Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Article

Land suitability for establishing espadín agave plantations in the state of Oaxaca

Ramiro Pérez-Miranda[1]
Antonio González-Hernández[1]
Martín Enrique Romero-Sánchez[2,§]
Efraín Velasco-Bautista[1]
Fernando Carrillo-Anzures[2]
Miguel Acosta-Mireles[2]

1 CENID-COMEF INIFAP Ave. Progreso, Alcaldía Coyoacán, Ciudad de México, 04010 México

2 Campo Experimental Valle de México INIFAP Carretera Los Reyes-Texcoco km 13.5, Coatlinchán, Texcoco, Estado de México, 56250 México

Autor para correspondencia: romero.martin@inifap.gob.mx .

Abstract

The growth in mezcal consumption has created an oversupply of agave, reducing prices for producers. This study evaluated land suitability for planting *Agave angustifolia* Haw. in Oaxaca using two distribution models: MaxEnt and Random Forest. Climatic, edaphological and topographic variables were used, along with records of the species' presence. Both models showed AUC values >0.85, with 16 relevant variables. The area with high potential was 4 475.89 km^2(MaxEnt) and 28 700.97 km^2(Random Forest); the ideal areas for plantations were 6 671.12 $km^2$and 9 618.67 km^2, respectively. Despite the differences, the models provide useful information for the sustainable planning of espadín agave.

Keywords:

Agave angustifolia Haw., MaxEnt, potential distribution, Random Forest.

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Introduction

Of the 211 agave species worldwide, 159 are found in Mexican territory, representing 75% of the species distributed on the planet (García-Mendoza, 2018). Agaves grow preferentially in the arid and semi-arid climates of central and northern Mexico; they are plants that tolerate low rainfall and high temperatures; however, they also grow, to a lesser extent, in temperate zones at altitudes ranging from zero to 3 400 masl (García-Mendoza, 2007).

The state of Oaxaca has the greatest diversity of agaves in Mexico; 35 of the 159 species reported in the country are found there (García-Mendoza, 2018); likewise, it is considered the largest producer of mezcal. Therefore, in order to achieve sustainable ecological and economic management of agave production and reduce anthropogenic pressure on wild populations, it is necessary to identify the ideal areas for establishing espadín maguey plantations in Oaxaca (Torres *et al*., 2013).

Species distribution models identify suitable areas under different environmental conditions. Nonetheless, the results vary with the algorithm, producing uncertainty in the delimitation of crop areas and territorial decisions (Serra *et al* ., 2025).

The most commonly used models are MaxEnt and Random Forest, with key differences: MaxEnt maximizes entropy based on presence data and environmental variables, whereas Random Forest uses decision trees to model complex, nonlinear relationships (Gutiérrez-Hernández *et al*., 2018). These differences yield opposing suitability estimates, with spatial underestimation or overestimation. Comparing them reduces uncertainty and provides robust information for sustainable planning of *Agave angustifolia* cultivation (Haile *et al*., 2024).

This work aimed to compare the territorial suitability for establishing *Agave angustifolia* in Oaxaca using the MaxEnt and Random Forest algorithms and to evaluate the spatial discrepancies between the two models and their implications for the sustainable planning of the crop. It is proposed that the choice of algorithm can significantly alter the estimate of suitable areas, generating uncertainty in decision-making related to territorial planning and the productive expansion of mezcal agave.

Materials and methods

Study area

The study was conducted in the state of Oaxaca, which is bordered to the north by Puebla and Veracruz, to the south by the Pacific Ocean, to the east by Chiapas, and to the west by Guerrero. Its extreme coordinates are 98° 33’ 9” W - 93° 52’ 4” W and 18° 39’ 32” N - 15° 39’ 25” N (INEGI, 2018) (Figure 1).

Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Figure 1. Location of the study area.

Data on the presence of the mezcal agave species

A database was compiled with 775 records of *A. angustifolia* presence from GBIF (2023) and field observations. The records used corresponded to *Agave angustifolia* Haw. in a broad sense, because the available databases did not allow consistent differentiation of the infraspecific varieties. Data cleansing included removing duplicates, erroneous coordinates, records located in water bodies or urban areas, and observations with inconsistent geographic information. Subsequently, the records were reviewed to ensure spatial and environmental coherence before modeling.

