


Article

Stepwise Multidimensional Climate Envelop Modeling of Pitch Pine (*Pinus rigida*)

Olga Rumyantseva and Nikolay Strigul * 

Department of Mathematics and Statistics, Washington State University Vancouver,
14204 NE Salmon Creek Avenue, Vancouver, WA 98686, USA

* Correspondence: nick.strigul@wsu.edu

Abstract: Understanding the intricate relationships between climate and vegetation remains a fundamental challenge in contemporary ecology. The ability to anticipate the specific climatic factors affecting different tree species and understand how they respond is crucial for mitigating the impacts of climate change on forested ecosystems. Additionally, quantitatively assessing habitat loss resulting from anthropogenic activities is essential for informed conservation efforts. Our objective is to evaluate the potential distribution of pitch pine (*Pinus rigida*) in North America and assess the associated habitat loss. To achieve this, we employ a stepwise multidimensional climate envelope modeling approach, comparing two data-intensive models—the Variable Interaction Model (VIM) and the Variable Non-Interaction Model (VNM). These models discern the influence of diverse combinations of climatic characteristics on the distribution of the species. Both VNM and VIM employ Shapley values for factor ranking during construction. VNM assumes independent effects, resulting in a hyperrectangle-shaped climate envelope, while VIM considers interactions, yielding a complex, data-driven multidimensional envelope. Data integration involves mining the US Forest Inventories and climatic data, encompassing 19 parameters. The results unequivocally highlight the superior predictive accuracy of VIM compared to the Variable Non-Interaction Model, VNM. The modeling approach developed in this study has the potential to enhance species distribution models for various tree species in the context of evolving climatic conditions.

Keywords: species distribution models; habitat loss; climate envelopes; pitch pine; Shapley values; variable interaction model; stepwise model



Citation: Rumyantseva, O.; Strigul, N. Stepwise Multidimensional Climate Envelop Modeling of Pitch Pine (*Pinus rigida*). *Forests* **2024**, *15*, 819. <https://doi.org/10.3390/f15050819>

Academic Editor: Juan A. Blanco

Received: 15 March 2024

Revised: 29 April 2024

Accepted: 29 April 2024

Published: 7 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding the intricate relationships between climate and vegetation remains a pivotal challenge in contemporary ecology. From a practical standpoint, the crucial need to quantitatively evaluate habitat loss for tree species under varying climate change and land use scenarios is imperative for preserving biodiversity, ecosystem services, and the overall well-being of the planet. Species distribution modeling aims to tackle this challenge [1,2]. Within this field, diverse quantitative methodologies are employed to unravel the ecological determinants influencing plant species distributions, predict potential changes in response to various climate change scenarios, and assess assorted conservation strategies [1,3–5]. Correlative Species Distribution Models (SDMs) rely on statistical relationships between environmental parameters and species occurrence, employing statistical techniques encompassing correlation and regression analyses, various distance-based metrics, entropy measures, and machine learning methods [2,6–8]. These models possess the capacity to anticipate species distributions when an extensive corpus of observational data is available [9]. The inherent limitations of both correlative and mechanistic species distribution models (SDMs) are well established [8].

The data-intensive approach enables the quantitative assessment of relationships within complex ecological systems [10,11]. Employing data mining and data-driven analyses of spatially explicit climatic datasets and individual-based forest inventories serves as

an effective tool for investigating the connections between various quantitative features of climate and vegetation [12,13]. For example, data-intensive modeling allowed us to investigate multidimensional forest dynamics [14], succession [15], and tolerance patterns [12,15]. In another data-driven study, we employed multivariate statistics and machine learning to rank climatic factors by their effects on forest basal area in different ecoregions [13].

The objective of this research is to evaluate the relative importance of climatic factors regarding the distribution of pitch pine in North America. We develop multidimensional climate envelope models of pitch pine distribution using a data-intensive paradigm. Pitch pine, with its unique habitat preferences and native geographic distribution [16,17], serves as a convenient model species for species distribution modeling (Section 4.1). Its range, spanning across the Eastern United States provides a diverse set of environmental conditions [18–20]. Understanding the factors influencing its presence within these varied habitats can offer valuable insights into the ecological requirements and adaptations of pitch pine [21]. Moreover, as a species well known for its preference for fire-prone ecosystems [16,20,22], studying its distribution can contribute significantly to refining and validating species distribution models, enhancing our comprehension of broader ecological patterns and climatic influences on species's ranges.

We constructed climate envelope models by ranking individual climatic factors according to their average expected marginal contribution using the methodology derived from cooperative game theory, specifically leveraging the concept of Shapley values. We conceptualized climatic factors as individual players within a cooperative game framework, wherein the Shapley values serve to quantify the impact of each climatic factor. The use of Shapley values in this context allows us to quantify the specific impact of individual climatic factors on the overall model's predictions. Shapley values are broadly employed in the realm of machine learning. They aim to explain the prediction of a machine learning model by quantifying the contribution of each feature to the prediction. Shapley values consider all possible coalitions of features and calculate the marginal contribution of each feature to the prediction. In general, this methodology allows for a more nuanced understanding of how each climatic factor contributes to the overall model, potentially enabling the identification of which factors are most influential or critical in shaping the climate envelope models.

We developed multidimensional climate envelope models for the pitch pine species using 19 distinct climatic parameters (Figure 1) in conjunction with spatial distribution data obtained from the USDA Forest Inventories dataset. Our study involved the comparative analysis of two distinct methodologies for climate envelope modeling: the Variable Interaction Model (VIM) and the Variable Non-Interaction Model (VNM). These modeling approaches are founded on differing assumptions regarding the interplay between climatic factors and species distributions. The Variable Non-Interaction Model operates under the assumption that each factor independently influences species distribution and formulates a multidimensional climate envelope resembling a hyperrectangle. Conversely, the Variable Interaction Model is structured to consider potential collective impacts resulting from the interaction of various climatic factors, resulting in the creation of a multidimensional climate envelope characterized by intricate, data-driven geometries. The construction of these climate envelope models involved the ranking of individual climatic factors based on their average expected marginal contribution. Overall, this method provides a better understanding of how individual climatic factors contribute to the overall model, thereby facilitating the identification of the most influential or critical factors in shaping the climate envelope models.

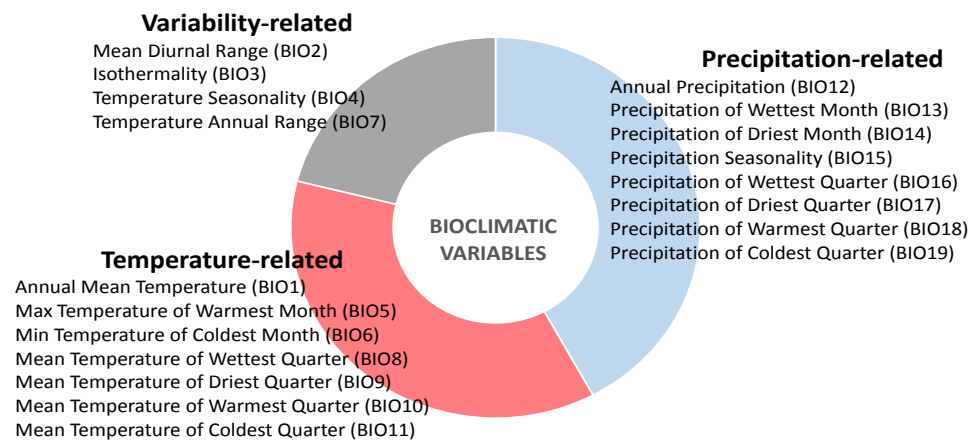


Figure 1. Bioclimatic variables extracted from the WorldClim dataset (see Table 1 for definitions).

Table 1. Definitions of bioclimatic variables. Notation: T_i is the average temperature for the month i ; $T_i = \frac{T_{max_i} + T_{min_i}}{2}$; T_{max_i} is the monthly mean of daily maximum temperatures for the month i ; T_{min_i} is the monthly mean of daily minimum temperatures for the month i . SD is a standard deviation. The wettest quarter is determined while comparing the total precipitations of each quarter: $PPT_{max} = \max\{P_{Q_1}, P_{Q_2}, \dots, P_{Q_{12}}\}$, and the values $P_{Q_1} = \sum_{i=1}^3 PPT_i$, $P_{Q_2} = \sum_{i=4}^6 PPT_i$, $P_{Q_3} = \sum_{i=7}^9 PPT_i$, \dots , $P_{Q_{10}} = \sum_{i=10}^{12} PPT_i$, $P_{Q_{11}} = \sum_{i=1}^3 PPT_i$, $P_{Q_{12}} = \sum_{i=4}^6 PPT_i$ are total precipitations of 12 quarters (PPT_i is a precipitation of the month i). The driest quarter is determined while comparing the total precipitations of each quarter: $PPT_{min} = \min\{P_{Q_1}, P_{Q_2}, \dots, P_{Q_{12}}\}$. The warmest quarter is determined comparing the total temperatures for each quarter: $T_{max} = \max\{T_{Q_1}, T_{Q_2}, \dots, T_{Q_{12}}\}$, where $T_{Q_1} = \sum_{i=1}^3 T_i$, $T_{Q_2} = \sum_{i=4}^6 T_i$, $T_{Q_3} = \sum_{i=7}^9 T_i$, \dots , $T_{Q_{10}} = \sum_{i=10}^{12} T_i$, $T_{Q_{11}} = \sum_{i=1}^3 T_i$, $T_{Q_{12}} = \sum_{i=4}^6 T_i$ -total temperatures of quarters (T_i is an average temperature of the month i). The coldest quarter is determined comparing the total temperatures for each quarter: $T_{min} = \min\{T_{Q_1}, T_{Q_2}, \dots, T_{Q_{12}}\}$.

Code	Appellation	Formula
BIO1	Annual Mean Temperature	$BIO1 = \frac{\sum_{i=1}^{12} T_i}{12}$
BIO2	Mean Diurnal Range	$BIO2 = \frac{\sum_{i=1}^{12} (T_{max_i} - T_{min_i})}{12}$
BIO3	Isothermality	$BIO3 = \frac{BIO2}{BIO7} * 100$
BIO4	Temperature Seasonality	$BIO4 = SD\{T_1, T_2, \dots, T_{12}\} * 100$
BIO5	Max Temperature of Warmest Month	$BIO5 = \max\{T_{max_1}, T_{max_2}, \dots, T_{max_{12}}\}$
BIO6	Min Temperature of Coldest Month	$BIO6 = \min\{T_{min_1}, T_{min_2}, \dots, T_{min_{12}}\}$
BIO7	Temperature Annual Range	$BIO7 = BIO5 - BIO6$
BIO8	Mean Temperature of Wettest Quarter	$BIO8 = \frac{\sum_{k=1}^3 T_k}{3}$, where T_k is an average temperature of the month k belonging to the wettest quarter.
BIO9	Mean Temperature of Driest Quarter	$BIO9 = \frac{\sum_{k=1}^3 T_k}{3}$, where T_k is an average temperature of the month k belonging to the driest quarter.
BIO10	Mean Temperature of Warmest Quarter	$BIO10 = \frac{\sum_{k=1}^3 T_k}{3}$, where T_k is an average temperature of the month k belonging to the warmest quarter.
BIO11	Mean Temperature of Coldest Quarter	$BIO11 = \frac{\sum_{k=1}^3 T_k}{3}$, where T_k is an average temperature of the month k belonging to the coldest quarter.
BIO12	Annual Precipitation	$BIO12 = \sum_{i=1}^{12} PPT_i$

Table 1. Cont.

