5-7-2020

Modeling species distribution and habitat suitability of American ginseng (Panax quinquefolius) in Virginia

Jacob Peters

Modeling species distribution and habitat suitability of American ginseng (*Panax quinquefolius*) in Virginia

Jacob D. J. Peters

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Science

Department of Biology

May 2020

FACULTY COMMITTEE:

Committee Chair: Dr. Heather P. Griscom

Committee Members:

Dr. Bruce A. Wiggins

Dr. Patrice M. Ludwig

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

American ginseng (*Panax quinquefolius*) is a well-known and sought-after medicinal plant native to North America that is facing increased threat of extinction due to overharvesting, herbivory, and habitat loss. Species distribution and habitat suitability models may be valuable to landowners interested in sustainable harvest or to institutions interested in the conservation and restoration of the species. With unequal sampling efforts across a region of interest, it is likely that some locations with appropriate habitat may be misrepresented in model predictions. This study refined a state-derived species distribution model for ginseng through increased sampling effort across the Cumberland Plateau of Virginia and experimental manipulation of model parameters using the machine learning method Random Forest. Through many iterations, sixteen final models were constructed with various parameters such as spatial partitioning, removal of correlated variables, and limiting the spatial extent for background point generation in an effort to reduce overfitting and increase accuracy. Models were evaluated using partial dependence plots, area under the curve (AUC), and out-of-bag error (OOB error). Of those models, this study determined that various methods may be used depending on the goal of the project—resulting in more accurate and realistic species distribution and habitat suitability models than were previously available. This study concludes that, although various model parameters can be altered to change the product thereby increasing accuracy or reducing overfitting, the most effective means of reducing the impact of sampling deficiency is to balance sampling effort across the region of study.

INTRODUCTION

Highlighting the Herbaceous Layer

The herbaceous layer of a forest, though often overlooked, comprises a startling amount of biodiversity found therein. Forest herbs provide food, shelter, substrate, and materials to many other species—and may be equally important to researchers and landowners attempting to understand temperate forests. In the average forest, 80% of plant species are found within the herbaceous layer, and the extinction rate of forest herbs is several times greater than that of canopy species (Gilliam 2007; Levin and Wilson 1976). This alone should encourage discussion of the herbaceous layer when considering how to preserve biodiversity in forest ecosystems, though the topic typically revolves around tree species that comprise the majority of biomass therein.

In addition to their diversity, herbaceous plants also have an important role in soil nutrient cycling within forests. A study by Welch and colleagues (2007) found that herbaceous plants contribute significant amounts of available nutrients such as phosphorus, potassium, and calcium to the soil. They also noted that the litter generated from herbaceous plants has a lower carbon-nitrogen ratio, causing them to decompose more rapidly and thus contribute those nutrients to the soil more promptly than tree leaf litter (Welch et al. 2007). As concluded by Gilliam (2007), the herbaceous layer may contain only one percent of the total biomass of a forest but can comprise the majority of the plant species present and contributes up to a fifth of the overall leaf litter, which typically contains more nutrients than that of tree-foliage. If forest herbs are as critical to regeneration and nutrient cycling of forest ecosystems as these studies suggest, then their importance to forests should be given due consideration by land-managers or government agencies when proposing treatments or harvests (Erikson 1995; Gilliam 2007; Welch et al. 2007).

Studies of herbaceous plants frequently occur in greenhouses, laboratories, or other experimental settings. While those settings may allow researchers to control influences and potentially confounding variables, there may be an increased risk to misinterpret results or overlook processes and influences that may occur within a working ecosystem (Gibson et al. 1999). Therefore, it is important for studies to be conducted on forest herbs in their natural setting and native range—with all influencing parameters present.

Land-managers must become aware of what might influence the health of their land and the ecosystems within. Species distribution and habitat suitability models are tools that may be valuable to forest farmers or others interested in sustainably harvesting non-timber forest products (NTFP). American ginseng (*Panax quinquefolius* (L.)), is an exemplary NTFP whose populations may depend on understanding the habitat in which they thrive. Landowners will be better able to make informed decisions regarding harvesting and planting locations if environmental influences on ginseng presence is understood and suitable habitat can be reckoned.

Morphology and Ecology of American Ginseng

American ginseng is a slow growing, long-lived, herbaceous, perennial plant native to North America that may become capable of flowering as early as their second year or as late as their eighth year in some cases (Carpenter and Cottam 1982; McGraw et al. 2013). Initially only having one small "prong" or leaf, after several years they will reach reproductive maturity and develop two to four prongs consisting of several whorled leaves at the apex of the stem (Carpenter and Cottam 1982; Lewis and Zenger 1982). The fruits are two-seeded drupes approximately one centimeter in diameter and bright red in color once mature, indicating that birds may be a means of dispersal (Howe 1986; Lewis and Zenger 1982). After the seeds have been dispersed, they may require anywhere from 18 months to several years to germinate (Lewis and Zenger 1982; McGraw

et al. 2013; Snow 2009). Ginseng populations tend to form clusters varying from just a few to over one hundred plants within several square-meters (Cruse-Sanders and Hamrick 2004; McGraw et al. 2010; Mooney and McGraw 2007a). One study by Elza et al. (2016) suggests that wood thrushes, also in decline, might have served an important role in dispersing seeds but are often absent near ginseng.

Ginseng appears to be somewhat more of a generalist than its rarity and current folklore would suggest—with a broad geographic range spanning from the American Midwest to Maine and from Canada to southern Appalachia and inhabiting different soil textures, moisture levels, and pH (McGraw et al. 2013). Although it appears to be more common in mixed mesophytic forests, ginseng can be found in a wide range of forest types (Albrecht 2009; McGraw et al. 2003; McGraw et al. 2013). According to many harvesters, ginseng is typically found on moist, north-facing slopes or in protected hollows. If ginseng is more of a generalist than currently thought, it may be that it is usually found in these conditions because that is where they are harvested and replanted—creating a feedback loop sustaining their population there while also reinforcing inaccuracies about ginseng's distribution.

To further complicate the mythos surrounding ginseng's habitat requirements, some studies suggest that certain soil minerals may be significantly correlated with plant health—stating that slightly acidic soils with higher levels of iron, sulfur, and aluminum with lower levels of magnesium may be most appropriate for increasing the levels of medicinal compounds within the root (Li and Mazza 1999; Lee and Wudge 2013). However, other studies such as one conducted by Lee and Wudge (2013) indicate that gypsum (and the resulting increase in calcium) may lower shoot to root ratio while still being correlated with an increase in certain concentrations of medicinal compounds. Lastly, study by Konsler et al. (1990) noted that soil chemistry may also

play a role in synthesis and storage of those compounds within the leaves rather than the roots—which may be worthy of future investigation, as most research and the use of ginseng focuses on the root.

Compounding the lack of concrete knowledge on factors that lead to appropriate habitat for ginseng, this species is facing increased threat of extinction due to climate change, overharvesting, herbivory, and habitat loss (Case et al. 2007; Farrington et al. 2009; Furedi and McGraw 2004; McGraw et al. 2003; McGraw et al. 2013; Souther and McGraw 2014). Using simulations, Souther and McGraw (2014) found that climate change, when coupled with harvest by humans, creates a 65% extinction risk for American ginseng over the next 70 years. Current regulations on ginseng harvest must be revised to mitigate the predicted effects of changing climate, and a pre-emptive style of management may increase the potential to conserve this species as well as others (Souther and McGraw 2014). Furthermore, a study by Furedi and McGraw (2004) found that deer browse may be having a large impact on American ginseng as seeds are most likely destroyed by the digestive process. Farrington and colleagues (2009) suggest that ginseng must be managed both to control harvest by humans and herbivory by deer to preserve and restore its populations. Lastly, as is the case with most rare species today, habitat loss is another important influence on species' decline—and ginseng is no exception. Understanding the myriad of threats to ginseng populations will enable land-managers to gain a perspective on how its habitat, characteristics of growth, and distribution may be positive as well as negative influences on its potential to be restored and sustainably harvested.

American Ginseng as a medicinal NTFP

Agroforestry and non-timber forest products are growing industries in many parts of the world including the central-Appalachian region (Chamberlain 2009; Peri et al. 2017; USDA 2017).

Medicinal plants make up the largest portion of the non-timber forest product industry, and it is estimated that nearly three-quarters of people worldwide depend mostly on herbal medicine for their healthcare (Priya 2017). A report by Chamberlain (1998) stated that the worldwide market for herbal medicines valued at almost $8 billion and that number was expected to grow to more than $12 billion by the year 2000. A more recent report estimated the herbal market to be worth $50 billion (Nirmal et al. 2013). This is clearly a growing industry that necessitates further research into medicinal plants, their ecology, and implications for economically stressed regions in which forest farming may be more practical.

Ginseng's medicinal properties are mostly due to chemical compounds called ginsenosides: more than sixty types of these compounds are found within the roots and leaves of ginseng (Fuzzati 2004; Mazza 1996; Qi 2011). American ginseng has been shown to have many medicinal uses, from being high in antioxidants and relieving stress to inhibiting cancer cell growth (Duda et al. 1999; Kim et al. 2010; Kitts et al. 2000; Qi et al. 2011). The concentrations of different ginsenosides may vary depending on geographic location and habitat conditions—although the effects remain mostly consistent and products are not processed separately based on varying chemistry (Carlson 1986, Hu 1976, Wang 2005). These potent and time-tested medicinal properties of ginseng have allowed natural, high-quality roots to command a price of several hundred dollars per pound in Chinese herbal markets for generations—making it one of the most valuable herbaceous plants native to North America (McGraw et al. 2013, Snow 2009). Due to the extent of research that has gone into the phytochemistry of ginseng, it stands to reason that it should remain one of the most sought-after medicinal herbs across the globe.

Harvest of American Ginseng

Shortly after European colonization of the New World, settlers recognized the relationship between Chinese ginseng (*Panax ginseng* C.A. Mey.) and American ginseng and began exporting the plant to China (Carlson 1986; Case et al. 2007). According to Carlson (1986), China's demand for American ginseng nearly led to its extinction due to overharvesting—just as it had for their own Chinese ginseng (as cited in McGraw et al. 2013). The demand for ginseng has been reduced, but ginseng still remains one of the most valuable plants in North America (McGraw et al. 2013, Snow 2009). Today, harvest by humans remains a large influence on the abundance and distribution of American ginseng in the United States (Farrington et al. 2009; McGraw et al. 2013; Van der Voort and McGraw 2006).

American ginseng is listed on Appendix II of Convention of International Trade in Endangered Species on Wild Fauna and Flora (CITES). This means that the plant is not currently threatened with extinction, although it may soon become so if trade is not directly controlled or regulated. Some states, such as Wisconsin, have attempted to protect ginseng with legislation. In 1905, Wisconsin passed a law prohibiting digging ginseng between January 1 and August 1 to ensure that ginseng is less likely to be harvested before it has had a chance to fruit—although this law now exempts landowners (Carlson 1986).

Issues involving ginseng harvest are worsened due to the fact that, while it can be grown and harvested in a commercial setting, buyers typically prefer wild ginseng which are distinguished by their forked roots and many annual scars indicating old age (Carlson 1986). Wild ginseng has also been shown to have differences in ginsenoside content, which suggests that there may be pharmacological and therapeutic advantages to using wild ginseng (Wang et al. 2010). This preference only exacerbated problems with overharvesting of wild ginseng—demonstrating the need to regulate wild ginseng harvest and poaching.

Contrary to what this may suggest, ginseng cultivation still occurs in the United States. In the 19th century, nearly thirty million pounds of ginseng were exported to Asia (Carlson 1986). Since the 1960's, Wisconsin has led the country in exports of American ginseng, and it is very important to their economy even today (Carlson 1986). Some farmers grow ginseng on their own in a wild-simulated manner, sowing pre-stratified seeds in mulched beds with synthetic shade to simulate forest canopy cover (Oelbermann and Milburn, 1994).

Today, the many regulations regarding ginseng harvest may not be adequate to protect wild populations. Van der Voort and McGraw (2006) suggest that a "stewardship-oriented harvester" who collects ginseng responsibly by delaying harvest, plants seeds, and limits the intensity of their harvest can improve population growth rates. Since ginseng produces seeds in the summer months, it is recommended that harvesters delay their digging until September and allow for the fruit to ripen before harvesting (Carpenter and Cottam 1982).

Given the growing preference for environmentally friendly and sustainable natural products, the market would likely benefit from a transition to this model of ginseng harvesters. There are several factors limiting the amount of ginseng that the average landowner may try to grow, such as little-understood habitat preferences coupled with long germination times, slow growth, and the risk of poaching (Carpenter and Cottam 1982; Snow 2009). As mentioned previously, using habitat models to identify environmental factors that may be important for the growth of ginseng could be a means of enabling landowners to more easily grow and harvest ginseng on their property.

Species distribution models for conservation efforts

When attempting to conserve a threatened species such as ginseng, land-managers must understand the habitat in which it best competes. Many agencies' funding is too limited to conduct in-depth or long-term field surveys and as a result land-managers can be unsure of where threatened species may be located. This issue can be alleviated using species distribution models to identify areas where the probability of wild populations being present is high (Guisan and Thuiller 2005; van Manen et al. 2005). When sampling is not equal throughout ecologically diverse areas such as Virginia, it may be more practical to create separate models for each region—as geology, soil conditions, and topography can vary across physiographic provinces. Evaluating models to find the most appropriate methods to predict either occurrence or suitable habitat could facilitate the location of rare plants or animals, help recovery efforts and field surveys, and allow land managers to protect those species (Odom and McNab 2000; van Manen et al. 2005). To that end, this study utilized random forest, a machine-learning algorithm that learns from a sample of data to improve the performance of predicting an outcome when provided with new data (Breiman 1999; Dietterich 2000; Ho 1998; Kohavi and Provost 1998; Liaw et al. 2002).

The advantages of using random forests include being reasonably simple to understand and interpret while being able to perform well with large data sets acquired through geographic information systems (GIS) and mirroring human decision making more closely than other approaches (Breiman 1996b; Breiman 1999; Dietterich 2000; Pal 2005). These models can ideally generalize from a trend in data rather than "memorizing" the training data and therefore becoming too rigid in its predictions. Specifically, random forest uses decision trees to classify data based on features. It creates a tree by first splitting the data, so information gain is highest (i.e., in a way that allows the algorithm to learn the most; Breiman 1999). The algorithm then begins to split the dataset in the largest and easiest way, then continues splitting it until entropy or randomness in the

dataset is zero, or the outcome is the same within each group of data (Fig 1a, Fig 1b). It also applies ensemble learning, using multiple learning algorithms to improve predictive performance further than could be obtained by a single algorithm (Breiman 1996a; Breiman 1996b; Dietterich 2000). In other words, random forests construct hundreds of trees and weighs them all equally to secure "votes" about the outcome. In this study, the outcome will be ginseng presence or absence. Data will be split based on all features and the model will learn how to use those features to predict the probability of presence or habitat suitability. With this process, quantifying the importance of certain variables and their influence on presence and growth will be simplified for researchers and potentially invaluable to landowners interested in harvesting ginseng.

