



The *Phaseolus vulgaris* L. landraces distribution patterns in Madeira: a modelling approach with maxent

Fabrizio Lopes Macedo · Carla Ragonezi · Carla Gouveia ·
Nuno Nunes · Filipe Ganança · Gregório Freitas · Humberto Nóbrega ·
Miguel Ângelo Almeida Pinheiro de Carvalho

Received: 8 August 2025 / Accepted: 28 September 2025 / Published online: 7 January 2026
© The Author(s) 2025, modified publication 2026

Abstract *Common bean* (*Phaseolus vulgaris* L.) is a widely cultivated legume and a key source of dietary protein globally. This study employed Max-Ent modelling to predict the potential distribution of *P. vulgaris* landraces on Madeira Island, Portugal, where local landraces are predominant. A dataset of 247 occurrence records (2000–2010) was integrated with environmental variables, including altitude, temperature (average, minimum, and maximum), precipitation, evapotranspiration parameters, and soil types. The model demonstrated strong predictive performance (AUC=0.843) with no evidence of overfitting. Jackknife analysis identified altitude as the most influential variable, contributing 76.7% to the model's

gain. Although temperature—particularly maximum temperature—was also important, its effects were closely correlated with altitude, which drives temperature gradients across the island. While precipitation alone had a limited influence, the most suitable areas aligned with zones of higher accumulated rainfall, suggesting a complementary role when thermal and altitudinal conditions are favourable. Response curves indicated that *P. vulgaris* favours lower altitudes, warmer climates (16–23 °C), and regions with elevated real evapotranspiration. Suitability maps confirmed that the northern coastal areas, characterized by low elevations, moderate temperatures, and higher precipitation, offer the most favourable conditions for

F. L. Macedo · C. Ragonezi (✉) · C. Gouveia ·
N. Nunes · F. Ganança · G. Freitas · H. Nóbrega ·
M. Â. A. P. de Carvalho
Center for Sustainable Agriculture and Food Technology,
ISOPlexis, University of Madeira, 9020-105 Funchal,
Portugal
e-mail: carla.ragonezi@staff.uma.pt

F. L. Macedo
e-mail: fabrizio.macedo@staff.uma.pt

C. Gouveia
e-mail: csgouveia@staff.uma.pt

N. Nunes
e-mail: nuno.nunes@staff.uma.pt

F. Ganança
e-mail: jofitei@staff.uma.pt

G. Freitas
e-mail: jgrfreitas@staff.uma.pt

H. Nóbrega
e-mail: humberto.nobrega@staff.uma.pt

M. Â. A. P. de Carvalho
e-mail: miguel.carvalho@staff.uma.pt

F. L. Macedo · C. Ragonezi · C. Gouveia · N. Nunes ·
M. Â. A. P. de Carvalho
Centre for the Research and Technology
of Agro-Environmental and Biological Sciences (CITAB),
Inov4Agro—Institute for Innovation, Capacity Building
and Sustainability of Agri-Food Production, University
of Trás-Os-Montes and Alto Douro, 5000-801 Vila Real,
Portugal

M. Â. A. P. de Carvalho
Faculty of Life Sciences, University of Madeira, Campus
da Penteada, 9020-105 Funchal, Portugal

cultivation. These findings provide valuable insights for the conservation and sustainable management of *P. vulgaris* agrobiodiversity in Madeira.

Keywords Common Bean · Modelling Training · Species Distribution Models · Agrobiodiversity · Climatic Conditions · Environmental variables

Introduction

The daily diet and protein intake of humanity rely heavily on legumes, with the common bean (*Phaseolus vulgaris* L.) being the most cultivated and consumed grain legume (Bitocchi et al. 2017). Its worldwide production has been estimated at 28.5 million tonnes (Mt) of dry grain by 2023 according to FAO (2025). The crop was harvested from a total area of 37.8 million hectares (Mha). String beans or green beans had a total production of 1.4 Mt cultivated across 0.14 Mha. In Europe, dry bean production, with 0.34 Mt in 0.2 Mha, accounts for only 1.2% of the total world production, whereas string bean production, with 1.1 Mt, contributes 78.1% of the world's production (FAO 2025). Portugal, with a production of dry and string beans of 3.6 and 9.9 thousand tonnes respectively, accounts for 1.1% and 0.9% of European common bean production (INE 2024). The Madeira Archipelago, an outermost region of Portugal, with an estimated production of 1.2 and 2.4 thousand tonnes, contributes 34.4% and 25.0% to the Portuguese production of dry and string beans, respectively (DREM 2024). It is important to highlight that bean production in Madeira is predominantly based on local cultivars, such as *landraces* or acclimated cultivars, cultivated under low-input conditions in a territory characterized by highly variable agroecological conditions and a diverse landscape (Pinheiro de Carvalho et al. 2019).

The common beans belong to the *Phaseolus* genus, comprising 70 identified species, including six domesticated crops: *P. vulgaris* L., *P. coccineus* L., *P. dumosus* Macfad [*P. polyanthus* Greenman], *P. lunatus* L., *P. acutifolius* A. Gray, and *P. ritensis* M.E. Jones (Hanelt 2001; Broughton et al. 2003; Bitocchi et al. 2017). In this study, we use the Mansfield classification system for beans (<https://mansfeld.>

[ipk-gatersleben.de/](https://mansfeld.ipk-gatersleben.de/)), which recognizes *P. polyanthus* and *P. ritensis* as local or neglected cultivated species. For the common bean crop, ten taxonomic and agronomic entities are recognized, including two subspecies, two varieties, and the group known as *nuñas* (Hanelt 2001). *P. vulgaris* was initially domesticated in the New World between 8,000 to 10,000 years ago (Debouck et al. 1993). Two distinct domestication events have been documented in the Americas: one in Mesoamerica and the other in the Andes, giving rise to two major domesticated gene pools that can be recognized by their specific *phaseolin* patterns (Debouck et al. 1993; da Silva et al. 2010). The adaptive capacity of *P. vulgaris* to a broad range of environments has led to widespread cultivation and the breakdown of spatial isolation between the Mesoamerican and Andean gene pools, facilitating spontaneous hybridization and increasing the likelihood of novel genotypes and phenotypes (Bitocchi et al. 2017).

The infraspecific diversity of a crop species can typically be expressed by several taxonomic and agronomic entities, namely subspecies, varieties, cultivars, and landraces (Zeven 1998; Hammer and Morimoto 2011). Landraces are the result of crop adaptation to local agroecological conditions and years of natural and human selection, aimed at improving adaptation and yield quality. They are defined as dynamic, heterogeneous populations of a crop species that have evolved within a specific eco-geographical region, adapted to local soil and climate conditions, traditional management practices, and uses (Bareke 2018).

A previous study assessing the morphological diversity of common beans on Madeira Island, including 50 known germplasm accessions from the ISOPlexis Germplasm Bank, characterized 58 morphological traits and identified 15 distinct groups encompassing both traditional *landraces* and local cultivars (Freitas et al. 2011). Due to the predominance of acid andosols and unsaturated cambisols in Madeiran soils, aluminium (Al) toxicity under acidic conditions can be a limiting abiotic factor for crops (Barceló and Poschenrieder 2002). An Al-tolerance assessment of 15 *P. vulgaris* accessions from Madeira showed high tolerance to Al-induced stress, demonstrating their adaptation to inhospitable soils (Domingues et al. 2013). Additionally, nutritional and mineral analyses of 52 local *P.*

vulgaris accessions cultivated under identical conditions revealed significant variability, confirming their inherent genetic diversity (Gouveia et al. 2014). In addition, most of these landraces have developed tolerance to specific abiotic and biotic stresses, showing resistance to pathogens, pests, and various environmental conditions (Newton et al. 2010). Each landrace population can be identified as a subgroup within a category or name used by farmers to distinguish different cultivars (Benz et al. 2007).

To identify the most suitable areas for *P. vulgaris* cultivation on Madeira Island and to understand the limiting factors for its growth, two modelling approaches are available: mechanistic and correlative. Mechanistic models integrate physiological constraints that limit a species' ability to tolerate environmental stress but require comprehensive knowledge of how the species responds to environmental factors, making them difficult to construct (Pearson 2010). In contrast, correlative models determine the influence of environmental conditions on species distribution by linking known occurrence records with environmental variables (Pearson 2010).

