Maximum Entropy Species Distribution Modeling for the spring ephemeral herb Bloodroot (Sanguinaria canadensis) in Eastern North America

Velan Manivannan
vsm11@uakron.edu

Follow this and additional works at: https://ideaexchange.uakron.edu/honors_research_projects

Part of the Environmental Health Commons, Other Ecology and Evolutionary Biology Commons, and the Population Biology Commons

Please take a moment to share how this work helps you through this survey. Your feedback will be important as we plan further development of our repository.

Recommended Citation

This Dissertation/Thesis is brought to you for free and open access by The Dr. Gary B. and Pamela S. Williams Honors College at IdeaExchange@UAkron, the institutional repository of The University of Akron in Akron, Ohio, USA. It has been accepted for inclusion in Williams Honors College, Honors Research Projects by an authorized administrator of IdeaExchange@UAkron. For more information, please contact mjon@uakron.edu, uapress@uakron.edu.
Maximum Entropy Species Distribution Modeling for the spring ephemeral herb Bloodroot

(Sanguinaria canadensis) in Eastern North America

Velan S. Manivannan

Williams Honors College
Abstract

The spring ephemeral plant Bloodroot (*Sanguinaria canadensis*) has a widespread native range in North America, spanning much of the eastern United States and Canada. While its current NatureServe conservation status is designated as ‘secure’ (NatureServe, 2023), its status as a spring ephemeral places it at a heightened risk for climate change-induced phenological mismatch with advancing forest canopy closure. Additionally, under continued anthropogenic climate change, Bloodroot may also experience range shifts or contractions as the edges of its present range warm past physiological thresholds. To determine the potential for range shifts and contractions under future warming, I generated a species distribution model (SDM) for Bloodroot. The model was built under contemporary (1960-1990) climate variables and then projected to a future climate scenario (2050s) using maximum entropy modeling (Maxent). 17,203 georeferenced and spatially rarefied occurrences with an 80-20% split were used for training and testing the model, respectively. I used the 19 Bioclim variables from WorldClim for both contemporary and future scenarios. Comparisons of thresholded range maps from the contemporary and future predictions demonstrated a northward shift in Bloodroot’s distribution, with a small overall decrease in the predicted suitable area, primarily at the trailing southern edge. These results will inform management and conservation for this important native species under climate change, with ramifications for other aspects of Bloodroot life history, including pollination and seed dispersal.
Introduction

Bloodroot (*Sanguinaria canadensis*) is an eastern North American native perennial herbaceous understory forest plant. It is recognized for the blood-red sap characteristic of its rhizome, and it has traditionally been used by Indigenous peoples for a variety of medicinal uses (Predny, 2005). Bloodroot is a spring ephemeral plant, meaning it emerges early in the growing season and completes most of its above-ground life cycle prior to full canopy closure to maximize access to sunlight. Like many other spring ephemerals, it relies heavily on mutualistic interactions with insects for pollination (e.g., bees, flies) and seed dispersal (e.g., ants). Bloodroot emerges with a single leaf and flower arising on separate stems. At first, the leaf completely enwraps the flower bud. As the leaf unravels, it can reach a height of 12-24 in. The leaves are large and round with deep clefts, and the single flower is white (Fig. 1; Flora of North America, 2019). Bloodroot grows in clumps with flowers and leaves produced from shallow-growing, branching rhizomes. The rhizome is about a half-inch thick and up to 4 in long, and grows slowly, eventually branching to form large colonies (Mahr, 2024).

While it is not considered to be at risk due to any one particular global change driver, as evidenced by its ‘secure’ status (NatureServe, 2023), concerns about its conservation status have nevertheless drawn the attention of researchers and conservationists in recent years. Since it is used for ancient medicinal and dye-producing uses, bloodroot is especially vulnerable to overharvesting (Predny, 2005), but also may be at risk due to habitat loss, habitat degradation, and climate change, among other threats. Maximum Entropy Species Distribution Modeling (Maxent) is a useful approach for forecasting how abiotic environmental changes may affect the distribution of species such as Bloodroot. This modeling approach relies solely on presence data, making it ideal for species with limited or inaccurate absence data. Maxent has been useful in
ecological modeling, delivering solid forecasts under varied environmental situations, and has been confirmed to perform adequately through high Area Under the Curve (AUC) values in species distribution studies (Phillips et al., 2006).

