Taylor & Francis
Taylor & Francis Group



Journal of Sustainable Forestry

ISSN: 1054-9811 (Print) 1540-756X (Online) Journal homepage: www.tandfonline.com/journals/wjsf20

Generalized Linear Models to Estimate the Probability of Occurrence of *Cinchona officinalis* L., *Cinchona pubescens* Vahl, and *Cinchona calisaya* Wedd in Peru

Elton Tafur, Seily Cuchca, Ligia García, Nilton B. Rojas-Briceño & Jaris Veneros

To cite this article: Elton Tafur, Seily Cuchca, Ligia García, Nilton B. Rojas-Briceño & Jaris Veneros (2025) Generalized Linear Models to Estimate the Probability of Occurrence of *Cinchona officinalis* L., *Cinchona pubescens* Vahl, and *Cinchona calisaya* Wedd in Peru, Journal of Sustainable Forestry, 44:1-5, 103-125, DOI: 10.1080/10549811.2025.2513230

To link to this article: https://doi.org/10.1080/10549811.2025.2513230

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 03 Jun 2025.

Submit your article to this journal

Article views: 1142

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=wjsf20

JOURNAL OF SUSTAINABLE FORESTRY
2025, VOL. 44, NOS. 1–5, 103–125
https://doi.org/10.1080/10549811.2025.2513230

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Generalized Linear Models to Estimate the Probability of Occurrence of *Cinchona officinalis* L., *Cinchona pubescens* Vahl, and *Cinchona calisaya* Wedd in Peru

Elton Tafur[a], Seily Cuchca[a], Ligia García[a,b], Nilton B. Rojas-Briceño[c], and Jaris Veneros[a]

[a]Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Perú; [b]Facultad de Ingeniería Zootecnista, Agronegocios y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Perú; [c]Grupo de Investigación en Ciencia de la Información Geoespacial, Escuela Profesional de Ingeniería Ambiental, Facultad de Ingeniería y Arquitectura, Universidad Nacional de Moquegua, Pacocha, Perú

ABSTRACT

Cinchona species inhabit disturbed and threatened ecosystems. To address this, Peru's Ministry of Agrarian Development and Irrigation approved a repopulation plan, focused on conservation and repopulation. However, their spatial distribution patterns remain largely unknown. Thus, Generalized Linear Models (GLM) were developed in this research to estimate the presence of *C. officinalis*, *C. pubescens*, and *C. calisaya* in Peru. For this purpose, the presence-absence of these three species was modeled with 26 environmental predictor variables using forward regression for GLMs. The accuracy of the resulting models was measured by the area under the curve (AUC), the true skill criterion (TSS), and the Kappa index. The results showed that the model selected to estimate the presence-absence of *C. pubescens* obtained values of AUC, TSS, and kappa index of 0.90, 0.72, and 0.71 respectively. While, for *C. officinalis*, the selected model got values of 0.93, 0.74, and 0.74 for the AUC, TSS, and Kappa index respectively. For *C. calisaya*, the selected model obtained values of 0.97, 0.92, and 0.91 for the AUC, TSS, and kappa index respectively. Finally, the San Martin Region showed the highest probability of suitability for these species. Our results detail their distribution in Peru, identifying the most suitable areas for their repopulation or conservation. These findings will allow for more efficient management of economic and natural resources, ensuring the care of the three most important Cinchona species in the country. This research also contributes to the 2030 Agenda in Goal 15, halting the loss of biodiversity within terrestrial ecosystems in the forest area.

KEYWORDS

Cinchona; GLM; repopulation; reforestation; bioclimatic variables

SUSTAINABLE DEVELOPMENT GOALS

SDG 11: Sustainable cities and communities; SDG 13: Climate action; SDG 15: Life on land

Introduction

Historically, species of the genus *Cinchona* were the only medicine for the treatment of malaria, due to their high alkaloid content (Albán-Castillo et al., 2020; Coronel et al., 2024). For this reason, the cinchona tree is the emblematic genus of Peru's natural resources (Coronel et al., 2024). The genus *Cinchona* has about 25 types of alkaloids, the most

CONTACT Jaris Veneros jaris.veneros@untrm.edu.pe Instituto de Investigación Para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, 342 higos Urco, Chachapoyas 01001, Perú

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

important being quinine and quinidine (Remuzgo et al., 2021). This genus is distributed from Costa Rica to the Bolivian Andes, with greater distribution in northern Peru and southern Ecuador (Perez, 2021). However, the species of the *Cinchona* genus are currently highly threatened by agriculture, cattle ranching, and burning, among others (Coronel et al., 2024). As a result, these species have come to be considered as endangered species or are present on lists of threatened species (García et al., 2022). For this reason, the Peruvian government approved the Cinchona repopulation plan in Peru (Albán-Castillo et al., 2020). However, there are very few studies on the spatial distribution of the genus *Cinchona* in Peru (García et al., 2022). In this sense, the use of Species Distribution Models (SDMs) in the management and conservation planning of endemic species is of vital importance (Frans et al., 2022). Since they allow inferring the ecological niche of the species and estimating its presence in other places with similar environmental conditions, favoring the repopulation of the species (Bystrova et al., 2021).

Modeling species distributions in space and time has been a major challenge in ecology, as many biotic and abiotic factors influence species distributions (Alatawi et al., 2020). Currently, species distribution models are based on the correlation of observed geolocated presence data of the species and environmental variables (Ghareghan et al., 2020; Rew et al., 2021). Moreover, SDMs are essential tools that allow investigation of how species respond to environmental conditions (Bystrova et al., 2021; Lausch et al., 2022). It is worth mentioning that the most used environmental conditions for SDMs are climatic, edaphic, and topographic variables (Ndlovu et al., 2018; Safaei et al., 2018). Different statistical and mathematical methods have been applied and studied to estimate the geographic distribution of species at different scales (Aghajanlou et al., 2021). Among the most widely used methods are the Maximum Entropy (MaxEnt) method (Fitzgibbon et al., 2022), Artificial Neural Networks (Aghajanlou et al., 2021), Random Forest (RF) (Rew et al., 2021), Generalized Linear Models (GLM) (Ghareghan et al., 2020), Generalized Additive Models (Rew et al., 2021), among others.

Generalized Linear Models (GLM) are commonly used and are easier to use than other SDMs (Breiner et al., 2015). GLMs represent a special case of multiple linear regression that allows obtaining non-normal responses (Ghareghan et al., 2020). GLMs for the case of species distribution, usually use a logit link function and the binomial error term (Prasetyo et al., 2020). Since most of the response variables are of the binary type, i.e., presence and absence (Rew et al., 2021; Safaei et al., 2018). This method has been widely applied and accepted in the literature (Ghareghan et al., 2020).

Worldwide, numerous investigations have made use of GLM to model species distribution, for example, in the European Alps it was applied to 2616 vascular plant species (Chauvier et al., 2021), in Germany, it was applied for *Bolboschoenus maritimus* (L.) *Palla and Phragmites australis* (Cav.) Trin. ex Steud (Carus et al., 2017), in Iran for *M. persica* (Ghareghan et al., 2020), in Ethiopia for *Coffea arabica* L (Amiri et al., 2020), in Canada for *Vaccinium myrtilloides* Michx (Nielsen et al., 2020), in Uruguay for *Eucalyptus grandis* W. Hill y *Eucalyptus dunnii* Maiden (Resquin et al., 2020) and in Ethiopia for *Prosopis juliflora* (Sw.) DC (Sintayehu et al., 2020).

At the Peruvian level, focused studies for Cinchona modeling are at the genus level (García et al., 2022). Or at the species level (*C. capuli* L. Anderson, *C. macrocalyx* Pav. Ex DC. y *C. pubescens* Vahl.) but only at the level of one region (Coronel et al., 2024), or other researchers evaluate modeling at the historical and current level (García et al., 2022). On the

other hand, studies that used GLMs in the distribution of flora species are very scarce; however, the use of GLMs was identified for the distribution of the species *Polylepis* spp. (Sevillano-Ríos & Morales, 2021). and the distribution of ferns in the Manú forests (Centeno, 2023). However, the spatial distribution of any species of the *Cinchona* genus has not been modeled using a GLM in Peru. Considering that in Peru the repopulation plan for the genus Cinchona was approved, prioritizing the species *Cinchona officinalis, Cinchona pubescens*, and *Cinchona calisaya* (Albán-Castillo et al., 2020). Therefore, the objective of this study was to develop GLMs to estimate the presence of these three species, and to contribute as a tool for decision-making on the most suitable areas for the repopulation of these three species in Peru.

Materials and methods

Study area

The study was carried out in Peru, which covers a territorial area of 1,285,215.60 km^2 (Figure 1). Peru has one of the greatest biodiversity on the planet, as it has eight natural regions, which were classified according to their altitudinal level in Costa, Yunga, Quechua, Suni, Puna, Cordillera, High Jungle and Low Jungle (Pulgar Vidal, 2014). Peru is also the country with the greatest diversity of *Cinchona* in the world, as it is home to 19 of the 25 species of the genus *Cinchona* present on the planet (Albán-Castillo et al., 2020; Perez, 2021). Likewise, García and collaborators in 2022 reported that Peru has 18 species of Cinchona out of 24 species registered worldwide (García et al., 2022). In Peru, the greatest

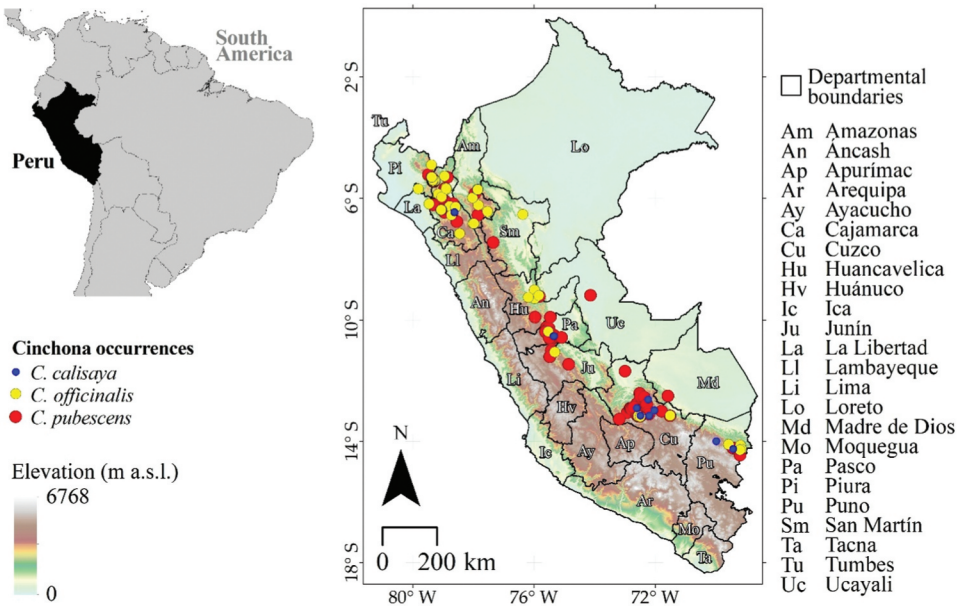


Figure 1. Observed presence map of *C. pubescens, C. calisaya*, and *C. officinalis* in Peru.

observed presence of Cinchona is found in the North Andean Cordillera (Coronel et al., 2024).