Edaphoclimatic variables

Twenty-two edaphoclimatic variables, adjusted to the same extent and spatial resolution (~900 m), were used. The 19 climatic variables come from WorldClim, while the nine edaphic variables are obtained from SoilGrids (Table 1).

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Table 1. Edaphological variables.

Variables	Description
Bdod	Bulk density of the fine earth fraction
Cec	Cation exchange capacity
Clay	Proportion of clay particles (<0.002 mm) in the fine earth fraction
Nitro	Total nitrogen (N)
Ph	Soil pH
Sand	Proportion of sand particles (>0.05 mm) in the fine earth fraction
Silt	Proportion of silt particles (>0.002 mm and #0.05 mm) in the fine earth fraction
Soc	Soil organic carbon content in the fine earth fraction
Mostpro	International Soil Classification System

The Lang humidity index (Sánchez and Garduño, 2008) was also included. The topographic variables (altitude, slope and orientation) were obtained from the National Digital Elevation Model (INEGI, 2018). The Pearson correlation test was applied and multicollinearity was assessed using NicheToolBox (Osorio *et al*., 2020) in RStudio 4.3, with a threshold of 85%.

Spatial modeling

MaxEnt

MaxEnt 3.4.4 was used, with logistic output and cross-validation (25% for testing, 10 000 background points). Low-contribution variables were iteratively removed. The final model was classified into four levels of suitability: null (0-0.25), low (0.25-0.50), medium (0.50-0.75), and high (0.75-1). The intervals were defined based on equivalent ranges of the models' logistic outputs in order to standardize the spatial comparison between algorithms and facilitate the ecological interpretation of the results.

Random Forest

The CART algorithm was implemented in RStudio 4.3, using the Gini index to select attributes. Irrelevant variables were removed, and the model was evaluated using the area under the curve (AUC) and the Kappa index. The categories generated were the same as in MaxEnt.

For the modeling process, the territory of the state of Oaxaca was defined as the accessible area (M), as this region encompasses the historical and current distribution of *Agave angustifolia* , which is associated with traditional and commercial mezcal production systems. The delimitation was ecologically justified by the existing environmental continuity in the state, the extensive documented presence of the species, and the climatic and edaphic conditions compatible with its establishment and management. This definition allowed the modeling to be restricted to the space potentially accessible to the species, reducing the risk of extrapolation to ecologically irrelevant environments.

Model validation

Validation was performed using comparable metrics across models to ensure uniformity in evaluating predictive performance (Mi *et al*., 2017). A total of 100 sites (georeferenced via Garmin™ 65sx GPS) were selected through spatial random sampling in the central region of the state of Oaxaca, considering field accessibility and the representation of the different suitability categories predicted by the models.

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Identification of suitable areas for agave plantations