In this project, I used Maximum entropy (Maxent) species distribution modeling (SDM) to generate predicted geographic distributions for the U.S. native spring ephemeral plant Bloodroot (*Sanguinaria canadensis*) under both current and future climate scenarios. As climate change continues to influence plant species distributions around the world, it is becoming increasingly important to understand how these changes will affect individual plant species, including Bloodroot. Understanding Bloodroot's distribution can help improve conservation initiatives, particularly in the face of climate change and habitat fragmentation. Maxent can assist in predicting where Bloodroot may thrive or decline by combining environmental variables such as temperature, precipitation, and elevation with occurrence data, directing conservation efforts to locations most likely to sustain its future survival. This particular modeling effort will contribute to our understanding of the range and ecology of the native *S. canadensis*, which may be at conservation risk due to the combined effects of a changing climate and other anthropogenic environmental changes.

**Methods**

This project involved generating a species distribution model (SDM) for *S. canadensis* under contemporary climate conditions using maximum entropy (Maxent) modeling, and then projecting the model under a future climate scenario to determine how its range might shift and/or contract under climate change. SDM is a method to project a species’ predicted geographic distribution across the landscape from its ecological niche. SDMs rely on statistical
correlations between the known occurrences of a species and the values of a variety of environmental data layers from the corresponding grid cells to project a potential distribution beyond the area of inference used to train the model. Generation of Maxent SDMs involves four major steps, described below.

Spatial extent

First, the spatial extent of the study must be determined. Bloodroot is fairly common throughout deciduous forests of eastern North America in both the U.S. and Canada. This area thus comprises the region of interest. After specifying the geographic bounds of the study (i.e., eastern North America north of Mexico and east of the Mississippi river), an appropriate spatial resolution was selected. Typically, this ranges from 1-10 km² grid cells. In this case, we used a spatial resolution of 5 minutes, or ~9 km². Further information about the spatial extent will be discussed under Environmental data below.

Occurrence data

Second, occurrence records in the form of GPS localities (latitude/longitude) were collected in a manner that reasonably represented the species’ entire known range from two publicly accessible databases. These records consisted of presence-only data (as opposed to presence/absence data), as that is the most commonly accessible form of occurrence data. Maxent has been shown to perform particularly well with presence-only data, especially compared to other SDM methods (Phillips et al., 2006; Elith et al., 2006; Rebelo & Jones, 2010; Elith et al., 2011). In this study, I located, downloaded, cleaned, and in some cases georeferenced occurrence records for Bloodroot from publicly accessible online databases including the Global Biodiversity Information Facility (GBIF) and SERNEC (SouthEast Regional Network of
Expertise and Collections). I collected 40,684 total records with latitude/longitude coordinates from GBIF and a total of 966 records from SERNEC.

The records collected from SERNEC were generally observed prior to the widespread availability of handheld GPS units, up to and around the year 1990. The location of each of these records was instead specified via written description. Since these records were important to include in the model due to their historic relevance and the probability that they were high-quality observations, given that they were often accompanied by verified herbaria records and/or specimens, I worked to georeference these occurrences whenever possible. I did so using the GeoLocate software program (GeoLocate, 2022), which allows a user to input country, state, and county data and then a descriptive locality string. The program identifies a number of potential sites that can then be assessed by a user for accuracy. Once I settled on a likely site, I set a centroid of uncertainty around the occurrence based on the program recommendation and my own assessment of the veracity of the record. I georeferenced 966 records from SERNEC, comprising 2% of the overall occurrence data, with the GBIF records contributing the other 98%. Any SERNEC records with centroids of uncertainty greater than 5,000 m were ultimately discarded from the dataset. Finally, all occurrence records were spatially rarefied such that only one record was included per grid cell in the study; this resulted in 41,650 total occurrence records for Bloodroot. From this final set of Bloodroot occurrences, I randomly split the records into training (80%) and testing (20%) data.