Presence and absence data

Presence data for *C. officinalis*, *C. pubescens*, and *C. calisaya* were downloaded from the Figshare repository of the Quina.pe Project – National University Toribio Rodríguez de Mendoza (UNTRM) (https://bit.ly/2UKojb8). From this database, we obtained 281 records for *C. pubescens*, 116 records for *C. officinalis*, and 39 records for *C. calisaya* (García et al., 2022).

Absence data for the three Cinchona species were randomly generated in the RStudio program (Alemayehu et al., 2024). The number of pseudo-absences generated for each of the three species was 1000 records, from which the same number of absence and presence records were randomly taken for each of the three species (Liu et al., 2019; Rew et al., 2021).

Climate data

The following were downloaded from the Worldclim database version 2 (http://worldclim.org/version2): a set of 19 bioclimatic variables at a resolution of 1 km^2. These bioclimatic variables represent annual, seasonal, extreme, or limiting averages of precipitation and temperature variables between 1970 and 2000 (Alatawi et al., 2020).

Topographical data

The data for the topographic variables (altitude and slope) were obtained from a mosaic of 44 digital elevation models (DEM), obtained from the Ministry of the Environment's geoserver. This DEM has a spatial resolution of 30 m, with a horizontal (WGS 1984) and vertical (EGM96) coordinate system. The DEM was used to generate a slope map in sexagesimal degrees for all of Peru using ArcGIS 10.8.

Edaphic data

Five edaphic variables (organic carbon, pH, CEC, and clay) were used, which were downloaded from the SoilGrids Version 2.0 database belonging to the International Soil Reference and Information Center (ISRIC) (https://soilgrids.org/). This database is in raster format with a resolution of 250 m, containing the physicochemical properties of the soil (Hengl et al., 2017).

Variable selection

First, data on climatic, topographic, and soil variables were extracted for each of the points of presence and absence of the three Cinchona species, using the RASTER package of RStudio. Secondly, uncorrelated variables were selected to avoid problems of collinearity and overfitting of the models (Lausch et al., 2022). For this purpose, the variance inflation factor (VIF) of the SDM package in R Studio was used

(Alatawi et al., 2020; Alemayehu et al., 2024). Finally, variables with VIF < 10 were selected for model building (Alatawi et al., 2020; Resquin et al., 2020; Rew et al., 2021).

GLM modeling

GLMs are used to model discrete dependent variables, which present a nonlinear relationship with the independent variables (Amiri et al., 2020). For this reason, it is widely used to model species distribution (Ghareghan et al., 2020; Meynard et al., 2019). The GLM models the presence and absence of species as a function of environmental predictor variables using the logit function and the binomial error term (Ghareghan et al., 2020; Prasetyo et al., 2020; Sousa-Guedes et al., 2021).

The presence and pseudo-absence data were labeled with a value of 0 and 1, where 0 represents the pseudo-absence of the species and 1 represents the presence of the species (Ghareghan et al., 2020). They were then randomly divided into model training data (70%) and model validation data (30%) in RStudio (Alatawi et al., 2020; Sintayehu et al., 2020).

Subsequently, with the training data, a GLM was generated for each of the three *Cinchona* species, using the automatic selection of variables with the forward selection method (Prasetyo et al., 2020). This type of regression uses the Akaike information criterion (AIC) to evaluate the inclusion or exclusion of the variable to the model (Resquin et al., 2020; Schwager & Berg, 2021). For this, the GLM command indicating the binomial distribution family was used in RStudio (Alemayehu et al., 2024).

Likewise, interactions between variables were added to this model, i.e. the multiplication between two variables with all possible combinations; these interactions were included in the model already generated using the procedure described above, to improve the predictive accuracy of the model. The resulting models will be selected according to the principle of parsimony, i.e., the model should have the lowest number of possible variables and high significance, but the one that explains the dependent variable the most (Dutang & Guibert, 2022; Guisset et al., 2019).

Validation and accuracy

Model validation consisted of measuring the predictive ability of the models with the validation or test data. A total of 281 presence data were used, 70% of these data were used to generate the models and 30% of these records were used as part of the validation of the models. Predictive ability was determined by threshold-independent (AUC) and threshold-dependent (Kappa index and TSS) measures of accuracy (Ndlovu et al., 2018). These criteria measure the discrimination between correct and incorrect prediction (Resquin et al., 2020). For the evaluation of the models, the Kappa Index and TSS were used, and values approaching the value of 1 mean that the values estimated by the model fit observed values of the presence of the species under study, thus ensuring that the training, test, and validation data are optimal for this type of research (Naimi & Araújo, 2016). Thus, they have been widely used in species distribution modeling (Rew et al., 2021).

Presence-absence locality patterns

Once the GLMs for the three *Cinchona* species were obtained, they were applied to the entire Peruvian territory to discriminate the areas suitable for the repopulation of the three *Cinchona* species in Peru. Likewise, for better visualization of the probability of occurrence (presence-absence) of the species for all of Peru, the probability data were categorized into absence (high and low) and presence (high and low) according to the classification threshold for each species (Resquin et al., 2020).

Results

Selection of environmental variables

The results of variables selection show that 12 of the 26 initial variables do not present collinearity problems (VIF <10) for *C. pubescens* and *C. officinalis*. While for *C. calisaya* only 10 of the 26 initial variables are not collinear (Table 1).

Modelling

For *C. pubescens*, Model 1 (without interaction between variables) consisted of three edaphic variables and three bioclimatic variables, which were highly significant at 95% with a *p-value* < 0.03 and an AIC = 310.57. In Model 2 (with interaction between variables), the inclusion of the interactions Carbon*CEC and PH*Bio15 contributed significantly to the model at 95% with a *p-value* < 0.001, an AIC = 264.36, and a delta of AIC = 64.21 concerning Model 1. This indicates that Model 2 is the best model for *C. pubescens* in terms of AIC. The addition of these interactions caused the variable Bio 19 to be non-significant at 95% with a p-value = .314 for Model 2 (Table 2). However, the variable Bio 19 was retained as part of Model 2, since considering eliminating this variable would cause a high increase in the model deviance.

For *C. calisaya*, Model 1 consisted of two edaphic variables and two bioclimatic variables, which were highly significant at 95% with a *p-value* < 0.05 except for the Bio 15 variable and an AIC of the model equal to 36.18. In Model 2, the inclusion of the interaction

Table 1. Selected variables as results of the variance inflation factor test.

Variables	*Cinchona pubescens*	*Cinchona officinalis*	*Cinchona calisaya*
Bio 2	3.78	4.23	—
Bio 3	2.62	2.87	2.60
Bio 8	5.58	5.73	3.80
Bio 13	2.98	2.97	3.51
Bio 15	3.83	3.27	2.96
Bio 19	3.60	3.44	2.92
Organic carbon	6.24	5.77	5.33
Nitrogen	3.25	2.45	3.44
Arcilla	2.17	2.57	3.59
pH	2.53	2.45	3.05
CEC	6.07	5.99	—
Slope	1.93	1.47	1.62

Table 2. Model for *C. pubescens*.

	Model 1		Model 2	
Variables	Coefficient	*p-value*	Coefficient	*p-value*
Intercept	−10.7521	0.01	−20.9569	0.009
Organic carbon	0.3397	0.006	−2.5036	<001
Bio 3	0.2397	<001	0.2184	<001
Bio 19	−0.0079	<001	−0.0015	0.314
pH	−2.7185	<001	2.6986	0.035
CEC	0.1748	<001	−1.0115	<001
Bio 15	0.0291	0.029	0.4945	<001
Carb*CEC			0.1610	<001
pH*Bio15			−0.0791	<001

Model 1 = Model without interaction between variables, Model 2 = Model with interaction between variables, Carb = Organic carbon.

For *C. officinalis*, Model 1 consisted of three edaphic variables and two bioclimatic variables, which were highly significant at 95% with a *p-value* <0.01 and an AIC = 91.13. In Model 2, the inclusion of the interaction Carbon*Bio8 contributed significantly to the model at 95% with a *p-value* = 0.018, an AIC = 86.13, and a delta of AIC = 5 concerning Model 1, which indicates that Model 2, is slightly the best model for *C. officinalis* in terms of AIC. Alternatively, the addition of this interaction did not generate the loss of significance at 95% of any of the other variables of Model 2, since these presented a *p-value* <0.02 (Table 3).

Table 3. Model for *C. officinalis*.

	Model 1		Model 2	
Variables	Coefficient	*p-value*	Coefficient	*p-value*
Intercept	−14.7971	<001	−32.2226	0.001
Organic carbon	1.9416	<001	5.1348	0.001
Nitrogen	−3.3730	<001	−4.4246	<001
Bio 13	−0.0124	0.009	−0.0156	0.004
CEC	0.3319	0.004	0.2889	0.017
Bio 8	0.2033	0.01	1.0660	0.007
Carb*Bio8			−0.1374	0.018

Model 1 = Model without interaction between variables, Model 2 = Model with interaction between variables, Carb = Organic carbon.

Table 4. Model for *C. calisaya*.

	Model 1		Model 2	
Variables	Coefficient	*p-value*	Coefficient	*p-value*
Intercept	−21.4605	0.03	8.2287	0.63
Arcilla	0.1037	0.00	0.0953	0.02
Bio 19	−0.0500	0.00	−0.0585	0.01
Nitrogen	1.1879	0.04	−10.5974	0.04
Bio 15	−0.1160	0.07	−0.6136	0.02
Nitr*Bio15			0.2128	0.03

Model 1 = Model without interaction between variables, Model 2 = Model with interaction between variables, Nitr = Nitrogen.