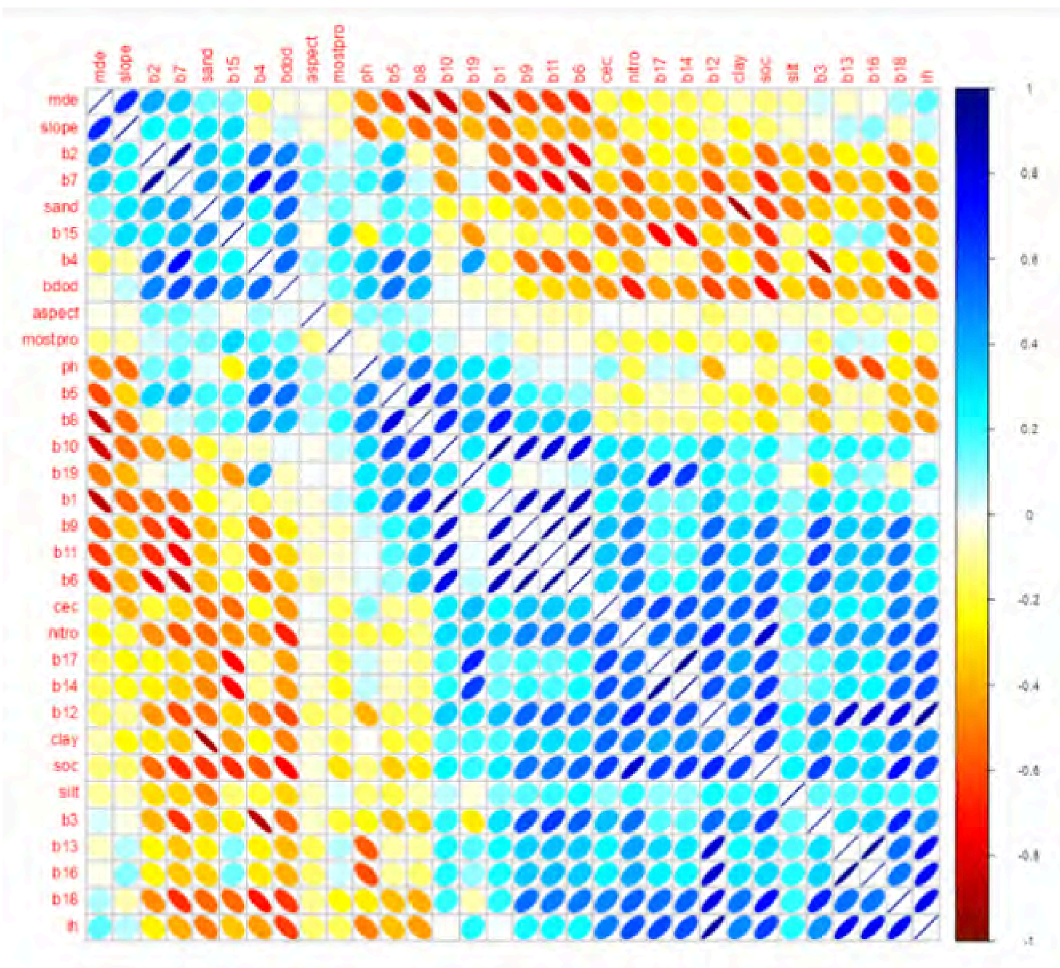
Areas classified as having medium or high ecological suitability by both models were considered ideal areas, as they represent favorable environmental conditions for the potential establishment of *Agave angustifolia* . In order to reduce spatial overestimation errors and increase the territorial applicability of the results, these areas were subsequently intersected with land-use layers compatible with agricultural and restoration activities. The delimitation was based on layers from INEGI (2018) , SNIGF-CONAFOR, and CONANP (2023) , processed in ArcMap™ 10.5.

Results and discussion

Pearson correlation

The Pearson correlation test identified 21 non-redundant variables to model the distribution of *A. angustifolia* , with correlation coefficients ranging from -0.85 to 0.85 (p< 0.0001), which allowed us to remove highly collinear variables (Dormann *et al*., 2013) (Figure 2).

Figure 2. Pearson correlation tests among edaphoclimatic variables.



DOI: https://doi.org/10.29312/remexca.v17i5.4002

Model accuracy metrics

The AUC values were 0.898 for MaxEnt and 0.931 for Random Forest, which are considered 'good' and 'excellent', respectively (Romo *et al* ., 2013). Both models showed adequate predictive power based on the AUC values obtained; however, this metric was interpreted only as a general indicator of performance, not as conclusive evidence of superiority between the algorithms.

Table 2 presents the contributions of the edaphoclimatic variables used to generate the potential distribution models of espadín maguey, using MaxEnt and Random Forest.

Table 2. Contribution of each variable to the modeling of *A. angustifolia* .

No.	MaxEnt		Random Forest	
	Variables	(%)	Variables	(%)
1	B13	35.5	B1	21.6
2	B1	27	B12	11.7
3	B3	11.3	Ph	11.5
4	B19	4.8	B8	10.2
5	Ph	4	B13	9
6	B15	3.3	Nitro	7
7	B18	2.9	B3	6.5
8	B14	2.4	B5	5.6
9	Silt	1.8	Mostpro	5
10	B8	1.6	B14	3.7
11	Nitro	1.4	B18	3.5
12	Mostpro	1.3	B19	2.5
13	Clay	1	Clay	1.1
14	Slope	0.8	B15	0.6
15	B12	0.6	Slope	0.3
16	B5	0.3	Silt	0.2

Five variables with low contribution (<5%) were discarded. In MaxEnt, climatic variables related to precipitation and temperature, particularly B13, B1, B3 and B19, had the greatest influence, suggesting that the algorithm primarily restricts the potential distribution of *Agave angustifolia* to specific climatic conditions.