**Environmental data**

The third step in modeling the distribution of Bloodroot was to select the environmental data to be used in the model. In this case, I selected the 19 BioClim variables from WorldClim (Table 1; worldclim.org) to predict both the present and future distribution of *S. canadensis*. The
19 BioClim variables are the most-commonly used environmental data layers for prediction of species distributions across the globe (Dinnage, 2023). These variables include measures such as the annual mean temperature of the region; the mean diurnal range; isothermality; temperature seasonality; maximum and minimum temperatures of the warmest and coldest months, respectively; annual precipitation; and more (Table 1).

I downloaded the 19 BioClim variables at 5 min resolution (0.08333333 Arc degrees) for both the contemporary climate (1960-1990) and under a future climate scenario, RCP 6.0 (2050s). Each dataset was downloaded as a set of raster layers. As mentioned above, the spatial extent of the study was eastern North America. To crop, mask and visualize the environmental data layers to the area of the spatial extent, I used the statistical program RStudio (RStudio, 2020). I downloaded a shapefile of the world (naturalearthdata.com) and then derived and plotted only the United States and Canada in order to illustrate the spatial extent. To narrow down the area of interest to that which is likely to be accessible via the dispersal of the study species, a spatial extent buffer with an area of 500,000 m surrounding the training occurrences of each species was generated. All 19 climate variable layers were then cropped and masked to the shape of the spatial buffer; these buffered climate variables were ultimately included in the Maxent model.

**Maxent model**

The final step in generating the species distribution model for Bloodroot was the Maxent model itself. The Maxent algorithm is a modeling approach that uses machine learning and is based on the maximum entropy algorithm. Given environmental constraints, Maxent estimates the probability distribution for a species’ occurrence, as represented by the statistical combination of climate variables that are used in the model (Phillips et al. 2006). To implement
Maxent, the program was run using Java (Philips et al., 2006). I uploaded the training and testing occurrence data as separate comma separated values files, along with the 19 contemporary (e.g., 1960-1990) BioClim environmental data layers. From there, three datasets were created by the program. The first dataset contained every training occurrence and the extracted, corresponding cell values of each of the 19 climate variables. Second, the program generated a pseudo-absence dataset consisting of 10,000 background points. The program generated this dataset by assigning each row in the dataset a value of 0 if it corresponded with the background point data, or a value of 1 if it corresponded with the training data. Finally, the background point and training datasets were combined to create a data frame.

I ran two model iterations to generate the current predicted distribution for Bloodroot. The first iteration of the model used all 19 climate variables. After the first model run, I consulted the table listing the percent contribution of each variable to overall model gain to determine the variables that had the most significant contribution to the model; i.e., those that added up to comprise 95% of model contribution. In ecological modeling, testing the covariance of environmental variables can be complex and sometimes unnecessary, so when the main goal is to understand the relative importance of each variable, it is best to use the percent contribution table. This method focuses on the impact of individual variables and is particularly useful in large datasets where multicollinearity might obscure the effects of individual variables if analyzed through covariance tests (Miller et al. 2021).

Following the first iteration of the model, a second model was developed which included only the five climate variables that contributed to 95% of model gain. These included bio5 (Max Temp of Warmest Month), bio6 (Min Temp of Coldest Month), bio12 (Annual Precip (mm)), bio10 (Mean Temp of Warmest Quarter), and bio1 (Annual Mean Temp) (Table 2). Including the
default settings, I specified the Maxent program to include threshold features, a logistic output format, creating response curves, and creating the proper directories. In the advanced settings, I wrote the plot data, turned off MESS analysis, and applied the threshold of 10 percentile training presence.

Following the second iteration model run for the contemporary distribution of Bloodroot, I consulted two model fit statistics to determine whether the model was adequate. These included the value of the area under the curve (AUC) of the receiver operating characteristic (ROC), which is the most frequently used threshold-independent measure for SDM model performance (Lui et al. 2009). For AUC, a number closer to 1 is the most accurate, and any value greater than 0.7 is considered adequate. A value of 0.5 would indicate that the model does not perform any better than would be expected from a random model. Along with AUC, I examined omission error. A model with a testing omission error of less than 0.3 is considered reliable (Lui et al. 2009). To visualize the predicted suitable distribution for Bloodroot based on the second iteration of the model, a Maxent logistic value that coincided with 0.1 emission error of training data was used as a threshold to convert the raw model suitability value to binary form, in which a value of 0 is considered to be unsuitable habitat, while a value of 1 considered to be suitable habitat. Finally, the total study area was plotted against the thresholded predicted suitable distribution with the remaining area being regarded as unsuitable.