Code	Appellation	Formula
BIO13	Precipitation of Wettest Month	$BIO13 = \max\{PPT_1, PPT_2, \dots, PPT_{12}\}.$
BIO14	Precipitation of Driest Month	$BIO14 = \min\{PPT_1, PPT_2, \dots, PPT_{12}\}$
BIO15	Precipitation Seasonality	$BIO15 = \frac{SD\{PPT_1, PPT_2, \dots, PPT_{12}\}}{1 + BIO12/12} * 100$
BIO16	Precipitation of Wettest Quarter	$BIO16 = \max\{P_{Q_1}, P_{Q_2}, \dots, P_{Q_{12}}\}.$
BIO17	Precipitation of Driest Quarter	$BIO17 = \min\{P_{Q_1}, P_{Q_2}, \dots, P_{Q_{12}}\}.$
BIO18	Precipitation of Warmest Quarter	$BIO18 = \sum_{i=1}^3 PPT_i, \text{ month } i \text{ is in the warmest quarter.}$
BIO19	Precipitation of Coldest Quarter	$BIO19 = \sum_{i=1}^3 PPT_i, \text{ month } i \text{ is in the coldest quarter.}$

2. Materials and Methods

2.1. Data Mining and Climate Envelop Modeling

The data mining procedure employed in our study closely aligns with the methodology utilized in our previous research endeavors [12–14]. We utilized two datasets: the USDA Forest Inventory and Analysis (FIA) dataset and the WorldClim dataset covering climate data for 1970–2000, which can be accessed at the WorldClim data website. The USDA FIA program, an ongoing forest inventory initiative initiated in the late 1960s, is designed to survey forested ecosystems across all ecological domains in the United States. The database is accessible from USDA Forest Service. Prior to statistical analysis, the data were meticulously scrutinized to identify and eliminate any conspicuous outliers, following the previously published data preparation procedures. We exclusively considered living tree species within the forest plots for which detailed climatic characteristics were recorded. This selection criteria led to a dataset comprising a total of 271,846 forest stands distributed across the contiguous United States [15]. The inventory design ensures that each plot samples a hexagonal area of approximately 6000 acres. Consequently, SDM statistics can be computed directly based on plot information [12].

To predict the spatial distribution of pitch pine, we utilized 19 bioclimatic variables accessible within the WorldClim dataset. The complete list of these variables is presented in Figure 1 and further detailed in Table 1. FIA dataset includes the geographical coordinates of FIA plots, designated by the fields LAT and LON, along with the associated values for the 19 bioclimatic attributes at these specific locations, which are identified as BIO1 through BIO19. FIA deliberately alters certain plot locations to comply with privacy requirements; however, this procedural adjustment does not introduce a substantial error in SDMs [12,23]. The information concerning the presence of pitch pine was obtained by using the Tree Species Code 126, which corresponds to this particular species. Consequently, each forest plot corresponds to a 19-dimensional vector representing the values of these climatic factors, denoted as $\mathbf{v} = (v_1, v_2, \dots, v_{19})$.

To outline our methodology for modeling the spatial distribution of pitch pine, we introduce a distinction between two pivotal concepts: the *realized distribution* and the *potential distribution*. We define the *realized distribution* as the collection of forest plots (i.e., FIA plots in this study) where the tree species is observed to be present according to the available dataset. In contrast, the *potential distribution* encompasses the set of forest plots where the tree species could potentially exist based on the model's predictions.

2.2. Multivariate Statistical Analysis

To explore the statistical relationships between climatic variables and pitch pine distribution, we utilized two standard methods: correlation analysis and Principal Component Analysis (PCA). The multidimensional statistical analysis was performed in R using standard packages for correlational and principal component analyses (FactoMineR, factoextra, and corrplot). All used software, detailed information, and supporting

documentation are accessible from the Comprehensive R Archive Network, CRAN repository (<https://cran.r-project.org/>) (downloaded in 10 April 2023).

While these analyses are widely used and their details can be found elsewhere [24], it is important to note that both approaches are linear dimensionality reduction methods. Correlation analysis constructs a matrix of correlation coefficients between all variables, while PCA computes a set of orthogonal variables called principal components that sequentially capture the largest variations.

2.3. Variable Interaction Model (VIM)

The Variable Interaction Model (VIM) proposed in this study discerns geographic locations demonstrating climatic conditions akin to those conducive for the growth of pitch pine. Notably versatile, this model is applicable across diverse tree species and geographical regions, facilitating the identification of areas with the potential for supporting the growth of a specific tree species.

The VIM discerns forest plots characterized by climatic factor values identical to those observed in the realized distribution of pitch pine (Figure 2). Put differently, a plot is deemed part of the potential distribution of pitch pine if the vector $\mathbf{v} = (v_1, v_2, \dots, v_{19})$ at that location corresponds to the combination of variables identified in the realized distribution:

$$\mathbf{v} = \mathbf{r}, \quad (1)$$

where $\mathbf{r} = (r_1, r_2, \dots, r_{19})$ represents the vector of climatic characteristic within the realized distribution. Since vector equality is determined componentwise, the following logical conjunction is upheld:

$$v_1 = r_1 \ \& \ v_2 = r_2 \ \& \ \dots \ \& \ v_{19} = r_{19}. \quad (2)$$

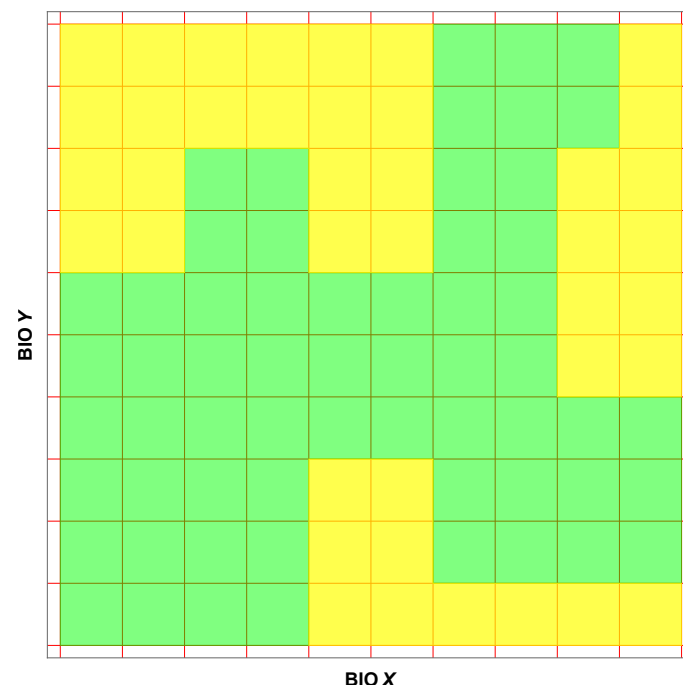


Figure 2. Hypothetical example illustrating the distinction between the Variable Non-Interaction Model (VNM) and the Variable Interaction Model (VIM). The spatial distribution of the focal tree species is depicted by green squares within the two-dimensional parameter space defined by two climatic characteristics, BIOX and BIOY. The VNM is constructed to encompass the entire two-dimensional region, represented as the Cartesian product of the intervals of BIOX and BIOY, inclusive of all green and yellow squares. In contrast, the VIM is exclusively trained on the subset of locations occupied by the green squares.

2.4. Variable Non-Interaction Model (VNM)

The Variable Non-Interaction Model (VNM) is formulated as a traditional climate envelope model. This model identifies geographic locations where climatic characteristics align with the values encapsulated within the hyper-rectangle (hyperbox) that encompasses all the observed values in the realized distribution. For instance, in the scenario where the dimensionality of climatic variables is $N = 19$, a forest plot is considered part of the potential distribution if the vector of climatic variable values, denoted as $\mathbf{v} = (v_1, v_2, \dots, v_{19})$, at that specific plot satisfies the following criterion:

$$\text{for every } i = \overline{1, 19} \quad v_i = r_i^1 \text{ or } r_i^2 \text{ or } \dots \text{ or } r_i^N, \quad (3)$$

where $\mathbf{r}^1, \mathbf{r}^2, \dots, \mathbf{r}^N$ represent vectors containing values of climatic characteristics within the realized distribution of pitch pine. In essence, for a forest plot to be considered part of the potential distribution, each component v_i ($1 \leq i \leq 19$) of the vector \mathbf{v} must align with the corresponding component of some vector from the realized distribution.

The VNM is similarly crafted to recognize locations with climates resembling those observed within the realized distribution. Notably, this model imposes fewer restrictions on geographical locations (Figure 2), resulting in a more expansive predicted potential distribution (See Section 3.2).

2.5. Ranking of Bioclimatic Variables Using Shapley Values

The Shapley value is a concept developed in cooperative game theory to fairly distribute the total payoff of the coalition among its members [25]. It captures the average marginal contribution of a player across all feasible sequences in which players might join the coalition. Calculated with fairness in mind, the Shapley value ensures that each player receives recognition in proportion to their individual contributions to the group. Denoted as player i , the Shapley value represents a numerical measure quantifying the significance of this player within the broader coalition of players [26]:

$$\sum_{S \subseteq N \setminus \{i\}} K_s \left(v(S \cup \{i\}) - v(S) \right), \quad (4)$$

where v represents a function that assigns real numbers to subsets of players, N is the total coalition consisting of n players, S is a specific subset of players within the coalition N , $v(S)$ denotes the value or worth of the coalition S , $v(S \cup \{i\})$ represents the value of the coalition formed by adding player i to the subset S , and K_s is a combinatorial coefficient, calculated as follows: $K_s = \frac{|S|!(n-|S|-1)!}{n!}$.