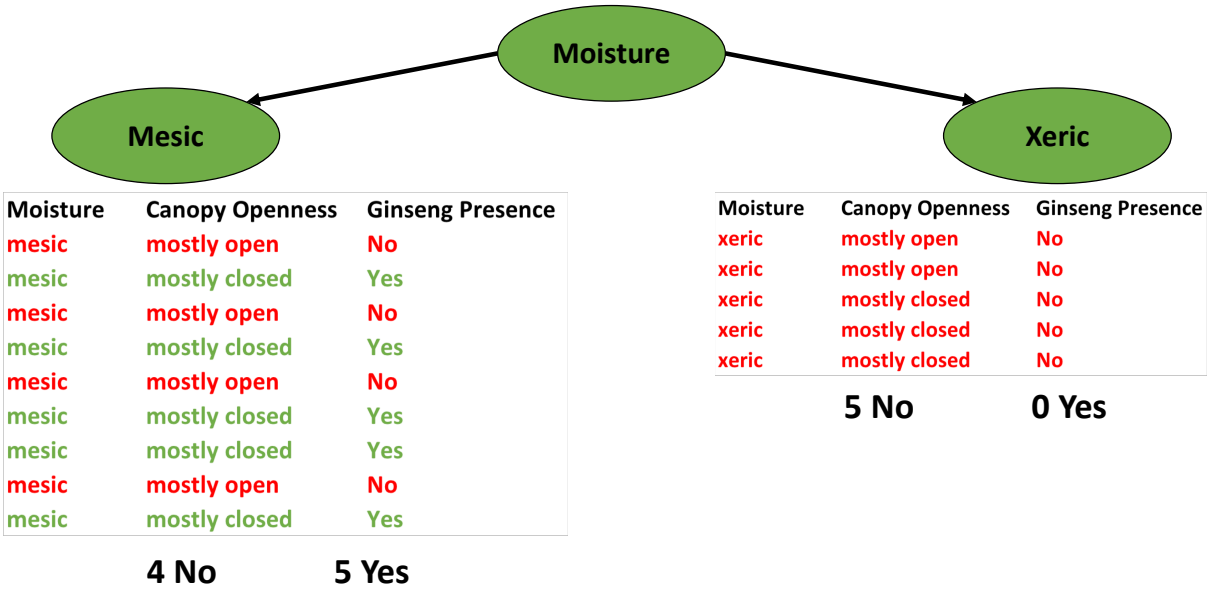

Figure 1a initial branch of a simulated data set containing soil moisture, canopy openness, and ginseng presence; decision tree first split these data by soil moisture, and then observes the output

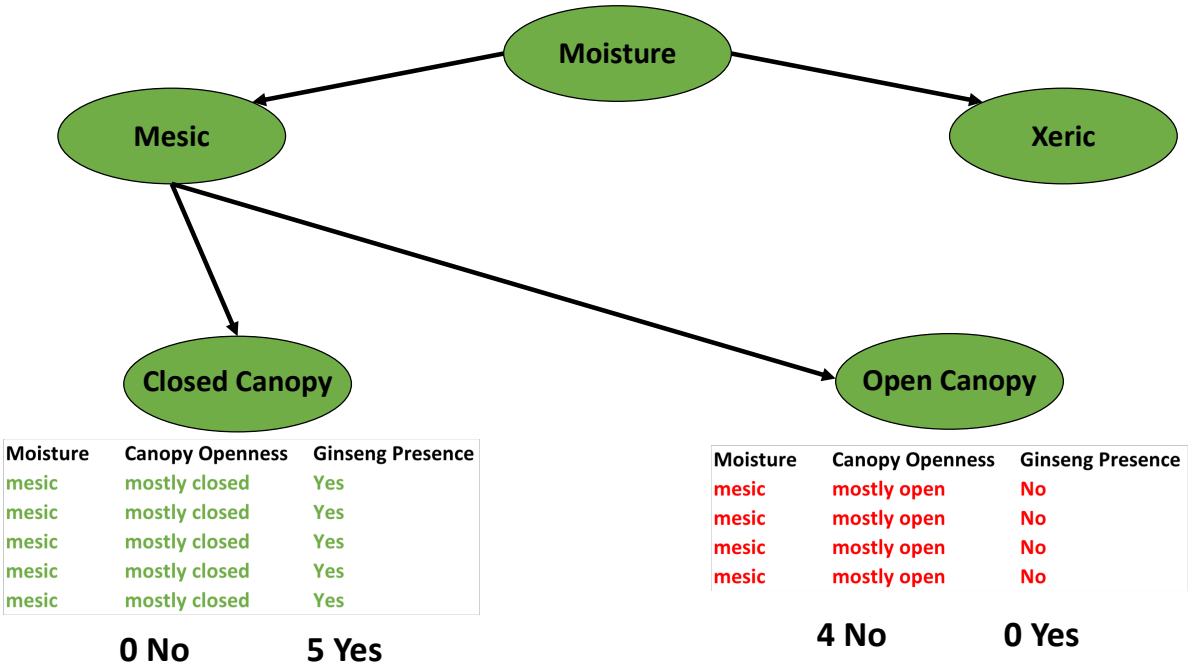

Figure 1b second branch of a hypothesized data set containing soil moisture, canopy openness, and ginseng presence; decision tree has observed that canopy openness at mesic sites can be a good predictor of ginseng presence

The Virginia Division of Natural Heritage is creating species distribution models using these methods. Recently, a model was constructed for American ginseng, of which this study examined and utilized presence points, a suite of 78 environmental variables, and model output (Virginia Natural Heritage Program 2017; see methodology). Ginseng presence data incorporated into the model reveal a sampling bias toward the Shenandoah National Park area—having over 500 documented populations while only 142 were documented in a comparably sized area of southwestern Virginia and 16 documented within the Cumberland Plateau (Virginia Natural Heritage Program 2017). Uneven sampling effort is a common issue across large areas, and although random forest and machine learning methods in general have been shown to be robust to low sample sizes, geographical sampling biases of this nature could have serious impacts on the predictive models constructed from those data (Hui et al. 2011; Syfert et al. 2013; Stockwell and Peterson 2002; Phillips et al. 2009). Ecologists suspect that, in Virginia, ginseng would be most

abundant in the Cumberland Plateau (or more specifically the Cumberland Plateau) as that region typically contains more productive mesophytic forests.

The Cumberland Plateau

The High Knob Landform of southwestern Virginia lies in the Clinch River Watershed: one of the richest biodiversity hotspots for rare, threatened, and endemic species in North America (Nature Conservancy 2016; NatureServe 2013; Wilson and Tuberville 2003). Due to its unique location and topography, High Knob and the surrounding landscape of the Cumberland Plateau contain more microclimates and species assemblages than ecologists had predicted (Braun 1950; Browning 2018). According to The Nature Conservancy's climate resiliency map, this region of southwestern Virginia is less susceptible to climatic variation and therefore may serve as refugia for many species in the coming decades as climate change progresses (Anderson 2016). However, due to its remote location and resulting distance from large research institutions, this area is not well known by researchers. This has led to a biodiversity sampling deficiency—making this a noteworthy region for ecological research.

Southwestern Virginia has suffered a decline in its economy over several decades; therefore, ensuring that sustainable harvest can be conducted by local forest-farmers and landowners may provide a valuable means of long-term economic gain (Maggard 1994; Taylor et al. 2017). With regional developments such as local forest farming groups and growing interest in sustainable harvesting, understanding this region's ecology as it pertains to forest herbs and soils may prove essential for success of those practices (Maggard 1994). The quantification of intelligible influences on ginseng's presence, health, and growth should be critical for landowners attempting to grow and harvest for a profit.

It was hypothesized that American ginseng may be present in the Cumberland Plateau of southwestern Virginia at aspects, elevations, slope inclinations, and soil conditions that current models have not predicted due to sampling deficiencies coupled with the unique physiography and regional microclimates. The major goals of this study were to (1) use occurrence data provided by the state of Virginia's Natural Heritage Program to test the efficacy of the model in southwestern Virginia through the alteration of model parameters and additional sampling, thereby improving the models' accuracy and efficiency, (2) attempt to increase predictive power in specific regions of Virginia by limiting the spatial extent of model training data, (3) quantify relationships between the presence of ginseng and environmental variables and (4) establish long-term seed plots to be monitored for survival and growth by citizen scientists.

It was predicted that documenting additional populations in addition to changing various methods such as removing correlated variables, thinning presence points (spatial partitioning), and reducing the extent from which background points are generated would result in more accurate predictions for species distribution and habitat suitability. Improved models and the knowledge generated from them could facilitate the location of wild populations, their preservation, and the use of caution when conducting intensive land-management projects. Identifying locations in which ginseng best competes could also enable easier, low-risk planting and more economical harvest of wild or wild-simulated ginseng. Finally, conserving ginseng in this region will lead to easier and more sustainable harvesting in the future and could provide a boost to declining economies in regions where ginseng is most prevalent.

METHODOLOGY

Site selection

Using ArcMap (version 10.7.0.10450; ESRI Inc., California), 130 random points were generated throughout the USDA Forest Service, Clinch Ranger District. These points facilitated surveys in and around the City of Norton as well as Wise and Buchanan counties in Virginia. They were used as a guide to select general survey locations, rather than strictly for the establishment of transects or small study plots. For any point, the entire locale would be surveyed, and any ginseng found while travelling to or returning from that location would be documented. This method was predicted to increase documentation of plant occurrence as opposed to strictly surveying along transects, as the area surveyed was greatly increased. No surveys took place on private property without permission from the landowner.

In addition to randomly generated points, communication with local citizens and professionals was used to identify additional ginseng populations. Establishing a rapport with locals and understanding folklore surrounding ginseng was determined to be an efficient means of identifying appropriate habitat or areas in which the presence of ginseng may be more likely. However, surveys were not exclusively focused on those areas. Habitat types ranging from rich cove forests to recently burned ericaceous habitats were surveyed. The USDA Forest Service uses a suite of ten species to determine appropriate ginseng habitat (Kauffman 2006). These species include maidenhair fern (*Adiantum pedatum* L.), rattlesnake fern (*Botrypus virginianus* (L.) Michx.), bloodroot (*Sanguinaria canadensis* L.), black cohosh (*Actaea racemosa* L.), blue cohosh (*Caulophyllum thalictroides* (L.) Michx.), hairy sweet cicely (*Osmorhiza claytonii* (Michx.) Britton), bedstraw (*Galium triflorum* Michx.), Canadian violet (*Viola canadensis* L.), dutchman's-pipe (*Aristolochia macrophylla* Lam.), and goldenseal (*Hydrastis canadensis* L).

To reduce sampling bias toward habitats containing companion species, habitats that contained no companions were also surveyed. Furthermore, since absence points are not surveyed and counted in this study, presence points in the region should be considered relatively unbiased. For example, a heavily harvested area that may be appropriate habitat—but harbors no ginseng—would not necessarily be considered an absence point. Using background data (pseudo-absence points generated across a study region) rather than true absence points enables the negative response values to represent the average habitat over the study area rather than focusing on a few select points strictly identified as being devoid of ginseng (Barbet-Massin et al. 2012).

Species distribution models

All GPS data underwent a differential correction process using Trimble GPS Pathfinder Office (version 5.90, 2018) to increase accuracy and reliability. Data were then exported to an ESRI shape file and projected using the NAD 1983 UTM Zone 17N coordinate system. Points were organized in ArcMap and then transferred to an R programming environment where they were managed from that point forward (R development core 2007). All raster data were opened within R Studio (version 3.6.0, 2019) using the rgdal package and stacked using the raster package to create a single object representing all 78 environmental variables (Bivand et al. 2019; Hijmans 2019). These rasters represented precipitation, temperature, topography, geology, hydrography, and land cover data. Using a raster stack, values from all variables could be extracted for each presence and background point to generate the data set used for modeling ginseng presence. Background points were used in lieu of absence points for this project to enable the models to learn the average background of the study region. For baseline models, the number of background points used was roughly equal to three times the number of presence points in the data set (1955 presence points and 5664 background points). Background points were generated using two different

methods to evaluate the efficiency of either method. The simplest method of background point creation is to produce a given number of points spread randomly but evenly throughout the study area, Virginia. Alternatively, Phillips and colleagues (2009) recommend that fewer background points should be generated and with similar bias as presence points (i.e., closer to roads or locations that are more accessible). Based on those guidelines, some models used roughly the same number of background points as presence points, generated exclusively within a 10-kilometer radius around presence points (Fig 2). This second method was used as an effort to alleviate bias such that areas that were not sampled would contain neither presence nor background points.

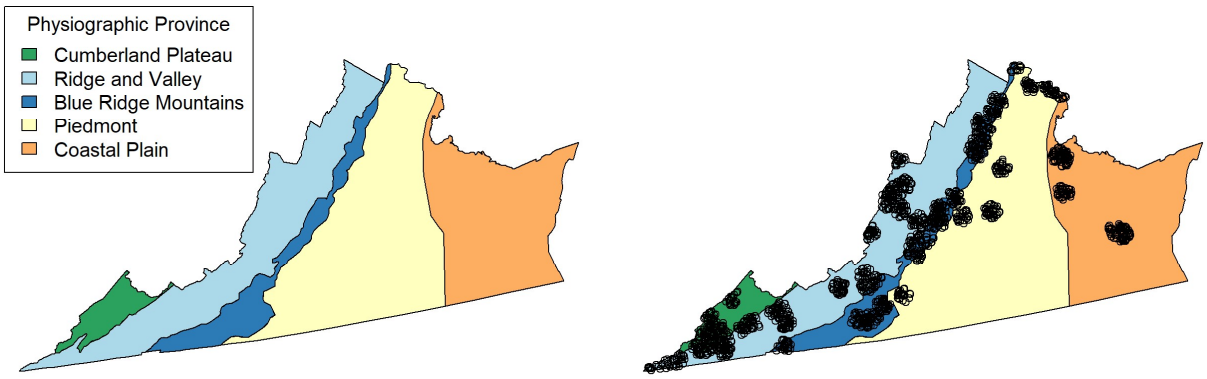

Figure 2 Virginia physiographic provinces (left); Background point generation using a 10-kilometer buffer around presence points (right)

Each forest constructed in this study consisted of 2000 trees (k), with each node of a tree containing a number of variables (m) equal to the square root of the total number of variables present in the overall dataset. This varied depending on the model but was usually either 7 or 9, meaning as many variables were tried, or examined, at each node (Breiman 2001; Liaw et al. 2002). This method creates many trees where each tree consists of a random subset (of a defined sample size) of the data with regard to the randomly selected variables in each branch. The tree continues to branch until the response (ginseng presence or absence) is the same result at each tip—allowing

the model to learn which variables may be the most important in predicting the presence of ginseng. The sample size used for stratification in the models were equal to 20% of the number of total presence points (391 for 1955 presence points). Using this stratification method enables each tree to contain an equal number of presence and background points.

In an attempt to reduce bias or noise in the modeling process, correlation tests were conducted among variables. Due to the large amount of processing power required, it was impractical to run a single test with all 79 variables—which would also generate a large correlation matrix prone to human error when reviewed. Variables were grouped logically by categories such as climate or geology and a test was conducted for each group. This allowed for more efficient processing and greater ease of analysis. Any correlation values over 0.6 were considered correlated for this study. Using the importance rankings from initial random forests, variables were selected to be used in refined models. If a group of variables were correlated, only the most important variable per the model would be selected out of that group for use in further models.

Rigorous comparison of model iterations must take place to determine which changes may or may not be necessary to improve accuracy—as each change to the modeling process requires more code, time, processing power, and storage space. This was particularly important, as a major goal of this study was to identify the simplest and most efficient means of creating species distribution models using random forests. Models were evaluated using the out-of-bag error (OOB error) as well as the area under the receiver operating characteristic curve (AUROC or AUC). These metrics were selected for evaluating models due the OOB error being calculated within random forest, and AUC being a common approach (Breiman 1996b; Fawcett 2006). Within each forest, one-third of the data are withheld for testing and used to calculate the OOB error (Breiman 1996b). Modeling runs were repeated 1000 times and the median OOB error was documented. The

average AUC was calculated from five-fold cross validation repeated 200 times for a total of 1000 iterations of each model. These evaluation metrics serve as both means of interpreting error in predictive power as well as overfitting. For example, a model with an incredibly high AUC (>0.99) might be very good at predicting presence points in the data but could potentially be too conservative when making a prediction using environmental variables. When model parameters were changed, such as using different methods for generating background points, it necessitated additional cross validation. A total of ten statewide models were tested. Each random forest model was trained and used to create a prediction in the form of a new raster where each pixel represents the probability of ginseng presence.