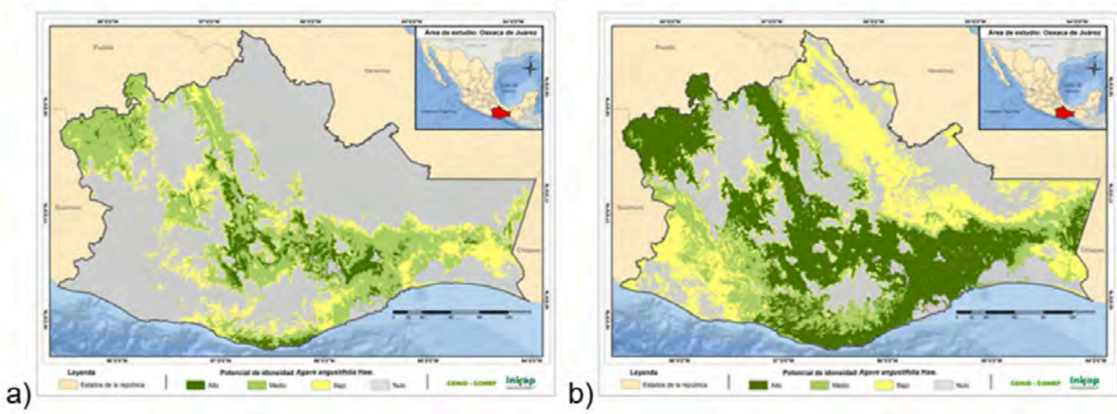
In contrast, Random Forest assigned greater importance not only to climatic variables but also to edaphic factors such as soil pH and nitrogen, indicating greater sensitivity to local environmental conditions and to complex interactions among variables.

Spatial distribution of espadín agave

MaxEnt estimated that 39.3% of the state was suitable, but only 4.77% had high suitability. On the other hand, Random Forest indicated a suitable area of 70.2%, of which 30.6% was classified as highly suitable (Figure 3).

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Figure 3. Distribution of the potential of *A. angustifolia* in the state of Oaxaca: a) MaxEnt; and b) Random Forest.



These differences reflect the greater sensitivity of Random Forest, as supported by other studies (Chunrong *et al*., 2017; Valavi *et al*., 2021). In the 'high' category, Random Forest identified 25.8% more area than MaxEnt.

The results of the MaxEnt model in Oaxaca highlight that high suitability is primarily concentrated in the central region of the state, in the southeast and, to a lesser extent, in the south. In contrast, Random Forest indicates high suitability across a broad area throughout the state (Table 3).

Table 3. Areas by suitability categories for *A. angustifolia* .

Potential	MaxEnt		Random Forest	
	(km^2)	(%)*	(km^2)	(%)*
High	4 475.89	4.77	28 700.97	30.61
Medium	20 890.95	22.28	14 870.79	15.86
Low	11 516.76	12.28	22 252.72	23.73
Total	36 883.61	39.34	65 824.49	70.21

*= % with respect to the total area covered by the state (93 757.6 km^2).

Based on Table 3 , the MaxEnt model estimated that approximately 40% of the state of Oaxaca is suitable for cultivating *Agave angustifolia* ; however, only 4.77% of this area (4 475.89 km^2) was classified as highly suitable. In contrast, the Random Forest model identified 70.21% of the state territory as suitable, of which 30.61% corresponds to areas of high suitability.

These results show notable differences between the two algorithms in the area potentially suitable for cultivating mezcal agave. The differences observed between MaxEnt and Random Forest can be explained by the methodological characteristics of each algorithm. MaxEnt uses a maximum-entropy-based approach that restricts predictions to environmental conditions similar to those of the presence records used during model training, resulting in more spatially conservative estimates.