The Shapley value of player i gauges their importance within the coalition N by considering their contributions to the value of various coalitions formed when adding player i to different subsets S of the total coalition. We consider all possible permutations of players and calculate their marginal contributions to each coalition. The Shapley value for a player is the average of their marginal contributions over all possible orderings. It is calculated as a weighted average, with weights determined by the combinatorial coefficients K_s . This approach ensures equitable distribution of the coalition's total value among its members by accounting for their respective contributions across all possible arrangements [26].

We employ Shapley scores to assess the climatic characteristics that make the most significant contributions to the potential distribution of pitch pine, taking into account both the Variable Interaction Model (VIM) and the Variable Non-Interaction Model (VNM). In our context, the Shapley inclusion score of an individual climatic variable, denoted as i , concerning a combination of climatic variables in set S (where set S does not include i), is determined as the scaled difference adjusted by the combinatorial coefficient K_s between the "model score" of S with the inclusion of i and the "model score" of S (as shown in Equation (4)). The "model score" in the Variable Interaction Model (VIM) is determined by the number of FIA plots where all values in a specific set of climatic variables match

exactly with at least one plot from the realized distribution. In the Variable Non-Interaction Model (VNM), the “model score” for a specific set of climatic variables is determined by the number of plots falling within the hyper-rectangle that is constructed from the values in the realized distribution.

In particular, to calculate the Shapley scores for the 19 climatic variables, we create a combination matrix with 18 columns and rows that represent all possible subsets of the set of 18 variables. Utilizing this combination matrix, we generate a 19-column matrix containing the Shapley values. The entry in this matrix at the i -th row and j -th column represents the Shapley inclusion score of the j -th climatic variable concerning the combination of climatic variables corresponding to the i -th row of the combination matrix. It is important to note that we do not consider the Shapley inclusion score of a climatic variable with respect to an empty set, which leads to the Shapley matrix having one fewer row compared to the combination matrix. Please refer to Appendix A for a comprehensive example illustrating the calculation of Shapley scores.

3. Results

3.1. Climatic Conditions in Pitch Pine Habitats: Multivariate Statistical Analysis

The correlation analysis unveiled clusters of closely interlinked climatic factors. The first cluster, identified as the largest cluster of positively correlated factors (Figure 1), encompasses nearly all precipitation-related variables (BIO12, BIO13, BIO14, BIO16, BIO17, BIO18, and BIO19), with the sole exception of BIO15, representing Precipitation Seasonality. These factors exhibit moderate-to-strong positive correlations among themselves. Notably, they demonstrate minimal correlation with temperature-related and variability-related factors.

The second-largest cluster in Figure 1 comprises five temperature-related factors (BIO1, BIO5, BIO6, BIO10, and BIO11). These factors exhibit moderate-to-strong positive correlations with each other and display no significant correlation with the factors from the first cluster of precipitation-related factors. However, two temperature-related factors do not fall into this temperature-related cluster. One of them, BIO9 Mean Temperature of Driest Quarter, demonstrates small-to-moderate positive correlations with many precipitation-related factors of the first cluster and temperature-related factors of the second cluster. The second exceptional factor in temperature-related group, BIO8 Mean Temperature of Wettest Quarter, demonstrates a modest negative correlation with the precipitation-related factor BIO19 Precipitation of Coldest Quarter but exhibits no significant correlations with any other factors.

Two variability-related variables, BIO4 Temperature Seasonality and BIO7 Temperature Annual Range, can be regarded as the third and smallest cluster. This association is distinctive, given that almost all significant correlations between the factors illustrated in Figure 3 are positive. However, BIO4 and, to a lesser extent, BIO7 show negative correlations with several precipitation- and temperature-related factors. An exceptional factor from the precipitation-related group, BIO15 Precipitation Seasonality, is somewhat more connected to this cluster, as BIO15 displays a small correlation with both BIO4 and BIO7.

Some of these observed associations can be readily explained by their respective formulas (Table 1 and Figure 1), while certain outcomes are unexpected and pose challenges for interpretation. In particular, a close positive correlation between BIO2 and BIO3 is obviously expected; however, BIO3 is not correlated with BIO7, which is also involved in the BIO2 definition (Table 1). Similarly, BIO7 shows a moderate negative correlation with BIO6, which one would anticipate given how BIO7 is computed. However, counterintuitively, BIO7 is not correlated with BIO5 but demonstrates another modest negative correlation with BIO11 and a positive correlation with BIO4. The significant negative correlation between BIO8 and BIO19 is a rare example of the association between temperature- and precipitation-related factors, but a rational explanation of this association is not obvious.

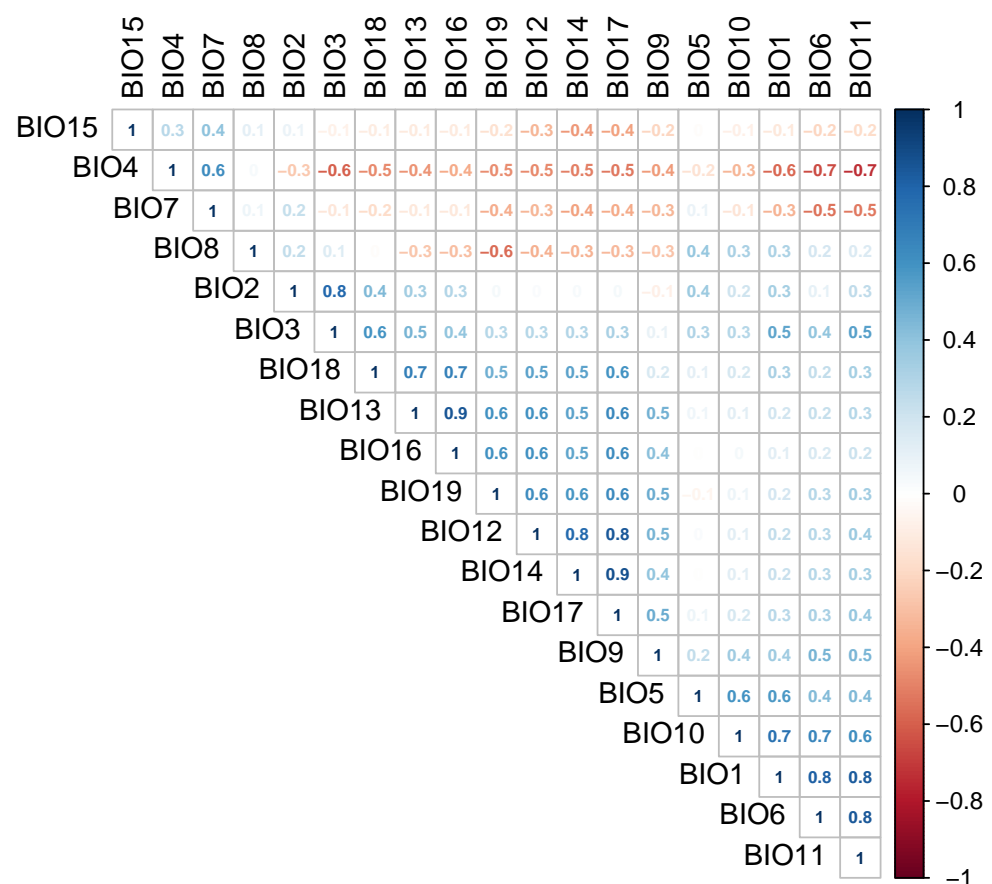


Figure 3. Climatic characteristics, correlation diagram.

The application of Principal Component Analysis (PCA) indicates a significant potential for dimensionality reduction within the dataset (Figures 4–9). The initial three principal components account for approximately 70% of the variance, while the first six components explain around 90%. We expect the PCA results to mirror the findings of the correlation analysis, revealing three clusters roughly associated with three groups of climatic variables: precipitation, temperature, and variability-related factors. This expectation stems from the statistical similarities between correlation analysis and PCA, as both methods estimate linear relationships and operate with covariances and correlations [27]. However, no climatic factor contributes more than 10% to the first principal component (Figure 5). Counterintuitively, the first principal component (Figure 5) is primarily influenced by uncorrelated factors (BIO17, BIO4, BIO11, BIO12, and BIO14) representing all three groups of climatic variables, along with seven other variables. The second and third principal components (Figures 6 and 7) align more closely with our expectations. It seems that temperature-related factors—BIO8, BIO1, BIO5, BIO10, BIO6, BIO11—predominantly contribute to the second principal component (Figure 6), while variability-related factors—BIO2, BIO7, and BIO3—have the most substantial impact on the third component (Figure 7). In summary, PCA indicates that the dimensionality of the set of climatic variables can be significantly reduced through orthogonal linear transformation, with six principal components explaining approximately 90% of the variance. It is noteworthy that these principal components do not closely align with specific climatic variables.

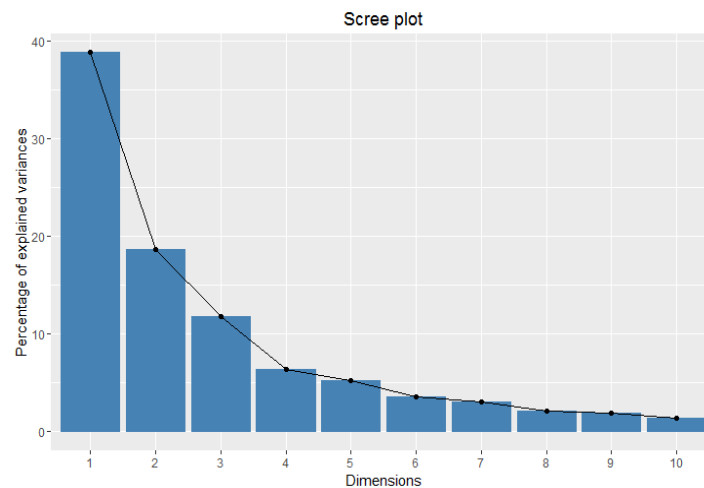


Figure 4. The percentage of explained variances of the principal components.

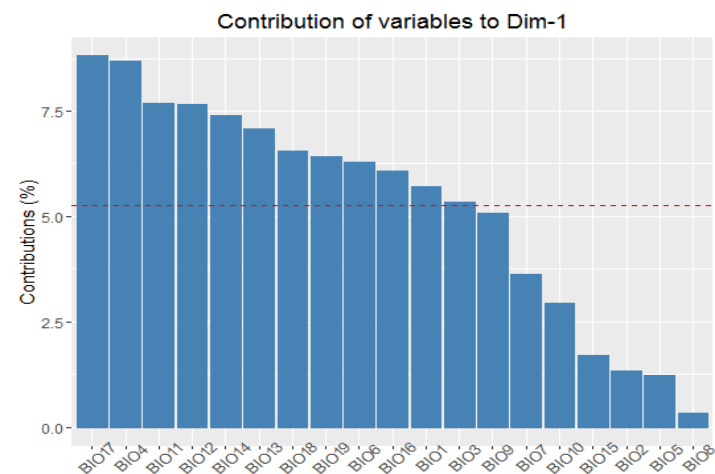


Figure 5. The contribution of the bioclimatic variables to the first principal component. The red dashed line illustrates the anticipated average contribution calculated under the assumption of equal contribution from all variables.