Initial models, including the model constructed by the Virginia Natural Heritage Program (2017), consisted of a final prediction calculated for the entire state. Ten separate statewide models (SM1–10) were constructed using various methods such as limiting the area from which background points are generated, spatial partitioning, selecting only non-correlated, important variables, and adding additional presence points documented in this study. The initial model, SM1, was constructed using basic methods (i.e., none of the previously described methods were used, and additional points were not yet added) to represent a baseline model for comparison. SM2 was made using spatial partitioning. Spatial partitioning methods reduced the number of presence points that are stacked in a small area, leading to multiple points within a single pixel. Points were rasterized and thinned such that only one presence point was remaining in a given pixel, which typically reduced the number of presence points by around 30%. SM3 was made using basic methods but underwent a variable selection process. SM4 was made using both spatial partitioning and selected variables. SM5 used only selected variables and limited background point generation. SM6 was constructed using a combination of methods from SM2–5. SM7 was made using base

methods from model 1, while SM8 used only selected variables and limited background points and SM9 used selected variables, limited background points, and underwent spatial partitioning on all presence points. Lastly, SM10 was an experimental model constructed using the same methods from SM9, but spatial partitioning was only conducted on the original 1955 presence points.

To investigate whether models generalized for the entire state are biased to the regions with increased sampling effort, models with limited spatial extent were constructed to focus solely on the Cumberland Plateau (CPM1–4). All rasters and presence points were loaded into an R programming environment and masked such that only variable and presence point data representing that region would be incorporated into the model. Two of these provincial models (CPM1, CPM3) were constructed without new occurrence points, using baseline or variable selection and limited background point generation, respectively. CPM2 and CPM4 included newly documented occurrence points with baseline or variable selection and limited background point generation, respectively.

Partial dependence plots (PDPs) of important variables from the models that appeared to be the most useful for predicting habitat suitability or species distribution (i.e., low indications of over-fitting, high accuracy, and ecologically sensible model predictions) were constructed using the pdp package in R (Greenwell 2017). Variables that frequently appeared as important according to the different models, while being intelligible in the field were selected for PDPs. These plots illustrate the marginal effect that a certain variable has on the outcome of the model in question. For example, one partial dependence plot could show how the probability of ginseng presence is affected by changes in precipitation levels across the study area. Following methods similar to those of the Virginia DCR division of natural heritage, PDPs will also have density plots showing the distribution of presence and absence superimposed above (Virginia Natural Heritage Program

2017). This is to visualize both how the variable impacts ginseng presence, in addition to being transparent with the distribution of ginseng presence across the range of variable.

Field surveys

Individual plants were documented using a Trimble GeoExplorer 6000 series (model 88950; Trimble Inc., California) unit. Plant height, leaf length and width, the number of prongs (leaves) present, reproductive capability, and chlorophyll content of the leaves were quantified. Plant height was measured in centimeters from the ground or top of leaf litter if a layer was present to the top of the most prominent leaf. If a plant was growing in an irregular way (with a curved or damaged stem), the height of the plant remained the distance between the ground to the top of the leaf, not the entire length of the stem. Individuals were considered reproductive if they had either evidence of a flower stalk early in the season or flowers and fruit later in the season. The "greenness" of each plant was recorded using an index of relative chlorophyll content (ranging from -9.9–199.9) with a Soil-Plant Analyses Development (SPAD) 502 chlorophyll meter (Konica Minolta Sensing, Inc., Japan). Leaf area and chlorophyll measurements were taken from one leaf typically at random or from the most-intact leaf if there was damage to the plant.

The abundance of each companion species was ranked using an ordinal scale (1–5) in the field when ginseng was present (Table 1). A rank of one was used to characterize species that were present in the same habitat or physiographic position (i.e., same slope or cove) but were not growing within two meters of any ginseng plant. A rank of two was to classify a species that was present in small numbers within two meters of ginseng, with one to five nearby neighbors. A rank of three indicated that the companion species was moderately abundant near a ginseng population, with six to nine nearby neighbors. A rank of four indicated that the companion species was abundant near a ginseng population, with 10–19 nearby neighbors. Lastly, a rank of five was used

to classify companion species that were very abundant and dominated the herbaceous layer, with 20 or more individuals in the immediate area.

Table 1 Companion species abundance rankings

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

All companion species were ranked using this scale with the exception of goldenseal and bedstraw. Goldenseal was predicted to be much rarer than the other species and therefore its abundance was not ranked initially. However, it was anecdotally noted that goldenseal was more abundant in this region than previously assumed and should be documented when conducting ginseng surveys in the future. Bedstraw is a small plant which spreads from creeping rhizomes, and often forms dense mats—making it difficult to estimate the number of individuals in an area. Because of this, bedstraw's presence was noted but an attempt to count and rank this species' abundance was not made. Additionally, since it may be easily confused with other members of *Galium* sp., this study was effectively noting presence of *Galium* sp. rather than *G. triflorum*.

Soil samples were collected near each ginseng population. Single populations were assumed if plants were directly surrounding each other, such as multiple small plants growing under a reproductive individual, or if they were in the same physiographic position, such as being located on the same slope, exposure, or cove. No pre-determined distance was used as a cutoff for

populations to be considered separate, as unrelated plants may be on opposite sides of a ridge but closer together than a parent located at the top of a slope with offspring far downslope. It should be noted that depending on dispersal in a given area, some groups of ginseng that were considered separate could be one population. The only method of accurately determining whether populations were truly separate would be through genetic analysis (not performed in this study).

Soil from randomly chosen samples (N = 22) was sent to the University of Georgia Extension's Agricultural and Environmental Services Laboratories (AESL) to obtain data on pH, phosphorus, potassium, calcium, magnesium, zinc, and manganese. Soil data were used only for exploratory analyses to summarize conditions where ginseng was present.

Long-term monitoring sites

In an effort to set up future studies on the germination and growth of American ginseng in the Cumberland Plateau, long-term monitoring sites were established at 15 locations in seemingly suitable or unsuitable habitat as interpreted by researchers (based on present site conditions or land use history) on property managed by the USDA Forest Service, Clinch Ranger District or the University of Virginia's College at Wise. Ginseng seeds were planted in relatively undisturbed Appalachian cove forest within national forest (logged >100 years ago), under 30-year old autumn olive (*Elaeagnus umbellata* Thunb.) on a reclaimed mine site, and in a 100-year forest that was mined in the 1800s. At these locations, sprouting, wild-sourced seeds were sown and marked by GPS. Seeds were obtained from the North Carolina Goldenseal and Ginseng Company in Marshall, NC. This seed source was from populations located in proximity to the study area such that the likelihood of introducing individuals from populations with major genetic differences was reduced.

Plots consisted of five subplots, with one center and four additional subplots established 10 meters from the center. Ten ginseng seeds were planted 2–3 cm deep spaced 5–10 cm apart in two rows of five (Fig 3). A GPS point and four photos were taken from each plot's center, facing each of the cardinal directions with plot number, a site description or name, and the direction recorded as well (Fig 3). GPS points and plot photos will all need to be used to find plots for follow-up studies.

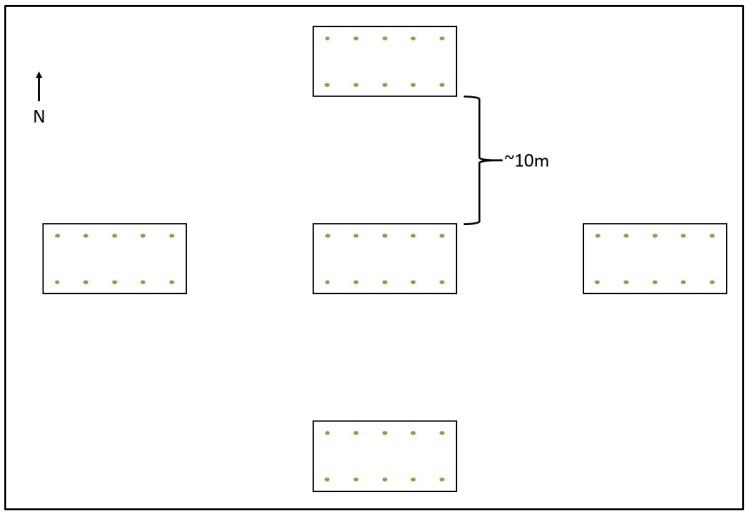

Figure 3 Long-term study plot design (N = 15) indicating the placement of subplots and arrangement of seeds within each; seeds were spaced 5–10 cm apart

RESULTS

Ginseng occurrence documentation

All newly documented ginseng plants were added as new occurrence points for modeling ginseng habitat and distribution. This represents a 1200% increase in documented plants over an area of ~1,500 square kilometers (around half the total area of the Cumberland Plateau). The association between ginseng and companion species quickly became apparent and thereafter the presence of companion species became the most commonly used method to locate wild ginseng

populations. Ginseng populations found in this study ranged from a single plant to upwards of 40 plants.

Species distribution models

The statewide models (SM) of most importance were determined to be SM1, SM7, and SM9. The models with the highest and lowest area under the curve (AUC) values were SM7 and SM9, respectively (Table 2). The models with the lowest and highest out-of-bag (OOB) error were SM1 and SM9, respectively (Table 2). The mean predicted probability of presence of ginseng in SM1, SM7, and SM9 were 0.099, 0.104, and 0.212, respectively (Fig 4). With the addition of new occurrence points from this study (N = 198), the mean predicted probability of presence of ginseng in the Cumberland Plateau increased from 0.170 (SM1) to 0.278 (SM7; Fig 5). These values are contrasted in models restricted to the Cumberland Plateau (CPM), which were made using either methods similar to those used in SM7 or SM9—where the mean predicted probability of presence of ginseng ranged from 0.143–0.660 (Table 3, Fig 6).

Table 2 Modeling method, number of presence and background points, number of variables, out-of-bag (OOB) error, and area under the ROC curve (AUC) for state models; highlighted models emboldened; var=variable, bg=background, pres=presence, gen=generation

Modeling method	n presence	n background	n variables	OOB Error	AUC
(**SM 1**) baseline	1955	5664	78	3.13%	0.9926
(SM 2) spatial partitioning	1419	5664	78	4.32%	0.9880
(SM 3) var selection	1955	5664	39	3.89%	0.9924
(SM 4) spatial partitioning and var selection	1419	5664	39	5.56%	0.9881
(SM 5) var selection and limited bg point gen	1955	1765	39	6.08%	0.9857
(SM 6) var selection, limited bg point gen, and spatial partitioning	1419	1264	39	7.55%	0.9808
(**SM 7**) baseline with add pres	2153	5664	78	3.33%	0.9927
(SM 8) var selection limited bg point gen, and add pres points	2153	1904	39	6.41%	0.9871
(**SM 9**) var selection, limited bg point gen, add pres, and spatial partitioning	1475	1311	39	8.08%	0.9792
(SM 10) var selection, limited bg point gen, add pres, and spatial partitioning on state-documented points	1617	1432	39	8.03%	0.9828

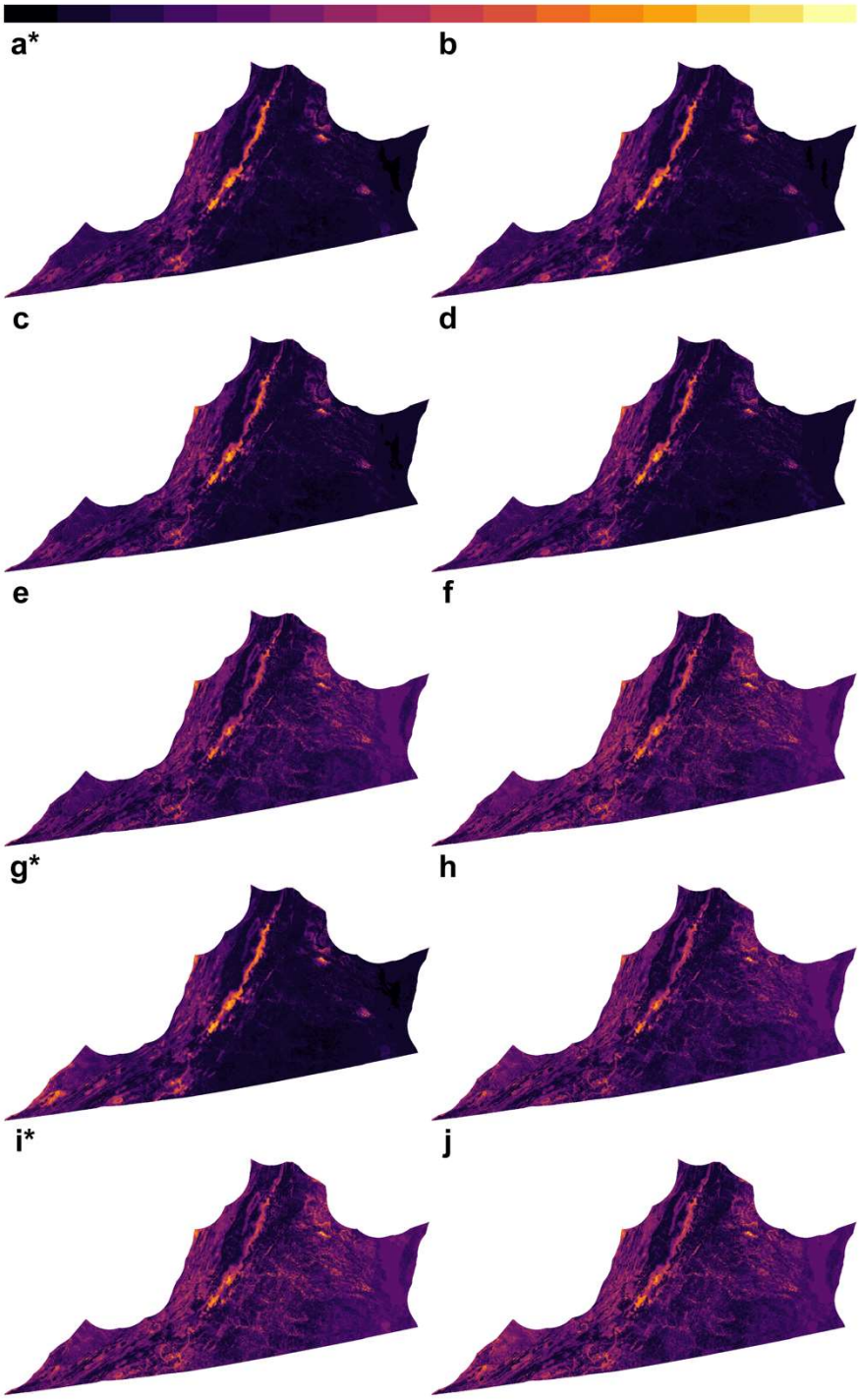

Figure 4 Comparison of statewide species distribution models (SM) for American ginseng; a–j represent SM1–10, respectively; the Cumberland Plateau includes the southwestern portion of each model; * indicates selected models; probability of ginseng presence ranging from 0–1 is shown above (darker to lighter)

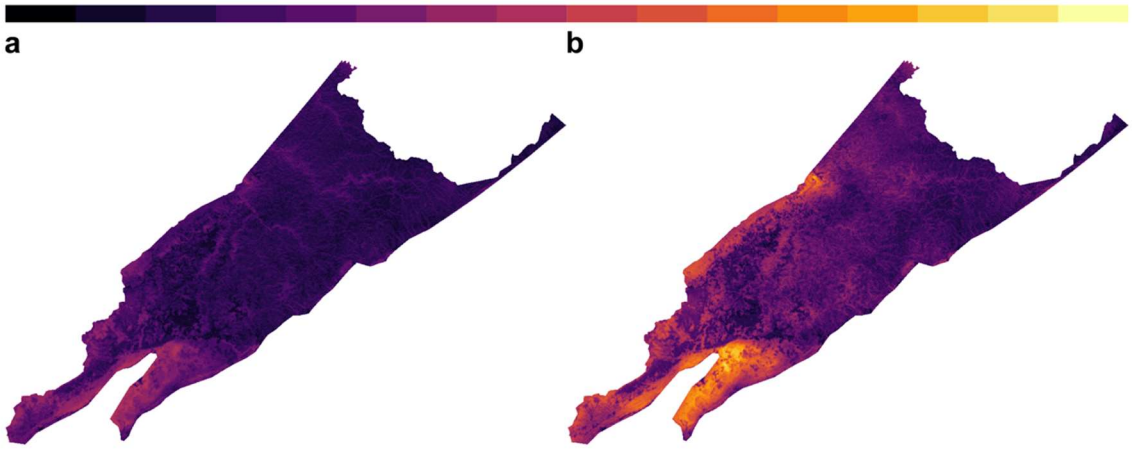

Figure 5 Comparison of selected statewide models cropped to the Cumberland Plateau in Southwestern Virginia; **a** SM1 **b** SM7; probability of ginseng presence ranging from 0–1 is shown above (darker to lighter)