The final step in this project was to project the model for Bloodroot against a future climate scenario to identify potential range shifts/contractions for this species. To do so, I used the downscaled general circulation model (GCM) between 2046 and 2055 to incorporate the Representative Concentration Pathway (RCP) 6.0 from the 5th assessment of the IPCC/CMIP (Van Vuuren et al., 2011), as mentioned above. To implement this prediction, I re-ran the second
iteration of the SDM for Bloodroot, but added the 19 future BioClim variables as projection
layers, which resulted in the predictions from the second model being projected onto the layers
for the future climate.

To compare the predicted distributions for Bloodroot under current and future climate
scenarios, the area of predicted suitable distribution (in km²) was calculated for each. To
calculate the areas, I used the following equation: 5 arc minutes resolution = 0.0833333 decimal
degrees = 9.297 km, resulting in ~66.2 km² for each grid cell (Miller et al., 2021). This value was
then multiplied by the number of grid cells comprising the suitable area predicted by the
contemporary and future models to obtain the total area in km².

Results

The performance metrics for the Maxent SDM for Bloodroot proved to be reliable. The
training and testing AUC of the ROC were both above 0.7 (0.706 and 0.738, respectively) (Fig.
2). The omission error was low, at 0.1. The predicted suitable distribution for Bloodroot under
the contemporary and future climates are illustrated in Figures 3 and 4. In the prediction map for
Bloodroot’s contemporary distribution (Fig. 3), most of the range is located in the eastern United
States, extending far down into the south and extending north into southeastern Canada and as
far east as Nova Scotia. Most of the occurrence points used to train the model are located within
the range predicted to be suitable for Bloodroot (e.g., above 1 in the thresholded prediction;
green on the map). This indicates that the prediction for Bloodroot’s current geographic range is
relatively accurate. The current distribution of Bloodroot covers an area of about 130,000,000
km², as calculated via the equation specified above (Fig. 3).
For the future distribution under the RCP 6.0 climate scenario (2050s; Fig. 4), a northward and eastward shift in the range of Bloodroot is evident. With this shift, the bulk of the distribution is now projected to be in southern Canada, rather than the U.S. The range also extends further east than is specified by the contemporary range prediction. In terms of the training occurrences, many now lie outside of the future distribution, especially those to the west and south. This suggests that these presently occupied habitats may become uninhabitable for Bloodroot by the 2050s due to climate change. The future distribution of Bloodroot has an area of 100,000,000 km² (Fig. 4).

Comparison of the ranges predicted to be occupied by Bloodroot in the current and future map projections shows a decrease in area of about 30,000,000 km². In addition, the predicted range for Bloodroot based on this model is expected to shift approximately 2 decimal degrees north at its southernmost edge and ~5 decimal degrees east at its westernmost edge (Fig. 3; Fig. 4).

Discussion

Climate change is causing plant species distributions to experience latitudinal and elevational shifts (Yadav et al. 2023). Species distribution models are widely used to predict species distribution shifts over both space and time. These are essential tools that can be used to examine how climate change will shift the distributions of spring ephemeral plants, including Bloodroot. Among the plant species that are expected to be disproportionately affected by climate change are spring ephemerals. Not only are there relatively few studies on how the distribution of spring ephemerals will shift under climate change, but there is also concerning evidence that the phenology of spring ephemerals may shift under climate change (Austin et al. 2024). Thus, it is important to determine whether these species will experience range shifts as
well, given that combined shifts in phenology and range could result in unexpected and/or negative consequences for plant fitness or persistence (Yadav et al. 2023).