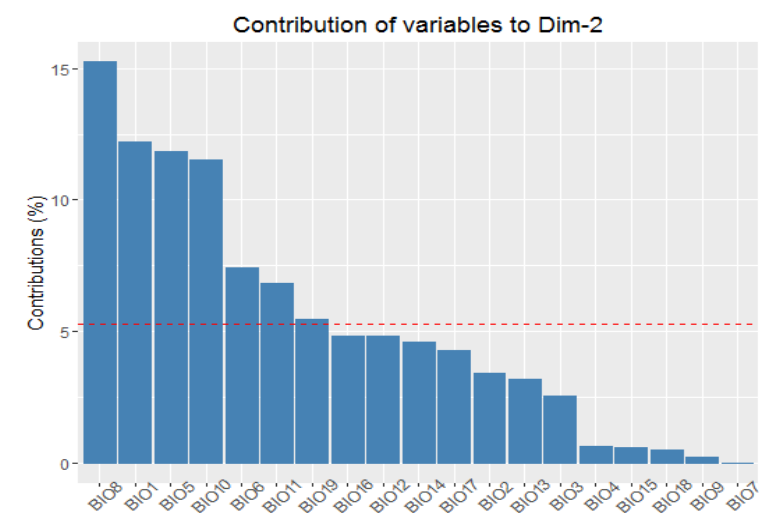


Figure 6. The contribution of the bioclimatic variables to the second principal component. The red dashed line illustrates the anticipated average contribution calculated under the assumption of equal contribution from all variables.

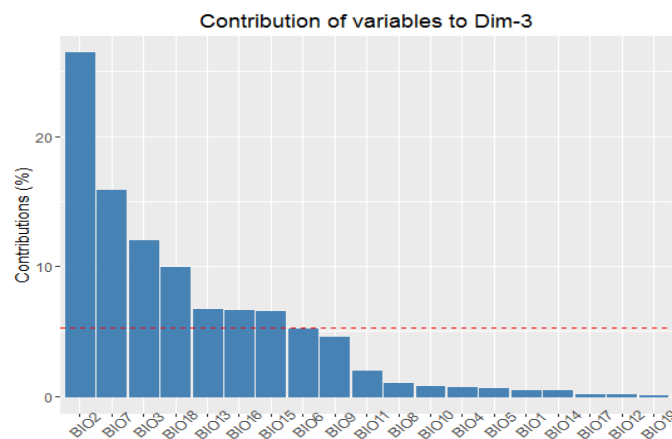


Figure 7. The contribution of the bioclimatic variables to the third principal component. The red dashed line illustrates the anticipated average contribution calculated under the assumption of equal contribution from all variables.

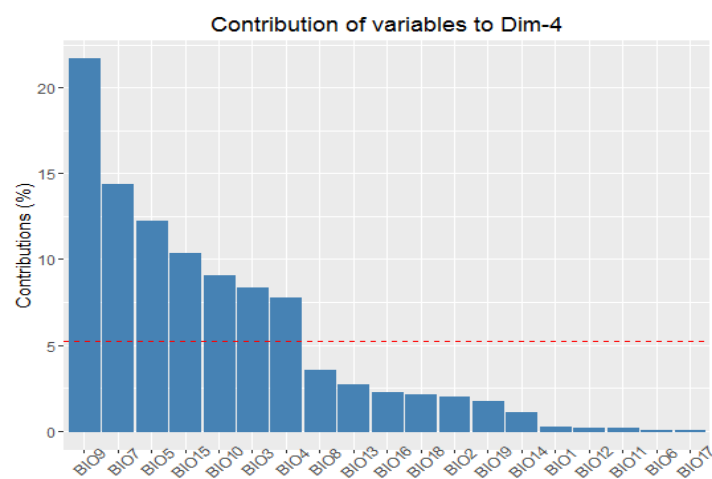


Figure 8. The contribution of the bioclimatic variables to the fourth principal component. The red dashed line illustrates the anticipated average contribution calculated under the assumption of equal contribution from all variables.

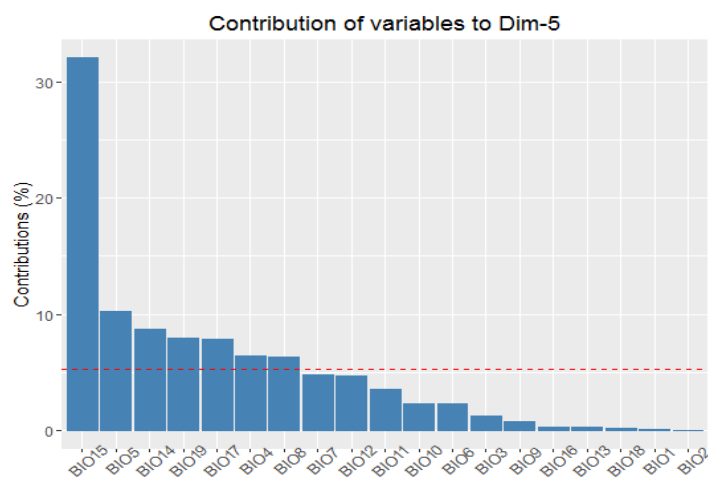


Figure 9. The contribution of the bioclimatic variables to the fifth principal component. The red dashed line illustrates the anticipated average contribution calculated under the assumption of equal contribution from all variables.

3.2. Stepwise Climate Envelop Modeling: VIM and VNM

Shapley scores were computed for both the Variable Interaction Model (VIM) and the Variable Non-Interaction Model (VNM) to establish a ranking of climatic factors influencing the distribution of the pitch pine (Figures 10–12). Figures 13 and 14 depict the alterations in the potential area for pitch pine during the stepwise simulations conducted with VIM and VNM. Ultimately, Figures 15 and 16 illustrate maps delineating the potential areas for pitch pine across the conterminous United States, calculated through the VIM and VNM stepwise models.

The Variable Interaction Model (VIM) demonstrates significant efficacy over the Variable Non-Interaction Model (VNM) in terms of mitigating the potential area for pitch pine. The initial potential area, representing the expanse covered by the network of USDA FIA permanent plots across the conterminous United States, is considered as the baseline. Following the execution of only the first two steps, VIM (VIM,2) reduces this original potential area by 50% (see Figure 11), effectively eradicating pitch pine from the western part of the USA (Figure 15a, FIM,2). Conversely, VNM achieves a comparable outcome only upon incorporating the first six factors (see Figures 12 and 15r, FNM,6). Upon the integration of all 19 climatic variables, the outcomes of both models also exhibit significant disparities. VIM,19 projects a potential area that, while larger than the realized area, does not deviate significantly (Figure 10a,b). In contrast, VNM anticipates a potential distribution area approximately four times greater than the observed area (Figure 10a,c).

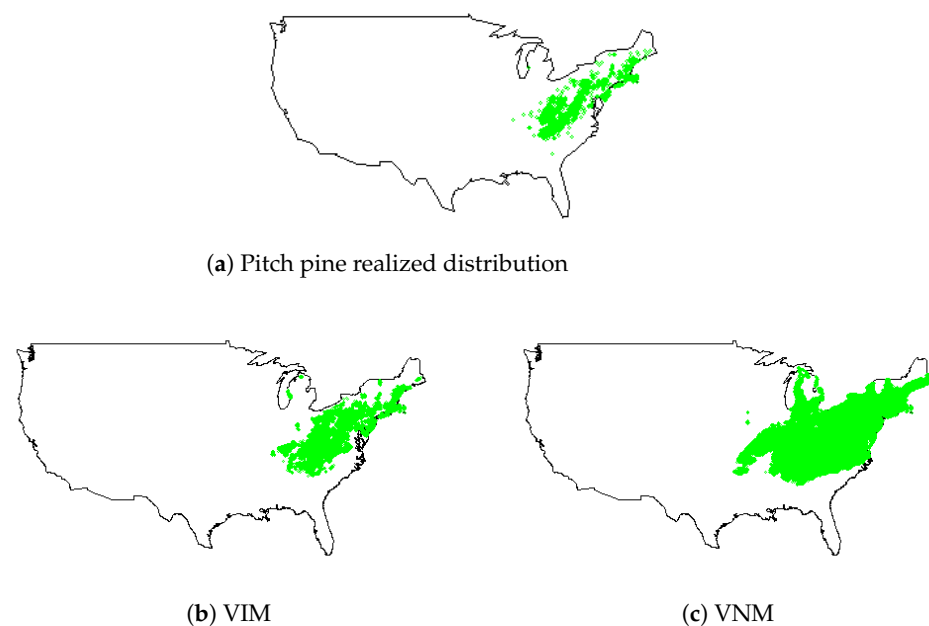


Figure 10. The pitch pine realized distribution is depicted in (a), while the potential distributions under the VIM and VNM models are represented in (b) and (c), respectively. These distributions are based on the set of 19 climatic characteristics BIO1 through BIO19.

The dynamics of potential area reduction, observed through the sequential addition of more regressors, manifests in three discernible stages: (1) an initial rapid decrease during the initial iterations, succeeded by (2) a more gradual reduction over subsequent iterations as the area converges towards a final state, and (3) the attainment of a stationary area during the concluding iterations, demonstrating minimal alteration (Figures 11 and 12). Notably, the initial stage of rapid decrease spans approximately six iterations, aligning broadly with the findings of the Principal Component Analysis (PCA) as detailed in Section 3.1. Indeed, we can regard the evolution of our stepwise climate envelope models guided by regressor selection through Shapley values as an alternative manifestation of dimensional analysis. In this context, the disparity in predictions between VIM and VNM underscores the pivotal

significance of the relationships that pitch pine shares not only with individual climatic variables but also with their collective and interactive effects.