In contrast, Random Forest employs multiple assembled decision trees capable of identifying complex, nonlinear relationships among climatic, edaphic, and topographic variables. This capability increases the algorithm's sensitivity to detect potentially suitable environments, which can result in a considerable expansion of areas classified as suitable.

In the present study, these methodological differences led to substantial discrepancies in the estimation of areas with high suitability for *Agave angustifolia* , particularly in environmentally intermediate or marginal regions. This suggests that the choice of algorithm is a relevant source of spatial uncertainty in studies of potential distribution (Chunrong *et al*., 2017; Valavi *et al*., 2021).

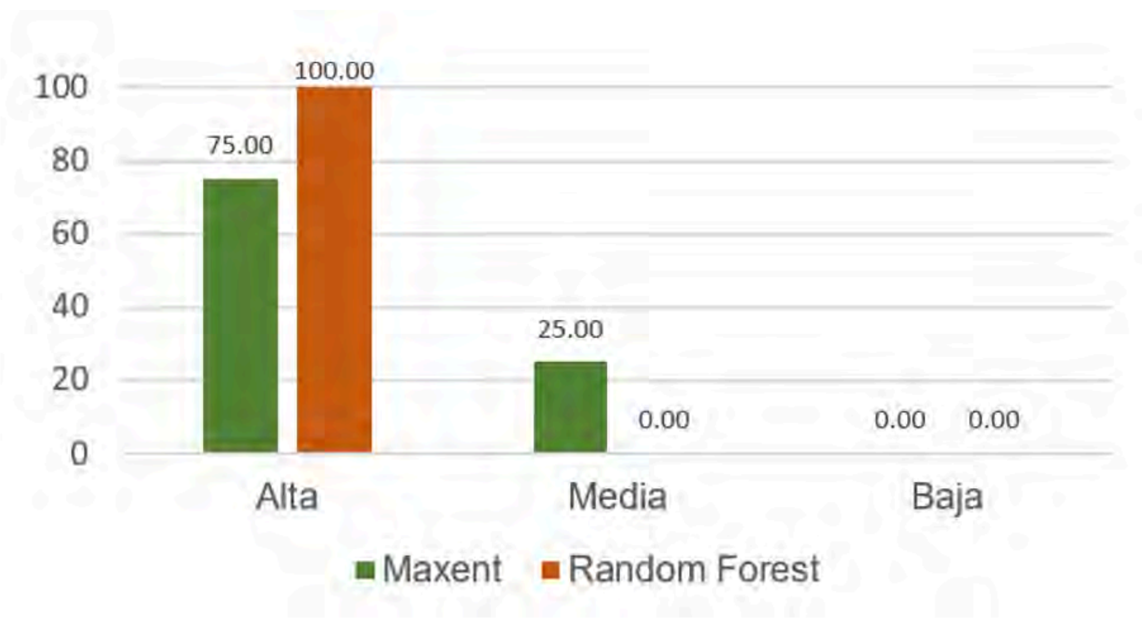
DOI: https://doi.org/10.29312/remexca.v17i5.4002

In the specific case of *A. angustifolia* , Random Forest estimated 25.84% more area as highly suitable than MaxEnt. By contrast, MaxEnt outperformed Random Forest in the medium-suitability category by 6.42%, while Random Forest estimated 11.45% more in the low category. These differences between models align with findings reported in the literature, which show that the choice of algorithm significantly influences the results.

Verification of agave suitability maps

The MaxEnt model ranked 75% of the validation points in the high-suitability category and 25.0% in the medium-suitability category. For its part, Random Forest located 100% of the verified sites in areas classified as highly suitable (Figure 4).

Figure 4. Potential distribution areas of *A. angustifolia* .