Nitrogen*Bio15 contributed significantly to the model at 95% with a *p-value* = 0.03, an AIC = 31.84, and a delta of AIC = 4.34 concerning Model 1, which indicates that Model 2, is slightly the best model for *C. calisaya* in terms of AIC. Also, the addition of this interaction generated that all the variables of Model 2 are significant at 95% since they presented a *p-value* < 0.05 (Table 4).

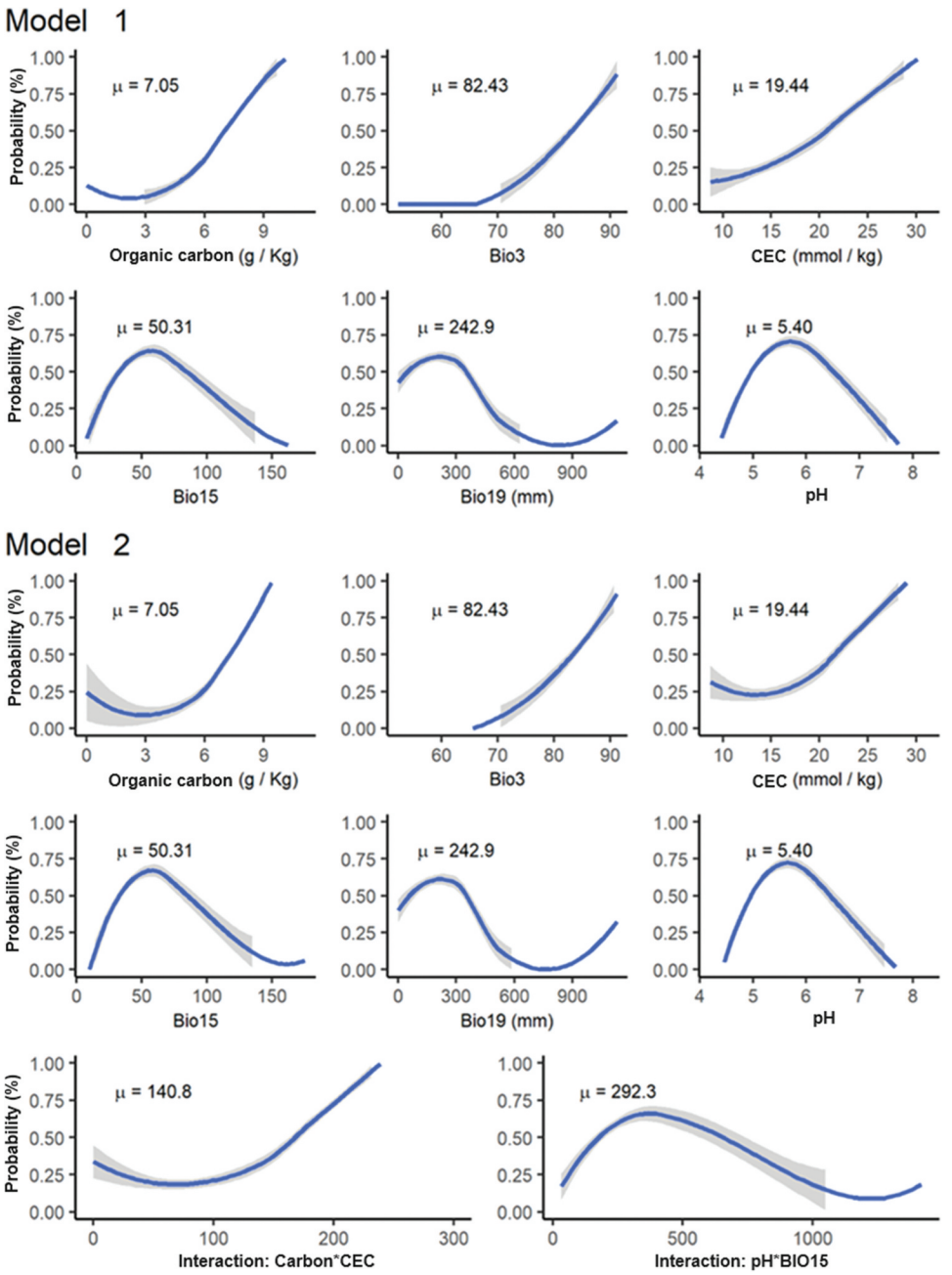


Figure 2. Response curves for *C. pubescens.*

Response curve analysis

The response curves of Model 1 and Model 2 visually show us the behavior of the variables of the model, according to the probability estimated by the model. The analysis of the response curves for *C. pubescens* shows that the increase above the mean (μ) of the variables organic carbon, Bio

3, and CEC, increases the probability of the estimated presence of this species in both models. On the other hand, the values of the variables Bio 15, Bio 19, and PH must be around the mean to obtain a high probability of the estimated presence of the species in both models. On the other hand, the increase above the mean (μ) of the interaction Carbon*CEC in model 2, increases the probability of the estimated presence of this species, while the values of the interaction pH*Bio15 must be around the mean to obtain a high probability of estimated presence in model 2 (Figure 2).

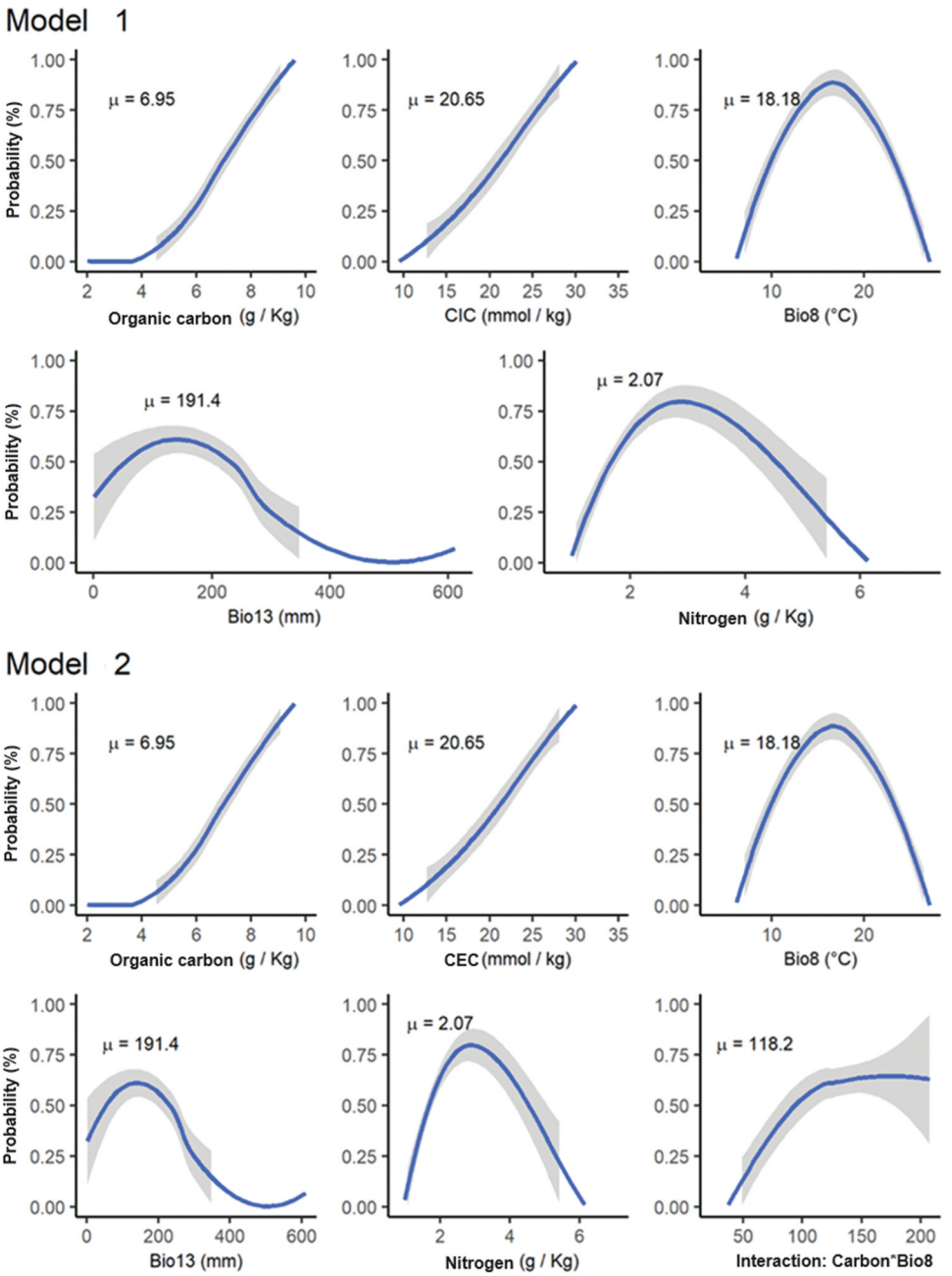


Figure 3. Response curves for *C. officinalis*.

The analysis of the response curves for *C. officinalis* shows that the increase above the mean of the organic carbon and CEC variables increases the probability of the estimated presence of this species in both models. On the other hand, the values of the variables Bio 8, Bio 13, and Nitrogen must be around the mean to obtain a high probability of the estimated presence of this species in both models. As for the interaction Carbon*Bio 8 in model 2, values above the mean (μ) of this interaction increase the probability of estimated presence for this species (Figure 3).

The analysis of the response curves for *C. calisaya* shows that the values of the variable Nitrogen must be around the mean to obtain a high probability of the estimated presence of this species in both models. On the other hand, the increase above the mean (μ) of the variable Clay increases the probability of the estimated presence of this species in both models. On the other hand, the values of the variables Bio19 and Bio15 must be around the mean to obtain a high probability of the estimated presence of this species in both models. As for the interaction Nitrogen*Bio 15 in model 2, the increase above the mean of this interaction decreases the probability of the estimated presence of *C. calisaya* (Figure 4).

Model accuracy and validation

The evaluation of the accuracy of the models for the training data shows that, for *C. pubescens*, Model 2 showed better results for the criteria AUC, TSS, and kappa index, in comparison to Model 1 (Table 5). For *C. officinalis* and *C. calisaya*, Model 1 and Model 2 showed similar results in the criteria AUC, TSS, and Kappa Index (Table 5).