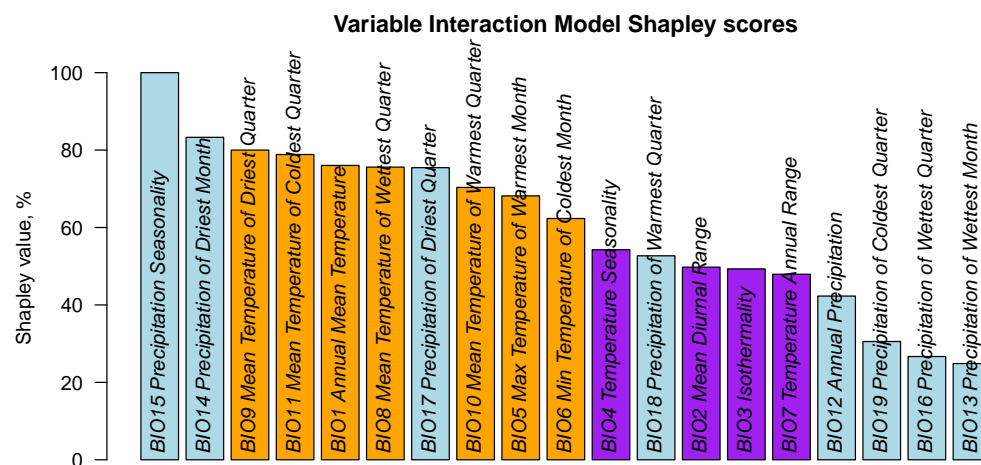


Figure 11. Shapley scores (in percentage from the maximal one) for the set of 19 climatic characteristics according to the VIM.

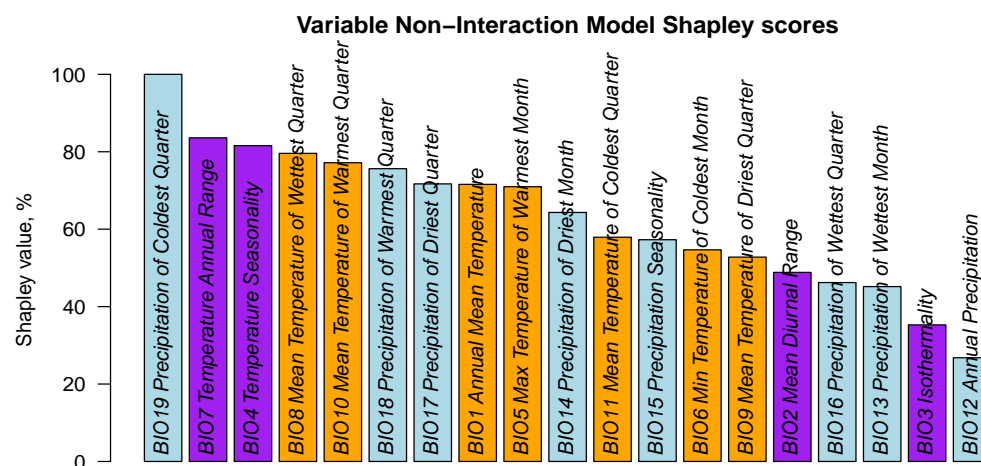


Figure 12. Shapley scores (in percentage from the maximal one) for the set of 19 climatic characteristics according to the VNM.

Upon examining the group representation, it becomes apparent that the Variable Interaction Model (VIM) predominantly features precipitation- and temperature-related factors as the foremost six primary climatic variables (BIO15, BIO14, BIO09, BIO11, BIO01, and BIO08, as indicated in Figure 11). Theoretically, the predominant inclusion of average temperature and precipitation-related factors in the Variable Interaction Model (VIM) aligns with and potentially justifies the conventional selection of these variables as pivotal in species distribution modeling. Nevertheless, a more nuanced analysis reveals that the scenario is inherently intricate and multifaceted.

The foremost variable, BIO15 Precipitation Seasonality, although officially categorized within precipitation-related variables, is, in reality, more accurately characterized as a variability-related quantity (Table 1). Furthermore, correlation analysis indicates that BIO15 is an exceptional variable within the precipitation-related group, displaying no significant correlation with other variables in this group (Figure 3). Instead, it aligns with the third cluster, comprised of variability-related characteristics (see Section 3.1). The second variable, BIO14 Precipitation of the Driest Month, falls within the category of common precipitation-related variables, aggregating in the first cluster as revealed by correlation analysis (see Figure 3 and Section 3.1).

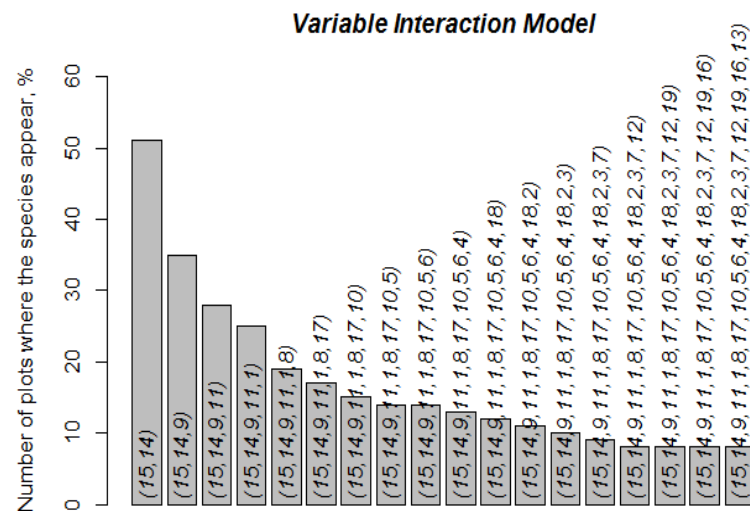


Figure 13. Predicted potential area for pitch pine (expressed as a percentage of FIA plots) utilizing the Variable Interaction Model (VIM). Histogram columns depict simulations ranging from VIM2 to VIM19. The vectors in brackets display climatic variables (BIO1–BIO19) ordered by Shapley values (refer to Figure 11). Distribution maps for pitch pine in each VIM model are showcased in Figures 15 and 16.

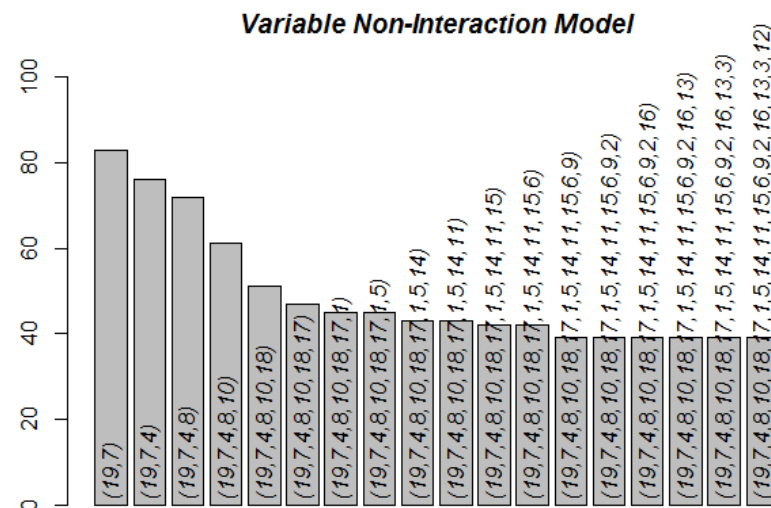


Figure 14. Predicted potential area for pitch pine (expressed as a percentage of FIA plots) utilizing the Variable Non-Interaction Model (VNM). Histogram columns depict simulations ranging from VNM2 to VNM19. The vectors in brackets display climatic variables (BIO1–BIO19) ordered by Shapley values (refer to Figure 12). Distribution maps for pitch pine in each VIM model are showcased in Figures 15 and 16.

The third-ranked variable, BIO9 Mean Temperature of Driest Quarter, stands out as an exceptional variable that does not align with any of the three clusters unveiled by correlation analysis. Instead, it exhibits modest correlations with both temperature- and precipitation-related factors (Figure 3). Subsequently, the variable BIO11 Mean Temperature of Coldest Quarter aligns as a typical representative from the third cluster encompassing temperature-related characteristics. It demonstrates high-to-moderate correlations with other variables within this cluster. The fifth-ranked variable, BIO1 Annual Mean Temperature, represents another quantity from the third cluster. Significantly, this represents the first occurrence where the algorithm selects a variable with substantial correlation to its predecessor, given that BIO11 and BIO1 exhibit a high degree of correlation. The sixth variable, BIO8 Mean Temperature of Wettest Quarter, stands out as another exceptional

quantity within the temperature-related group. It displays no significant correlation with other variables and exhibits a closer association with the third cluster of variability-related factors (Figure 3). The VIM model, employing these six variables (VIM,6), results in a highly significant reduction in the potential area relative to the addition of subsequent variables. The inclusion of the next two variables leads to progressively smaller decreases in potential area. Unsurprisingly, it seems that the algorithm prioritizes uncorrelated factors for the selection of primary climatic variables.