Correlative models, also referred to as species distribution models (SDMs), can be analysed using software such as Maximum Entropy (MaxEnt). MaxEnt estimates the probability distribution of maximum entropy based on a set of constraints reflecting limited knowledge about the species' distribution (Phillips et al. 2006). MaxEnt has been used to predict crop yield potential and to support precision agriculture, including the analysis of relationships between soil properties and topography (Norberto et al. 2023). Crops like rice, wheat, maize, and soybean have been modelled using MaxEnt to simulate the distribution of planting suitability areas concerning climate, ecology, and human activity (Taba-Morales et al. 2020; Yang et al. 2022; Fitzgibbon et al. 2022; Ali et al. 2023).

In this study, we developed correlative models using the MaxEnt software based on the correlation between *P. vulgaris* occurrence records on Madeira Island and physical and climatic variables retrieved from several sources, to better understand the potential geographic limits of its distribution.

Methodology

Sampling of *Phaseolus vulgaris* in Madeira

The ISOPlexis—Center for Sustainable Agriculture and Food Technology (University of Madeira, Portugal) has systematically collected information and seed samples from local farmers across Madeira Island over several years. This study is based on data obtained from *Phaseolus vulgaris* cultivars collected during the most intensive and comprehensive sampling period, spanning from 2000 to 2010 (Fig. 1). During this decade, the presence of *P. vulgaris* cultivars was recorded at 247 distinct locations across the island. These geo-referenced occurrence records serve as the foundational dataset for the current analysis and modelling approach. This period (2000–2010) was specifically chosen because it coincides with the most extensive and systematic collection effort of our research group across Madeira Island. The comprehensive nature of this campaign, which covered the entire territory, ensures that the dataset used is the most robust and representative currently available for modelling purposes. It is also important to note that all occurrence records correspond to cultivated stands of landraces maintained under traditional human management. Collections were carried out directly in farmers' fields, ensuring that the dataset reflects both the environmental suitability and the long-standing agricultural practices that have shaped the distribution of common bean in Madeira.

Environmental variables

Given the limited geographical area of Madeira Island (741 km²) and the relatively coarse resolution (~1 km²) of global environmental layers such as those from WorldClim, directly applying these layers could lead to pixel overlap and visual distortion in the images generated by MaxEnt. To mitigate this issue, localized climate data (Table 1) were used from eight meteorological stations across the island, managed by the Portuguese Institute of the Sea and Atmosphere (IPMA). These stations covered the same period as the sampling (2000–2010). The climatic variables used in this study included average, minimum, and maximum temperatures (°C), as well as accumulated precipitation (mm), following Macedo et al. (2023). Monthly values were initially collected for

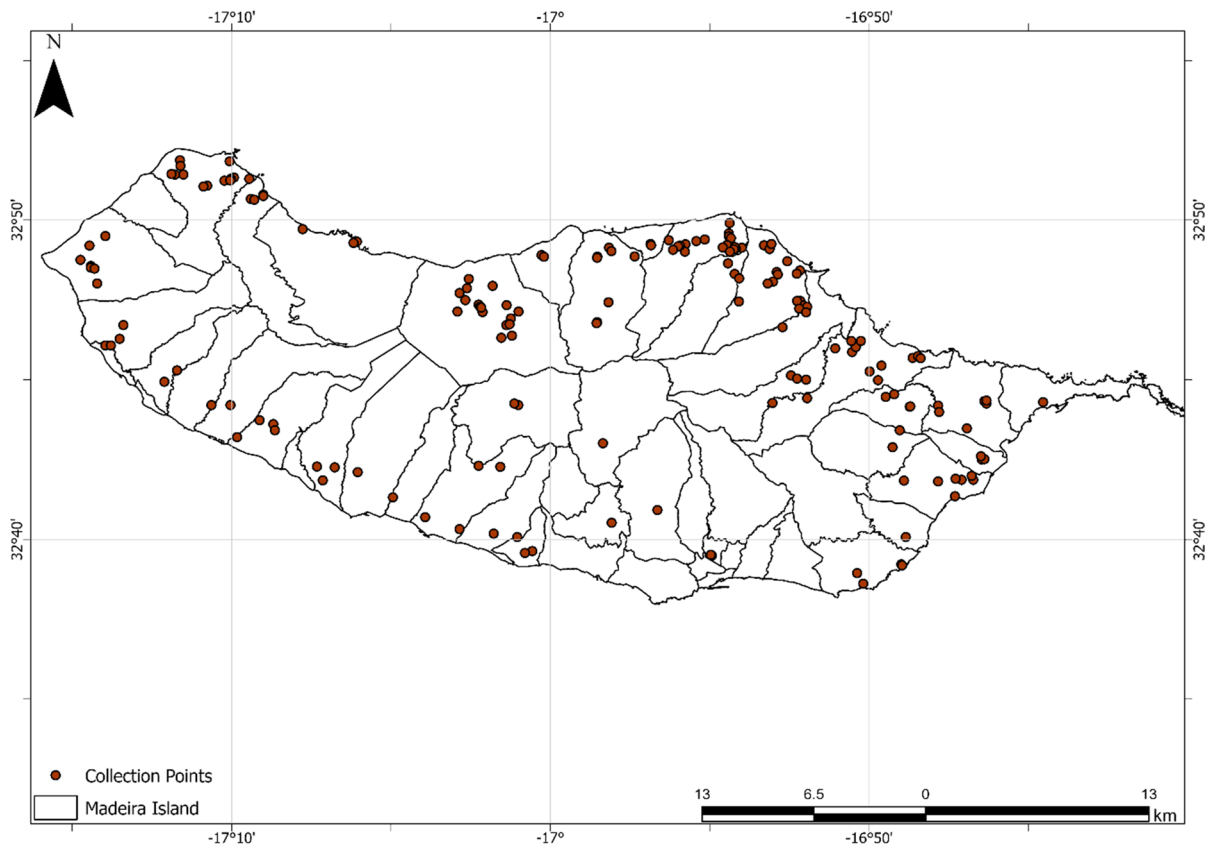


Fig. 1 Sampling sites for *Phaseolus vulgaris* on Madeira Island

Table 1 Spearman's correlation matrix between the study variables

	Tmed	Tmax	Tmin	Alt	Acc_Prec	ETP	ETR	Def
Tmed	—	—	—	—	—	—	—	—
Tmax	0,976**	—	—	—	—	—	—	—
Tmin	0,952**	0,905**	—	—	—	—	—	—
Alt	−0,952**	−0,905**	−1,000**	—	—	—	—	—
Acc_Prec	−0,524	−0,571	−0,381	0,381	—	—	—	—
ETP	0,952**	0,929**	0,881**	−0,881**	−0,476	—	—	—
ETR	−0,405	−0,333	−0,286	0,286	0,619	−0,429	—	—
Def	0,952**	0,929**	0,881**	−0,881**	−0,476	1,000**	−0,429	—

**Correlation is significant at the 0.01 level (2-tailed). Ave_Temp: average temperature, Min_Temp: minimum temperature, Max_Temp: maximum temperature, Alt: altitude, Acc_Prec: accumulated precipitation, ETP: potential evapotranspiration, ETR: real evapotranspiration, Def: water deficit

each station and variable, from which annual averages were subsequently calculated.

Spatial analysis of temperature was performed using annual data on average temperature

(Ave_Temp), minimum temperature (Min_Temp), and maximum temperature (Max_Temp), and was processed according to the methodology described by Santos et al. (2015). Spatial interpolation of

accumulated precipitation (Acc_Prec), Potential Evapotranspiration (ETP), Real Evapotranspiration (ETR), and Water Deficit (Def) was conducted using the Inverse Distance Weighting (IDW) method in ArcGIS Pro 3.4.2 software (ESRI 2022). Altitude data were obtained from the Space Shuttle Radar Topographic Mission (SRTM) v.4.1, provided by the CGIAR-CSI (Farr and Kobrick 2000). Soil type information was provided as a shapefile by the Regional Directorate for Spatial Planning (DROTA) and was converted into numerical values, with each number corresponding to a specific soil class. The soil types included: 1—Rocky Terrain (TRe); 2—Vitric Andosols (Anz); 3—Dystric Rocky Terrain (TRd); 4—Dystric Cambisols (CMd); 5—Dystric Hilly Terrain (TAd); 6—Humic Cambisols (CMu); 7—Chromic Cambisols (CMx); 8—Haplic Phaeozems (PHh); 9—Eutric Hilly Terrain (TAe); 10—Vertic Cambisols (CMv); 11—Eutric Vertisols (VRe); 12—Eutric Fluvisols (FLe); 13—Eutric Cambisols (CMe); 14—Umbric Andosols (Anu); 15—Beach Deposits (Dp); 16—Dystric Fluvisols (FLd); 17—Saline Accumulations (S); 18—Haplic Phaeozems (Phh); 19—Haplic Calcisols (Clh); 20—Calcaric Arenosols (Arc); 21—Urban Area.

