Distribution model of the Iberian Desman (*Galemys pyrenaicus*) and forecasting the impact of climate change on its ecological niche

Sara Pinto

NOVA University of Lisbon – NOVA IMS, 20230915@novaims.unl.pt

Abstract: Through this work, a distribution model of the iberian desman (*Galemys pyrenaicus*), a species endemic to the Iberian Peninsula and the French Pyrenees, was developed to analyze its current distribution and project the future impacts of climate change. Using historical bioclimatic data and future projections, as well as anthropogenic variables, MaxEnt software was used to model the ecological niche of the species. The results indicated that only a small part of the current habitat is suitable for the species, with future projections suggesting a further decrease in habitat suitability due to rising temperatures and changes in the summer water balance. The distribution models produced are important tools for predicting the impacts of climate change and for guiding conservation efforts, ensuring that effective measures are implemented to protect critical habitats and mitigate future threats to the survival of the iberian desman.

Key words: iberian desman; distribution model; MaxEnt; Climate change

1. Introduction

The iberian desman (*Galemys pyrenaicus*) is a semi-aquatic insectivorous mammal endemic to the Iberian Peninsula and the French Pyrenees (Mathias et al., 2023). Adapted to cold climates, it occurs mainly in clean and oxygenated mountain water lines, feeding on aquatic insects, and seeking shelter and refuge in the contiguous riparian galleries (Mathias et al., 2023).

In addition to the distribution of the species being dependent on these freshwater habitats, it is very sensitive to their quality. For this reason, the gradual degradation and fragmentation of these ecosystems has caused a sharp contraction in their range and population (Quaglietta et al., 2018).

Although the historical distribution (determined through field sampling) of the iberian desman overlaps with a large part of the district of Bragança, Vila Real, Braga and Viseu, and partially with the districts of Guarda and Castelo Branco) (Figure 1), recent studies indicate that in the last 10 years, its distribution area has decreased by 30% (Quaglietta et al., 2018)). According to the latest estimate, there are fewer than 10,000 mature individuals in the entire national territory (Mathias et al., 2023)

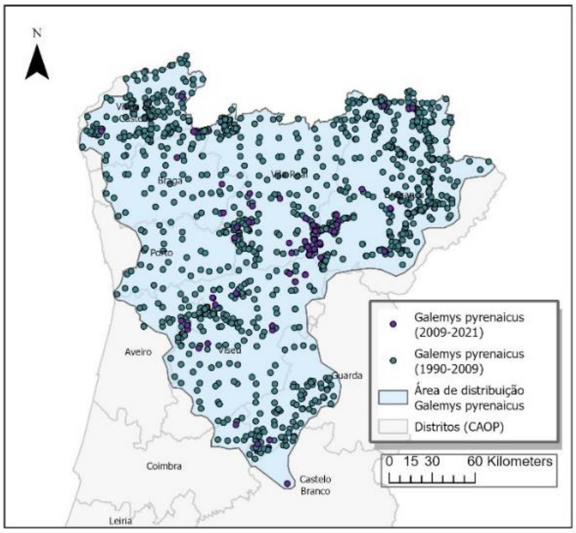


Figure 1 - Distribution area and confirmed points of presence of iberian desman (Galemys pyrenaicus) collected between 1990 and 2021. It should be noted that the most recent data (collected between 2009 and 2021) point to a drastic reduction in the species' range

For these reasons, in addition to other criteria, the conservation status of the iberian desman in the national territory is considered ENDANGERED (Mathias et al., 2023) and its distribution in Portugal is currently limited to river basins located in the north of the country.

1.1 Objectives

The first objective of this work is to create a predictive model of Iberian desman distribution using historical bioclimatic data (1970-2000) and projected data for climate change scenarios (2020-2040). In this sense, the second objective of this work is to compare the two models generated and investigate whether climate change will have impacts on the ecological niche of the species in the near future.

2. Study area

The study area (6598.34 km2) is located in the northern part of mainland Portugal and partially overlaps with the districts of Vila Real and Bragança (Figure 2). Although the species has very particular habitat requirements, the study area is very heterogeneous. Altitude varies between 35 and 1456 m ($\bar{x}$=603.8 ± 201.9 m), the average annual temperature is 12.6 ± 0.9 ^{Oc} and the human density is approximately 32.48 inhabitants/km^2 (INE, 2021).