Table 3 Modeling method, number of presence and background points, number of variables, out-of-bag (OOB) error, area under the curve (AUC), and mean predicted probability of presence for Cumberland Plateau models (CPM)

Modeling method	n presence	n background	n variables	OOB error	AUC	Probability
(CPM1) baseline	16	48	78	14.06%	0.9955	0.5990
(CPM2) baseline with add pres	209	627	78	3.48%	0.9890	0.1432
(CPM3) var selection and limited bg point gen	16	10	40	12.50%	0.9954	0.6602
(CPM4) var selection, limited bg point gen, and add pres	209	150	40	5.29%	0.9899	0.2476

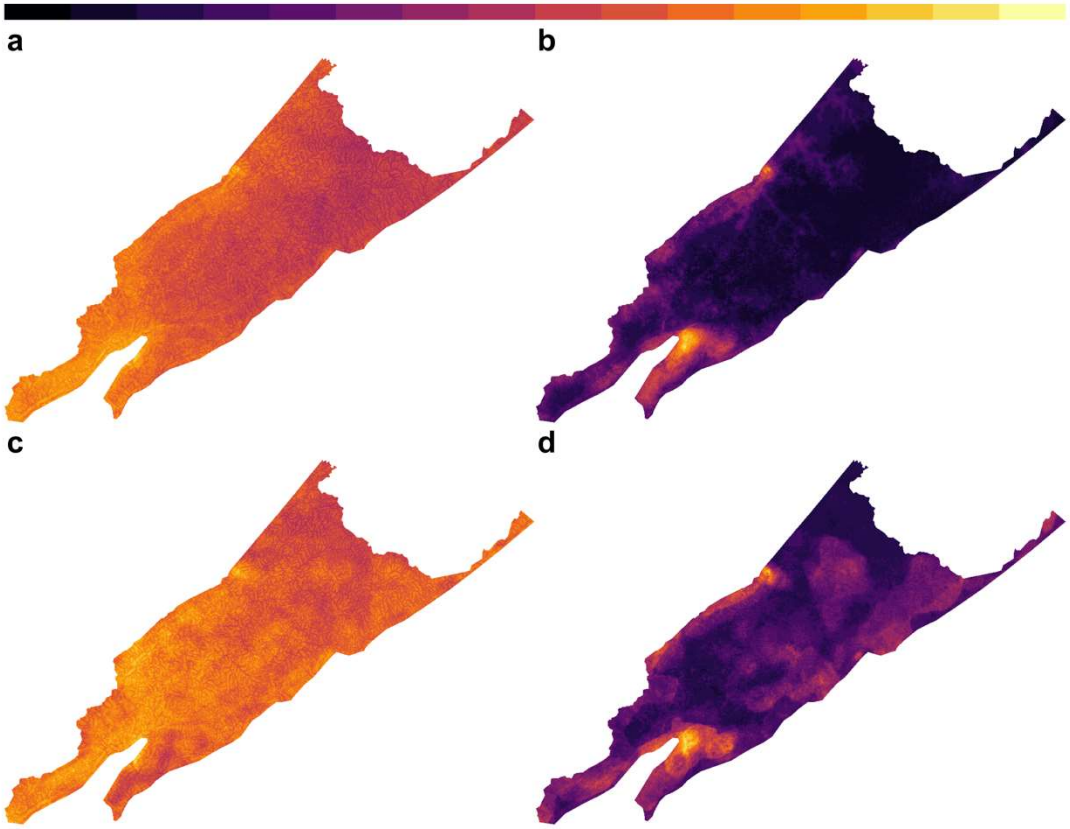

Figure 6 Comparison of provincial models for the Cumberland Plateau; **a** CPM1 **b** CPM2 **c** CPM3 **d** CPM4; probability of ginseng presence ranging from 0–1 is shown above (darker to lighter)

In all ten statewide models deciduous forest cover was the most important variable per random forest for predicting either ginseng presence or suitable habitat. This is displayed in the selected models SM1, SM7 and SM9 (Fig 7, Fig 8, Fig 9). Other variables frequently quantified as highly important were mean diurnal range, canopy cover and openness, and distance to streams (Fig 7, Fig 8, Fig 9). In SM7, as deciduous forest cover and general canopy cover increases, the influence of these variables in predicting the probability of ginseng presence increases—while increasing distance to rivers decreased predictive power (Fig 10). Other important variables in SM7, 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 (Fig 10). For example, SM7 experienced a

particular improvement to predictive power where the mean diurnal range (the mean difference in monthly maximum and minimum temperature) ranged from 10–12 °C (degrees Celsius; Fig 10). In SM9 these relationships between important variables and predictive power are mostly similar, with few exceptions. For example, roughness—a variable not displayed in a partial dependence plot for SM7—had larger impact on this model than in previous iterations. The relationship between the model's power and roughness was similar in its complexity to those variables in SM7, where median values (2000–4500 cm) had the greatest impact on predicting suitable habitat (Fig 10).

The provincial models with the highest OOB error and probability of presence were those which only included the 16 original presence points: CPM1 and CPM3 (Fig 6, Table 3). Pearson correlation values between provincial and statewide models (cropped to the plateau) were inconsistent. For example, CPM2 and SM7 were made using the same methods and had a correlation coefficient of 0.848—however, CPM4 and SM9 were also made using the same methods and had a correlation coefficient of 0.423. Additional provincial models are not presented, as there was not an apparent visual or statistical improvement for the plateau over using predictions from statewide models (Table 3, Fig 6).

In most cases, limited background points and variable selection (when combined with other methods) were shown to increase OOB error and decrease AUC (Table 2, Table 3). Variable selection reduced the number of variables used in statewide models from 78 to 39. Although spatial partitioning was not conducted in provincial models, it was shown to increase OOB error and decrease AUC in statewide models (Table 2). In statewide models where new occurrences were included, spatial partitioning decreased the number of presence points from 2153 to 1475 and the number of background points from 1904 to 1311 (Table 2).

In additional statewide predictions that were constructed from models (N = 2) that were trained exclusively on data representing either the Cumberland Plateau or the Blue Ridge Mountains, the mean statewide probability of presence of ginseng was 0.257 and 0.204, respectively (Fig A1). In particular, the model trained only on data representing the Cumberland Plateau predicted an average probability of presence of 0.143 for the Cumberland Plateau and of 0.239 for the Blue Ridge Mountains (Fig A1). The model trained only on data representing the Blue Ridge Mountains predicted an average probability of presence of 0.265 for the Cumberland Plateau and 0.232 for the Blue Ridge (Fig A1).

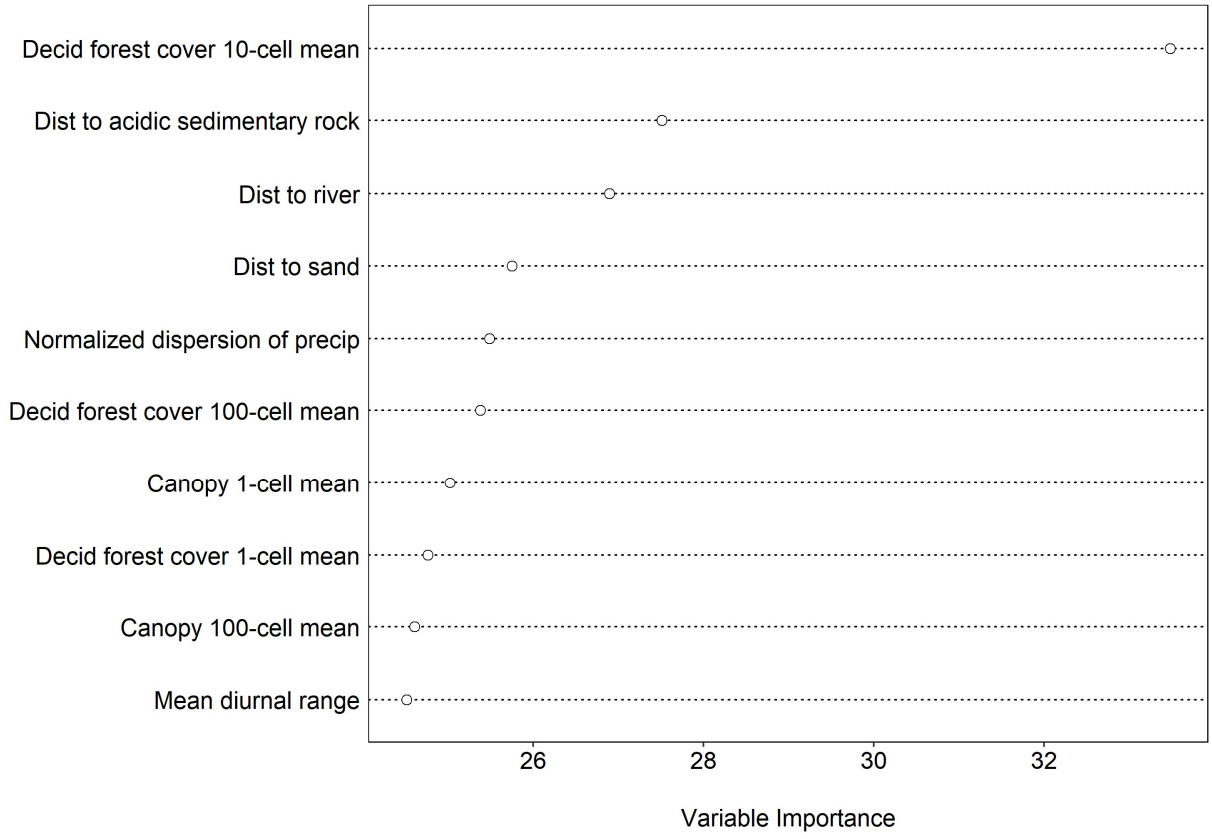

Figure 7 Importance rankings of top ten variables from SM1; importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, precip=precipitation

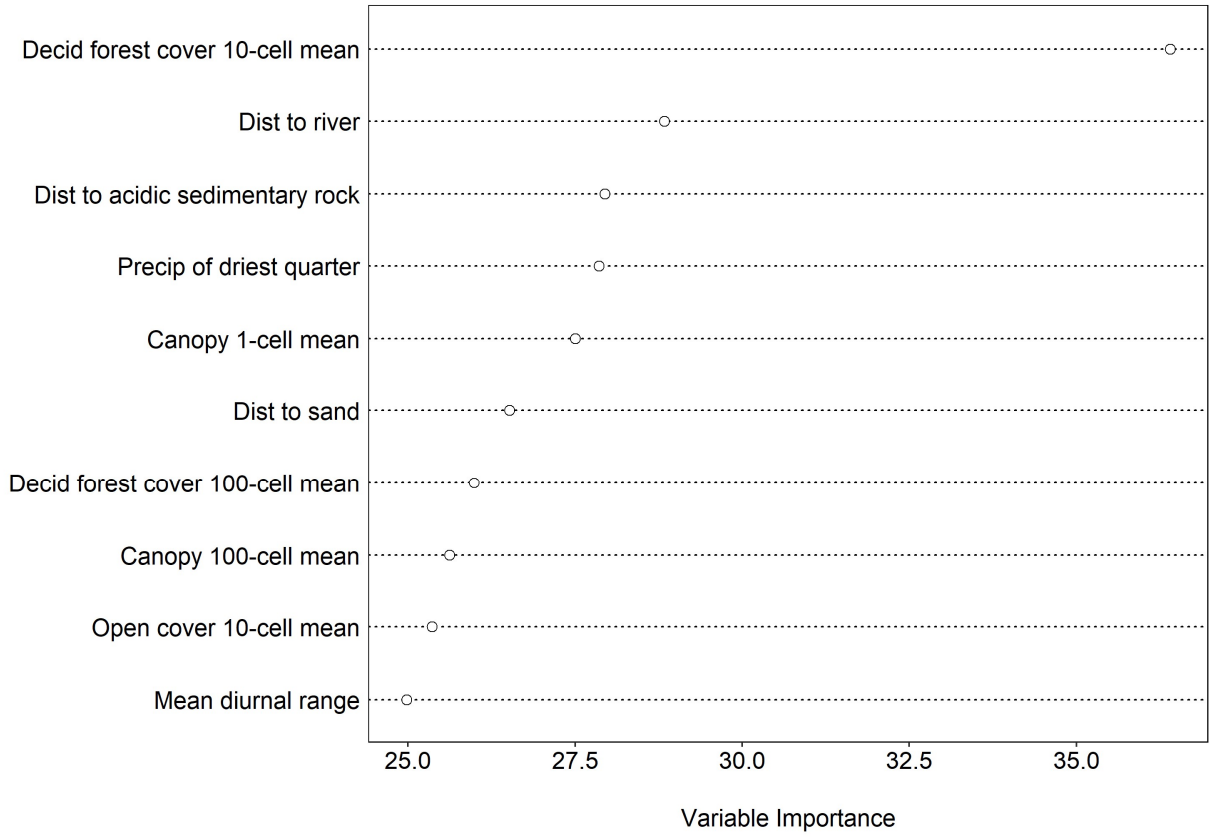

Figure 8 Importance rankings of top ten variables from SM7; importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, precip=precipitation

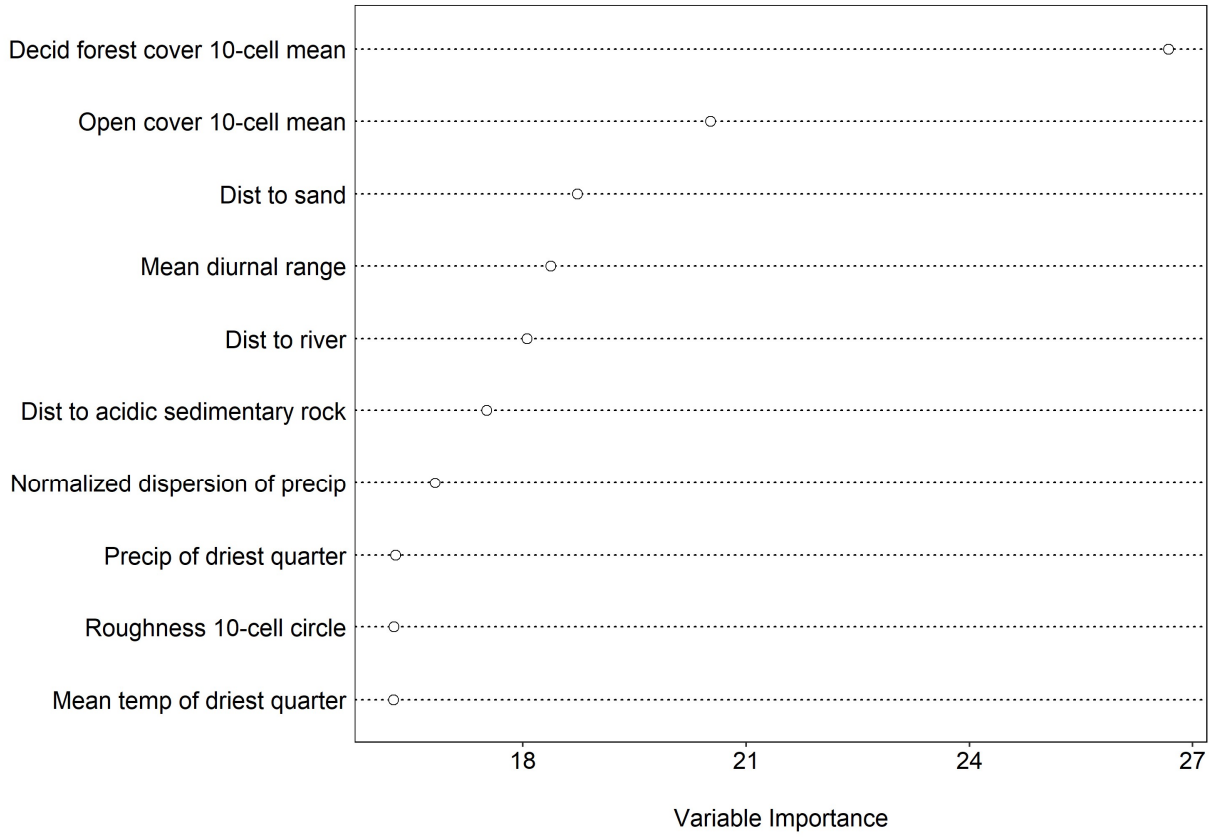

Figure 9 Importance rankings of top ten variables from SM9; importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, precip=precipitation, temp=temperature