All environmental variables were resampled to a 90-m spatial resolution to match the SRTM data. These raster layers were then converted into ASCII format using ArcGIS Pro to ensure compatibility with MaxEnt version 3.4.1 (Phillips et al. 2006).

Data modelling

Species distribution modelling was conducted using MaxEnt software due to its intuitive interface and its proven ability to produce robust predictions using *P. vulgaris* presence-only data combined with environmental variables. MaxEnt operates by estimating the probability distribution of maximum entropy, constrained by the environmental conditions at the presence locations, and comparing it with randomly sampled background points. In this study, 70% of the data was employed to simulate the model, whereas the remaining 30% was used for testing. The following modelling parameters were applied: linear, quadratic, and product features were enabled; auto features were activated; logistic output format was selected; random seed was activated to ensure replicability; the regularization multiplier was set to 3 to avoid overfitting; 10

replicates were performed; the convergence threshold was set to 10^{-5} ; the maximum number of iterations was kept at the default value of 500; and the option to add samples to the background was activated to improve model reliability.

MaxEnt was deliberately selected as the sole modelling approach in this study, as it provides robust results with presence-only datasets, is straightforward to implement, and is widely recognized for its reliability. The objective of this work was not to compare different modelling algorithms, but rather to establish a consistent and representative baseline prediction of the potential distribution of *P. vulgaris* landraces in Madeira. For this reason, alternative methods involving pseudo-absence generation or presence-absence models were intentionally kept outside the scope of this article.

Model evaluation

The species distribution model developed using MaxEnt was evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC) and the Jackknife test. Additionally, the model's output was used to classify potential habitat suitability into five distinct categories. In this study, we chose AUC as the sole evaluation metric because it is threshold-independent, widely used, and well suited for presence-only modelling approaches such as MaxEnt. Alternative statistics, such as TSS or Kappa, require the definition of classification thresholds, which may introduce additional subjectivity. While these complementary metrics could be considered in future research, AUC was deemed sufficient to evaluate the discriminatory capacity of the model in the present work.

The AUC is a key metric for assessing the model's discriminatory power, indicating its ability to distinguish between presence and absence (true positives and true negatives) of the species (Phillips et al. 2006). The values were interpreted according to standard classification thresholds and categorized into five performance levels: failure (0.5–0.6), poor (0.6–0.7), regular (0.7–0.8), good (0.8–0.9), and excellent (0.9–1.0). This framework allows for a clear and objective evaluation of model performance (Krzanowski and Hand 2009).

MaxEnt also produces a logistic output representing the probability of species occurrence, which was used to categorize habitat suitability into five levels: unsuitable (0–0.2), barely suitable (0.2–0.4), suitable (0.4–0.6), highly suitable (0.6–0.7), and very highly suitable (0.7–1.0) (Chang et al. 2020). This classification aids in identifying key areas for the conservation and sustainable management of *P. vulgaris* landraces.

To determine the relative importance of each environmental variable, the Jackknife test was applied. This analysis provides insight into which variables most strongly influence the distribution of *P. vulgaris* by measuring the gain of the model when each variable is used in isolation and when omitted.

Correlation analysis

Correlation analysis was performed using IBM SPSS Statistics, version 29 (IBM Corp 2022). The analysis included eight environmental variables: seven climatic variables collected from eight weather stations across Madeira Island, managed by IPMA, and altitude. The data covered the sampling period from 2000 to 2010. The assumption of normality for each variable was evaluated using quantile–quantile (Q–Q) plots. Visual inspection of these plots revealed that some variables did not follow a normal distribution. Therefore, the non-parametric Spearman's rank-order correlation was applied to assess relationships among variables. The strength of the correlations was interpreted following Akoglu (2018), categorized as: very weak (0.00–0.19), weak (0.20–0.39), moderate (0.40–0.69), strong (0.70–0.89), and very strong (0.90–1.00). These categories apply to both positive and negative values, as the classification is based on the magnitude of the coefficient, while the sign indicates the direction of the relationship.

It is important to note that the correlation analysis was conducted after the modelling process, not as a pre-selection step. This choice was deliberate, as the aim was to interpret the relationships among variables highlighted by MaxEnt rather than to reduce collinearity. The potential effects of correlated predictors were mitigated by the regularization procedures applied in MaxEnt, which are designed to minimize overfitting.

Results and discussion

Modeling results

When analyzing the Area Under the Receiver Operating Characteristic Curve (AUC) value from MaxEnt for the test dataset, the model was classified as having good discrimination (0.8–0.9) (Fig. 2). An AUC of 0.843 indicates a robust model with strong predictive accuracy and high reliability in predicting species presence based on environmental variables (Fielding and Bell 1997). The closer the AUC value is to 1.0, the stronger the model's predictive accuracy.

The training AUC (0.821), indicated by the red line in Fig. 2, was slightly lower than the test AUC (0.843), indicated by the blue line. This difference suggests that the model did not suffer from overfitting. Typically, an overfitted model performs exceptionally well on training data but poorly on test data. The minimal gap between the training and test AUC values demonstrates that the model effectively generalizes to unseen data. The black diagonal line in Fig. 2, representing an AUC of 0.5, signifies a model with no predictive capability. Because both the training and test curves (red and blue) lie well above this line, this confirms that the model performs much better than random chance (Peterson et al. 2007). Furthermore, the model's strong performance on the test data, with an AUC above 0.80, suggests that it effectively captures the relationships between species landrace occurrence and environmental variables. This implies that the model can be confidently used to predict the potential distribution areas of *P. vulgaris*.

Importance of environmental variables

The Jackknife test is commonly used in species distribution research to identify which individual environmental variables contribute most to model performance (Esfanjani et al. 2018; Mahatara et al. 2021). Developed within the MaxEnt framework as described by Phillips et al. (2006), the Jackknife analysis is essential for evaluating the relative importance of environmental variables in predicting species distribution. The graph (Fig. 3) uses color coding to illustrate variable significance: cyan bars indicate model performance when a specific variable is excluded, blue bars represent performance when only one variable is used, and the red bar shows

performance when all environmental variables are included (Elith et al. 2011).

In Fig. 3, the percentage contribution reflects the extent to which a variable directly contributes to the model's gain, while the permutation importance indicates the overall impact of a variable on model performance, based on the effect of its removal. The results show that Alt is the most influential variable, as its exclusion leads to a substantial decrease in predictive accuracy, and it retains strong predictive power when used independently. Its high percentage contribution (76.7%) and permutation importance (78.8%) confirm that Alt is the primary factor influencing habitat suitability. This suggests that *P. vulgaris* exhibits a strong preference for specific elevation ranges, making altitude a critical determinant of suitable habitat.

Temperature-related variables such as Ave_Temp, Max_Temp, and Min_Temp also play important roles. Their removal results in a decline in training

gain, and each demonstrates meaningful predictive value when considered individually. Max_Temp, for example, had a modest contribution (2.8%) but a relatively high permutation importance (15.8%). This indicates that while it is not the dominant factor, changes in maximum temperature significantly affect habitat suitability—an observation consistent with the findings of Elith et al. (2011). In contrast, Min_Temp had a lower contribution (1.8%) and permutation importance (2.4%), while Ave_Temp showed the least relevance, with a contribution of only 1.2% and no permutation importance, suggesting it is not a key determinant.