The evaluation of the accuracy of the models for the validation data shows that, for *C. pubescens* and *C. officinalis*, Model 2 showed better results in the criteria AUC, TSS, and Kappa index, in contrast to Model 1 (Table 6). For *C. calisaya*, Model 1 showed the best results in the TSS and Kappa index criteria, in comparison to Model 2; however, the AUC criterion was slightly higher in Model 2 (Table 6).

On the other hand, no significant differences were found in the TSS, Kappa index, and AUC criteria in the training and validation data for Models 1 and 2. This shows that Models 1 and 2 adequately predict the presence-absence of *C. pubescens*, *C. officinalis*, and *C. calisaya* since they present a good fit in known data (training) and unknown data (validation). Likewise, the analysis of the ROC curves for *C. pubescens*, *C. officinalis*, and *C. calisaya* shows no significant differences in the ROC curves of the training and validation data (Figure 5).

Patterns of locality of cinchona occurrence in Peru

It was shown that the patterns of localities with a high probability of finding the occurrence of *C. pubescens* in Peru cover a territorial area of 138,819.62 km^2 and 97,195.04 km^2 for Model 1 and Model 2, respectively. Model 1 covered a larger territorial area than Model 2, with a considerable difference of 41,624.57 km^2. This change can be attributed mainly to the significant difference between Models 1 and 2, as well as the variance between their cutoff points, which cause a distinction in the classification of presence-absence of *C. pubescens*. In addition, this change was visually observed in the maps of the presence-absence locality patterns predicted by Models 1 and 2 for this species (Figure 6).

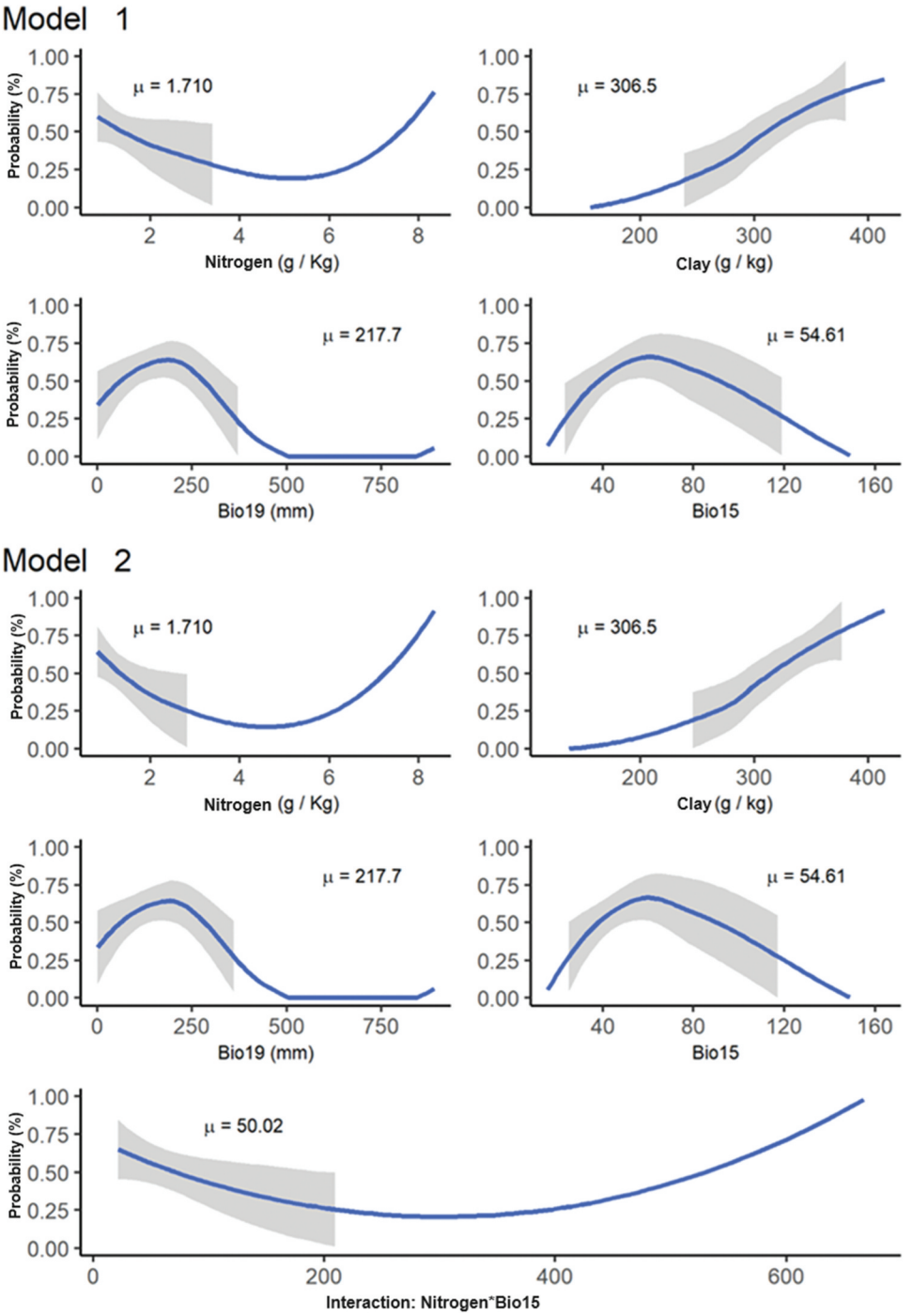


Figure 4. Response curves for *C. calisaya*.

Table 5. GLM accuracy metrics on training data.

Specie	Model	Cut	Sensitivity	Specificity	TSS	AUC	Kappa index	AIC
C. pubescens	M1	0.4	0.91	0.82	0.74	0.91	0.73	310.57
	M2	0.5	0.93	0.87	0.80	0.94	0.79	264.36
C. officinalis	M1	0.5	0.92	0.93	0.84	0.96	0.84	91.13
	M2	0.5	0.91	0.94	0.84	0.97	0.84	86.13
C. Calisaya	M1	0.3	1.00	0.88	0.88	0.96	0.86	36.18
	M2	0.3	0.96	0.90	0.86	0.98	0.86	31.84

M1 = Model without interaction between variables, M2 = Model with interaction between variables. AUC = Area under the curve, AIC = Akaike's information criterion.

Table 6. GLM accuracy metrics on validation data.

Specie	Model	Cut	Sensitivity	Specificity	TSS	AUC	Kappa index	AIC
C. pubescens	M1	0.4	0.91	0.82	0.72	0.90	0.71	310.57
	M2	0.5	0.89	0.87	0.76	0.92	0.76	264.36
C. officinalis	M1	0.5	0.90	0.84	0.74	0.93	0.74	91.13
	M2	0.5	0.88	0.88	0.76	0.94	0.76	86.13
C. Calisaya	M1	0.3	0.92	0.91	0.92	0.97	0.91	36.18
	M2	0.3	0.96	0.90	0.86	0.98	0.86	31.84

M1 = Model without interaction between variables, M2 = Model with interaction between variables. AUC = Area under the curve, AIC = Akaike's information criterion.

Then, it was also calculated that the locality patterns with a high probability of finding *C. officinalis* in Peru cover a territorial area of 107,229.01 km^2 and 109,211.8667 km^2 for Model 1 and Model 2, respectively. Model 2 covered a larger territorial area than Model 1, with a relatively low difference of 1982.85 km^2, which is consistent with similar precision metrics for Models 1 and 2, with the same cutoff point (Figure 7).

Later, it was also shown that the locality patterns with a high probability of finding *C. calisaya* in Peru cover a territorial area of 196,149.87 km^2 and 182,455.6195 km^2 for Model 1 and Model 2, respectively. Model 1 covered a larger territorial area than Model 2, with a moderately low difference of 13,694.25 km^2, which is consistent with similar precision metrics and the same cutoff point for Models 1 and 2 (Figure 8).

The distribution of presence with high probability was also carried out at the departmental level, where it was identified that the departments of San Martín, Cuzco, Cajamarca, Junín, Ucayali, Amazonas, Huánuco, and Loreto have the highest distribution with high probability for the presence of *C. calisaya*, *C. pubescens*, and *C. officinalis* predicted by Models 1 and 2. Likewise, for *C. calisaya* the largest distribution of presence with high probability was found in the department of Ucayali, with a total area of 30,818.98 km^2 for Model 1 (Figure 9) and 29,714.74 km^2 for Model 2 (Figure 9). Meanwhile, for *C. officinalis* the largest distribution of presence with high probability was found in the department of San Martín, with a total area of 26,200.76 km^2 for Model 1 (Figure 9) and 24,868.20 km^2 for Model 2 (Figure 9). On the other hand, for *C. pubescens* the largest distribution of presence with high probability was found in the department of San Martín, with a total area of 19,836.31 km^2 for Model 1 (Figure 9), while for Model 2 the department of Huánuco presented the largest distribution of presence with high probability, with a total area of 13,076.07 km^2, followed by the department of San Martín, with a total area of 11,363.83 km2 (Figure 9).

