam, F.S. 2007. The ecological significance of the herbaceous layer in temperate forest ecosystems. BioScience 57:845–858.

Gilliam, F.S., and M.R. Roberts. 2003. The Herbaceous Layer in Forests of Eastern North America. Oxford University Press, New York, NY.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: Offering more than simple habitat models. Ecology Letters 8:993–1009.

Hackett, R.A., H.D. Enander, and P.J. Higman. 2020. Huron-Manistee National Forest: Ginseng Restoration Project. United States Forest Service MNFI Report No. 2020-29.

Hall, D.K. 2021. Ginseng companion plants. Accessed 1 August 2021 from <https://herbspeak.com/ginseng-companion-plants/>.

Hanley, J.A., and B.J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36.

Hsu, P.C. 2002. Commercial production of American ginseng (*Panax quinquefolius* L.). Native Plants Journal 3:106–108.

Hui, C., L.C. Foxcroft, D.M. Richardson, and S. MacFadyen. 2011. Defining optimal sampling effort for large-scale monitoring of invasive alien plants: A Bayesian method for estimating abundance and distribution. Journal of Applied Ecology 48:768–776.

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use

Kauffman, G. 2006. Conservation assessment for American ginseng (*Panax quinquefolius*) L. USDA Forest Service, Eastern Region. Unpublished report, National Forests in North Carolina, Asheville, NC.

Lee, J.C., S.K. Cheon, Y.T. Kim, and J.S. Jo. 1980. Studies on the effect of shading materials on the temperature, light intensity, photosynthesis and the root growth of Korean ginseng (*Panax ginseng* C.A. Meyer). Journal of Korean Society of Medicinal Crop Science 25:91–98.

Levin, D.A., and A.C. Wilson. 1976. Rates of evolution in seed plants: Net increase in diversity of chromosome numbers and species numbers through time. Proceedings of the National Academy of Sciences 73:2086–2090.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2(3):18–22.

Liphadzi, K.B., and C.F. Reinhardt. 2006. Using companion plants to assist *Pinus patula* establishment on former agricultural lands. South African Journal of Botany 72:403–408.

McGraw, J.B., and J.L. Chandler. 2018. Demographic hallmarks of an overbrowsed population state in American ginseng. Global Ecology and Conservation 15:e00435.

McGraw, J.B., A.E. Lubbers, M. Van der Voort, E.H. Mooney, M.A. Furedi, S. Souther, J.B. Turner, and J. Chandler. 2013. Ecology and conservation of ginseng (*Panax quinquefolius*) in a changing world. Annals of the New York Academy of Sciences 1286:62–91.

McGraw, J.B., S.M. Sanders, and M. Van der Voort. 2003. Distribution and abundance of *Hydrastis canadensis* L. (Ranunculaceae) and *Panax quinquefolius* L. (Araliaceae) in the Central Appalachian Region. Journal of the Torrey Botanical Society 130:62–69.

Merow, C., M.J. Smith, T.C. Edwards, G. Antoine, S.M. McMahon, S. Normand, W. Thuiller, R.O. Wuest, N.E. Zimmermann, and J. Elite. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography 37:1267–1281.

Mohapatra, J., C.P. Singh, M. Hamid, A. Verma, S.C. Semwal, B. Gajmer, A.A. Khuroo, A. Kumar, M.C. Nautiyal, N. Sharma, and H.A. Pandya. 2019. Modelling *Betula utilis* distribution in response to climate-warming scenarios in Hindu-Kush Himalaya using Random Forest. Biodiversity and Conservation 28:2295–2317.

Mooney, E.H., and J.B. McGraw. 2007. Effects of self-pollination and outcrossing with cultivated plants in small natural populations of American ginseng, *Panax quinquefolius* (Araliaceae). American Journal of Botany 94:1677–1687.

Narouei-Khandan, H.A., C.L. Harmon, P. Harmon, J. Olmstead, V.V. Zelenev, W. Van der Werf, S.P. Worner, S.D. Senay, and A.H.C. Van Bruggen. 2017. Potential global and regional geographic distribution of *Phomopsis vaccinii* on *Vaccinium* species projected by two species distribution models. European Journal of Plant Pathology 148:919–930.