Other variables, such as Acc_Prec and soil, contributed minimally to the model. Specifically, Acc_Prec accounted for 2.1% of the contribution, and soil only 0.6%. Their low permutation importance values—0.9 for precipitation and 0.2 for soil—further indicate that changes in these variables have little impact on habitat suitability predictions. The limited contribution of soil type also supports its minor role in shaping species distribution. By contrast, altitude has consistently been identified as a major factor

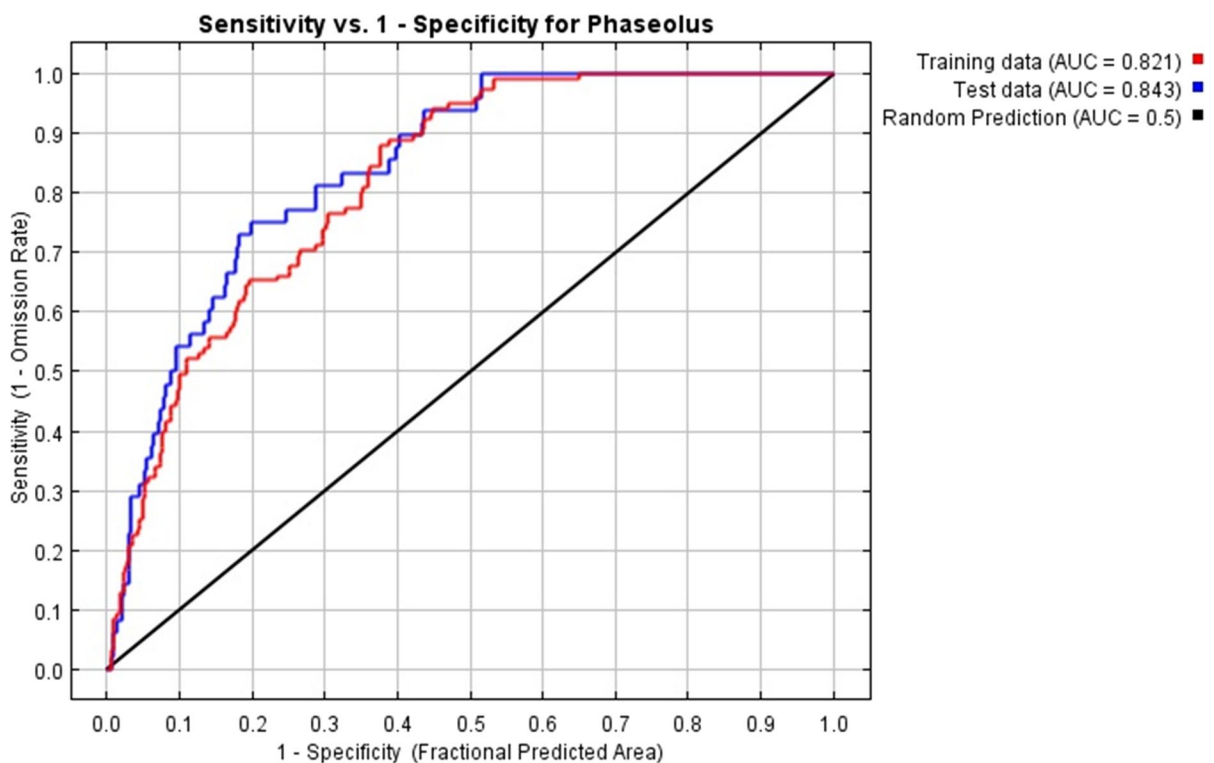


Fig. 2 Receiver Operating Characteristic (ROC) curve of the MaxEnt model

shaping the distribution of *Phaseolus* species, with wild beans in Mexico occurring across a wide altitudinal range, from 8 to more than 3000 m above sea level, reflecting strong adaptation to elevation-driven climatic gradients (Cerdeira-Hurtado et al. 2018). These findings align with previous research emphasizing the predominance of topographic and climatic factors in species distribution modeling (Merow et al. 2013). Evapotranspiration-related variables, including ETP, ETR, and Def, also influenced the model, particularly when combined with other variables. While the removal of these variables decreased the training gain, they had a limited impact individually. For example, Def contributed 7.7% to the model but had a permutation importance of only 1.5%, indicating a relatively weak standalone influence. Similarly, ETR contributed 4.6% but had a very low permutation importance (0.3), suggesting it is not a primary driver of distribution.

The Jackknife test confirmed that multiple variables interact to enhance model performance, with altitude contributing more than 70% to the probability of *P. vulgaris* landrace distribution. This analysis underscores altitude, followed by maximum temperature, as the most influential environmental factors. These results highlight the importance of focusing on these variables in species management and conservation strategies (Merow et al. 2013). Although temperature variables were strongly correlated with altitude and therefore partially redundant, they were deliberately retained in the model because of their interpretative value in showing how elevation-driven thermal gradients influence the suitability of *P. vulgaris*. Their inclusion did not compromise the model's performance, as MaxEnt incorporates regularization mechanisms to mitigate multicollinearity, and the model showed no evidence of overfitting.

Once altitude was identified as the most significant variable, a correlation analysis was performed to examine relationships among the environmental variables (Table 1).

There were strong and statistically significant negative correlations between altitude and average, maximum, and minimum temperatures ($p < 0.01$). As terrain altitude increased, all three temperature variables analyzed in this study decreased accordingly. In contrast, the correlation between altitude and accumulated precipitation was weak, suggesting a limited relationship between elevation and rainfall in the

study area. Additionally, there were strong, significantly negative correlations between altitude and both potential evapotranspiration (ETP) and water deficit ($p < 0.01$). This indicates that as altitude increases, both potential evapotranspiration and water deficit tend to decrease. However, real evapotranspiration (ETR) did not show any significant correlation with altitude, implying that actual water use by vegetation may not be directly influenced by elevation.

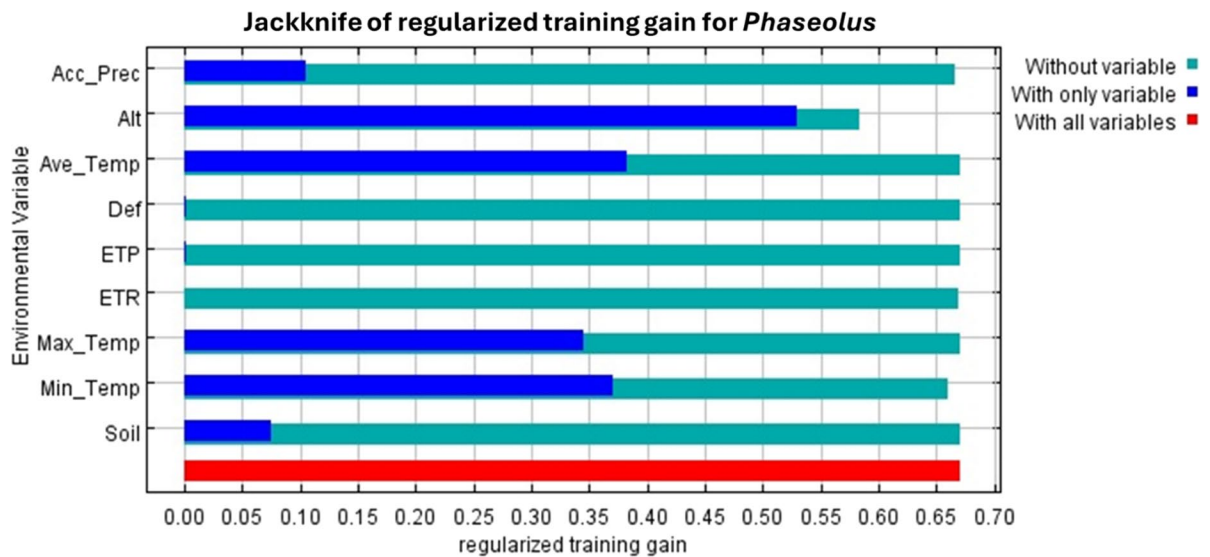
Individual response curves

The response curves in Fig. 4 illustrate how the presence probability of *P. vulgaris* varies concerning each environmental predictor variable used in the MaxEnt model. These curves provide insights into the species' ecological preferences by showing the influence of individual variables on habitat suitability, while keeping all other variables at their average value.

The response to altitude (Fig. 4b) revealed a steep decline in habitat suitability, with the highest suitability occurring at low elevation. As altitude increased, suitability decreased sharply, indicating that *P. vulgaris* favours lower altitudes and is less likely to thrive at higher elevations (> 500 m a.s.l.).

These curves are useful for assessing how individual environmental factors influence species distribution by illustrating changes in habitat suitability as each variable varies, while all others are held constant. The response curve for accumulated precipitation (Acc_Prec; Fig. 4a) shows a negative correlation with habitat suitability, indicating that higher precipitation levels lead to reduced suitability. This suggests that *P. vulgaris* prefers regions with moderate to low precipitation. The response to Alt revealed a sharp decline in habitat suitability as elevation increases, with the highest suitability observed at lower altitudes. Suitability decreased significantly above 500 m a.s.l., indicating that *P. vulgaris* favours lower elevations and is less likely to thrive at higher altitudes.