In this study, I demonstrate via Maximum entropy species distribution modeling that the geographic range of Bloodroot (*Sanguinaria canadensis*) is likely to shift further north and east and to experience a contraction of up to 30,000 km$^2$. Such a shift will likely render presently occupied habitats, especially those in the Midwest and southeast (outside of the Appalachian mountain range) uninhabitable by Bloodroot in the near future (2050s) due to climate change. This is a concerning discovery for the preservation of woodland understory plant communities and spring ephemeral plants alike, given that Bloodroot is generally recognized as a widespread spring ephemeral in eastern North American deciduous forests, and also is not presently considered at particular risk due to anthropogenic activities (NatureServe, 2023). That this species may experience such a drastic reduction in range due solely to anticipated changes to climatic conditions (excepting any other shifts due to habitat loss, fragmentation, or biotic effects) does not bode well for spring ephemeral plants as a whole.

That being said, this kind of species distribution model may not be entirely accurate, for a number of reasons. First, correlational SDMs do not include information about biotic interactions in predicting potential species distributions. Biotic interactions, including disease, predation, competition, and mutualism, are important factors determining species distributions and abundance; the absence of precise knowledge about these interactions in models of current and future species ranges can cloud our understanding of the potential for range shifts or contractions. For example, diseases can manage populations; predators can control the spread of prey; and mutualistic connections, such as those between pollinators and plants, can improve species survival and dispersal. However, data on these interactions at the scale and of the
frequency that would be required for SDM are typically lacking due to the difficulty of accurately collecting them in natural contexts.

Furthermore, dispersal mode and capacity has a considerable impact on a species’ geographic spread. Species with low dispersion potential may be unable to colonize physiologically appropriate habitats, whereas those with improved dispersal mechanisms, such as interactions with dispersing agents like ants, can more efficiently broaden their ranges. Thus, parsing the complexities of biotic interactions, particularly dispersal mechanisms, is critical for accurately forecasting changes in species distributions in response to environmental changes (Lui et al. 2021).

Ant seed dispersal, known as myrmecochory, is a fascinating ecological interaction in which ants play a vital role in seed dispersal. Many plant species, especially spring ephemerals, have developed seeds that are particularly appealing to ants, often with a nourishing seed-coat appendage known as an elaiosome. Within the myrmecochory mutualism, ants transport seeds to their nests, consume the elaiosome, and then discard the seed, often in a pre-determined site such as a midden. This mutualistic relationship benefits both the ants—via a nutritious food reward—and the plants, by allowing for more effective seed dissemination to nutrient-rich microsites, which are often less competitive and provide shelter from seed predators (Handel et. al 1990). This relationship also contributes to ecological balance by facilitating a form of seed dispersal that decreases both predation and competition (Kim et. al 2017).

Bloodroot, as a myrmecochore, has consequently developed a dispersal mechanism that permits its seeds to be distributed over a large region and planted in new areas that lead to germination and growth. The dispersal of Bloodroot seeds by ants is an excellent example of
coevolution, in which both the plant and the ant species evolved features that benefit one another. This type of ecological interaction is critical for the preservation of biodiversity in wooded environments, allowing a wide range of plant and animal species to flourish (Kim et. al 2017). Understanding its seed dispersal mechanism can assist in determining the ecological niche of Bloodroot. Directly influenced by its geographic distribution and vice versa, the ecological niche of Bloodroot can reveal habitat needs, life cycle timing, and interactions with other species. This relationship lets us see how specific ecological requirements and interactions help define the areas where a species can thrive, contributing to its overall pattern of distribution (Pulliam et. al 2019).

The results from this study show that Bloodroot is likely to experience a northward and eastward shift in its geographic range, as well as a contraction in its overall distribution. These shifts could have major, adverse effects for the future prospects of this species. As the rate of warming increases, it may become harder for species such as Bloodroot to continue shifting north or adapting to the future climate because they will presumably meet a threshold at which there will no longer be anywhere for them to go (Hijmans et al. 2007). Such a result would not only bode poorly for Bloodroot, but may also bring trouble for species that depend on these plants, such as pollinators like Andrena spp., who rely on pollen resources from plants such as Bloodroot. Mutualistic species could face extinction if they cannot keep up with geographic and/or phenological shifts in their partner species, and vice versa. There is already evidence that the relationship between pollinators and plants has been disrupted due to climate change (Kearns et al. 1997). The results of this and similar studies can provide direction for future research on climate change-related range shifts.
Bibliography


Kim HoiJin, K. H., & Kim GabTae, K. G. (2017). Diaspore, seed dispersion and seed germination characteristics of two myrmecochrous spring ephemerals-Jeffersonia dubia and Corydalis remota-.