Oppel, S., A. Meirinho, I. Ramírez, B. Gardner, A.F. O'Connell, P.I. Miller, and M. Louzao. 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biological Conservation 156:91–104.

Parmenter G., and R. Littlejohn. 2000. Effect of shade on growth and photosynthesis of *Panax ginseng*. New Zealand Journal of Crop and Horticultural Science 28:255–269.

Phillips, S.J., M. Dudik, J. Elith, C. Graham, A. Lehman, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications 19:181–197.

Pritts, K.D. 2010. Ginseng: How to Find, Grow, and Use America's Forest Gold. Stackpole Books, Mechanicsburg, PA.

Proctor, J.T.A. 1988. The micrometeorological requirements for the culture of ginseng (*Panax* sp.). Pp. 129–132 *in* Proceedings of the 5th International Ginseng Symposium, Seoul, South Korea. Korean Society of Ginseng.

Proctor, J.T.A., and J.W. Palmer. 2017. Optimal light for greenhouse culture of American ginseng seedlings. Journal of Ginseng Research 41:370–372.

Proctor, J.T.A., J.C. Lee, and S.S. Lee. 1990. Ginseng production in Korea. HortScience 25:746–750.

Reynolds, J., K. Wesson, A.L.J. Desbiez, J.M. Ochoa-Quintero, and P. Leimgruber. 2016. Using remote sensing and Random Forest to assess the conservation status of critical cerrado habitats in Mato Grosso do Sul, Brazil. Land 5:12.

Souther, S., and J.B. McGraw. 2014. Synergistic effects of climate change and harvest on extinction risk of American ginseng. Ecological Applications 24:1463–1477.

Stockwell, D.R.B., and A.T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modeling 148:1–13.

Suh, H., S.M. Seo, S.Y. Woo, and D. Lee. 2011. Forest cultivated ginseng in Korea: All cure medicinal plants. Journal of Medicinal Plants Research 5(22):5331–5336.

Syfert, M.M., M.J. Smith, and D.A. Coomes. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLoS ONE 8(2):e55158.

Tringovska, I., Y. Vinelina, D. Markova, and M. Mihov. 2015. Effect of companion plants on tomato greenhouse production, Scientia Horticulturae 186:31–37.

Van Manen, F.T., J.A. Young, C.A. Thatcher, W.B. Cass, and C. Ulrey. 2005. Habitat models to assist plant protection efforts in Shenandoah National Park, Virginia, USA. Natural Areas Journal 25:339–350.

Virginia Botanical Associates. 2020. Digital Atlas of the Virginia Flora. Virginia Botanical Associates, Blacksburg, VA. <http://www.vaplantatlas.org>

Virginia Department of Conservation and Recreation. 2021. DCR-DNH Vegetation Plots Database. Division of Natural Heritage, Richmond, VA. Data exported on 1 August 2021.

Virginia Natural Heritage Program. 2017. Species distribution model for American Ginseng (*Panax quinquefolius*). Virginia Department of Conservation and Recreation - Division of Natural Heritage, Richmond, VA.

Wagner, A., and J.B. McGraw. 2013. Sunfleck effects on physiology, growth, and local demography of American ginseng (*Panax quinquefolius* L.). Forest Ecology and Management 291:220–227.

Williams, J.N., C. Seo, J. Thorne, J.K. Nelson, S. Erwin, J.M. O'Brien, and M.W. Schwartz. 2009. Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions 15:565–576.

Wilson, I.T., and T. Tuberville. 2003. Virginia's Precious Heritage: A Report on the Status of Virginia's Natural Communities, Plants, and Animals, and a Plan for Preserving Virginia's Natural Heritage Resources. Natural Heritage Technical Report 03-15. Virginia Department of Conservation and Recreation, Division of Natural Heritage, Richmond, VA.

Woods, M. 2015. American Ginseng & Companions. Wild Ozark, LLC, Kingston, AR.

Zhang, C., Y. Chen, B. Xu, Y. Xue, and Y. Ren. 2020. Improving prediction of rare species' distribution from community data. Scientific Reports 10:12230.

Downloaded From: https://bioone.org/journals/Natural-Areas-Journal on 26 Apr 2022
Terms of Use: https://bioone.org/terms-of-use