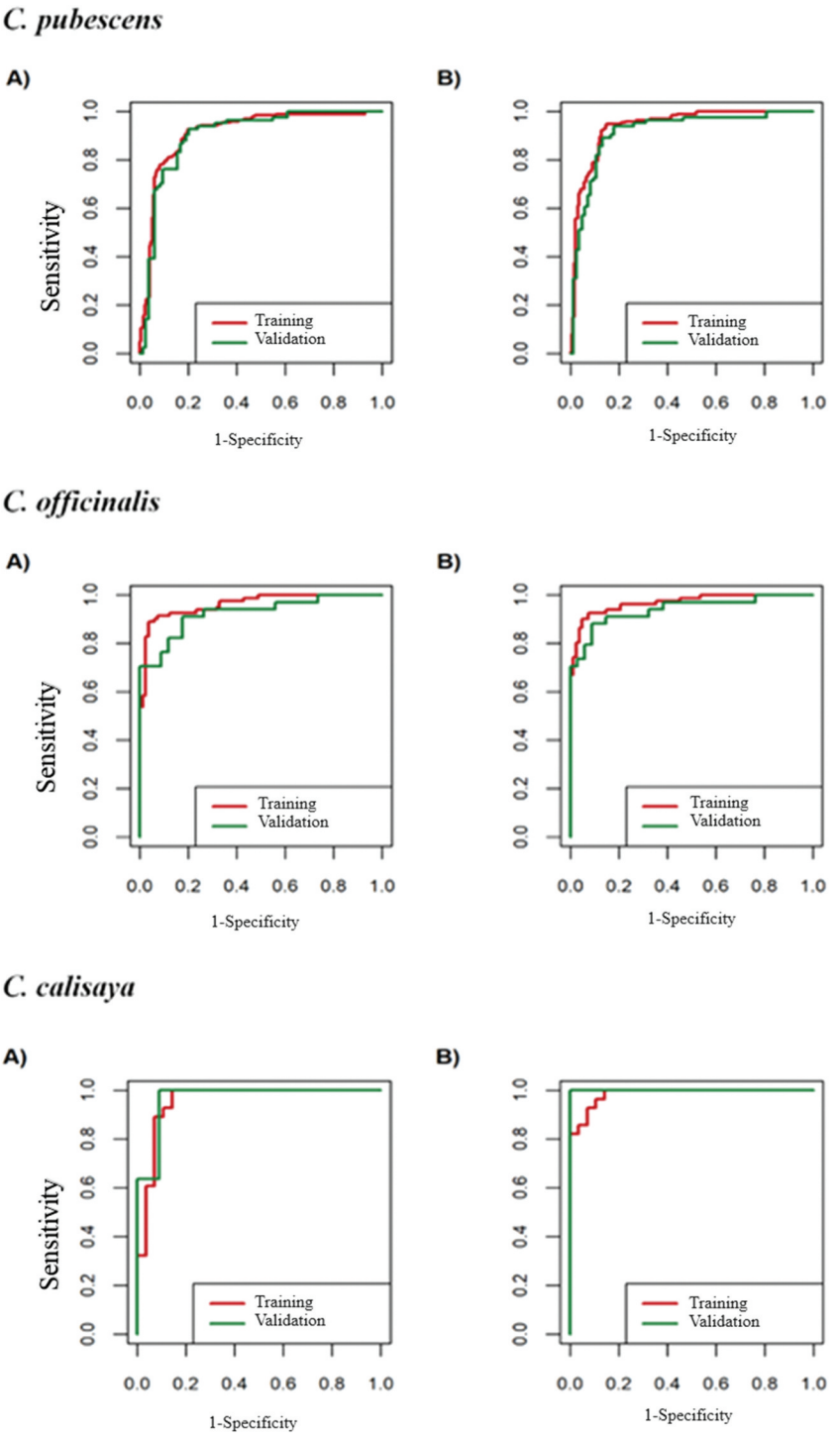


Figure 5. ROC curve plot for Model 1 (A) and Model 2 (B).

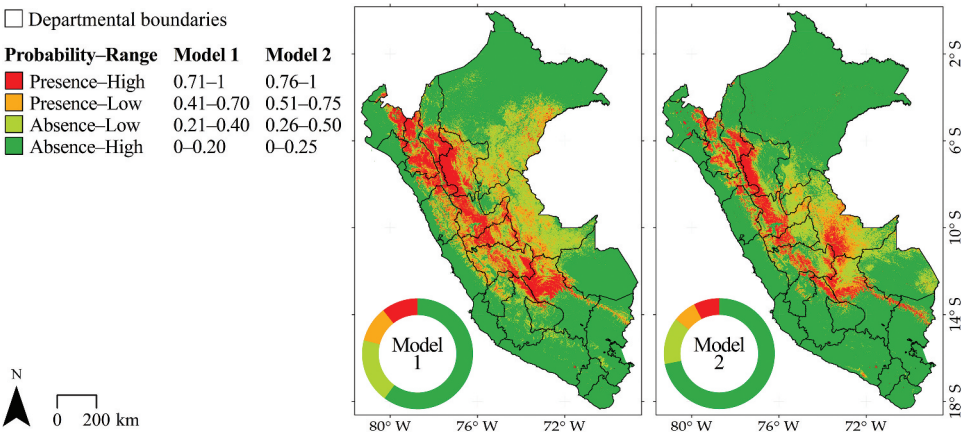


Figure 6. Maps of the potential distribution of *C. pubescens* in Peru estimated by Model 1 (A) and Model 2 (B) in Peru.

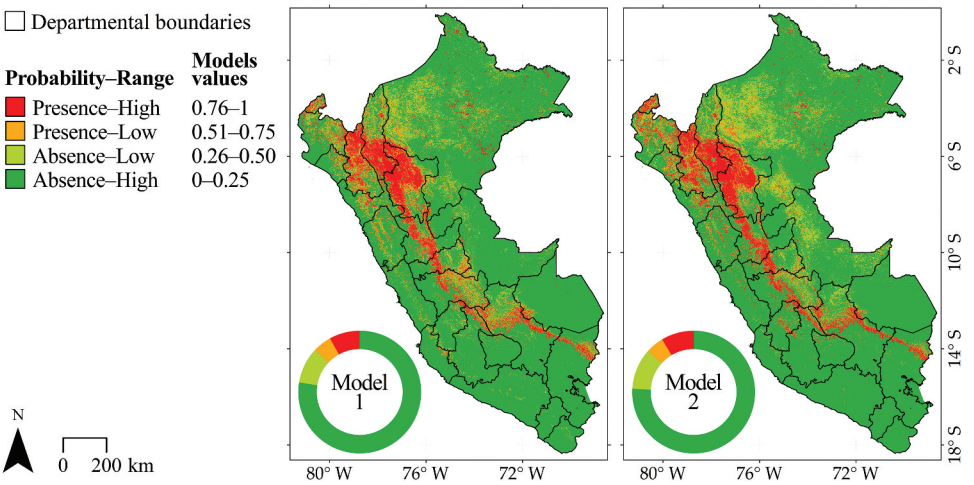


Figure 7. Maps of the potential distribution of *C. officinalis* in Peru estimated by Model 1 (A) and Model 2 (B) in Peru.

Discussions

The number of predictors in the SDM has great relevance from an ecological point of view (Lausch et al., 2022). However, a high number of predictor variables lead to problems of model overfitting and multicollinearity among variables (Arabameri et al., 2021; Avand et al., 2022). Therefore, the selection of predictor variables is a key step in SDM (Liang et al., 2020).

Our results show that more than half of the initial variables presented multicollinearity problems (VIF >10) (Kaky et al., 2020). Therefore, in this study, collinear variables were excluded for model building, regardless of the ecological effect of variables on species (Arabameri et al., 2021). Therefore, in this study, collinear variables were excluded for model building, regardless of the ecological effect of variables on species (Chen et al., 2020;

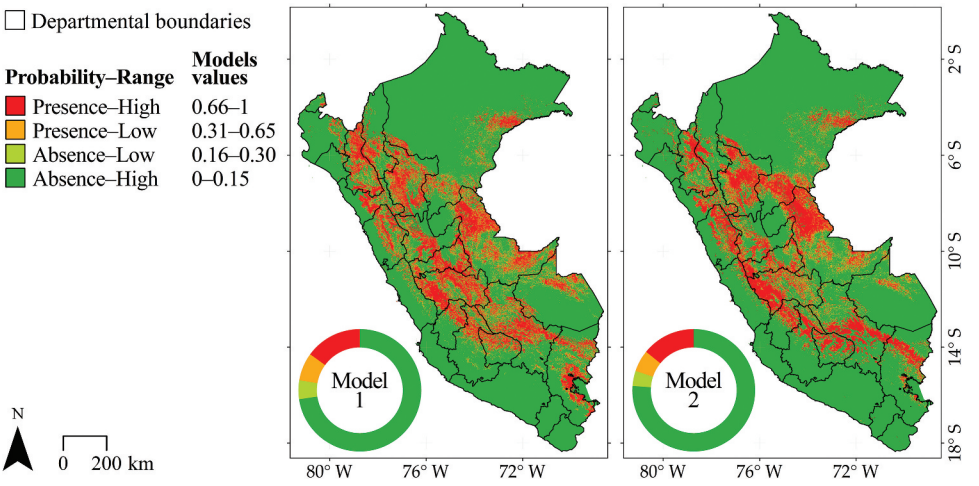


Figure 8. Maps of the potential distribution of *C. calisaya* in Peru are estimated by Model 1 (A) and Model 2 (B) in Peru.

Kaky et al., 2020; Santini et al., 2021). Which is supported by many previous studies (Alatawi et al., 2020; Carus et al., 2017; Ghareghan et al., 2020; Rew et al., 2021; Zielińska et al., 2024).

The selection of the non-collinear predictor variables that make up the GLM should be correlated with the presence or absence of the species with a *p-value* < 0.05 (Monnier-Corbel et al., 2023). They should also minimize the AIC of the resulting models for a good model fit on the training data (Santini et al., 2021; Valavi et al., 2022). Our results of automatic variable selection using the forward method based on AIC and *p-value* showed a good fit for model creation for each of the three *Cinchona* species.

Frequently, SDM studies have used only bioclimatic variables to predict species distribution (Amiri et al., 2020). However, some research has shown that edaphic and topographic variables are important drivers in the SDM (Chauvier et al., 2021; Kaky et al., 2020). Our results show that half of the predictor variables for each GLM of the three species were edaphic variables. This demonstrates the importance of incorporating edaphic variables in the SDM (Dindaroglu et al., 2023; Pazmiño et al., 2024).

The most important edaphic variables (variables with recurrence in two or more models) found for the three *Cinchona* species in our study were organic carbon content, CEC, and amount of nitrogen. High soil organic carbon content increases the survival of plant species due to the improvement of the soil physical and biological properties (Dindaroglu et al., 2023). In addition, high organic carbon content in soils increases CEC (Soinne et al., 2024). Trees of the genus *Cinchona* grow in volcanic soils rich in organic carbon (Aslam et al., 2023; Zevallos, 1989). In our study, the organic carbon and CEC variables presented a direct correlation with the presence of *C. pubescens* and *C. officinalis* species. Similar studies support the use of organic carbon as an important predictor of plant species distribution (Chauvier et al., 2021; Dindaroglu et al., 2023).