The soil variable (Fig. 4i) exhibited a unimodal (hump-shaped) response, suggesting that habitat suitability increases up to an optimal soil condition and then declines. This indicates that *P. vulgaris* is best adapted to specific soil types, particularly Umbric Andosols (classification 14), with suitability decreasing beyond this optimal range. For potential evapotranspiration (ETP; Fig. 4e) and water deficit (Def; Fig. 4d), the response curves showed



Variable	Percent contribution	Permutation importance
Acc_Prec	2.1	0.9
Alt	76.7	78.8
Ave_Temp	1.2	0
Def	7.7	1.5
ETP	2.5	0
ETR	4.6	0.3
Max_Temp	2.8	15.8
Min_Temp	1.8	2.4
Soil	0.6	0.2

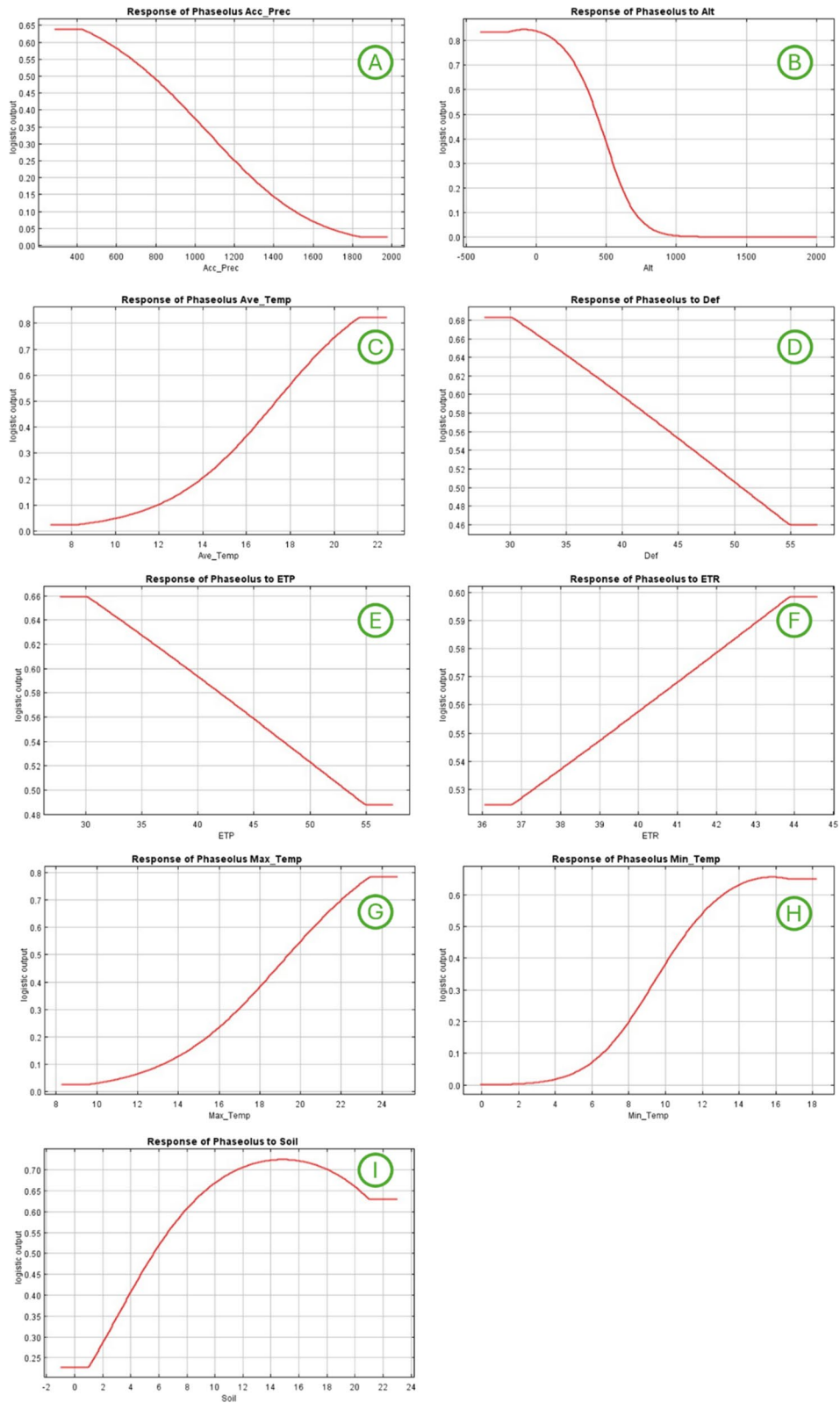
Fig. 3 Jackknife test for environmental variables, percentage contribution, and permutation importance of each variable. Environmental Variables: Acc_Prec—accumulated precipitation; Alt—altitude; Ave_Temp—average temperature; Def—

water deficit; ETP—potential evapotranspiration; ETR—real evapotranspiration; Max_Temp—maximum temperature; Min_Temp—minimum temperature

negative relationships... In contrast, the real evapotranspiration (ETR; Fig. 4f) response curve showed an increasing trend, indicating that higher real evapotranspiration values are associated with greater habitat suitability. It should be noted, however, that the relatively simple (linear-like) trends observed for ETP, Def, and ETR largely reflect the limited variability of these variables within the 2000–2010 dataset. Because the sampling period corresponds to relatively stable climatic conditions in Madeira, the model was unable to capture more complex or non-linear response patterns that might emerge under a wider range of climatic scenarios. Therefore, these relationships should be interpreted with caution, as they likely represent constrained responses within

the available data range rather than genuine linear ecological effects. This suggests that *P. vulgaris* thrives in environments where actual water use by vegetation is higher, potentially reflecting regions with more stable moisture availability.

Temperature-related variables, including average temperature (Ave_Temp; Fig. 4c), maximum temperature (Max_Temp; Fig. 4g), and minimum temperature (Min_Temp; Fig. 4h), all displayed positive correlations with habitat suitability. Suitability increased as temperatures rose from 16 °C to approximately 23 °C, indicating a preference for warmer climates. This trend was especially pronounced for Min_Temp, where very low minimum



◀**Fig. 4** The mean response curves for the tested predictor variables for the *Phaseolus vulgaris* model were created using MaxEnt. Environmental Variables graphs: A—Acc_Prec, accumulated precipitation; B—Alt, altitude; C—Ave_Temp, average temperature; D—Def, water deficit; E—ETP, potential evapotranspiration; F—ETR, real evapotranspiration; G—Max_Temp, maximum temperature; H—Min_Temp, minimum temperature; and I—Soil types

temperatures (< 10 °C) significantly reduced the likelihood of crop presence.

Overall, the response curves indicate that *P. vulgaris* is most suited to environments with moderate precipitation, lower altitudes, specific soil conditions (notably Umbric Andosols), and higher real evapotranspiration levels. The crop also prefers warmer temperatures—around 23 °C—as shown by the strong positive associations with average, maximum, and minimum temperature variables. These insights are valuable for ecological modelling and conservation planning, as they help identify the key environmental drivers shaping the species' distribution and may inform future research on potential local adaptations among landraces across different regions.

Potential distribution *Phaseolus vulgaris*

The hypsometric map (Fig. 5) illustrates the distribution of *P. vulgaris* across an altitudinal gradient, with altitude (Alt) represented using a colour scale. Green shades indicate lower elevations, while brown and red represent mid-altitude areas, and greyish-white denotes the highest elevations, reaching up to 1,842 m above sea level. Yellow dots indicate the specific locations where *P. vulgaris* accessions were collected. A noticeable concentration of accessions is observed in mid-to-lower elevation zones, particularly along the north coast, highlighting the species' preference for these altitudinal ranges.

The distribution of *P. vulgaris* across the island was uneven. The hypsometric map revealed that this crop is primarily found in mid- to lower-elevation regions, with limited or no presence in the highest mountainous zones. This spatial pattern aligns with the MaxEnt model outputs (Fig. 4), which identified altitude as the most influential environmental variable for predicting species presence. The absence or low occurrence of *P. vulgaris* in higher altitude areas may be attributed to specific microclimatic

conditions—such as lower temperatures, increased wind exposure, or shorter growing seasons—as well as edaphic (soil-related) factors that characterize harsher environments.