Appendix

Figure 1. Picture of Bloodroot

Figure 2. The receiver operating characteristic (ROC) curve for the Bloodroot SDM. Note that the specificity is defined using predicted area, rather than true commission. This implies that the maximum achievable AUC is less than 1.
Figure 3. Projection of the distribution of Bloodroot under the current climate. The area in green is predicted by the model to be, for Bloodroot, suitable habitat, while the grey area is predicted to be unsuitable. The blue dots are the training occurrences, while the red dots are the testing occurrences.

Figure 4. Projection of the distribution of Bloodroot under the future climate. The area in green is predicted by the model to be, for Bloodroot, suitable habitat, while the grey area is predicted to be unsuitable. The blue dots are the training occurrences, while the red dots are the testing occurrences.
Table 1. These bioclimatic variables represent annual seasonality and extreme or limiting environmental factors.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO1</td>
<td>Annual Mean Temp (C)</td>
</tr>
<tr>
<td>BIO2</td>
<td>Mean Diurnal Range (C)</td>
</tr>
<tr>
<td>BIO3</td>
<td>Isothermality (100 * BIO2 / BIO7)</td>
</tr>
<tr>
<td>BIO4</td>
<td>Temp Seasonality (100 * SD)</td>
</tr>
<tr>
<td>BIO5</td>
<td>Max Temp of Warmest Month (C)</td>
</tr>
<tr>
<td>BIO6</td>
<td>Min Temp of Coldest Month (C)</td>
</tr>
<tr>
<td>BIO7</td>
<td>Temp Annual Range (C) (BIO5-BIO6)</td>
</tr>
<tr>
<td>BIO8</td>
<td>Mean Temp of Wettest Quarter (C)</td>
</tr>
<tr>
<td>BIO9</td>
<td>Mean Temp of Driest Quarter (C)</td>
</tr>
<tr>
<td>BIO10</td>
<td>Mean Temp of Warmest Quarter (C)</td>
</tr>
<tr>
<td>BIO11</td>
<td>Mean Temp of Coldest Quarter (C)</td>
</tr>
<tr>
<td>BIO12</td>
<td>Annual Precip (mm)</td>
</tr>
<tr>
<td>BIO13</td>
<td>Precip of Wettest Month (mm)</td>
</tr>
<tr>
<td>BIO14</td>
<td>Precip of Driest Month (mm)</td>
</tr>
<tr>
<td>BIO15</td>
<td>Precip Seasonality (CV)</td>
</tr>
<tr>
<td>BIO16</td>
<td>Precip of Wettest Quarter (mm)</td>
</tr>
<tr>
<td>BIO17</td>
<td>Precip of Driest Quarter (mm)</td>
</tr>
<tr>
<td>BIO18</td>
<td>Precip of Warmest Quarter (mm)</td>
</tr>
<tr>
<td>BIO19</td>
<td>Precip of Coldest Quarter (mm)</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0028554.t001

Table 2. WorldClim variables that contributed the most to Bloodroot occurrence points.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corr_SDMs_2024bio1</td>
<td>33.2</td>
<td>16.8</td>
</tr>
<tr>
<td>Corr_SDMs_2024bio10</td>
<td>29.3</td>
<td>18.2</td>
</tr>
<tr>
<td>Corr_SDMs_2024bio12</td>
<td>25.7</td>
<td>34.2</td>
</tr>
<tr>
<td>Corr_SDMs_2024bio5</td>
<td>11.4</td>
<td>27.8</td>
</tr>
<tr>
<td>Corr_SDMs_2024bio6</td>
<td>0.4</td>
<td>3</td>
</tr>
</tbody>
</table>