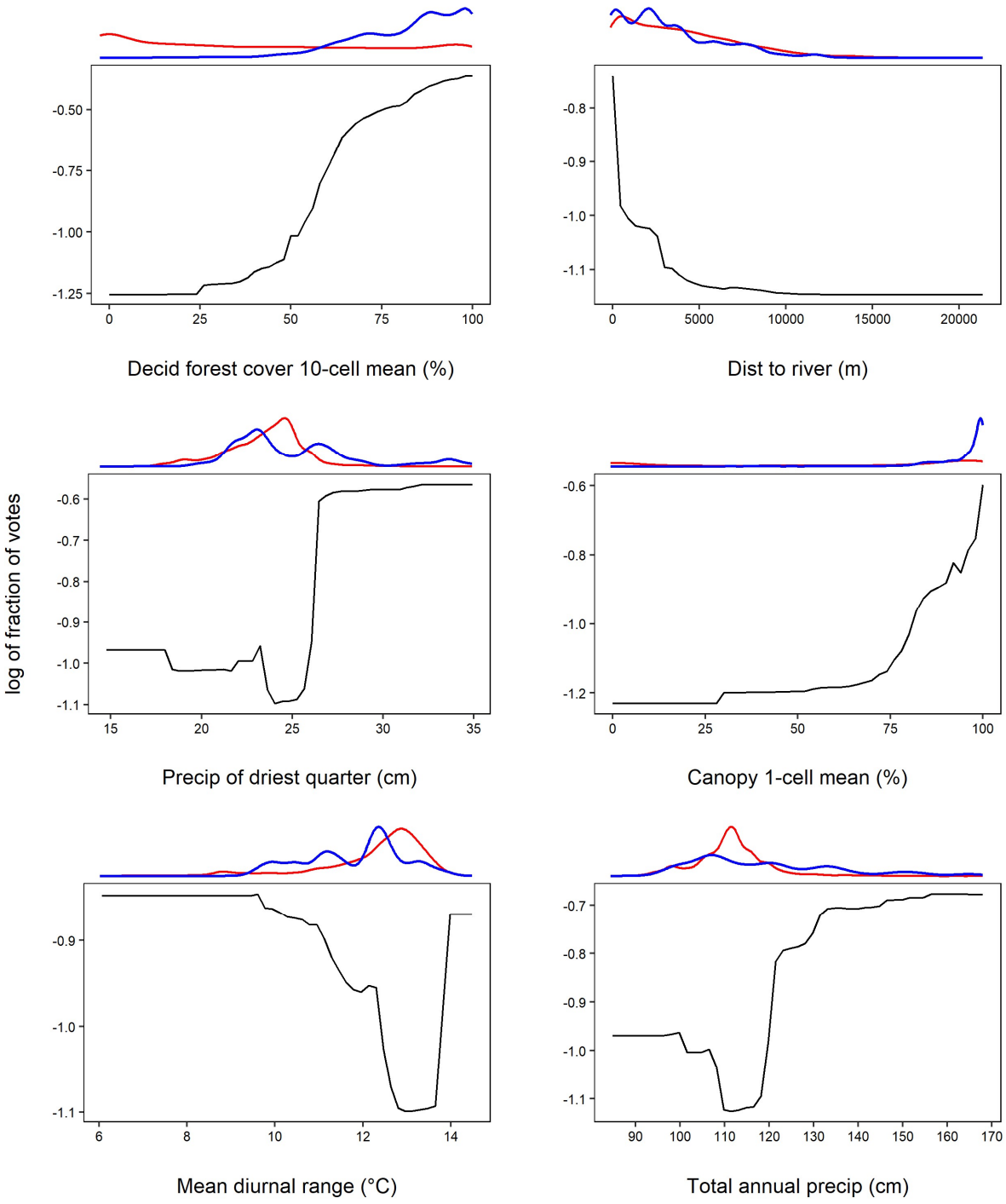

Figure 10 Partial dependence plots of six selected variables from SM7; y axes demonstrate the marginal effect that the variable has on predicting presence or suitable habitat for American ginseng; density plots of presence and background data are displayed above plot margins in blue and red, respectively; decid=deciduous, dist=distance, precip=precipitation

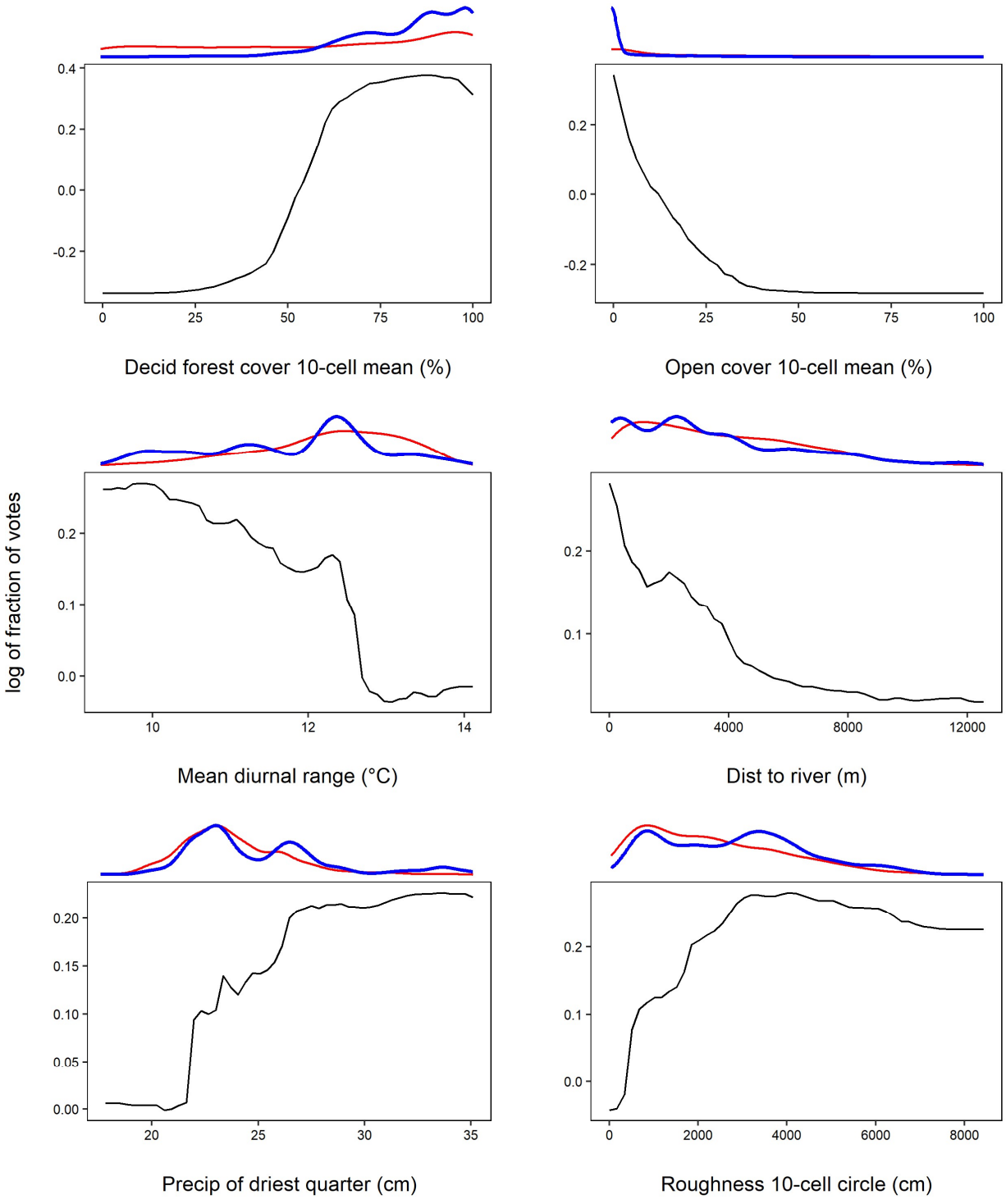

Figure 11 Partial dependence plots of six selected variables from SM9; y axes demonstrate the marginal effect that the variable has on predicting presence or suitable habitat for American ginseng; density plots of presence and background data are displayed above plot margins in blue and red, respectively; decid=deciduous, dist=distance, precip=precipitation

Ginseng health parameters

Although this area was previously predicted to be relatively low probability, robust populations were found with large reproductive individuals and offspring frequently present. Of the 192 plants measured in this study, 68 plants showed reproductive capability (Fig 12). The height of non-reproductive ginseng plants ranged from 4.7–26.0 centimeters (cm) with a median value of 9 cm, and their leaf area ranged from 17.5–792.0 cm^2 with a median value of 76 cm^2 (N = 124; Fig 12). Among reproductive plants, height ranged from 9.0–31.0 cm with a median value of 15.0 cm, and their leaf area ranged from 104.5–870.0 cm^2, with a median value of 360 cm^2 (N = 68; Fig 12).

The tallest plants were reproductive individuals with three prongs (Fig 12). The number of individuals found with one, two, and three prongs was 65, 80, and 47, respectively (Fig 12). Among reproductive plants, only one was documented as having a single prong, while 24 were found with two prongs and 43 with three prongs (Fig 12).

Reproductive plants were found to have significantly higher chlorophyll levels than non-reproductive plants ($p<0.01$; Fig 13). Chlorophyll content (relative greenness) among all plants ranged from 12.8–38.2, with an average value of 27.9 (Fig 13). Chlorophyll content among reproductive plants ranged from 21.4–38.2, with an average value of 29.9 (Fig 13).

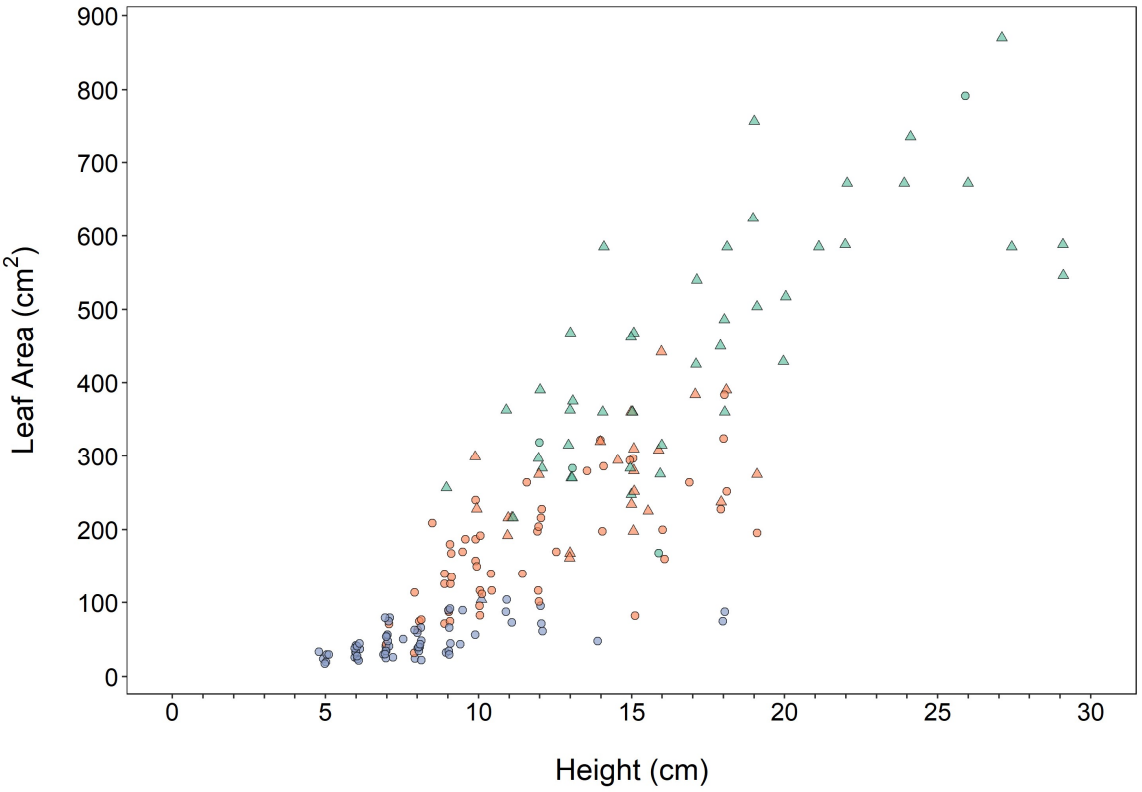

Figure 12 Height and leaf area per prong number and reproductive capability; circles and triangles represent non-reproductive and reproductive individuals, respectively; purple, orange, and green represent prong numbers of 1–3, respectively; R2 = 0.7491; $p < 0.01$

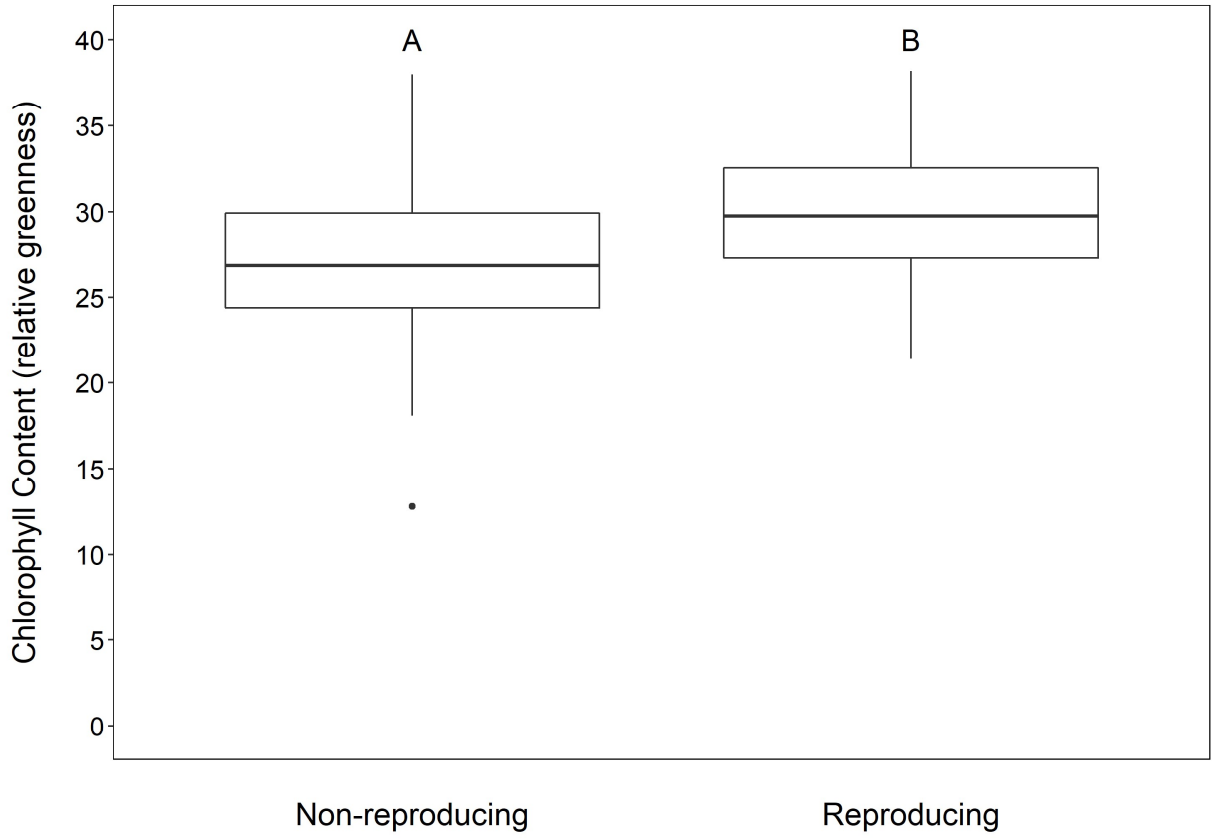

Figure 13 Chlorophyll content and reproductive capability; A and B denote significantly different median values ($p < 0.01$)

Ginseng habitat parameters

Herbaceous companion species were found with 191 of the documented ginseng plants, and a total of 538 occurrences of herbaceous companions were documented. The most common companions were *Galium* sp. (N = 165) and *Actaea racemosa* (N = 156) (Table 4). Other frequent companion species were *Botrypus virginianus*, *Caulophyllum thalictroides*, and *Osmorhiza claytonii* (Table 4). *Liriodendron tulipifera*, a tree species associated with high calcium uptake, was present at around 67% of locations where ginseng was found (Table 4).