The higher concentration of *P. vulgaris* at lower to mid-elevations likely reflects the plant's climatic and agronomic requirements, including warmer temperatures, adequate precipitation, and a sufficiently long growing season. As a cultivated species, its distribution is also influenced by human activities, particularly agricultural practices that are more prevalent along the North Coast. Additionally, ecological factors such as soil composition, water availability, and interspecific competition may contribute to the observed distribution patterns. The spatial clustering of occurrence points within specific altitudinal zones highlights the species' agroecological preferences, emphasizing the crucial role of altitude-related environmental gradients in shaping habitat suitability.

The habitat suitability map generated by the MaxEnt model (Fig. 6) classifies the island into five categories: unsuitable (red), barely suitable (yellow), suitable (light green), highly suitable (green), and very highly suitable (blue). This classification is mainly driven by altitude, identified as the most influential predictor in the model. The values of Ave_Temp, Min_Temp, Max_Temp, and Acc_Prec shown in Fig. 6 correspond to characteristic agroecological conditions associated with each suitability class, while the MaxEnt model itself was constructed using the full set of eight environmental and soil variables. The red dots representing actual collection points show a strong correspondence with areas predicted to have high to very high suitability, further validating the model's accuracy in identifying environmentally favourable areas for *P. vulgaris* growth.

Note. Ave_Temp, Min_Temp, and Max_Temp correspond to annual mean values of average, minimum, and maximum temperature, respectively, while Acc_Prec corresponds to the accumulated annual precipitation for each suitability class.

Figure 6 illustrates the spatial distribution of habitat suitability for *Phaseolus vulgaris* across Madeira Island. The central region of the island—encompassing the protected area of the Madeira Natural Park and largely occupied by *Laurisilva* rainforest—predominantly shows unsuitable conditions for *P. vulgaris*. Although this area may support traditional agroecosystems with common bean populations, its

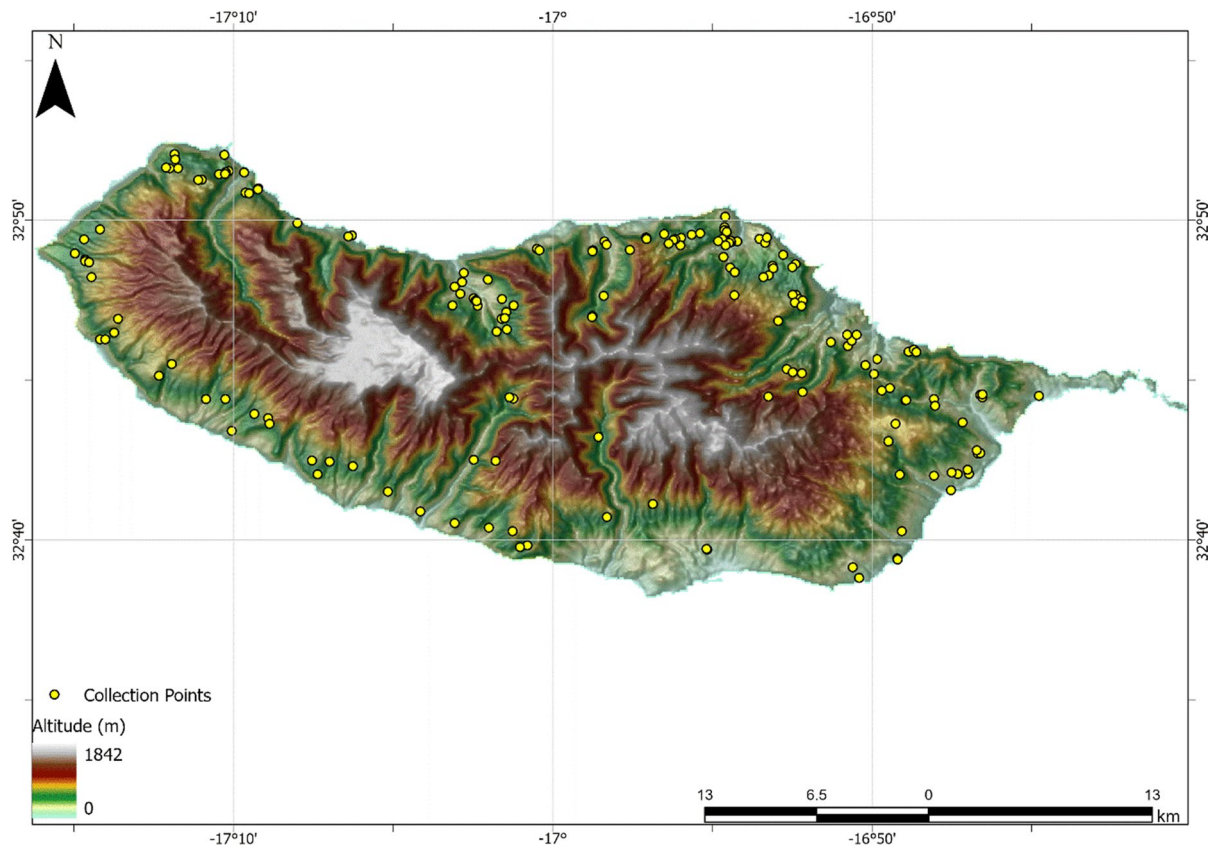


Fig. 5 Hypsometric map showing the presence of *Phaseolus vulgaris* (yellow dots) across an altitudinal gradient. Colour scheme: green indicates lower elevations, brown and red represent mid-altitude regions, and greyish-white denotes the highest elevations

generally lower temperatures and higher precipitation likely limit crop suitability. In contrast, the northern coastal areas exhibit very highly suitable agroecological conditions. These are attributed to a favourable combination of higher precipitation and moderate average temperatures. Transition zones—classified as barely suitable and suitable—are situated between the unsuitable inland and highly suitable coastal regions. These zones are characterized by relatively higher temperatures and reduced precipitation. Notably, several common bean accessions were collected in these barely suitable areas, indicating some level of adaptation or resilience.

Most of the collection points (represented by red dots) are located within the suitable, highly suitable, and very highly suitable zones, demonstrating strong alignment with the MaxEnt model's predictions. This supports the model's accuracy and reinforces the conclusion that altitude is the most critical factor

influencing habitat suitability. Temperature, which correlates strongly with altitude, also plays a key role. Additionally, precipitation contributes significantly, particularly in very highly suitable areas where accumulated rainfall is greatest. However, the presence of *P. vulgaris* in areas classified as barely suitable or even unsuitable suggests that some populations may be adapted to less favourable conditions, or that other influential factors—such as microclimatic variation, soil properties, or human cultivation practices—are not fully captured by the model. Overall, this habitat suitability map provides valuable insights into the optimal agroecological conditions for *P. vulgaris* cultivation on Madeira Island, emphasizing the roles of altitude, temperature, and precipitation. These findings are essential for guiding germplasm prospecting and conservation efforts, supporting agricultural planning, identifying useful adaptive traits, and assessing