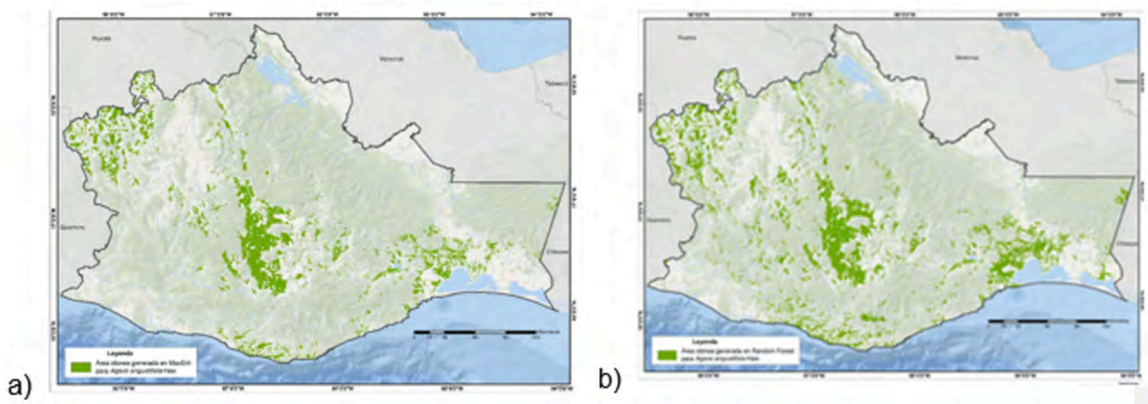
Therefore, although the validation percentages differed, both models performed well.

Identification of suitable areas for agave plantations

The maps of suitable areas for espadín maguey are presented in Figure 5 ; the derivatives of the algorithms: a) MaxEnt; and b) Random Forest.

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Figure 5. Potential area for *A. angustifolia* plantations: a) MaxEnt; and b) Random Forest.



The distribution of espadín maguey with both models is observed in a strip extending from northwest to southeast, with a wide presence in the central zone; for its part, the suitable area and its percentage relative to the state total are shown in Table 4 .

Table 4. Area suitable for establishing *A. angustifolia* in the state of Oaxaca.

Land use	MaxEnt		Random Forest	
	(km^2)	(%)*	(km^2)	(%)*
Agriculture	4 903.09	5.23	7 794.99	8.31
Pasture	1 661.85	1.77	1 698	1.81
Areas allocated for restoration	106 18	0.11	125.68	0.13
Total	6 671.12	7.12	9 618.67	10.26

*= % relative to the total area covered by the state (93 757.6 km^2).

The Random Forest model estimated that 10.25% of Oaxaca's territory is suitable for *Agave angustifolia* plantations, whereas MaxEnt estimated 7.11%. The difference between the two models amounts to 3.14 percentage points, equivalent to 2 947.55 km^2.

Due to the spatial differences observed between MaxEnt and Random Forest, the average of the two estimates was considered as an intermediate consensus approximation between models to represent more conservatively the area potentially suitable for the establishment of *Agave angustifolia* .

Conclusions

The differences observed between MaxEnt and Random Forest show that the delimitation of suitable areas for *Agave angustifolia* depends significantly on the algorithm used. This discrepancy is a relevant source of methodological uncertainty in studies of potential distribution and can directly influence processes of territorial planning and productive expansion of the crop.

While MaxEnt showed a more conservative spatial delimitation, Random Forest estimated larger areas with high suitability, likely due to its greater sensitivity to model complex environmental relationships. However, the results do not allow us to affirm an absolute superiority of one model over the other; rather, they reflect methodological differences in the way ecological suitability is represented.

DOI: https://doi.org/10.29312/remexca.v17i5.4002

Consequently, the complementary use of multiple algorithms can help generate more robust and useful territorial scenarios to support sustainable management strategies for mezcal agave in Oaxaca.

In addition to its usefulness for the territorial delimitation of suitable areas, this type of study can contribute to the design of sustainable management strategies for mezcal agave when integrated with participatory processes that involve local producers and other social actors linked to the mezcal production chain.

Bibliography

1 Chunrong, M.; Huettmann, F.; Guo, Y.; Han, X. and Wen, L. 2017. Why choose random forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ. 5:e2849. https://doi.org/10.7717/peerj.2849.

2 CONANP. 2023. Comisión Nacional de Áreas Naturales Protegidas. Información espacial de las áreas naturales protegidas. http://sig.conanp.gob.mx/website/pagsig/info-shape.htm.