Table 4 Companion species presence, abundance, and relative frequency; frequency is calculated as the proportion of companion occurrence (n) to ginseng occurrence (n=192); ¥Abundance of Galium sp. was not quantified; *Liriodendron tulipifera is included although it is not considered a companion species and abundance was not quantified

Companion	n presence	Median abundance rank	Relative frequency
¥*Gallium* sp.	165	NA	0.86
A. racemosa	156	3	0.81
**L. tulipifera*	128	NA	0.67
B. virginianus	87	1	0.45
C. thalictroides	85	2	0.44
O. claytonii	70	2	0.36
A. pedatum	66	1	0.34
V. canadensis	43	1	0.22
A. macrophylla	18	1	0.09
S. canadensis	13	1	0.07

A total of 22 soil samples representing different populations or clusters of plants across the Cumberland Plateau in Virginia were analyzed to quantify pH, calcium, magnesium, manganese, phosphorus, potassium, and zinc (Table 5). There was a weak, negative correlation between soil calcium and chlorophyll content ($p = 0.0144$, $R^2 = 0.0537$; Fig 14). That correlation is the exception, as all other soil ion concentrations as well as pH were not significantly correlated with any plant size or health variables ($p > 0.1$, $R^2 < 0.1$). The pH of soil collected near wild ginseng populations ranged from 4.55–6.34 with an average value of 5.38. Calcium levels in soil at locations of ginseng presence varied widely, ranging from 217–6540 parts per million (ppm), with an average value of 1564.3 ppm (Table 5). The median magnesium, potassium, manganese, and zinc were 178.9, 144.45, 66.45, 4.07, and 3.17 ppm, respectively (Table 5).

Table 5 Ion concentration from soil samples collected near wild ginseng populations; n = 22

Ion	Min (ppm)	Median (ppm)	Max (ppm)
Calcium	217	1005.50	6540
Magnesium	35.6	178.90	600.2
Potassium	79.3	144.45	343.4
Manganese	38.2	66.45	152.4
Phosphorus	0.15	4.07	40.61
Zinc	0.64	3.17	12.95

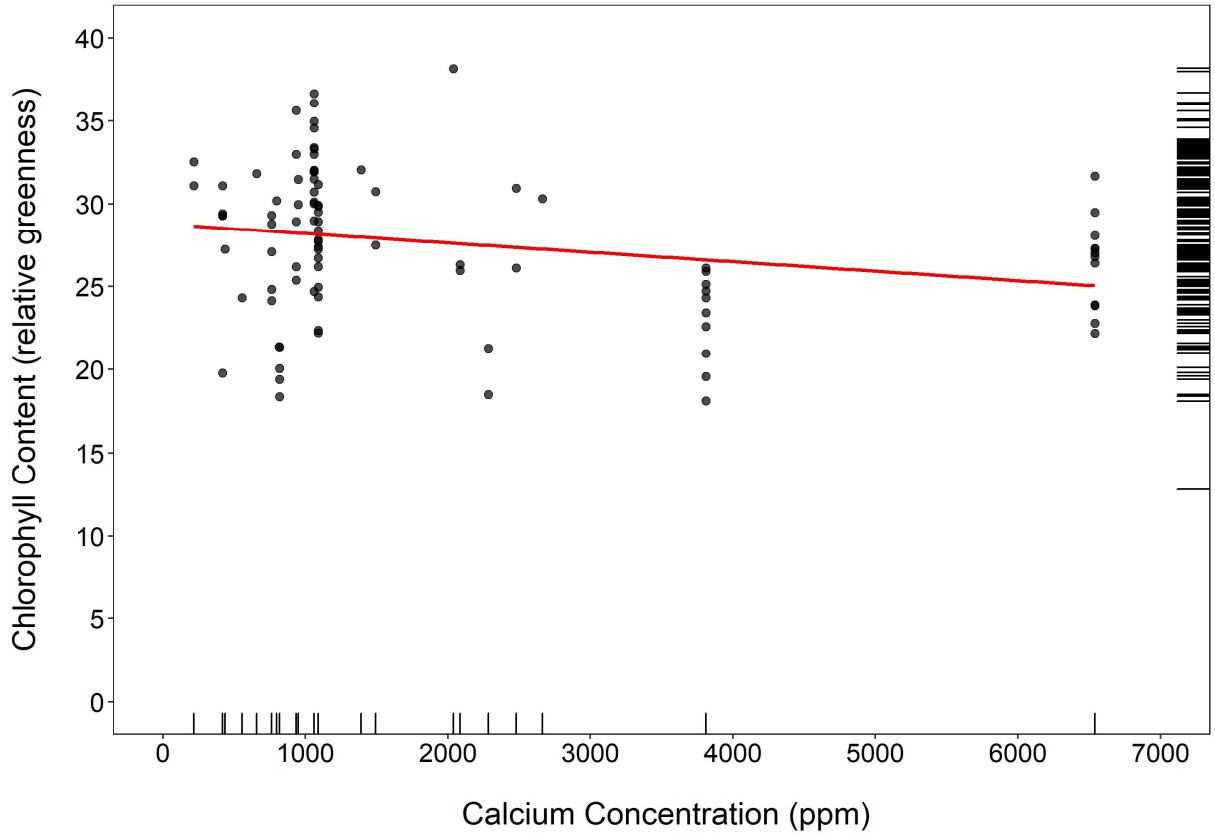

Figure 14 Chlorophyll content and soil ion concentration; red line displays linear regression, R2 = 0.0537; rugs demonstrate density of points along the y and axes (right and left rug, respectively)

DISCUSSION

This study increased the representation of a nearly unsampled region to identify another hotspot for American ginseng in Virginia. Through increased sampling and alteration of random forest parameters, the reliability of species distribution and habitat suitability models for American ginseng is expected to have been improved. Although the number of documented occurrence points in the region was increased by orders of magnitude, it is suspected that many more presence points could be documented elsewhere in the Cumberland Plateau. This study was restricted mainly to the USDA Forest Service, Clinch Ranger District, Breaks Interstate Park in the Cumberland Plateau, and a small number of private properties. Therefore, it is likely that most of the Cumberland Plateau is still under-surveyed, although significant progress has been made to alleviate sampling deficiency. With increased sampling in the Cumberland Plateau of Virginia, predictive outputs for species distribution and habitat suitability models in this province and Virginia overall are likely to improve.

Companion species were determined to be an incredibly 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 harvesting and reintroduction currently used by agencies and citizens alike (Kauffman 2006; Woods 2015; Davis and Persons 2014; Bonnabeaux 2016). Among those companion species, black cohosh (*Actaea racemosa* L.), blue cohosh (*Caulophyllum thalictroides* (L.) Michx.), and bloodroot (*Sanguinaria canadensis* L.) appeared to be some of the most reliable for indicating areas in which ginseng may be present. Although *S. canadensis* and dutchman's-pipe (*Artistolochia macrophyllum* Lam.) were not found as frequently as other companions, it was anecdotally noted that ginseng was almost always found on surveys if either of those companions were present. On the other hand, most other companion species could be

found in locations without ginseng. This does not necessarily indicate poor habitat, nor does it mean companion species are unreliable—but may point to areas in which ginseng was removed prior to the survey. Many ginseng harvesters recognize companion species (including those used here among others), and companion plants can be used to enhance growth of wild simulated ginseng as well (Pritts K.D. 2010; Tringovska et al. 2015). Companion species may alter 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.

Some species were noted to be frequent in locations where ginseng was found in this study 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 (T*rillium erectum* L.), and white trillium (*Trillium grandiflorum* (Michx.) Salisb.). Therefore, these plants warrant consideration for use as companions when surveying for American ginseng for the purpose of modeling species distribution and habitat suitability.

Random forest was shown to be an accurate means of creating species distribution models for American ginseng in Virginia. Among the statewide models, the most accurate—according to evaluation metrics (low OOB error and high AUC)—were the simpler models (SM1, SM7), using background points generated evenly across the state, all 78 environmental variables, and no spatial partitioning (Table 2). These two simpler models—both with OOB error below 3.5% and an AUC of over 0.99—are overfitting to some degree, and they may not be practical to use as a habitat suitability predictor, as an AUC approaching one is extremely high and likely does not represent the true accuracy of predicting occurrence (Table 2). However, SM1 and SM7 models may have

the greatest accuracy and reliability when predicting where wild populations currently occur. It is a common issue for species distribution models to display overfitting, only predicting presence in smaller patches closely resembling locations where training data were located—which will happen frequently in models where sample size may be small (Aguirre-Gutierrez et al. 2013; Stockwell and Peterson 2002). Although there was a high sample size in this study, the sample size in the Cumberland Plateau was not initially adequate for modeling species distribution or habitat suitability in the region, as evidenced by having only 16 occurrence points documented.

The overfitting seen in the initial model (SM1) is likely due to the ratio of background points (N = 5664) to presence points (N = 2153) combined with the high number of correlated variables (Table 2). The model may become very adept at correctly predicting presence points but may not provide a realistic assessment of habitat suitability. As Merow and colleagues (2014) point out in their study on the complexity of species distribution models, a researcher's goals may determine how complex or simple their model should be. These overfitting models are not necessarily incorrect or impractical but may have specialized uses. Models with this type of performance may be most useful for conservationists, organizations, and others looking to identify locations in which wild populations may be most likely to occur and would therefore allow them to focus search efforts and restoration projects.

Determining the modeling methods that most accurately predict habitat suitability rather than presence may be more complicated for ecologists than anticipated. It is paramount for investigators to not only select the proper algorithm for modeling species distribution, but an accurate evaluation metric as well (Aguirre-Gutierrez et al. 2013). Instinctively, researchers strive to create the most accurate model according to some evaluation metric—AUC and OOB error, in this case (Merow et al. 2014; Stockwell and Peterson 2002; Evans and Cushman 2009). However,

models to predict habitat suitability may perform most realistically when there is some amount of generalization—and thus higher error—than models that are deemed most accurate. Among statewide models, there was a noticeable reduction in overfitting when applying specific methods such as reducing the number of correlated variables, spatial partitioning, and limiting the generation of background points (Table 2). When this method was followed, AUC decreased to as low as 0.976 and OOB error increased to around 8.08% in SM9 (Table 2, Fig 4). The predictive models became visually different—although all models were still significantly correlated with one another. Typically, spatial partitioning caused the predictions to be much less conservative, and every province had locations where the probability of presence was predicted to be over 0.83, per SM9. These models are most likely generalizing more than others and may be of use as habitat suitability models rather than species distribution models. These may accurately predict locations where ginseng may grow well but might not contain wild populations. The models with the lowest overfitting (and highest error) may best serve those with a goal of establishing new populations—either for reintroduction or economic benefit.

Machine learning methods, including random forest, have been shown to reduce overfitting or prediction biases compared to more conventional methods (Breiman 1996b, Breiman 1999, Dietterich 2000, Pal 2005; Stockwell and Peterson 2002). In this study, different parameters within random forest were tried to produce varying levels of overfitting, to reduce overall error, and to alleviate any bias toward certain areas with the highest sampling effort. Limited background point generation and reducing the number of correlated variables resulted in models that were less likely to only recognize the region where the most sampling occurred (Fig 4f, Fig 4i, Table 2). It has been shown that removing correlated variables from species distribution models tend to increase accuracy and as such, it is becoming a standard step in data preparation for species distribution

modeling (Crawford and Hoagland 2010; Genuer et al. 2010; Narouei-Khandan et al. 2017). With these methods, similar trends were still observed in the predictive models (i.e., the Blue Ridge Mountains are still predicted to be appropriate habitat for ginseng), but this does not mean the model is biased toward a single area. A biased habitat suitability model does not always signify that an area shouldn't be predicted as suitable habitat. Rather, additional regions with appropriate habitat could get lost in the modeling process and not be accurately represented. When using the most basic methods (represented by SM7), there was a reduction in overfitting once additional points were added to statewide models, and from those models the Cumberland Plateau was predicted to contain more suitable habitat than before (Fig 4g). The Shenandoah region was indicated to be a ginseng hotspot within Virginia with many documented wild populations (Virginia Botanical Associates 2020; van Manen et al. 2005). However, the Cumberland Plateau was not predicted to be a ginseng hotspot until additional sampling was conducted there in this study. Sampling effort should be divided across all regions in such a diverse state to avoid overlooking potential populations, particularly hotspots. There are many modeling methods, such as the ones tested in this study, may be used to correct for unequal sampling. However, increased sampling effort is a better method of alleviating sampling bias (Foxcroft et al. 2011; Fithian et al. 2015; Syfert et al. 2013; Phillips et al. 2009).

If it is unclear whether a model is biased toward a specific region (such as the Shenandoah region), a model could be trained while excluding all data representing that region, followed by a prediction including that region. This was conducted in a tangential modelling run, where a model was trained exclusively on points from the Cumberland Plateau, and a statewide prediction was made. There were still similarities between the Shenandoah region (i.e., the Blue Ridge Mountains) and Cumberland Plateau, as the model predicted highest probabilities of ginseng presence in both

regions, but not elsewhere (Fig A1). This indicates that while the overall model may be biased toward Shenandoah and sampling efforts should be increased in other areas, it is not necessarily incorrect in its prediction.

In an experimental step to reduce bias, spatial partitioning was conducted on points only within specific areas. Specifically, SM10 was constructed using selected variables and limited background points, but spatial partitioning was performed on each province except the Cumberland Plateau (Fig 4j). When compared to SM9—the most similar model—the AUC increased and OOB error decreased slightly, indicating that the model was more specific in its predictions, and the Cumberland Plateau contained higher predicted probability values (Fig 4i, Fig 4j, Table 2). This approach effectively introduces a bias toward other regions and should be used with caution and only in instances where it is clear that there are biased or uneven sampling efforts such that species distribution or habitat suitability models would be severely impacted.

To further investigate the model predictions and representation of the Cumberland Plateau, four provincial models representing that region were made and evaluated. Provincial models constructed prior to adding newly documented occurrences would not be useful, as there was a very high OOB error, and generally incredibly high predicted probability (Table 3, Fig 6). Once additional points were added to provincial models, their predictive output aligned more with that from statewide models (although not significantly correlated). This study concludes that limiting the spatial extent to areas with limited sampling (relative to other regions) is not an effective means of alleviating sampling deficiency or misrepresentation in predictive outputs for species distribution and habitat suitability. It is more beneficial to have an increased data set, even if those data may be skewed towards other regions.

Deciduous forest cover was always the top variable in statewide models, no matter the model parameters. This is likely the most intelligible influence on ginseng presence but is not practical to researchers. Although it must be included in the modelling process, most growers or researchers would naturally begin their search for planting sites in deciduous forest. The most useful information is likely to be more specific variables such as the amount of cover, mean diurnal range, and precipitation levels in which ginseng will perform best (Fig 10, Fig 11).

Initial interpretations of some variables present in this project may not be intuitive. For example, roughness (10-cell circle) may be interpreted as more rugged terrain (understood as the term is used to describe areas like the Cumberland Plateau in Virginia). In reality, however, this variable is highlighting cliffs and other very steep areas. While roughness as one would interpret it may still be important for ginseng—rugged terrain contains many protected hollows that could harbor healthy ginseng populations—the data for these variables may be different than researchers anticipate.