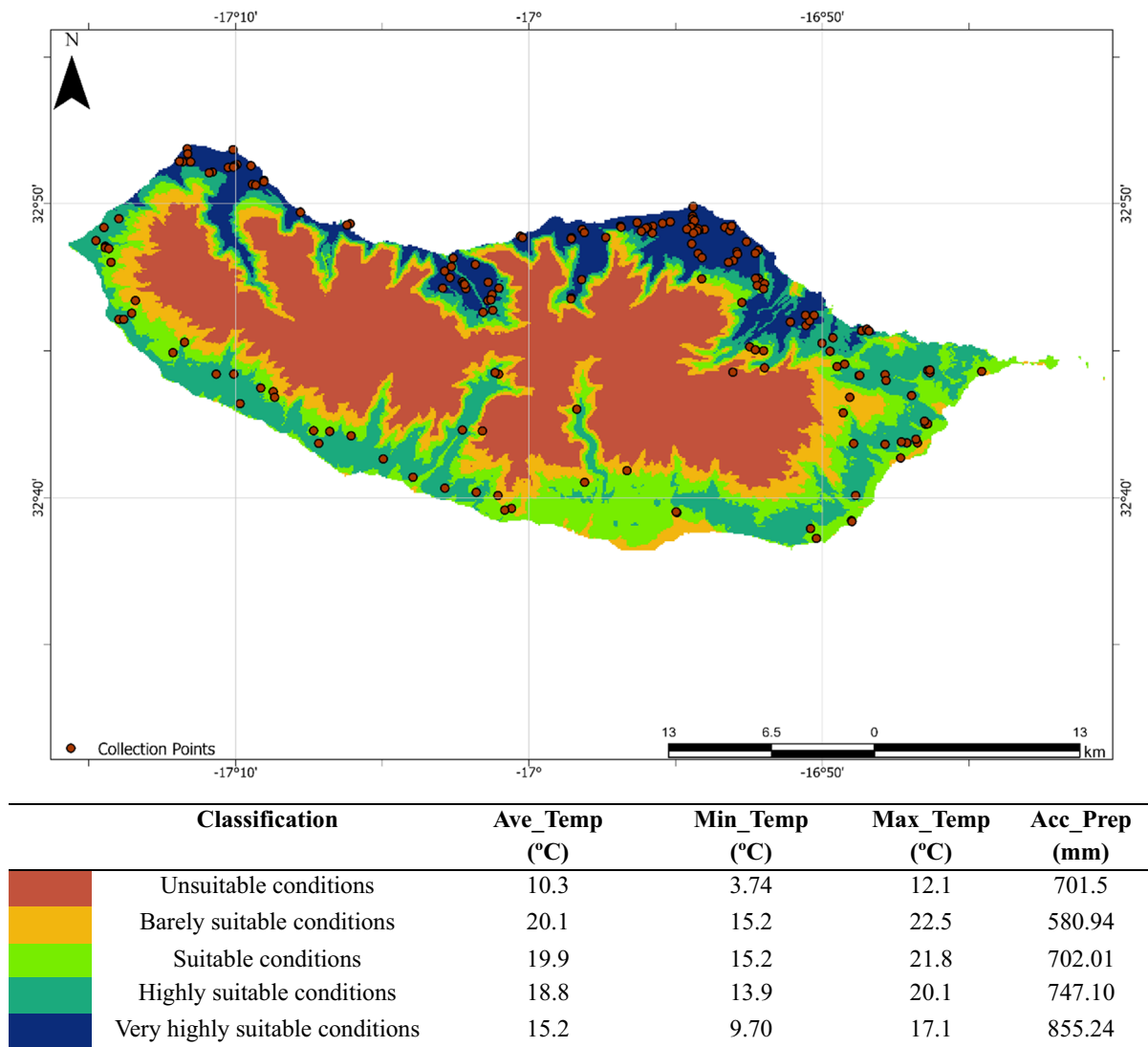


Fig. 6 Suitability of agroecological conditions for *Phaseolus vulgaris* occurrence probability

the potential impacts of climate change on common bean cultivation.

Conclusion

The model demonstrated good discriminatory ability, achieving an AUC of 0.843 on the test data, indicating high reliability in predicting the presence of *P. vulgaris* based on environmental variables and showing no evidence of overfitting. Jackknife analysis identified altitude as the most influential variable,

contributing 76.7% to the model and exhibiting a permutation importance of 78.8%. This highlights a strong preference of *P. vulgaris* for specific agroecological conditions determined primarily by elevation. Specifically, suitability is highest at low altitudes and declines sharply above 500 m above sea level. Correlation analysis supported this finding, revealing strong and significant negative correlations between altitude and temperature variables (average, maximum, and minimum temperatures). Therefore, temperature variables are important in the model largely because they are strongly linked to elevation, acting as secondary

factors mediated by the altitude-driven temperature gradient across Madeira Island.

Among temperature-related variables, maximum temperature (Max_Temp) showed particular importance. Although its effect is largely explained by its correlation with altitude, excluding Max_Temp caused a decrease in model training gain, and alterations to this variable substantially impacted agroecological suitability. The crops thrive in warmer climates, with suitability increasing as temperature rises from 16 °C to 23 °C, while very low minimum temperatures (< 10 °C) significantly restrict species presence. Evapotranspiration-related variables, such as ETP, ETR, and water deficit (Def), also contribute significantly, especially when combined with other predictors. Accumulated precipitation (Acc_Prec), while having limited influence alone, corresponds to regions identified as most suitable due to their higher precipitation levels, suggesting it acts as a complementary factor alongside temperature and altitude in defining optimal habitats.

The potential distribution maps confirmed that the most suitable habitats for *P. vulgaris* are in the northern coastal regions of Madeira, characterized by favourable temperature, altitude, and precipitation conditions. Most recorded presence points fall within "Suitable" or "Highly Suitable" areas, though some occurrences in "Barely Suitable" or "Unsuitable" zones may reflect influences not captured by the model, such as microclimate variation, soil properties, or human activity. In summary, this study confirms that altitude is the primary driver of *P. vulgaris* distribution on Madeira Island, affecting both direct habitat suitability and indirectly shaping temperature gradients. Temperature acts as a key secondary variable strongly correlated with elevation, while precipitation plays a complementary role in defining the most favourable agroecological zones. These findings not only help define the optimal environmental conditions for bean cultivation in Madeira but also highlight the potential of cultivars found outside the main distribution zones as candidates for adaptation to sub-optimal environments.

It is important to note that the occurrence records used in this study correspond to the most comprehensive and systematic sampling of *P.*

vulgaris landraces in Madeira (2000–2010), a period that also reflects relatively stable agricultural practices and the predominance of traditional landraces, ensuring the robustness and representativeness of the modelling results. The selection of this period was deliberate, as it coincides with the most complete prospection performed by our group across the island, providing consistency and reliability for the modelling outputs. While more recent occurrence data exist, they are fragmented and localized, and therefore were not included to avoid introducing bias.

Nevertheless, it should be acknowledged that the model does not provide a definitive estimation of the optimal environmental conditions for *P. vulgaris*, but rather a robust approximation based on the available data. This approximation is valuable to guide future, more detailed research that integrates additional ecological, agronomic, and climatic variables.

Acknowledgements . This work is supported by National Funds by FCT – Portuguese Foundation for Science and Technology, under the projects UID/04033/2025: Centre for the Research and Technology of Agro-Environmental and Biological Sciences (<https://doi.org/10.54499/UID/04033/2025>) and LA/P/0126/2020 (<https://doi.org/10.54499/LA/P/0126/2020>). The authors also acknowledge the farmers for the studied sites availability and the Secretaria Regional de Agricultura e Desenvolvimento Rural of Madeira for partnership and support.

Author contributions All authors contributed to the study conception and design. Conceptualization: All authors; Investigation: All authors; Writing—original draft preparation: F.L.M., C.R., C.G., N.N.; Writing—review and editing: F.L.M., C.R., C.G., N.N., and M.A.A.P.d.C.; Supervision: M.A.A.P.d.C.; Funding acquisition: M.A.A.P.d.C. All authors read and approved the final manuscript.

Funding This work was supported by the ClimAGRO4.0 project—Technological Innovation in Monitoring, Production, Sustainability and Climate Adaptation of Agri-food Systems, funded under the Madeira 2030 Program (Priority Axis 1A—Madeira + Smart and Competitive: Knowledge and Innovation), with co-financing from FEDER and the University of Madeira, ID: M2030-FEDER-03050600.