3 Dormann, C. F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; García-Márquez, J. R.; Gruber, B.; Lafourcade, B.; Leitão, P. J.; Münkemüller, T.; McClean, C.; Osborne, P. E.; Reineking, B.; Schröder, B.; Skidmore, A. K.; Zurell, D. and Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36(1):27-46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

4 Falk, H.; Yumin, G.; Xuesong, H. and Lijia, W. 2017. Why choose random forests to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ. 5:e2849. https://doi.org/10.7717/peerj.2849.

5 García-Mendoza, A. J. 2007. Los agaves de México. Revista Ciencias. Universidad Nacional Autónoma de México (UNAM). 87:14-23. https://www.revistas.unam.mx/index.php/cns/article/view/12113.

6 García-Mendoza, A. J. 2018. Los agaves de Oaxaca. Boletín UNAM-DGCS-045. https://www.dgcs.unam.mx/boletin/bdboletin/2018-045.html.

7 GBIF. 2023. Global Biodiversity Information Facility. https://www.gbif.org/.

8 Gutiérrez-Hernández, O.; Artigas, R. C.; González, J. M. S. y García, L. V. 2018. Modelos predictivos en biogeografía: aplicación para la modelización de nichos ecológicos en geografía física. Boletín de La Asociación de Geógrafos Españoles. 78:88-126. https://doi.org/10.21138/bage.2395.

9 Haile, A.; Desta, A. B. and Dejene, S. W. 2024. Modeling the distribution of *Aloe ankoberensis* and *A. debrana* under different climate change scenarios in North Shewa Zone, Amhara National Regional State, Ethiopia. Ecological Processes. 13(1):39. https://doi.org/10.1186/s13717-024-00511-x.

10 Hijmans, R. J. and Graham, C. H. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology. 12(12):2272-2281. https://doi.org/10.1111/j.1365-2486.2006.01256.x.

11 INEGI. 2018. Instituto Nacional de Estadística y Geografía. Conjunto de datos vectoriales de uso del suelo y vegetación. Escala 1:250 000. Serie VII. Conjunto nacional. https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463842781.

12 Romo, B. H.; Sanabria, P. and García-Barros, S. E. 2013. Predicción de los impactos del cambio climático en la distribución sobre las especies de Lepidoptera: el caso del género Boloria Moore, 1900 en la Península Ibérica (Lepidoptera: Nymphalidae). SHILAP Revista de Lepidopterología. 41(162):267-286.

DOI: https://doi.org/10.29312/remexca.v17i5.4002

13 Sánchez, N. S. and Garduño, R. L. 2008. Algunas consideraciones acerca de los sistemas de clasificación climática. Contactos. 68:5-10. https://www.divulgameteo.es/fotos/meteoroteca/Sistemas-clasificaci%C3%B3n-clim%C3%A1tica.pdf.

14 Serra, E.; Debolini, M.; Fraga, H.; Trabucco, A.; Mereu, V. and Spano, D. 2025. Machine learning and species distribution models for crops: A review. Ecological Informatics. 93:103563. https://doi.org/10.1016/j.ecoinf.2025.103563.

15 Torres, I.; Casas, A.; Delgado-Lemus, A. and Rangel-Landa, S. 2013. Aprovechamiento, demografía y establecimiento de *Agave potatorum* en el Valle de Tehuacán, México: aportes ecológicos y etnobiológicos para su manejo sustentable. Zonas Áridas. 15(1):92-109. https://doi.org/10.21704/ZA.V15I1.110.

16 Valavi, R.; Guillera-Arroita, G.; Lahoz-Monfort, J. J. and Elith, J. 2021. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Modelling. 92(1):e1486. https://doi.org/10.1002/ecm.1486.

Land suitability for establishing espadín agave plantations in the state of Oaxaca

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
ISSN (electronic): 2007-9230
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (México)

Article/Issue Information
Date received: 01 March 2026
Date accepted: 01 June 2026
Publication date: 01 August 2026
Electronic Location Identifier: e4002
DOI: 10.29312/remexca.v17i5.4002

Categories

Subject: Article

Keywords:

Keywords:

Agave angustifolia Haw.

MaxEnt

potential distribution

Random Forest

Counts

Figures: 5

Tables: 4

Equations: 0

References: 16