Soil nitrogen content is an important component in the distribution of plant species (Foltran & Lamersdorf, 2024). Previous studies have shown that the addition of *Pseudomonas* sp. strains to the root system of *C. officinalis* seedlings increases

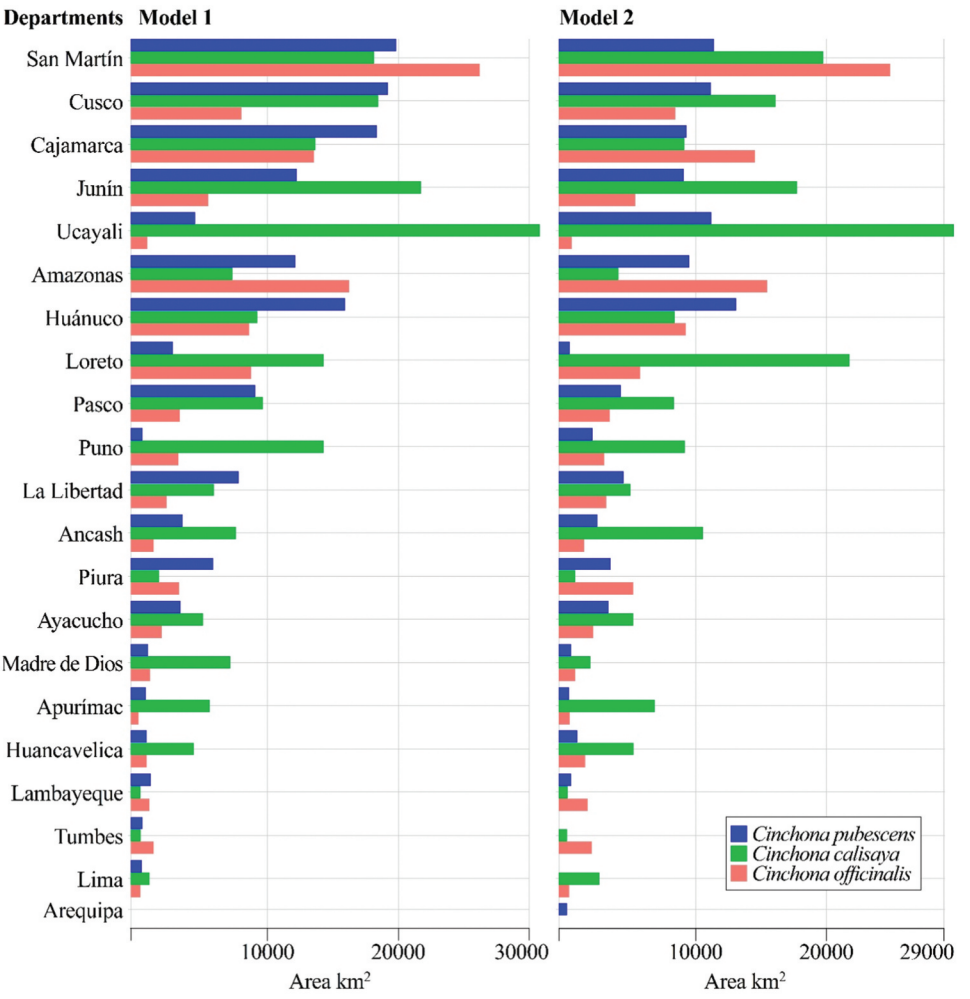


Figure 9. Areas of high probability of presence of the three species at the departmental level estimated by Model 1 and Model 2.

atmospheric nitrogen fixation, which allows for increased vegetative growth (Quesquén Esquen, 2020). The increase of nitrogen in the soil also increases the content of alkaloids present in *Cinchona* barks (Aslam et al., 2023). In our study, the nitrogen variable presented specific ranges around the mean (1.7–2.7 g/kg) to obtain a high probability of the presence of *C. officinalis* and *C. calisaya* species. The edaphic variables pH and clay content did not occur in the models but were part of the models for *C. pubescens* and *C. calisaya*, respectively. From an ecological point of view, it is expected that *C. pubescens* trees are correlated with pH since the presence of this species requires soils with an acid pH between 5 and 6.5 for its development (Aslam et al., 2023), which agrees with our results of the high probability of the presence of this species in this range. We also found the existence of a direct correlation between clay content and the presence of *C. pubescens*. The genus *Cinchona* is commonly distributed in clay soils of medium to heavy texture (Zevallos, 1989). Soils with high clay content have high

moisture content, thus providing greater availability of nutrients (Dindaroglu et al., 2023).

We found that for Peru *C. pubescens* is distributed in areas with a mean annual temperature between 14.2°C and 17.9 °C and a mean annual precipitation between 1041 mm and 1150 mm. Similar results were also obtained for *C. officinalis*, which is distributed in areas with a mean annual temperature between 13.73 °C and 16.56 °C and a mean annual precipitation between 1151 mm and 1444 mm. These characteristics classify the distribution of these two species in humid temperate and mountain climates according to the Köppen-Geiger climate classification (Beck et al., 2018). As for *C. calisaya*, our results for precipitation and mean annual temperature were spatially biased, as more than 50% of our data (25 data) were distributed in the Ucayali Region. This region is not in the range of distribution documented for any species of the genus Cinchona in Peru (Zevallos, 1989). Therefore, we consider the problems of georeferencing or some taxonomic confusion (Perez, 2021).

Bioclimatic variables delimit climatic conditions at very large scales, so they are expected to have a similar influence on the occurrence of species with similar habitats (Schwager & Berg, 2021). Our results indicate that the highest probability of occurrence for *C. pubescens*, *C. officinalis*, and *C. calisaya* occurs at the mean (μ) of each of the bioclimatic variables of temperature and precipitation selected for each species, except for the variable Isothermality which showed a direct relationship with the probability of occurrence. Likewise, our results show only the recurrence of the variables Bio 15 and Bio 19 in the models of *C. pubescens* and *C. calisaya*. Due to the distribution of species of the genus *Cinchona* in humid temperate and mountain climates, it is expected that the distribution areas present low seasonality of precipitation and high Isothermality (Beck et al., 2018). Our results show concordance, since we found a high probability of the presence of *C. pubescens* and C. calisaya, in areas with low seasonality of precipitation, around 50 mm. Likewise, for *C. pubescens* we found results of high probability of presence in areas with high Isothermality, which suggests low-temperature variations between each month. Our results showed similarity with those obtained by Vargas Aguilar (2021), where he found that Isothermality or Bio3 was the most important variable in predicting the distribution of the genus *Cinchona*.

According to the classification of Rew et al. (2021) the accuracy of models 1 and 2 based on AUC was excellent (AUC >0.9) for *C. pubescens*, *C. officinalis*, and *C. calisaya*. Similarly, the accuracy ranking of both models based on TSS for *C. calisaya* was excellent (TSS >0.8), while for *C. pubescens* and *C. officinalis*, it was good (0.6 <TSS < 0.8). Similarly, the accuracy rating of both models based on the Kappa Index for *C. calisaya* was good (0.8 <Kappa Index < 0.9), while for *C. pubescens* and *C. officinalis* it was fair (0.6 <kappa index <0.8). The high Kappa Index values for these Cinchonas reveal that the potential distribution of the species studied in Peru follows certain growth rules due to the suitability of the technique and the closeness of the estimation maps to reality, as has been interpreted in other investigations (Isinkaralar & Varol, 2023).

Vargas Aguilar (2021) also obtained similar results based on AUC (0.928) using Maxent's algorithm for the genus *Cinchona*. Models 1 and 2 do not show significant differences in accuracy based on AUC, TSS, and Kappa Index for the three *Cinchona* species. Therefore, considering the principle of parsimony, model 1 was considered the greatest model for calculating the presence of *C. pubescens*, *C. officinalis*, and *C. calisaya.*

Despite efforts to establish priority areas for proper management in the conservation of species of global concern, such as in oak forests or *Crataegus azarolus* and *Crataegus monogyna* (Radha & Khwarahm, 2022),or other timber forest species (Rojas Briceño et al., 2020).Or, as in the case of animals such as the *Panthera pardus tulliana* (Majeed et al., 2023) or *Tremarctos ornatus* (Meza Mori et al., 2020); modeling techniques for these approaches have benefits and limitations that have not yet been fully explored. Among the benefits are the generation of data on the consequences of prospective environmental conditions on their distributions, identifying key factors influencing their distributions, and identifying the key factors influencing their distributions (Hama & Khwarahm, 2023), among others. Limitations include mainly the lack of more detailed species distribution information such as spatial data, record dates and availability of species, which would lead to uncertain and unverifiable impact assessment decisions (Carneiro et al., 2016).

Also, models with low sensitivity result in wasted or unnecessary effort and resources for species conservation (Yoon & Lee, 2023). Therefore, it is necessary to know the extent to which the model correctly classifies the presence and absence of a species. Given this, the TSS criterion is considered the most reliable metric to measure model performance (Resquin et al., 2020). In our study, we sought to obtain models with high sensitivity and specificity (Ghareghan et al., 2020). However, we prevailed in sensitivity over model specificity, from a conservation point of view (Liu et al., 2019). Since our results are intended to contribute to the Action Plan for Reforestation with Species of the genus *Cinchona* (Quina Tree) 2020–2022, in Peru, approved by RD N°067–2020-MINAGRI-DVDIAR-AGRO RURAL-DE (Albán-Castillo et al., 2020).

The models generated show that the areas of a high probability of presence (probability >0.75) are found in northern Peru along the Andes Mountains, with a percentage of 42%, 47%, and 68% for *C. pubescens*, *C. calisaya*, and *C. officinalis* respectively. These results agree with the high distribution recorded by Coronel et al. (2024), for *C. pubescens* and *C. officinalis* in this area. However, the results found for the distribution of *C. calisaya* in northern Peru differ from the results obtained by Zevallos (1989), which indicates that the distribution recorded for this species is mostly in southern Peru, specifically in the Puno Region. This difference could be due to the lack of spatial reliability (latitude and longitude data) and the low number of occurrence records (39 records) that we obtained for this species.

The results obtained in this study show that the area of high probability of occurrence (probability >0.75) calculated at the Peruvian level by model 1 was greater than the area calculated by model 2 by 8% and 42% for *C. calisaya* and *C. pubescens*, respectively. This generated differences at the departmental level, changing the position of the departments with the largest area of high probability of presence calculated for both models. Meanwhile, for *C. officinalis* the areas of high probability calculated by models 1 and 2 remained practically similar with a difference of only 2% at the Peruvian level. This explains that for *C. officinalis* the positions of the departments with the highest area of high probability of presence remained constant for both models.