Data availability The datasets generated during and/or analysed during the current study are not publicly available due to ongoing related research but are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akdoglu H (2018) User's guide to correlation coefficients. *Turk J Emerg Med* 8(3):91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>
- Ali S, Makanda TA, Umair M, Ni J (2023) MaxEnt model strategies to studying current and future potential land suitability dynamics of wheat, soybean and rice cultivation under climatic change scenarios in East Asia. *PLoS ONE* 18(12):e0296182. <https://doi.org/10.1371/journal.pone.0296182>
- Barceló J, Poschenrieder C (2002) Fast root growth responses, root exudates, and internal detoxification as clues to the mechanisms of aluminium toxicity and resistance: a review. *Environ Exp Bot* 48:75–92. [https://doi.org/10.1016/S0098-8472\(02\)00013-8](https://doi.org/10.1016/S0098-8472(02)00013-8)
- Bareke T (2018) Diversity of common bean (*Phaseolus vulgaris* L., Fabaceae) landraces in parts of southern and eastern Ethiopia. *APAR* 8:449–457. <https://doi.org/10.15406/apar.2018.08.00365>
- Benz B, Perales H, Brush S (2007) Tzeltal and Tzotzil farmer knowledge and maize diversity in Chiapas, Mexico. *Curr Anthropol* 48:289–300. <https://doi.org/10.1086/512986>
- Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia R, Santo D, Nanni L, Attene G, Papa R (2017) Beans (*Phaseolus* spp.) as a model for understanding crop evolution. *Front Plant Sci* 8:722. <https://doi.org/10.3389/fpls.2017.00722>
- Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. *Plant Soil* 252:55–128. <https://doi.org/10.1023/A:1024146710611>
- Cerda-Hurtado IM, Mayek-Pérez N, Hernández-Delgado S, Muruaga-Martínez JS, Reyes-Lara MA, Reyes-Valdés MH, González-Prieto JM (2018) Climatic adaptation and ecological descriptors of wild beans from Mexico. *Ecol Evol* 8:6492–6504. <https://doi.org/10.1002/ece3.4106>
- Chang YL, Xia Y, Peng MW, Chu GM, Wang M (2020) MaxEnt modelling for predicting impacts of climate change on the potential distribution of *Anabasis aphylla* in northwestern China. *Appl Ecol Env Res* 18:1637–1648. https://doi.org/10.15666/aeer/1801_16371648
- IBM Corp (2022) *IBM SPSS Statistics for Windows* (Version 29.0) [Software]. IBM Corp. Available from: <https://www.ibm.com/products/spss-statistics>. Accessed 10 May 2025.
- da Silva EM, Correia ASA, Lopes NAA, Nóbrega HGM, Ganança JFT, Domingues AM, Khadem M, Slaski JJ, Pinheiro de Carvalho MAA (2010) Phytogeographical origin of Madeiran common beans based on phaseolin patterns. *Pesq Agropec Bras* 45(8):863–871. <https://doi.org/10.1590/S0100-204X2010000800012>
- de Pinheiro Carvalho MAA, Ragonezi C, de Lopes Macedo F, Antunes G, Freitas G, Nóbrega H (2019) Contributo para o conhecimento da agrobiodiversidade no concelho de Santa Cruz, Madeira. *Rev Cienc Agr* 42(3):575–605. <https://doi.org/10.19084/rca.16658>
- Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* (Fabaceae) in northwestern South America. *Econ Bot* 47:408–423. <https://doi.org/10.1007/BF02907356>
- Domingues AM, Silva E, Freitas G, Ganança JF, Nóbrega H, Slaski JJ, de Pinheiro Carvalho MAA (2013) Aluminium tolerance in bean traditional cultivars from Madeira. *Rev Cienc Agr* 36:148–156. <https://doi.org/10.19084/rca.16292>
- DREM (2024) Estatísticas da Agricultura e Pesca da Região Autónoma da Madeira. Direção Regional de Estatística da Madeira, Funchal.
- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Divers Distrib* 17(1):43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- Esfanjan J, Ghorbani A, Zare Chahouki MA (2018) MaxEnt modeling for predicting impacts of environmental factors on the potential distribution of *Artemisia aucheri* and *Bromus tomentellus*–*Festuca ovina* in Iran. *Pol J Environ Stud* 27(3):1041–1047. <https://doi.org/10.15244/pjoes/76496>
- Esri (2022) *ArcGIS Pro* (Version 3.4.2) [Software]. Environmental Systems Research Institute. Available from: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>. Accessed 10 May 2025.
- FAO (2025) FAOSTAT (Food and agriculture data). Rome: FAO. Available from: <https://www.fao.org/faostat/en/#data/QCL>. Accessed 18 April 2025.
- Farr TG, Kobrick M (2000) Shuttle radar topography mission produces a wealth of data. *Eos Trans Amer Geo Union* 81(48):583–585
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24(1):38–49. <https://doi.org/10.1017/S0376892997000088>
- Fitzgibbon A, Pisut D, Fleisher D (2022) Evaluation of maximum entropy (MaxEnt) machine learning model to assess relationships between climate and corn

- suitability. *Land* 11(9):1382. <https://doi.org/10.3390/land11091382>
- Freitas G, Ganança JFT, Nóbrega H, Nunes É, Costa G, Slaski JJ, de Pinheiro Carvalho MAA (2011) Morphological evaluation of common bean diversity on the island of Madeira. *Genet Resour Crop Evol* 58:861–874. <https://doi.org/10.1007/s10722-010-9624-y>
- Gouveia CSS, Freitas G, de Brito JH et al (2014) Nutritional and mineral variability in 52 accessions of common bean varieties (*Phaseolus vulgaris* L.) from Madeira Island. *Agric Sci* 5:317–329. <https://doi.org/10.4236/as.2014.54034>
- Hammer K, Morimoto Y (2011) Classifications of infraspecific variation in crop plants. In: Guarino L, Ramanatha Rao V, Goldberg E (eds) *Collecting plant genetic diversity: technical guidelines*. Accessed 13 June 2025.
- Hanelt P (2001) *Mansfeld's Encyclopedia of Agricultural and Horticultural Crops*. Institute of Plant Genetics and Crop Plant Research, Springer.
- INE (2024) Estatísticas Agrícolas – 2023. Instituto Nacional de Estatística, Lisboa.
- Krzanowski WJ, Hand DJ (2009) *ROC curves for continuous data*, 1st ed. CRC Press, Taylor and Francis Group, 332 p. <https://doi.org/10.1201/9781439800225>
- Macedo FL, Ragonezi C, Reis F, de Freitas JGR, Lopes DH, Aguiar AMF, Cravo D, Carvalho MAAPd (2023) Prediction of the potential distribution of *Drosophila suzukii* on Madeira Island using the maximum entropy modeling. *Agriculture* 13(9):1764. <https://doi.org/10.3390/agriculture13091764>
- Mahatara D, Acharya AK, Dhakal BP, Sharma DK, Ulak S, Paudel P (2021) MaxEnt modelling for habitat suitability of vulnerable tree *Dalbergia latifolia* in Nepal. *Silva Fenn* 55(4):e10441. <https://doi.org/10.14214/sf.10441>
- Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36(10):1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Newton AC, Akar T, Baresel JP, Bebeli PJ, Bettencourt E, Bladenopoulos KV, Czembor JH, Fasoula DA, Katsiotis A, Koutis A, Koutsika-Sotitiou M, Kovacs G, Larsson H, de Pinheiro Carvalho MAA, Rubiales D, Russell J, Dos Santos TMM, Vaz Patto MC (2010) Cereal landraces for sustainable agriculture: a review. *Agric Sci Dev* 30:237–269. <https://doi.org/10.1051/agro/2009032>
- Norberto M, Sillero N, Coimbra J, Cunha M (2023) Filling the maize yield gap based on precision agriculture – a MaxEnt approach. *Comput Electron Agric* 211:107970. <https://doi.org/10.1016/j.compag.2023.107970>
- Pearson RG (2010) Network of Conservation Educators and Practitioners species distribution modeling for conservation educators and practitioners. *LiC* 3:54–89. <https://doi.org/10.5531/cbc.linc.3.1.3>
- Peterson AT, Papes M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30(4):550–560. <https://doi.org/10.1111/j.0906-7590.2007.05102.x>
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190(3–4):231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Santos AR, Ribeiro CAAS, Sediya GC, Peluzio JBE, Pez-zopane JBM, Bragança R (2015) Espacialização de dados meteorológicos no ArcGIS 10.3: passo a passo. CAUFES—Centro Agropecuária da Universidade Federal do Espírito Santo, Vitória, Brasil.
- Taba-Morales G, Hyman G, Mejía JR, Castro-Llanos F, Beebe S, Rubygo JC, Katungi E, Buruchara R (2020) Improving African bean productivity in a changing global environment. *Mitig Adapt Strateg Glob Change* 25:1013–1029. <https://doi.org/10.1007/s11027-019-09910-4>
- Tom G, Farr Mike, Kobrick (2000) Shuttle radar topography mission produces a wealth of data. *Eos Trans Amer Geoph Union* 81(48): 583–585. <https://doi.org/10.1029/EO081i048p00583>
- Yang J, Huang Y, Jiang X, Chen H, Liu M, Wang R (2022) Potential geographical distribution of the endangered plant *Isoetes* under human activities using MaxEnt and GARP. *Glob Ecol Conserv* 38:e02186. <https://doi.org/10.1016/j.gecco.2022.e02186>
- Zeven A (1998) Landraces: a review of definitions and classifications. *Euphytica* 104:127–139. <https://doi.org/10.1023/A:1018683119237>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.