When interpreting these models in the future, researchers must strive to understand and explain relationships between variables and the predictive output. Focusing on variables that are the most useful or unique may serve best for researchers or others in the field. This study demonstrates that areas within deciduous forest cover with increased canopy closure (at least 60%) are important—which would not be surprising to seasoned ginseng growers, harvesters, or researchers (Fig 10, Fig 11). However, when those conditions are present in areas within a few kilometers of water (rivers, streams) with a mean diurnal range of 10–12 °C, ginseng may be easier to locate (Fig 10, Fig 11). For most citizens or landowners, it is most important that they recognize if they are living in areas with a generally higher probability of ginseng presence. Generally, stakeholders in the Cumberland Plateau or Blue Ridge Mountains may want to focus their search

and restoration efforts in forested locations with ample moisture (total annual precipitation between 100–130 cm or 20–25 cm in winter) and relatively steep or rugged terrain, while searching for areas with appropriate companion species. It should be noted, however, that precipitation may be more difficult to use in the field—as there was a higher density of both presence and absence points in those ranges, indicating that those values may simply be typical for Virginia's climate (Fig 10, Fig 11). A combination of variables and their importance (using knowledge from models and field surveys) may best serve to aid location and conservation of American ginseng.

Once habitat and species distribution models have been utilized along with identification of important variables to identify an area for planting, one may wish to monitor the health and success of those plants. As reproductive plants tend to have increased chlorophyll levels in their leaves, it is reasonable to conclude that both reproductive capability and chlorophyll content may serve as accurate means of assessing plant health (Fig 12, Fig 13). This is further supported by plant size and prong data, where larger plants tended to be reproductive and have higher levels of chlorophyll. Therefore, plant health could be quantified in the future mainly by reproductive capability and an individual's size—without the need to measure chlorophyll. There was a significant correlation between leaf area and height ($p < 0.01$, $R^2 = 0.7491$), which may suggest that researchers need not document both when measuring ginseng growth. Nevertheless, this study recommends that researchers document both plant height and leaf area, as younger plants (those with only one leaf) do not exhibit as strong of a correlation between those variables, though it is statistically significant ($p < 0.01$, $R^2 = 0.35$).

Soil calcium concentrations ranged from 217–6540 parts per million (ppm; Table 5). Previous research indicates that ginseng may grow best at calcium ranges above 4000 ppm (Thyroff and Griscom 2019). This study did not find significant correlations between calcium

concentrations and plant health measured by leaf area, height, or reproductive capability ($p > 0.1$, $R^2 < 0.1$). There was, however, a slight correlation between calcium concentrations and chlorophyll content—although these data do not support a true correlation at the biological level ($p = 0.0144$, $R^2 = 0.0537$; Fig 14). It proved difficult for this study to confidently classify locations as absence points (poor habitat) due to the abundance of ginseng and its companions in this region, coupled with a history of heavy poaching. If ginseng was not present in a given location, there was a strong possibility that it could be due to poaching rather than habitat conditions. As a result, there are no soil samples classified as being collected from absence points. This study recommends that future work be focused on identifying soil conditions and their relationship to the growth and fecundity of wild American ginseng.

CONCLUSION

The Cumberland Plateau and Blue Ridge Mountains are hotspots for American ginseng in Virginia. Of those hotspots, the former became apparent in statewide predictive models only after the addition of nearly 200 additional occurrence points in that region. Although it is likely that the Cumberland Plateau's representation in these models could still be improved with continued sampling effort, this study has greatly increased the accuracy of species distribution and habitat suitability models with respect to the region. When constructing models such as these, researchers and agencies must consider the prediction they are seeking to construct prior to selecting model parameters within random forest. Through varying levels of complexity in data preparation and model parameters, random forest was shown to be a practical means of constructing realistic models for different purposes—which will inform agencies and allow them to more easily protect this species, while also aiding those interested in growing or reintroducing American ginseng throughout Virginia.

APPENDIX A

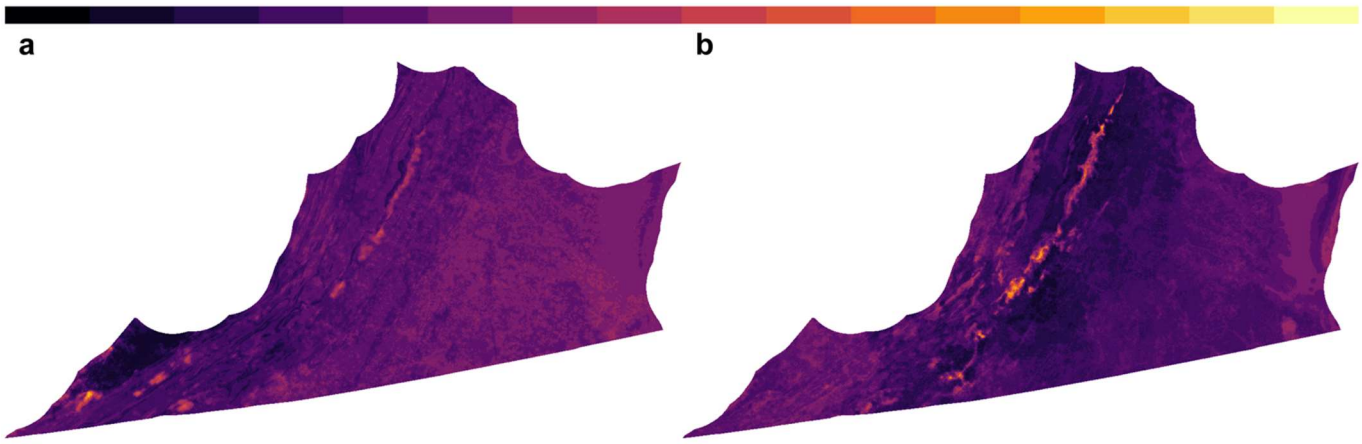

Figure A1 Provincial models' statewide predictions; **a** statewide model trained on data representing the Cumberland Plateau **b** statewide model trained on data representing the Blue Ridge Mountains; probability of ginseng presence ranging from 0–1 is shown above (lighter to darker)

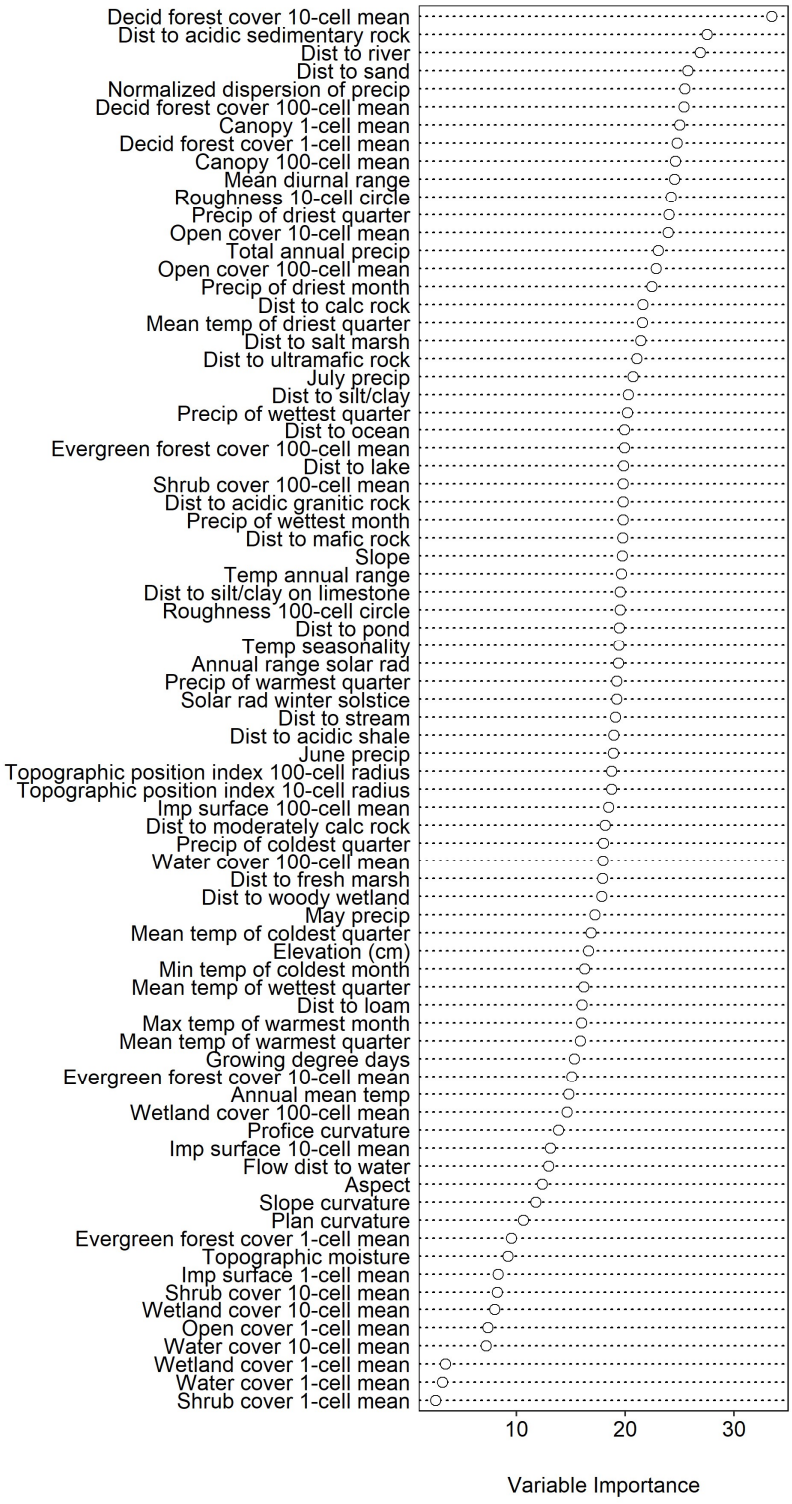

Figure A2 Variable importance from SM1; importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, imp=impervious, precip=precipitation, temp=temperature

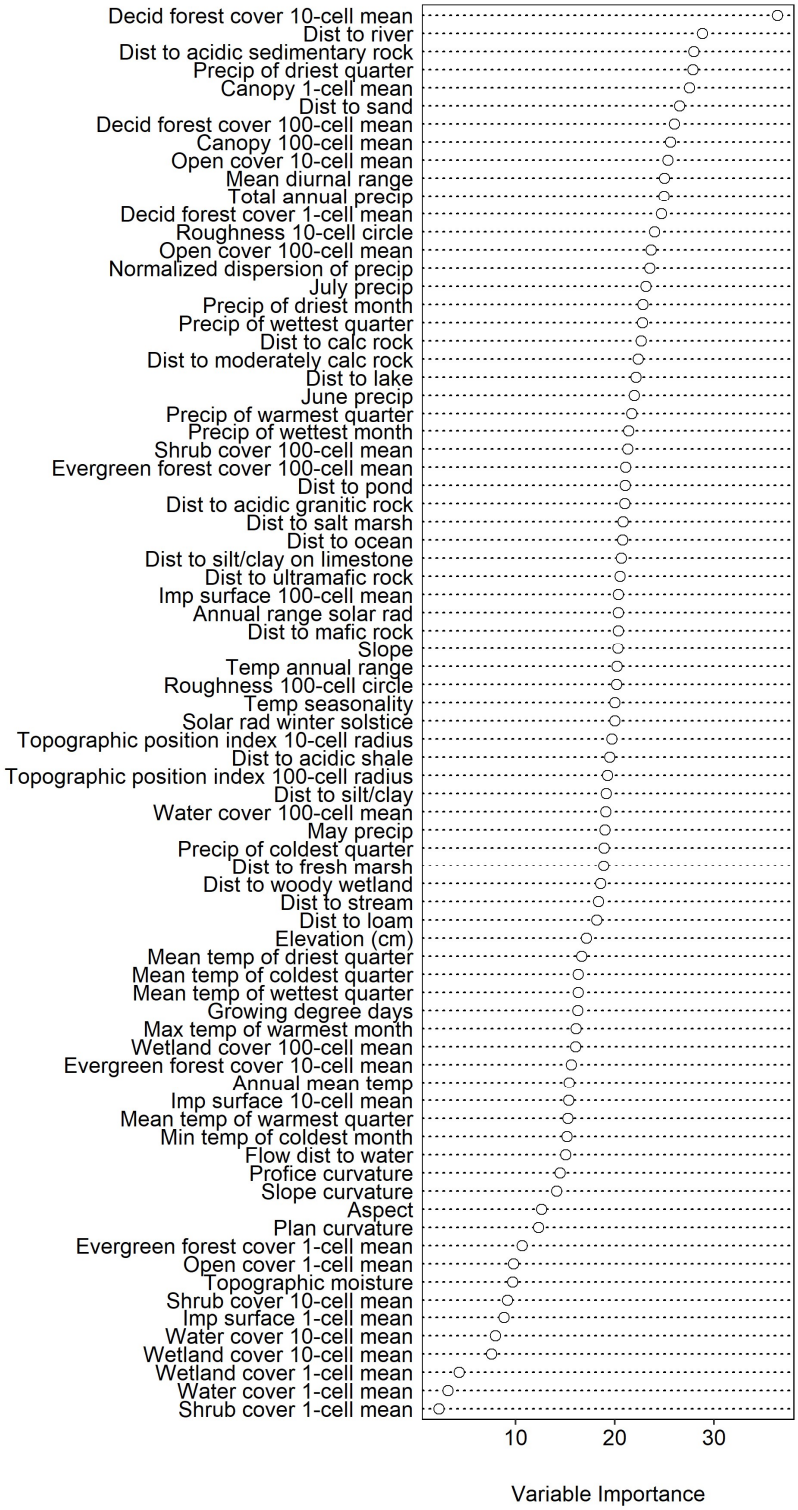

Figure A3 Variable importance from SM7; importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, imp=impervious, precip=precipitation, temp=temperature

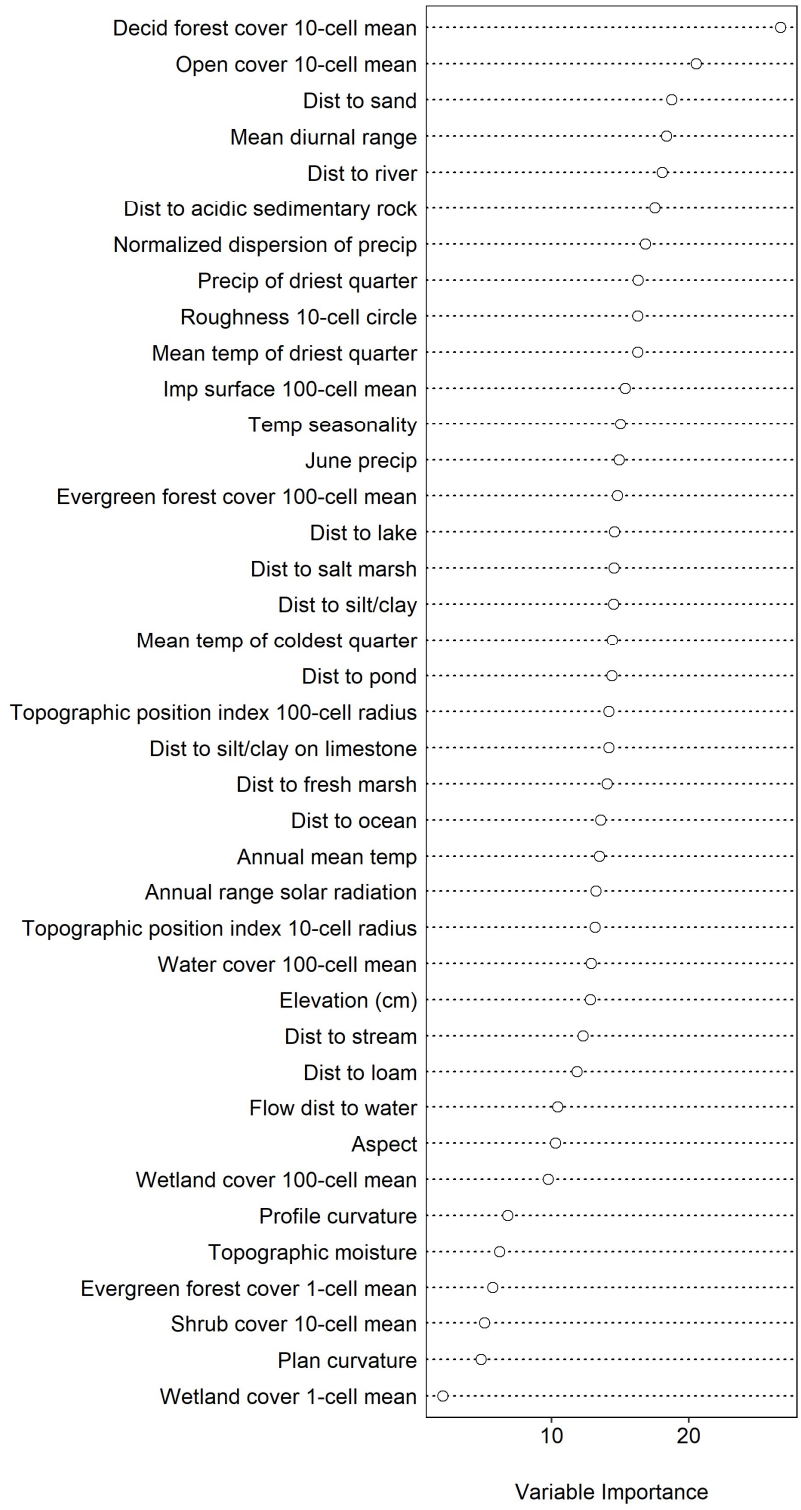

Figure A4 Variable importance from SM9; importance is quantified as decrease in accuracy (%IncMSE) for each variable; decid=deciduous, dist=distance, imp=impervious, precip=precipitation, temp=temperature

LITERATURE CITED

Aguirre-Guitierrez, J., Carvalheiro, L. G., Polce, C., van Loon, E. E., Raes, N., Reemer, M., & Biesmeijer, J. C. (2013). Fit-for-purpose: Species Distribution Model Performance Depends on Evaluation Criteria — Dutch Hoverflies as a Case Study. PLoS ONE, 8(5).