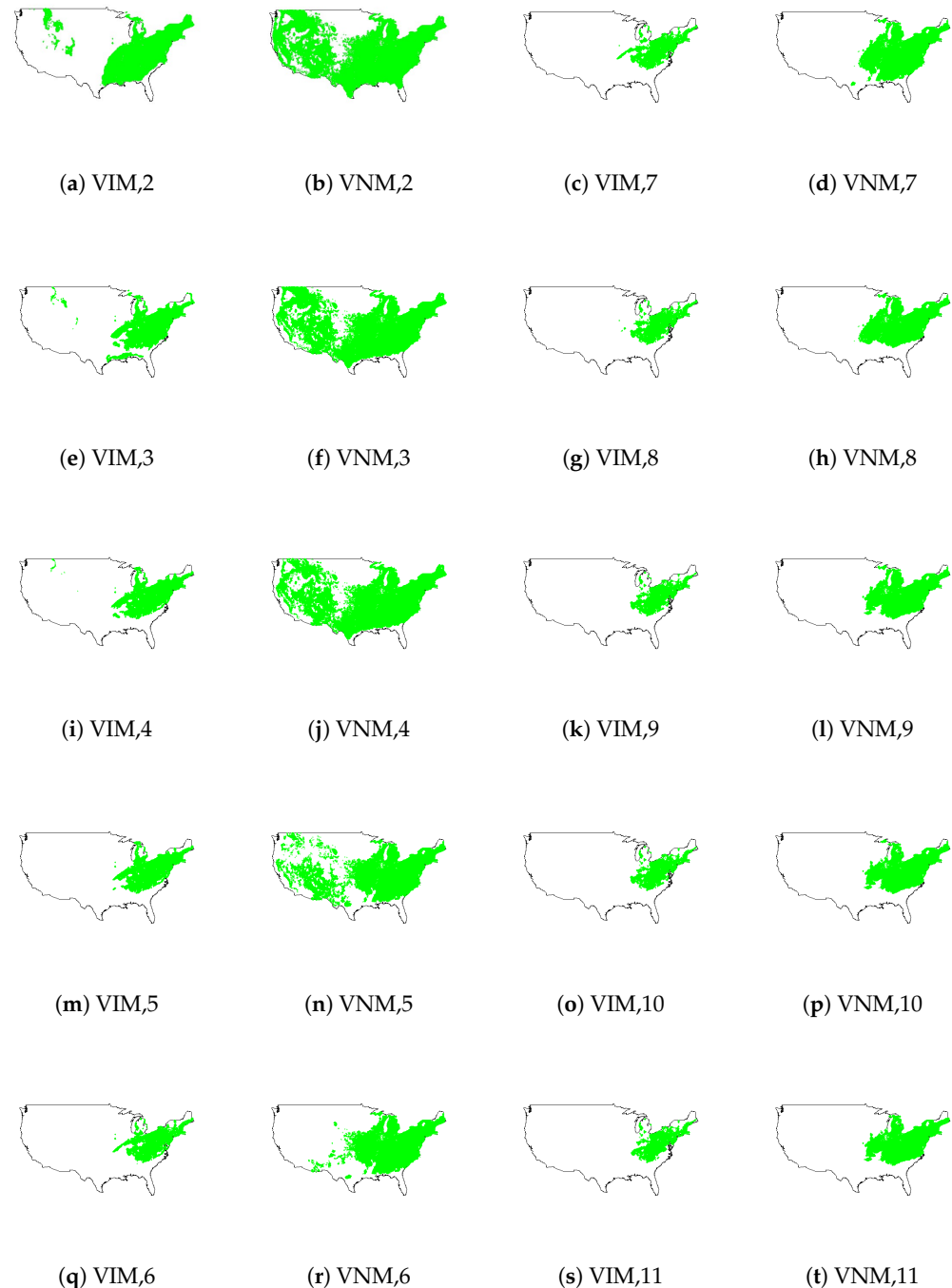


Figure 15. Potential distribution for pitch pine in the USA based on Variable Interaction Models (VIMs) and Variable Non-Interaction Models (VNMs) using sets of variables ranging from 2 to 11 (VIM2-11 and VNM2-11). Climatic variables (BIO1–BIO19) are arranged by Shapley values (Figures 11 and 12). The areas of each potential distribution are provided in Figures 13 and 14. Distribution maps for VIM12-19 and VNM12-19 are presented in Figure 16.

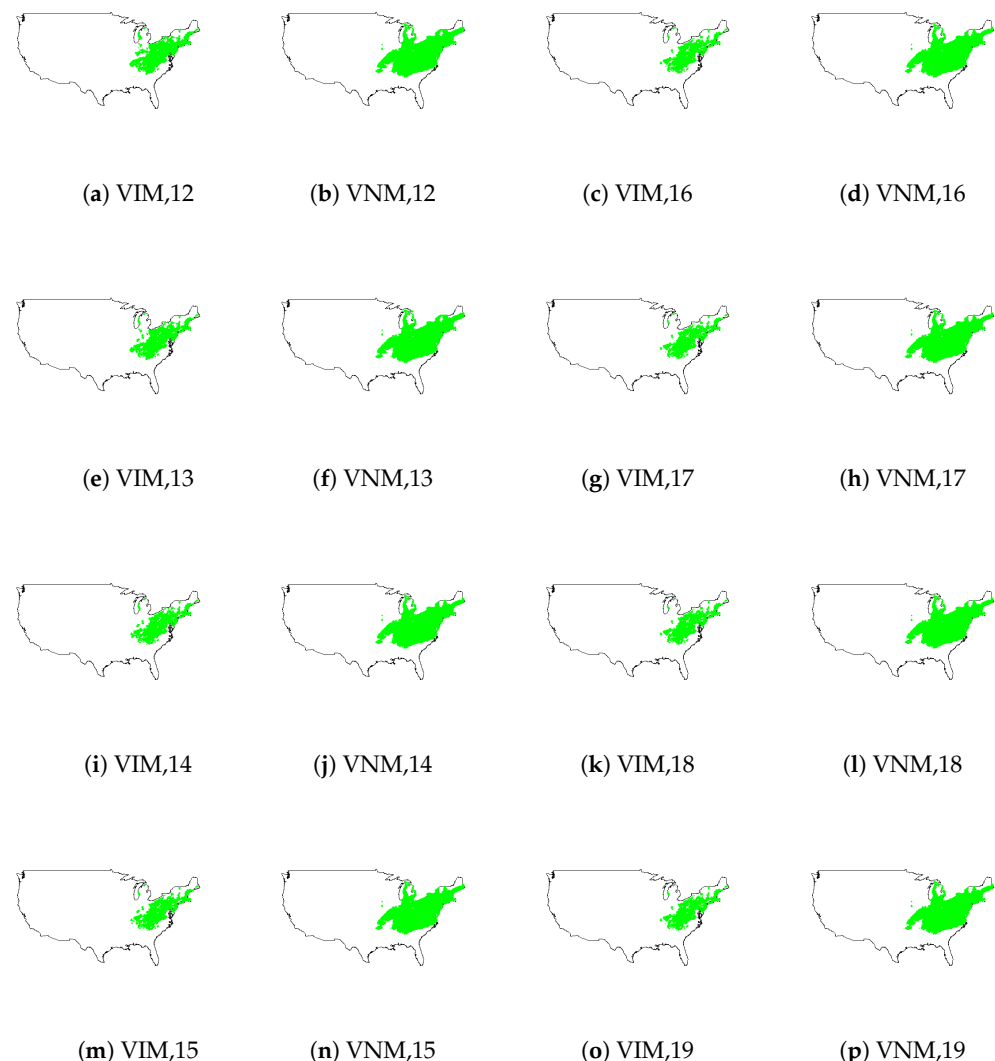


Figure 16. Potential distribution for pitch pine in the USA based on Variable Interaction Models (VIMs) and Variable Non-Interaction Models (VNMs) using sets of variables ranging from 12 to 19 (VIM12–19 and VNM12–19). Climatic variables (BIO1–BIO19) are arranged by Shapley values (Figures 11 and 12). The areas of each potential distribution are provided in Figures 13 and 14. Distribution maps for VIM2–11 and VNM2–11 are presented in Figure 15.

The comparison with PCA analysis reveals some similarities, despite the complexity of PCA outcomes, where many variables contribute almost equally to the leading principal components. Nevertheless, we attempt to identify the variables that contribute the most to the first five principal components (BIO17, BIO8, BIO2, BIO9, and BIO15) and compare them with the first variables selected by VIM (BIO15, BIO14, BIO9, BIO11, BIO1). This comparison results in two common variables (BIO15 and BIO9). Notably, both the sixth and seventh variables, namely BIO8 and BIO17, are included among the primary contributors to the top five principal components.

VIM manifests a notable proficiency in forecasting the distribution of pitch pine while concurrently facilitating the prioritization of climatic factors and dimensionality reduction within the system. Through a dimensional analysis involving PCA, VIM, and VNM, it is apparent that these systems may be considered five–six-dimensional. This estimation holds significance, particularly in addressing concerns of overfitting when constructing models with multiple dimensions. Multidimensional statistical models entail

a trade-off, necessitating the utilization of all available information to prevent underfitting, while avoiding the construction of an overly fitted model that functions effectively solely for a specific dataset. The convergence dynamics of VIM and VNM as summarized in Figures 13 and 14, indicate the potential for overfitting after the inclusion of the initial five–seven factors. However, this potential overfitting does not significantly impact the model predictions, as the predictions remain unchanged after 14 steps. The substantial contrast between VIM and VNM underscores the critical importance of the combined effects of diverse climatic factors, highlighting the inefficacy of models built on averaged aggregated climatic characteristics.

4. Discussion

4.1. Pitch Pine: Ecology and Modeling

To evaluate the efficacy of our species distribution modeling (SDM) methodology, we seek a tree species meeting specific criteria: (1) the tree species must exhibit a distribution confined exclusively within the contemporary United States, encompassing regions entirely covered by the Forest Inventory and Analysis (FIA) dataset; (2) the tree should function as a dominant species within certain habitats; and (3) the species should lack significant commercial or decorative value, precluding intentional cultivation or selective harvesting. Consequently, its range mirrors native habitats and is minimally impacted by silvicultural practices and land management activities. The selection of pitch pine aligns closely with our specified criteria, rendering it virtually ideal for our study.

The primary habitat of pitch pine (*Pinus rigida*) is predominantly situated in the eastern region of the United States, spanning from Southern Maine to Northern Georgia and Alabama, with sporadic occurrences in the Midwest. Although the native range of pitch pine is predominantly concentrated in North America, its distribution extends marginally into Canada. In Canada, pitch pine is primarily localized in Southern Ontario and Quebec, particularly around the vicinity of Lake Erie in environments characterized by well-drained sandy soils and ecosystems susceptible to fires [28]. The Canadian distribution, in contrast to its prevalence in the United States, is noticeably constrained and geographically concentrated. Notably, our model disregards this smaller Canadian population in its assessments.

The ecology of pitch pine is closely tied to its native habitat in the Eastern United States [16,19,28]. This species is well adapted to thrive in a variety of challenging environmental conditions and plays a significant role in the ecosystems where it occurs [18,29]. Pitch pines are medium-sized trees; typically, they attain maximum heights of approximately 20 m. They prefer well-drained, sandy soils and can tolerate nutrient-poor, acidic soils [30]. Their ability to grow in nutrient-poor soils also reduces competition from other tree species that may require more fertile soil conditions. Their range includes regions with a history of frequent wildfires, and they have adapted to these conditions, more specifically, the cones of pitch pines typically remain closed until exposed to the high temperatures of a fire [16,20,29]. This heat triggers the cones to open and release seeds, allowing for post-fire regeneration. Fires also help control competing vegetation, clear the forest floor, and create open areas for new seedlings to establish [22,29]. The dominance of pitch pine in certain ecosystems is a result of its unique ecological adaptations, and its ability to persist and regenerate after wildfires [30,31]. Pitch pine often serves as a dominant species in a range of fire-dependent ecosystems, encompassing pine barrens, sand plains, coastal dunes, early successional habitats, heathlands, and savannas.

Pitch pine has held historical economic significance, although its commercial value for timber and resin production has waned. Presently, its ecological and conservation importance, alongside its role in preserving specialized ecosystems, supersedes its economic utility [19,21,29,30]. Since the 1950s, pitch pine has been used to a limited extent in forest plantations in Korea [31,32]. The term “pitch pine” originates from the historical application of its resinous sap, referred to as “pitch”, prized for its diverse uses, notably in waterproofing and preserving wooden materials in shipbuilding. Previously, pitch pine wood was harvested for shipbuilding, construction, and fuel due to its renowned durability

and resistance to decay, ideal for outdoor applications. However, the ascendancy of other pine species with larger, straighter trunks in the commercial timber market led to a decline in the economic significance of pitch pine. Its ornamental use is also exceedingly rare in landscaping. Nevertheless, certain anthropogenic factors continue to impact its range. Historical land use practices like urbanization, agriculture, and fire suppression have notably diminished the extent of pitch pine habitats [30]. Fire suppression in particular has disrupted the natural fire regime crucial for pitch pines, leading to a decline in suitable habitats [21]. These factors should indeed be considered when evaluating the validity of modeling results. Nonetheless, the distribution of pitch pine, as a non-commercial species, remains primarily confined to its native habitat, making it an advantageous option for species distribution modeling.