Our results, like those obtained by Vargas Aguilar (2021) showed that the Lima and Lambayeque regions have very small areas (area <500 km^2) with a high probability of the presence of *C. pubescens, C. officinalis*, and *C. calisaya*; however, these two regions were listed in the Action Plan for the Reforestation with Species of the genus Cinchona (Quina Tree) 2020–2022. Therefore, we suggest that the results

obtained in this research serve and reinforce the areas prioritized in this Plan from an ecological and statistical point of view.

Methodological applications for new investigations of the potential distribution of the genus Cinchona, in Peru, allow for more accurate pixel-by-pixel prediction rates; however, there are still important questions for future analyses of predictive residuals to improve the models (Paegelow & Olmedo, 2005). Also, further research is suggested using digital mapping and integrating, for example, integrating scenarios into cellular automata (Isinkaralar & Varol, 2023; Isinkaralar et al., 2022).

Conclusions

Our results showed that GLMs presented an excellent performance (AUC >0.9) in predicting the presence-absence of *C. pubescens*, *C. officinalis*, and *C. calisaya*. On the other hand, the interpretation and analysis of the response curves showed that edaphic variables as well as bioclimatic variables are key predictors in modeling the distribution of these three species. It was also demonstrated that edaphic and bioclimatic variables must present a range of specific values around the mean, to obtain a high probability of presence for these three species. In addition, it was found that the department of San Martin has the largest area at the departmental level, with a high probability of finding the presence of *C. pubescens* and *C. officinalis*, with a value of 19,836.31 km^2 (19.68%) and 26,200.76 km^2 (24.44%) respectively. It also occupies the fourth place at the departmental level, with the largest area with a high probability of finding the presence of *C. calisaya*, with a value of 18,123.29 km^2 (10%). Therefore, the greatest reforestation efforts should be focused on this department. Finally, our results show that the department of Amazonas occupies the sixth place at the departmental level with high habitat suitability for the presence of the three *Cinchona* species. Therefore, this department should be considered as a priority area for the repopulation of these three species.

Acknowledgments

Thank to the Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), Quina.pe-UNTRM Project Contract: No [367-2019-FONDECYT-BM-IADT-SE] and the ApiGen-UNTRM Project Contract: No [PE501083491-2023-PROCIENCIA]; Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica [N° 367-2019-FONDECYT-BM-IADT-SE].

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Aghajanlou, F., Mirdavoudi, H., Shojaee, M., MacSweeney, E., Mastinu, A., & Moradi, P. (2021). Rangeland management and ecological adaptation analysis model for *Astragalus curvirostris* Boiss. *Horticulturae*, *7*(4), 67. https://doi.org/10.3390/horticulturae7040067

Alatawi, A. S., Gilbert, F., & Reader, T. (2020). Modelling terrestrial reptile species richness, distributions and habitat suitability in Saudi Arabia. *Journal of Arid Environments*, *178*, 104153. https://doi.org/10.1016/j.jaridenv.2020.104153

Albán-Castillo, J., Chilquillo, E., Melchor- Castro, B., Arakaki, M., León, B., & Suni, M. (2020). *Cinchona* L. "Árbol de la Quina". *Revista Peruana de Biología*, *27*(3), 423–426. https://doi.org/10.15381/rpb.v27i3.18697

Alemayehu, B., Suarez-Minguez, J., & Rosette, J. (2024). Modeling the spatial distribution of *Acacia decurrens* plantation forests using PlanetScope images and environmental variables in the northwestern highlands of Ethiopia. *Forests*, *15*(2), 277. https://doi.org/10.3390/f15020277

Amiri, M., Tarkesh, M., Jafari, R., & Jetschke, G. (2020). Bioclimatic variables from precipitation and temperature records vs. remote sensing-based bioclimatic variables: Which side can perform better in species distribution modeling? *Ecological Informatics*, *57*, 101060. https://doi.org/10.1016/j.ecoinf.2020.101060

Arabameri, A., Rezaie, F., Pal, S. C., Cerda, A., Saha, A., Chakrabortty, R., & Lee, S. (2021). Modelling of piping collapses and gully headcut landforms: Evaluating topographic variables from different types of DEM. *Geoscience Frontiers*, *12*(6), 101230. https://doi.org/10.1016/j.gsf.2021.101230

Aslam, S., Jabeen, T., Ahmad, M., & Al-Huqail, A. A. (2023). Cinchona. In Muhammad Zia-Ul-Haq, Arwa Abdulkreem AL-Huqail, Muhammad Riaz, & Umar Farooq Gohar (Eds.), *Essentials of medicinal and aromatic crops* (pp. 221–248). Springer International Publishing. https://doi.org/10.1007/978-3-031-35403-8_10

Avand, M., Kuriqi, A., Khazaei, M., & Ghorbanzadeh, O. (2022). DEM resolution effects on machine learning performance for flood probability mapping. *Journal of Hydro-Environment Research*, *40*, 1–16. https://doi.org/10.1016/j.jher.2021.10.002

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Breiner, F. T., Guisan, A., Bergamini, A., Nobis, M. P., & Anderson, B. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, *6*(10), 1210–1218. https://doi.org/10.1111/2041-210X.12403

Bystrova, D., Poggiato, G., Bektaş, B., Arbel, J., Clark, J. S., Guglielmi, A., & Thuiller, W. (2021). Clustering species with residual covariance matrix in joint species distribution models. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.601384

Carneiro, L. R. D. A., Lima, A. P., Machado, R. B., Magnusson, W. E., & Lumbsch, H. T. (2016). Limitations to the use of species-distribution models for environmental-impact assessments in the Amazon. *PLOS ONE*, *11*(1), e0146543. https://doi.org/10.1371/journal.pone.0146543

Carus, J., Heuner, M., Paul, M., & Schröder, B. (2017). Plant distribution and stand characteristics in brackish marshes: Unravelling the roles of abiotic factors and interspecific competition. *Estuarine, Coastal and Shelf Science*, *196*, 237–247. https://doi.org/10.1016/j.ecss.2017.06.038

Centeno, J. (2023). *Diversidad de helechos arbóreos en una gradiente de elevación - Reserva de Biosfera del Manu*. Universidad Nacional de San Antonio Abad del Cusco. https://repositorio.unsaac.edu.pe/handle/20.500.12918/7663

Chauvier, Y., Thuiller, W., Brun, P., Lavergne, S., Descombes, P., Karger, D. N., Renaud, J., & Zimmermann, N. E. (2021). Influence of climate, soil, and land cover on plant species distribution in the European Alps. *Ecological Monographs*, *91*(2). https://doi.org/10.1002/ecm.1433

Chen, Q., Yin, Y., Zhao, R., Yang, Y., Teixeira da Silva, J. A., & Yu, X. (2020). Incorporating local adaptation into species distribution modeling of *Paeonia mairei*, an endemic plant to China. *Frontiers in Plant Science*, *10*, 10. https://doi.org/10.3389/fpls.2019.01717

Coronel, E., Meza-Mori, G., Torres, J. M. C., Mondragón, E. P., Cotrina-Sanchez, A., Oliva Cruz, M., López, R. S., & Campo Ramos, R. E. (2024). Potential distribution and identification of critical areas for the preservation and recovery of three species of *Cinchona* L. (Rubiaceae) in Northeastern Peru. *Forests*, *15*(2), 321. https://doi.org/10.3390/f15020321

Dindaroglu, T., Tunguz, V., Babur, E., Menshov, O., & Battaglia, M. L. (2023). Determination of the relationship among compound topographic index (CTI), soil properties and land-use in karst ecosystems. *Physical Geography*, *44*(3), 307–329. https://doi.org/10.1080/02723646.2021.1961361

Dutang, C., & Guibert, Q. (2022). An explicit split point procedure in model-based trees allowing for a quick fitting of GLM trees and GLM forests. *Statistics and Computing*, *32*(1), 6. https://doi.org/10.1007/s11222-021-10059-x

Fitzgibbon, A., Pisut, D., & Fleisher, D. (2022). Evaluation of maximum entropy (maxent) machine learning model to assess relationships between climate and corn suitability. *The Land*, *11*(9), 1382. https://doi.org/10.3390/land11091382

Foltran, E. C., & Lamersdorf, N. (2024). Tree species identity drives soil carbon and nitrogen stocks in nutrient-poor sites. *Forest Ecology and Management*, *567*, 122090. https://doi.org/10.1016/j.foreco.2024.122090

Frans, V. F., Augé, A. A., Fyfe, J., Zhang, Y., McNally, N., Edelhoff, H., Balkenhol, N., & Engler, J. O. (2022). Integrated SDM database: Enhancing the relevance and utility of species distribution models in conservation management. *Methods in Ecology and Evolution*, *13*(1), 243–261. https://doi.org/10.1111/2041-210X.13736

García, L., Veneros, J., Chavez, S. G., Oliva, M., & Rojas-Briceño, N. B. (2022). World historical mapping and potential distribution of *Cinchona* spp. In Peru as a contribution for its restoration and conservation. *Journal for Nature Conservation*, *70*, 126290. https://doi.org/10.1016/j.jnc.2022.126290

Ghareghan, F., Ghanbarian, G., Pourghasemi, H. R., & Safaeian, R. (2020). Prediction of habitat suitability of *Morina persica* L. species using artificial intelligence techniques. *Ecological Indicators*, *112*, 106096. https://doi.org/10.1016/j.ecolind.2020.106096

Guisset, S., Martin, M., & Govaerts, B. (2019). Comparison of PARAFASCA, AComDim, and AMOPLS approaches in the multivariate GLM modelling of multi-factorial designs. *Chemometrics and Intelligent Laboratory Systems*, *184*, 44–63. https://doi.org/10.1016/j.chemolab.2018.11.006

Hama, A. A., & Khwarahm, N. R. (2023). Predictive mapping of two endemic oak tree species under climate change scenarios in a semiarid region: Range overlap and implications for conservation. *Ecological Informatics*, *73*, 101930. https://doi.org/10.1016/j.ecoinf.2022.101930

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., & Bond-Lamberty, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Isinkaralar, O., & Varol, C. (2023). A cellular automata-based approach for spatio-temporal modeling of the city center as a complex system: The case of Kastamonu, Türkiye. *Cities*, *132*, 104073. https://doi.org/10.1016/j.cities.2022.104073