Albrecht, M. A., & McCarthy, B. C. (2009). Seedling establishment shapes the distribution of shade-adapted forest herbs across a topographical moisture gradient. Journal of Ecology, 97(5), 1037–1049. https://doi.org/10.1111/j.1365-2745.2009.01527.x

Anderson, M. G., Barnett, A., Clark, M., Prince, J., Ollivero Sheldon, A., & Vickery, B. (2016). Resilient and Connected Landscapes for Terrestrial Conservation. Boston, MA: The Nature Conservancy, Eastern Conservation Science, Eastern Regional Office.

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

Bivand, R., Keitt, T., Rowlingson, B. (2019) rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-2. https://CRAN.R-project.org/package=rgdal

Bonnabeaux, M. (2016). Ginseng: the black market herb of the Appalachian. http://www.technicianonline.com/arts_entertainment/article_771f9b02-eb2b-11e5-936c-132b98d780d8.html

Braun, E.L., (1950). Deciduous Forests of Eastern North America. The Blackburn Press. ISBN-10: 1-930665-30X and ISBN-13: 978-1-930665-30-9.

Breiman, L. (1996a). Bagging predictors. Machine Learning, 24(2), 123–140.

Breiman, L. (1996b). Out-of-Bag Estimation.

Breiman, L. (2001) Random forests. Mach Learn 45:5–32.

Breiman, L. (1999). Random Forests. Journal of the Electrochemical Society, 129, 2865.

Browning, W., Peters, J. 2018. Unpublished data.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research, 33(2), 261–304. https://doi.org/10.1177/0049124104268644

Carpenter, S. G., & Cottom, G. (1981). Growth and reproduction of American ginseng (Panax quinquefolius) in Wisconsin, U.S.A. Canadian Journal of Botany, 60, 2692–2696.

Case, M. A., Flinn, K. M., Jancaitis, J., Alley, A., & Paxton, A. (2007). Declining abundance of American ginseng (Panax quinquefolius L.) documented by herbarium specimens. Biological Conservation, 134(1), 22–30. https://doi.org/10.1016/j.biocon.2006.07.018

Carlson, A. W. (1986). Ginseng: America's Botanical Drug Connection to the Orient. Economic Botany, 40(2), 233–249.

Chamberlain, J. L., Mitchell, D., Brigham, T., Hobby, T., Zabek, L., Davis, J., & Garrett, H. E. G. (2009). Forest Farming Practices, 1–38. http://doi.org/10.2134/2009.northamericanagroforestry.2ed.c9

Chamberlain, J., Bush, R., & Hammett, A. L. (1998). Non-timber forest products. Forest Products Journal, 48(10), 10–19.

Dietterich, T., & Michalski, R. S. (1983). A comparative review of selected methods of learning from examples. Palo Alto, California: Tioga publishing company.

Duda, R. B., Zhong, Y., Navas, V., Li, M. Z. C., Toy, B. R., & Alavarez, J. G. (1999). American ginseng and breast cancer therapeutic agents synergistically inhibit MCF-7 breast cancer cell growth. Journal of Surgical Oncology, 72(4), 230–239. https://doi.org/10.1002/(SICI)1096-9098(199912)72:4<230::AID-JSO9>3.0.CO;2-2

Elza, M. C., Slover, C., & McGraw, J. B. (2016). Analysis of wood thrush (Hylocichla mustelina) movement patterns to explain the spatial structure of American ginseng (Panax quinquefolius) populations. Ecological Research, 31(2), 195–201. https://doi.org/10.1007/s11284-015-1327-6

Erikson, O. (1995). Seedling recruitment in deciduous forest herbs: the effects of litter, soil chemistry and seed bank. Flora, 190(1), 65–70.

Farrington, S. J., Muzika, R. M., Drees, D., & Knight, T. M. (2009). Interactive effects of harvest and deer herbivory on the population dynamics of american ginseng. Conservation Biology, 23(3), 719–728. https://doi.org/10.1111/j.1523-1739.2008.01136.x

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. Methods in Ecology and Evolution, 6(4), 424–438. https://doi.org/10.1111/2041-210X.12242

Furedi, M. A. N. N., & Mcgraw, J. B. (2004). White-Tailed Deer: Dispersers or Predators of American Ginseng Seeds? The American Midland Naturalist2, 152(2), 268–276.

Fuzzati, N. (2004). Analysis methods of ginsenosides. Journal of Chromatography B, 812(1–2), 119–133.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using Random Forests. Pattern Recognition Letters, Elsevier, 31(14), 2225–2236. https://doi.org/10.1017/CBO9781107415324.004

Gibson, D. J., Connolly, J., Hartnett, D. C., & Weidenhamer, J. D. (1999). Designs for greenhouse studies of interactions between plants. Journal of Ecology, 87, 1–16. http://doi.org/10.1007/s10272-005-0161-4

Gilliam, F. S. (2007). The Ecological Significance of the Herbaceous Layer in Temperate Forest Ecosystems. BioScience, 57(10), 845–858.

Brandon M. Greenwell (2017). pdp: An R Package for Constructing Partial Dependence Plots. The R Journal, 9(1), 421--436. https://journal.r-project.org/archive/2017/RJ-2017-016/index.html.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. Ecology Letters, 8(9), 993–1009. http://doi.org/10.1111/j.1461-0248.2005.00792.x

Hijmans, R. J. (2019) raster: Geographic Data Analysis and Modeling. R package version 2.8-19. https://CRAN.R-project.org/package=raster

Ho, K. T. (1998). The Random Subspace Method for Constructing Decision Forests. Ieee Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832–844. https://doi.org/doi:10.1109/34.709601

Hui, C., Foxcroft, L. C., Richardson, D. M., & Macfadyen, S. (2011). Defining optimal sampling effort for large-scale monitoring of invasive alien plants: a Bayesian method for estimating abundance and distribution. Applied Ecology, 48, 768–776. https://doi.org/10.1111/j.1365-2664.2011.01974.x

Hu, S. Y. (1976). The Genus Panax (Ginseng) in Chinese Medicine. Economic Botany, 30(1), 11–28.

Kauffman, G. (2006) Conservation assessment for American ginseng (Panax quinquefolius) L. USDA Forest Service, Eastern Region. Unpublished report, National Forests in North Carolina, Asheville, North Carolina. 75 pp.

Kim, C. S., JO, Y. J., Park, S. H., Kim, H. J., Han, J. Y., Hong, J. T., … Oh, K. W. (2010). Anti-stress effects of ginsenoside Rg3-standardized ginseng extract in restraint stressed animals. Biomolecules and Therapeutics, 18(2), 219–225. https://doi.org/10.4062/biomolther.2010.18.2.219

Kitts, D. D., Wijewickreme, A. N., & Hu, C. (2000). Antioxidant properties of a North American ginseng extract. Molecular and Cellular Biochemistry, 203(1–2), 1–10. https://doi.org/10.1023/A:1007078414639

Kohavi, R., & Provost, F. (1998). Glossary of Terms. Journal of Machine Learning1.

Konsler, T. R., Zito, S. W., Shelton, J. E., & Staba, E. J. (1990). Lime and Phosphorus Effects on American Ginseng: II. Root and Leaf Ginsenoside Content and Their Relationship. J. Amer. Soc. Hort. Sci, 115(4), 575–580.

Lewis, W. H., & Zenger, V. E. (1982). Population Dynamics of the American Ginseng Panax quinquefolium (Araliaceae). American Journal of Botany, 69(9), 1483–1490.

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

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

Li, T. S. C., & Mazza, G. (1999). Correlations between leaf and soil mineral concentrations and ginsenoside contents in American ginseng. HortScience, 34(1), 85–87. https://doi.org/10.21273/hortsci.34.1.85

Lee, J., & Mudge, K. W. (2013). Gypsum effects on plant growth, nutrients, ginsenosides, and their relationship in American ginseng. Horticulture Environment and Biotechnology, 54(3), 228–235. https://doi.org/10.1007/s13580-013-0029-7

Maggard, S. W. (1994). From Farm to Coal Camp to Back Office and McDonald's: Living in the Midst of Appalachia's Latest Transformation. Journal of Appalachian Studies, 6, 14–38.

Mazza, G., C., C. A., & Gao, L. (1996). Ginsenosides in Roots and Leaves of American Ginseng, 717–720.

McGraw, J. B., Lubbers, A. E., Van der Voort, M., Mooney, E. H., Furedi, M. A., Souther, S., Chandler, J. (2013). Ecology and conservation of ginseng (Panax quinquefolius) in a changing world. Annals of the New York Academy of Sciences, 1286(1), 62–91. https://doi.org/10.1111/nyas.12032

McGraw, J. B., Sanders, S. M., & Van der Voort, M. (2003). Distribution and Abundance of Hydrastis canadensis L. (Ranunculaceae) and Panax quinquefolius L . (Araliaceae) in the Central Appalachian Region Author. Torrey Botanical Society, 130(2), 62–69.

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

Narouei-Khandan, H. A., Harmon, C. L., Harmon, P., Olmstead, J., Zelenev, V. V., van der Werf, W., Worner, S. P., Senay, S. D., & van Bruggen, A. H. C. (2017). Potential global and regional geographic distribution of Phomopsis vaccinii on Vaccinium species project by two species distribution models. Eur J Plant Pathol, 148, 919–930.

NatureServe and its Natural Heritage member programs. (2013). NatureServe Rarity-Weighted Richness Model of Critically Imperiled and Imperiled Species in the United States.

Nirmal, S. A., Pal, S. C., Otimenyin, S. O., Aye, T., Elachouri, M., Kundu, S. K., … Manda, S. C. (2013). Contribution of herbal products in global market. The Pharma Review, (November December 2013), 95–104.

Odom, R. H., & McNab, W. H. (2000). Using Digital Terrain Modeling to Predict Ecological Types in the Balsam Mountains of Western North Carolina. United States Department of Agriculture: Forest Service.

Oelbermann, M., & Milburn, M. (1994). Worth the Trouble. Nature Canada, 23(3), 10–11.

Peri, P. L., Caballé, G., Hansen, N. E., Bahamonde, H. A., Lencinas, M. V., von Müeller, A. R., Martinez Pastur, G., et al. (2017). Temperate Agroforestry Systems.

Phillips, S. J., Dudik, M., Elith, J., Graham, C., Lehman, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications, 19(1), 181–197.

Priya, M. S., Wesley, E. G., & Subramaniam, P. (2017). Conservation of Medicinal Plants in Bangladesh *. In Propagation of ethno verterinary herbal knowledge (pp. 28–35).

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

Qi, L.-W., Wang, C.-Z., & Yuan, C.-S. (2011). Ginsenosides from American ginseng: Chemical and pharmacological diversity. Phytochemistry, 72(8), 689–699. https://doi.org/10.1002/bmb.20244.DNA

R Development Core Team (2007) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-3-900051-07-0. http://www.R-project.org

Snow, M., & Snow, R. (2009). The Reestablishment of American Ginseng (Panax quinquefolius). WSEAS TRANSACTIONS on BIOLOGY and BIOMEDICINE, 6(2), 38–47.

Souther, S., & McGraw, J. B. (2014). Synergistic effects of climate change and harvest on extinction risk of American ginseng. Ecological Applications, 24(6), 1463–1477.

Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. PLoS ONE, 8(2). https://doi.org/10.1371/journal.pone.0055158

Taylor, B., Hufford, M., & Bilbrey, K. (2016). A GREEN NEW DEAL FOR APPALACHIA: ECONOMIC TRANSITION, COAL RECLAMATION COSTS, BOTTOM-UP POLICYMAKING (PART 1). Journal of Appalachian Studies, 23(1), 8–29.

Thyroff, E. C., & Griscom, H. P. (2019). Experimental Study of Soil and Aspect on American Ginseng in an Appalachian Cove Ecosystem. *Natural Areas Journal*, *39*(3), 378. https://doi.org/10.3375/043.039.0310

Tringovska, I., Vinelina, Y., Markova, D., & Mihov, M. (2015). Effect of companion plants on tomato greenhouse production, 186, 31–37.

USDA. (2017). Agroforestry: Enhancing Resiliency in U.S. Agricultural Landscapes Under Changing Conditions. Gen. tech. Report WO-96.

USDA, NRCS. (2019). The PLANTS Database (http://plants.usda.gov, 21 February 2019). National Plant Data Team, Greensboro, NC 27401-4901 USA.

Van der Voort, M. E., & McGraw, J. B. (2006). Effects of harvester behavior on population growth rate affects sustainability of ginseng trade. Biological Conservation, 130(4), 505–516. https://doi.org/10.1016/j.biocon.2006.01.010

Van Manen, F. T., Young, J. A., Thatcher, C. A., Cass, W. B., & Ulrey, C. (2005). Habitat models to assist plant protection efforts in Shenandoah National Park, Virginia, USA. Natural Areas Journal, 25(4), 339–350.

Virginia Botanical Associates. (2020). Digital Atlas of the Virginia Flora (http://www.vaplantatlas.org). c/o Virginia Botanical Associates, Blacksburg

Virginia Natural Heritage Program. (2017). Species distribution model for American Ginseng (Panax quinquefolius). Created on 25 Sep 2017. Virginia Department of Conservation and Recreation - Division of Natural Heritage, Richmond, VA.

Wang, A., Wang, C.-Z., Wu, J.-A., Osinski, J., & Yuan, C.-S. (2005). Determination of Major Ginsenosides in Panax quinquefolius (American ginseng) using High-performance Liquid Chromatography. Phytochemical Analysis, 16(4), 272–277.

Wang, J. R., Leung, C. Y., Ho, H. M., Chai, S., Yau, L. F., Zhao, Z. Z., & Jiang, Z. H. (2010). Quantitative comparison of ginsenosides and polyacetylenes in wild and cultivated American ginseng. Chemistry and Biodiversity, 7(4), 975–983. https://doi.org/10.1002/cbdv.200900264

Welch, N. T., Belmont, J. M., & Randolph, J. C. (2007). Sumer ground layer biomass and nutrient contribution to above-ground litter in an Indiana temperate deciduous forest. The American Midland Naturalist, 157(1), 11–26.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wilson, I. T., & Tuberville, T. (2003). Virginia's Precious Heritage. Richmond, Virginia.

Woods, M. (2015). American Ginseng & Companions. Wild Ozark, LLC.