4.2. General Discussion

The well-known limitations of mechanistic and correlative SDMs have led to extensive efforts to evaluate these model types. This evaluation often involves inter-comparisons utilizing well-studied model species [4,33]. Occasionally, researchers seek to integrate both approaches, aiming to develop hybrid models that alleviate the constraints associated with each model type [34]. In the context of SDMs that primarily consider climatic factors as predictors, often referred to as climate envelope models, the conventional choice for primary factors typically encompasses average temperature and precipitation [12,13,35]. However, a heightened degree of complexity arises when SDMs establish connections between species abundance and a multitude of climatic factors, leading to the notable issue of multicollinearity among these factors, which substantially exacerbates problems related to model overfitting [13].

The selection of average temperature and precipitation as primary climatic factors in SDMs is influenced by both historical conventions and practical considerations related to modeling convenience [13]. The roots of climate–vegetation modeling trace back to 19th-century biogeographical studies, with notable contributions from Alexander von Humboldt and Aimé Bonpland [36], and the pioneering climate–vegetation classification by Wladimir Köppen, developed between 1884 and 1936 [37–39]. Early studies recognized the complexity of climate–vegetation interactions. The initial Köppen classification relied solely on average temperature and precipitation [37]; however, subsequent refinements highlighted the importance of incorporating inter-seasonal changes for effective climate–vegetation classification [38,39]. This foundational work has been expanded upon [40–43] and applied in delineating US ecoregions [44]. The later developed Holdridge system [45–47] incorporates three primary variables: precipitation, biotemperature (annual average temperature adjusted according to the vegetation period’s duration), and the potential evapotranspiration ratio. While the intricate nature of climate–vegetation systems is widely acknowledged, many modern advancements, including Climate Envelope Models (CEMs), Correlative and Mechanistic Species Distribution Models (SDMs), and Dynamic Global Vegetation Models (DGVMs), still rely on averaged macroscopic climatic parameters [12,34].

5. Conclusions

Our original modeling methodology serves two main objectives: firstly, it facilitates the examination of the individual impact of each climatic variable on the distribution of pitch pine. Secondly, it enables the assessment of the significance of interactions between various climatic factors for this species and isolates the most influential ones. We compare two climate envelope modeling methods, the Variable Interaction Model (VIM) and the Variable Non-Interaction Model (VNM), to evaluate their effectiveness in ranking the 19 climatic factors. Both the VNM and VIM methodologies involve the utilization of a factor-ranking approach based on Shapley values during their construction processes. The VNM assumes independent effects of each climatic factor on species distribution, forming a hyper-rectangle-shaped climate envelope. In contrast, the VIM accounts for interactions between climatic factors, creating a complex, data-driven multidimensional

climate envelope. The findings unequivocally indicate a notable superiority of the Variable Interaction Model (VIM) in terms of predictive accuracy when compared to the Variable Non-Interaction Model (VNM). This underscores the significance of acknowledging the intricate structure of the climatic system and the interconnected nature of various climatic attributes that nonlinearly influence species distribution.

Author Contributions: Conceptualization, N.S.; methodology, O.R. and N.S.; software, O.R.; validation, O.R.; investigation, O.R. and N.S.; resources, N.S.; data curation, O.R.; writing—original draft preparation, O.R. and N.S.; writing—review and editing, N.S.; visualization, O.R.; supervision, N.S.; project administration, N.S.; funding acquisition, N.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Simons Foundation grant to N.S. number 965207.

Data Availability Statement: Data utilized in this study are readily accessible from the USDA Forest Service.

Acknowledgments: We would like to extend our gratitude to our colleagues John Harrison, Jean Lienard, and Keefe Koenig for their valuable suggestions and assistance in this study. Additionally, we appreciate the anonymous reviewers for providing numerous useful comments.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Calculation of Shapley Values: Numerical Example

Here is an illustrative example demonstrating the calculation of Shapley scores for VIM and VNM models. The dataset for consideration is as follows:

Plot ID	Clim_var_1	Clim_var_2	Clim_var_3	Presence
1	1	2	3	1
2	1	2	3	0
3	3	2	1	1
4	2	2	1	0
5	3	3	1	1
6	3	3	2	0
7	2	4	3	1
8	4	1	1	0
9	4	3	3	0
10	2	2	1	1

Figure A1. In the dataset: Plot ID—identifier of a forest plot; Clim_var_1, Clim_var_2 and Clim_var_3—climatic characteristics (having binned values); Presence—indicator of a tree species presence ('1'—if present; '0'—if not).

We make a combination matrix for two climatic variables. The logical-valued combination matrix is:

$$\begin{pmatrix} T & T \\ T & F \\ F & T \\ F & F \end{pmatrix}$$

Based on the combination matrix, we can see what combinations of climatic variables we need for Shapley values computation:

$$M = \begin{pmatrix} v_1, \{v_2, v_3\} & v_2, \{v_1, v_3\} & v_3, \{v_1, v_2\} \\ v_1, \{v_2\} & v_2, \{v_1\} & v_3, \{v_1\} \\ v_1, \{v_3\} & v_2, \{v_3\} & v_3, \{v_2\} \\ v_1, \{\emptyset\} & v_2, \{\emptyset\} & v_3, \{\emptyset\} \end{pmatrix}$$

Now, using matrix M , we can compute the Shapley scores.

As we know, a Shapley value of a climatic variable v_j is a scaled (by a combinatorial coefficient) difference between the model score of the ensemble of climatic variables without v_j and the model score of this ensemble including v_j .

Regarding the matrix M , its entry m_{ij} is $m_{ij} = (v_j, \{ \dots \})$, where v_j is a single climatic variable and $\{ \dots \}$ is the mentioned ensemble of climatic variables.

Appendix A.1. VIM Shapley Scores

Here, we calculate Shapley scores for the Variable Interaction Model (VIM).

Let S_{VIM} be the matrix having VIM Shapley scores: $S_{VIM} = (s_{ij})$.

The S_{VIM} matrix is calculated based on matrix M : its entry s_{ij} is the VIM Shapley score of the climatic variable v_j with respect to the ensemble of climatic variables from the entry m_{ij} of the matrix M .

For more detailed illustration of how the Shapley matrix is calculated, let us compute the entry s_{11} .

s_{11} is the VIM Shapley score of v_1 with respect to the set of climatic variables $\{v_2, v_3\}$ (see m_{11} of matrix M). So, s_{11} is a scaled difference between the VIM score of $\{v_2, v_3\}$ and VIM score of $\{v_1, v_2, v_3\}$.

Now, what are VIM scores of the sets $\{v_2, v_3\}$ and $\{v_1, v_2, v_3\}$? To answer this question, we need to look at the realized distribution (FIA plots where tree presents). In our case, the realized distribution are the plots 1, 3, 5, 7 and 10.

The VIM score of $\{v_1, v_2, v_3\}$ is the number of plots having exactly the same values of v_1, v_2 , and v_3 as in one of the plots from realized distribution:

- In plot 1, the values of v_1, v_2 , and v_3 are 1, 2, and 3 correspondingly. The second plot is the only one that has the same values, so the VIM score of $\{v_1, v_2, v_3\}$ is 2.
- In plot 3, the values of v_1, v_2 , and v_3 are 3, 2, and 1 correspondingly. It is the only plot having such values, so the VIM score of $\{v_1, v_2, v_3\}$ is $2 + 1 = 3$.
- In plot 5, the values of v_1, v_2 , and v_3 are 3, 3, and 1 correspondingly. It is the only plot having such values, so the VIM score of $\{v_1, v_2, v_3\}$ is $3 + 1 = 4$.
- In plot 7, the values of v_1, v_2 , and v_3 are 2, 4, and 3 correspondingly. It is the only plot having such values, so the VIM score of $\{v_1, v_2, v_3\}$ is $4 + 1 = 5$.
- In plot 1, the values of v_1, v_2 , and v_3 are 2, 2, and 1 correspondingly. Plot 4 is the only one that has the same values, so the VIM score of $\{v_1, v_2, v_3\}$ is $5 + 2 = 7$.

This way, we obtain that the VIM score of the collection $\{v_1, v_2, v_3\}$ is 7. Using similar considerations, we obtain that the VIM score of the set $\{v_2, v_3\}$ is 7.

Finally, s_{11} is a scaled difference between the VIM score of $\{v_2, v_3\}$ and the VIM score of $\{v_1, v_2, v_3\}$:

$$s_{11} = k(7 - 7) = 0,$$

where k is combinatorial coefficient from the definition of the Shapley score:

$$k = \frac{|S|!(n - |S| - 1)!}{n!}.$$

In our case, in total, we have three climatic variables v_1, v_2 , and v_3 , so $n = 3$, S is the set $\{v_2, v_3\}$, so the number of variables in this subset is $|S| = 2$ and

$$k = \frac{2!(3 - 2 - 1)!}{3!} = 0.$$

This way, we compute the whole Shapley matrix:

$$S_{VIM} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{pmatrix}$$

Next, the VIM Shapley score of climatic factor v_i is the average value of the i -th column of the matrix S_{VIM} . Hence, the climatic variables v_1 , v_2 , and v_3 have VIM Shapley scores (0.3, 0.7, and 0.5) correspondingly. The second climatic variable v_2 has the highest score, and therefore, v_2 plays the most important role in the VIM 'game' of predicting the potential area of a species.

Appendix A.2. VNM Shapley Scores

In this section, we calculate the Shapley scores for the Variable Non-Interaction Model (VNM).

Let S_{VNM} be a Shapley matrix with VNM Shapley scores: $S_{VNM} = (s_{ij})$.

The S_{VNM} matrix is also based on matrix M: its entry s_{ij} is the VNM Shapley score of the climatic variable v_j with respect to the ensemble of climatic variables from the entry m_{ij} .

Next, we compute entry s_{11} , which is a scaled difference between the VNM score of $\{v_2, v_3\}$ and the VNM score of $\{v_1, v_2, v_3\}$.

The VNM score of the set $\{v_1, v_2, v_3\}$ is the number of plots having the following property: v_1 , v_2 , and v_3 must be the same as in some of the plots from the realized distribution. In our case, v_1 must be 1 or 2 or 3; v_2 must be 2 or 3 or 4; and v_3 must be 1 or 3. The number of plots with this property is 7, so the VNM score of $\{v_1, v_2, v_3\}$ is 7. Similarly, we obtain that the VNM score of $\{v_2, v_3\}$ is 8.