Isinkaralar, O., Varol, C., & Yilmaz, D. (2022). Digital mapping and predicting the urban growth: Integrating scenarios into cellular automata-Markov chain modeling. *Applied Geomatics*, *14*(4), 695–705. https://doi.org/10.1007/s12518-022-00464-w

Kaky, E., Nolan, V., Alatawi, A., & Gilbert, F. (2020). A comparison between ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecological Informatics*, *60*, 101150. https://doi.org/10.1016/j.ecoinf.2020.101150

Lausch, A., Schaepman, M. E., Skidmore, A. K., Catana, E., Bannehr, L., Bastian, O., Borg, E., Bumberger, J., Dietrich, P., Glässer, C., Hacker, J. M., Höfer, R., Jagdhuber, T., Jany, S., Jung, A., Karnieli, A., Klenke, R., Kirsten, T., Ködel, U. & Baatz, R. (2022). Remote sensing of geomorphodiversity linked to biodiversity—part III: Traits, processes and remote sensing characteristics. *Remote Sensing*, *14*(9), 2279. https://doi.org/10.3390/rs14092279

Liang, P., Qin, C.-Z., Zhu, A.-X., Zhu, T.-X., Fan, N.-Q., & Hou, Z.-W. (2020). Using the most similar case method to automatically select environmental covariates for predictive mapping. *Earth Science Informatics*, *13*(3), 719–728. https://doi.org/10.1007/s12145-020-00466-5

Liu, C., Newell, G., & White, M. (2019). The effect of sample size on the accuracy of species distribution models: Considering both presences and pseudo-absences or background sites. *Holarctic Ecology*, *42*(3), 535–548. https://doi.org/10.1111/ecog.03188

Majeed, K. A., Khwarahm, N. R., & Ahmed, S. (2023). Predicting the geographical distribution of the Persian leopard, *Panthera pardus tulliana*, a rare and endangered species. *Journal for Nature Conservation*, *76*, 126505. https://doi.org/10.1016/j.jnc.2023.126505

Meynard, C. N., Leroy, B., & Kaplan, D. M. (2019). Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Holarctic Ecology*, *42*(12), 2021–2036. https://doi.org/10.1111/ecog.04385

Meza Mori, G., Barboza Castillo, E., Torres Guzmán, C., Cotrina Sánchez, D. A., Guzman Valqui, B. K., Oliva, M., Bandopadhyay, S., Salas López, R., & Rojas Briceño, N. B. (2020). Predictive modelling of current and future potential distribution of the spectacled bear (*Tremarctos ornatus*) in amazonas, Northeast Peru. *Animals*, *10*(10), 1816. https://doi.org/10.3390/ani10101816

Monnier-Corbel, A., Robert, A., Hingrat, Y., Benito, B. M., & Monnet, A.-C. (2023). Species distribution models predict abundance and its temporal variation in a steppe bird population. *Global Ecology and Conservation*, *43*, e02442. https://doi.org/10.1016/j.gecco.2023.e02442

Naimi, B., & Araújo, M. B. (2016). Sdm: A reproducible and extensible R platform for species distribution modelling. *Holarctic Ecology*, *39*(4), 368–375. https://doi.org/10.1111/ecog.01881

Ndlovu, P., Mutanga, O., Sibanda, M., Odindi, J., & Rushworth, I. (2018). Modelling potential distribution of bramble (*Rubus cuneifolius*) using topographic, bioclimatic and remotely sensed data in the KwaZulu-Natal Drakensberg, South Africa. *Applied Geography*, *99*, 54–62. https://doi.org/10.1016/j.apgeog.2018.07.025

Nielsen, S. E., Dennett, J. M., & Bater, C. W. (2020). Predicting occurrence, abundance, and fruiting of a cultural keystone species to inform landscape values and priority sites for habitat enhancements. *Forests*, *11*(7), 783. https://doi.org/10.3390/f11070783

Paegelow, M., & Olmedo, M. T. C. (2005). Possibilities and limits of prospective GIS land cover modelling-a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). *International Journal of Geographical Information Science*, *19*(6), 697–722. https://doi.org/10.1080/13658810500076443

Pazmiño, Y. C., de Felipe, J. J., Vallbé, M., & Pazmiño, Y. (2024). Modeling the influence of changes in the edaphic environment on the ecosystem valuation of the zone of influence of the ozogoche and Atillo Lake systems in Ecuador. *Applied Sciences*, *14*(6), 2249. https://doi.org/10.3390/app14062249

Perez, J. (2021). *Diversidad y distribución de las quinas (Cinchona, Rubiaceae) usando dna barcoding para para su conservación en Amazonas*. Universidad Nacional Toribio Rodriguez de Mendoza. https://hdl.handle.net/20.500.14077/2329

Prasetyo, R. B., Kuswanto, H., Iriawan, N., & Ulama, B. S. S. (2020). Binomial regression models with a flexible generalized logit link function. *Symmetry*, *12*(2), 221. https://doi.org/10.3390/sym12020221

Pulgar Vidal, J. (2014). Las ocho regiones naturales del Perú. *Terra Brasilis*, *3*(3). https://doi.org/10.4000/terrabrasilis.1027

Quesquén Esquen, N. J. (2020). *Influencia de bacterias promotoras del crecimiento vegetal en la aclimatación de plántulas de Cinchona officinalis*. Lambayeque–Perú. Universidad Nacional Pedro Ruiz Gallo. https://repositorio.unprg.edu.pe/handle/20.500.12893/9862#

Radha, K. O., & Khwarahm, N. R. (2022). An integrated approach to map the impact of climate change on the distributions of *Crataegus azarolus* and *Crataegus monogyna* in Kurdistan region, Iraq. *Sustainability*, *14*(21), 14621. https://doi.org/10.3390/su142114621

Remuzgo, J. R., Alvarez Melo, J. B., Sales Dávila, F., & Valdivieso Arenas, G. (2021). Taxonomic and phytochemical characterization of *Cinchona pubescens* and *Ladenbergia oblongifolia* in the upper huallaga valley area -Huanuco region. *REBIOL*, *40*(2), 242–255. https://doi.org/10.17268/rebiol.2020.40.02.11

Resquin, F., Duque-Lazo, J., Acosta-Muñoz, C., Rachid-Casnati, C., Carrasco-Letelier, L., & Navarro-Cerrillo, R. M. (2020). Modelling current and future potential habitats for plantations of *Eucalyptus grandis* hill ex maiden and *E. dunnii* maiden in Uruguay. *Forests*, *11*(9), 948. https://doi.org/10.3390/f11090948

Rew, J., Cho, Y., & Hwang, E. (2021). A robust prediction model for species distribution using bagging ensembles with deep neural networks. *Remote Sensing*, *13*(8), 1495. https://doi.org/10.3390/rs13081495

Rojas Briceño, N. B., Cotrina Sánchez, D. A., Barboza Castillo, E., Barrena Gurbillón, M. Á., Sarmiento, F. O., Sotomayor, D. A., Oliva, M., & Salas López, R. (2020). Current and future distribution of five timber forest species in Amazonas, Northeast Peru: Contributions towards a restoration strategy. *Diversity*, *12*(8), 305. https://doi.org/10.3390/d12080305

Safaei, M., Tarkesh, M., Bashari, H., & Bassiri, M. (2018). Modeling potential habitat of *Astragalus verus* Olivier for conservation decisions: A comparison of three correlative models. *Flora: Morphol Distrib Funct Ecol Plants*, *242*, 61–69. https://doi.org/10.1016/j.flora.2018.03.001

Santini, L., Benítez-López, A., Maiorano, L., Čengić, M., Huijbregts, M. A. J., & Fourcade, Y. (2021). Assessing the reliability of species distribution projections in climate change research. *Diversity and Distributions*, *27*(6), 1035–1050. https://doi.org/10.1111/ddi.13252

Schwager, P., & Berg, C. (2021). Remote sensing variables improve species distribution models for alpine plant species. *Basic and Applied Ecology*, *54*, 1–13. https://doi.org/10.1016/j.baae.2021.04.002

Sevillano-Ríos, C. S., & Morales, L. V. (2021). La temperatura y radiación solar explican diferencias en la distribución de dos árboles altoandinos (*Polylepis* spp.) localmente simpátricos en la Cordillera Blanca, Perú. *Neotropical Biodiversity*, *7*(1), 327–340. https://doi.org/10.1080/23766808.2021.1938887

Sintayehu, D. W., Dalle, G., & Bobasa, A. F. (2020). Impacts of climate change on current and future invasion of *Prosopis juliflora* in Ethiopia: Environmental and socio-economic implications. *Heliyon*, *6*(8), e04596. https://doi.org/10.1016/j.heliyon.2020.e04596

Soinne, H., Hyyrynen, M., Jokubė, M., Keskinen, R., Hyväluoma, J., Pihlainen, S., Hyytiäinen, K., Miettinen, A., Rasa, K., Lemola, R., Virtanen, E., Heinonsalo, J., & Heikkinen, J. (2024). High organic carbon content constricts the potential for stable organic carbon accrual in mineral agricultural soils in Finland. *Journal of Environmental Management*, *352*, 119945. https://doi.org/10.1016/j.jenvman.2023.119945

Sousa-Guedes, D., Franch, M., & Sillero, N. (2021). A spatial approach for modeling amphibian road-kills: Comparison of regression techniques. *ISPRS International Journal of geo-Information*, *10*(5), 343. https://doi.org/10.3390/ijgi10050343

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs*, *92*(1). https://doi.org/10.1002/ecm.1486

Vargas Aguilar, C. A. (2021). *Mapas de distribución geográfica histórica y actual de Cinchona* sp. *en la región Amazonas* (Universidad Nacional Toribio Rodriguez de Mendoza). https://hdl.handle.net/20.500.14077/2520

Yoon, S., & Lee, W.-H. (2023). Application of true skill statistics as a practical method for quantitatively assessing CLIMEX performance. *Ecological Indicators*, *146*, 109830. https://doi.org/10.1016/j.ecolind.2022.109830

Zevallos, P. (1989). *Taxonomía, distribución geográfica y status del género Cinchona en el Perú*.

Zielińska, K. M., Kiedrzyński, M., Tołoczko, W., Kiedrzyńska, E., & Mętrak, M. (2024). The edaphic niche of ploidy-different grasses in the light of the coarse-grained data modeling and direct soil sampling. *Ecological Indicators*, *158*, 111548. https://doi.org/10.1016/j.ecolind.2024.111548