Finally, s_{11} is a scaled difference between the VNM score of $\{v_2, v_3\}$ and the VNM score of $\{v_1, v_2, v_3\}$:

$$s_{11} = k(8 - 7) = 0,$$

where k is combinatorial coefficient calculated above (see the section about the VIM Shapley matrix); $k = 0$. This way, we can compute the whole VNM Shapley matrix:

$$S_{VNM} = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Next, the VNM Shapley score of climatic variable v_i is the average value of the i -th column of matrix S_{VNM} . Hence, climatic variables v_1 , v_2 , and v_3 have VNM Shapley scores (0.5, 0.3, and 0.7) correspondingly. The third climatic variable v_3 has the highest score, and therefore, v_3 plays the most important role in the VNM 'game' of predicting the potential area of a tree species.

References

1. Austin, M. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecol. Model.* **2007**, *200*, 1–19. [\[CrossRef\]](#)
2. Elith, J.; Leathwick, J.R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* **2009**, *40*, 677–697. [\[CrossRef\]](#)
3. Loiselle, B.A.; Howell, C.A.; Graham, C.H.; Goerck, J.M.; Brooks, T.; Smith, K.G.; Williams, P.H. Avoiding pitfalls of using species distribution models in conservation planning. *Conserv. Biol.* **2003**, *17*, 1591–1600. [\[CrossRef\]](#)
4. Buckley, L.B.; Urban, M.C.; Angilletta, M.J.; Crozier, L.G.; Rissler, L.J.; Sears, M.W. Can mechanism inform species' distribution models? *Ecol. Lett.* **2010**, *13*, 1041–1054. [\[CrossRef\]](#)
5. Merow, C.; Smith, M.J.; Edwards, T.C., Jr.; Guisan, A.; McMahon, S.M.; Normand, S.; Thuiller, W.; Wüest, R.O.; Zimmermann, N.E.; Elith, J. What do we gain from simplicity versus complexity in species distribution models? *Ecography* **2014**, *37*, 1267–1281. [\[CrossRef\]](#)
6. Araújo, M.B.; Guisan, A. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* **2006**, *33*, 1677–1688. [\[CrossRef\]](#)
7. Zimmermann, N.E.; Edwards, T.C., Jr.; Graham, C.H.; Pearman, P.B.; Svenning, J.C. New trends in species distribution modelling. *Ecography* **2010**, *33*, 985–989. [\[CrossRef\]](#)
8. Dormann, C.F.; Schymanski, S.J.; Cabral, J.; Chuine, I.; Graham, C.; Hartig, F.; Kearney, M.; Morin, X.; Römermann, C.; Schröder, B.; et al. Correlation and process in species distribution models: Bridging a dichotomy. *J. Biogeogr.* **2012**, *39*, 2119–2131. [\[CrossRef\]](#)
9. Wisz, M.S.; Hijmans, R.; Li, J.; Peterson, A.T.; Graham, C.; Guisan, A.; NCEAS Predicting Species Distributions Working Group. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **2008**, *14*, 763–773. [\[CrossRef\]](#)
10. Kelling, S.; Hochachka, W.; Fink, D.; Riedewald, M.; Caruana, R.; Ballard, G.; Hooker, G. Data-intensive Science: A New Paradigm for Biodiversity Studies. *BioScience* **2009**, *59*, 613–620. [\[CrossRef\]](#)

11. Michener, W.K.; Jones, M.B. Ecoinformatics: Supporting ecology as a data-intensive science. *Trends Ecol. Evol.* **2012**, *27*, 85–93. [[CrossRef](#)]
12. Liénard, J.; Harrison, J.; Strigul, N. US forest response to projected climate-related stress: A tolerance perspective. *Glob. Chang. Biol.* **2016**, *22*, 2875–2886. [[CrossRef](#)]
13. Rumyantseva, O.; Strigul, N. Data-Driven Analysis of Forest–Climate Interactions in the Conterminous United States. *Climate* **2021**, *9*, 108. [[CrossRef](#)]
14. Rumyantseva, O.; Sarantsev, A.; Strigul, N. Time series analysis of forest dynamics at the ecoregion level. *Forecasting* **2020**, *2*, 20. [[CrossRef](#)]
15. Liénard, J.; Florescu, I.; Strigul, N. An Appraisal of the Classic Forest Succession Paradigm with the Shade Tolerance Index. *PLoS ONE* **2015**, *10*, e0117138. [[CrossRef](#)] [[PubMed](#)]
16. Little, S.; Garrett, P.W. *Pinus rigida* Mill. pitch pine. *Silvics N. Am.* **1990**, *1*, 456–462.
17. Williams, C.E. History and status of Table Mountain pine–pitch pine forests of the southern Appalachian Mountains (USA). *Nat. Areas J.* **1998**, *18*, 81–90.
18. Bernard, J.M.; Seischab, F.K. Pitch pine (*Pinus rigida* Mill.) communities in northeastern New York state. *Am. Midl. Nat.* **1995**, *134*, 294–306. [[CrossRef](#)]
19. Motzkin, G.; Patterson Iii, W.; Foster, D.R. A historical perspective on pitch pine–scrub oak communities in the Connecticut Valley of Massachusetts. *Ecosystems* **1999**, *2*, 255–273. [[CrossRef](#)]
20. Brose, P.H.; Waldrop, T.A. Fire and the origin of Table Mountain pine pitch pine communities in the southern Appalachian Mountains, USA. *Can. J. For. Res.* **2006**, *36*, 710–718. [[CrossRef](#)]
21. Grand, J.; Buonaccorsi, J.; Cushman, S.A.; Griffin, C.R.; Neel, M.C. A multiscale landscape approach to predicting bird and moth rarity hotspots in a threatened pitch pine–scrub oak community. *Conserv. Biol.* **2004**, *18*, 1063–1077. [[CrossRef](#)]
22. Parshall, T.; Foster, D.R.; Faison, E.; MacDonald, D.; Hansen, B. Long-term history of vegetation and fire in pitch pine–oak forests on cape cod, massachusetts. *Ecology* **2003**, *84*, 736–748. [[CrossRef](#)]
23. Gibson, J.; Moisen, G.; Frescino, T.; Edwards, T.C., Jr. Using publicly available forest inventory data in climate-based models of tree species distribution: Examining effects of true versus altered location coordinates. *Ecosystems* **2014**, *17*, 43–53. [[CrossRef](#)]
24. Hair, J.E.; Anderson, R.E.; Tatham, R.L.; Black, W.C. *Multivariate Data Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
25. Roth, A.E. Introduction to the Shapley value. *Shapley Value* **1988**, *1*, 1–27.
26. Winter, E. The shapley value. *Handb. Game Theory Econ. Appl.* **2002**, *3*, 2025–2054.
27. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [[CrossRef](#)] [[PubMed](#)]
28. Meilleur, A.; Brisson, J.; Bouchard, A. Ecological analyses of the northernmost population of pitch pine (*Pinus rigida*). *Can. J. For. Res.* **1997**, *27*, 1342–1350. [[CrossRef](#)]
29. Welch, N.; Waldrop, T.A.; Buckner, E. Response of southern Appalachian table mountain pine (*Pinus pungens*) and pitch pine (*P. rigida*) stands to prescribed burning. *For. Ecol. Manag.* **2000**, *136*, 185–197. [[CrossRef](#)]
30. Jordan, M.J.; Patterson, W.A., III; Windisch, A.G. Conceptual ecological models for the Long Island pitch pine barrens: Implications for managing rare plant communities. *For. Ecol. Manag.* **2003**, *185*, 151–168. [[CrossRef](#)]
31. Ledig, F.T.; Smouse, P.E.; Hom, J.L. Postglacial migration and adaptation for dispersal in pitch pine (Pinaceae). *Am. J. Bot.* **2015**, *102*, 2074–2091. [[CrossRef](#)]
32. Hwang, J.; Son, Y. Short-term effects of thinning and liming on forest soils of pitch pine and Japanese larch plantations in central Korea. *Ecol. Res.* **2006**, *21*, 671–680. [[CrossRef](#)]
33. Kearney, M.R.; Wintle, B.A.; Porter, W.P. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conserv. Lett.* **2010**, *3*, 203–213. [[CrossRef](#)]
34. Talluto, M.V.; Boulangeat, I.; Ameztegui, A.; Aubin, I.; Berteaux, D.; Butler, A.; Doyon, F.; Drever, C.R.; Fortin, M.J.; Franceschini, T.; et al. Cross-scale integration of knowledge for predicting species ranges: A metamodeling framework. *Glob. Ecol. Biogeogr.* **2016**, *25*, 238–249. [[CrossRef](#)] [[PubMed](#)]
35. Whittaker, R. *Communities and Ecosystems*; Current Concepts in Biology; Macmillan Company, Collier-Macmillan Limited: London, UK, 1970.
36. von Humboldt, A.; Bonpland, A. *Essai sur la Géographie des Plantes*; Chez Levrault, Schoell et Compagnie: Paris, France, 1805.
37. Köppen, W. Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet. *Meteorol. Z.* **1884**, *1*, 5–226.
38. Köppen, W. Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geogr. Z.* **1900**, *6*, 593–611.
39. Köppen, W. Klassifikation der Klima nach Temperatur, Niederschlag und Jahreslauf. *Petermanns Geogr. Mitteilungen* **1918**, *64*, 193–203.
40. Geiger, R.; Pohl, W. Eine neue Wandkarte der Klimagebiete der Erde nach W. Köppens Klassifikation (A New Wall Map of the Climatic Regions of the World According to W. Köppen's Classification). *Erdkunde* **1954**, *8*, 58–61.
41. Rubel, F.; Kottek, M. Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen–Geiger climate classification. *Meteorol. Z.* **2010**, *19*, 135–141. [[CrossRef](#)]
42. Trewartha, G.; Horn, L. *An Introduction to Climate*, 5th ed.; McGraw-Hill Book Co.: New York City, NY, USA, 1980.

43. Belda, M.; Holtanová, E.; Halenka, T.; Kalvová, J. Climate classification revisited: From Köppen to Trewartha. *Clim. Res.* **2014**, *59*, 1–13. [[CrossRef](#)]
44. Bailey, R.G. *Ecosystem Geography: From Ecoregions to Sites*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
45. Holdridge, L.R. Determination of world plant formations from simple climatic data. *Science* **1947**, *105*, 367–368. [[CrossRef](#)]
46. Holdridge, L.R. *Life Zone Ecology*, Rev. ed.; Tropical Science Center: San Jose, Costa Rica, 1967.
47. Lugo, A.E.; Brown, S.L.; Dodson, R.; Smith, T.S.; Shugart, H.H. The Holdridge life zones of the conterminous United States in relation to ecosystem mapping. *J. Biogeogr.* **1999**, *26*, 1025–1038. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.