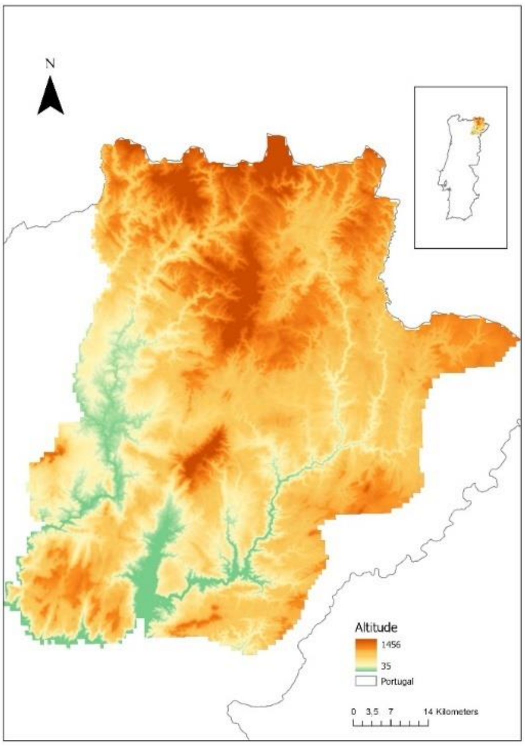


Figure 2 - Representation of the study area and altitude.

3. Data and Methods

3.1 Presence of the species

The Iberian desman presence points (n=1449) were downloaded from the page *Global Biodiversity Information Facility* ("Presence of the species: GBIF.org (21 February 2024) GBIF Occurrence Download," 2024) and consist of a dataset collected between 2010 and 2021. A pre-processing of the data was carried out, having eliminated all points that were located more than 500m from the nearest waterline, leaving 1335 points of presence of the species for model training and analysis.

3.2 Land Use and Environmental Variables

Altitude and 19 bioclimatic variables were extracted from the Worldclim database, with a resolution of 30" (approx. 1km), for the historical period (1970-2000) and for future projections (2021-2040).

The monthly and annual water balance was constructed from WorldClim data and data from the Consortium for Spatial Information, namely *Global High-Resolution Soil-Water Balance* (Trabucco and Zomer, 2019).

Given that the target species of this study is a semi-aquatic mammal, wetland habitats assume a special importance in the development of their ecological niche model. For this purpose, a model of the hydrological network was developed and the Strahler order of the water lines then created was calculated using the hydrology tools of the ArcGIS Pro software (Attachment A). The Strahler order is a river classification system, which assigns values on a scale based on the hierarchy of tributaries. When a water line has no tributaries, its order is 1; When two order 1 water lines come together, they form a segment of order 2 downstream, and so on (Strahler, 1957).

3.3 Anthropogenic variables

In order to reduce the number of variables under analysis, it was decided to use a single set of data that reflects the human impact on the landscape – the human footprint, or *human footprint* ("Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic)," 2005). This global dataset has been widely used in the construction of MDE's (Liu et al., 2020) (Jhala et al., 2021) (Dakhil et al., 2021) (Morueta-Holme et al., 2010) and aggregates data on population density, artificial land use, built-up areas, night lighting and accessibility (road network, railway, navigable rivers).

3.4 Selection of covariates

Multicollinearity occurs when highly correlated variables are included among the predictor co-variables of a model (Gunst and Webster, 1975). The inclusion of dependent variables can lead to a bias in the estimates, thus contributing to the final model being unreliable (Yoo et al., 2014). In this sense, several authors suggest a previous analysis of the covariables, eliminating those that present a high correlation (Sillero and Barbosa, 2021)

In order to determine which bioclimatic variables are relevant predictors for the species, the literature was analyzed (Quaglietta et al., 2018) (Morueta-Holme et al., 2010) (Barbosa et al., 2009), and 9 variables were selected.

The rasters of these variables were processed so that they had the same length and resolution, and were subsequently subjected to Pearson and Principal Component correlation analysis (
Attachment B).

Through this process, it was possible to eliminate 3 of the variables initially included. It should be noted that this analysis was performed only for continuous bioclimatic variables, as it would not make sense to do so for categorical or binary variables. In total, 9 variables were used (Table 1).

The variables of the future period used to project the climate change scenario were obtained from the WorldClim database, using the CMIP6 models and the SSP2 scenario. This scenario is a middle ground with regard to socio-economic development indicators, i.e., it maintains the values of the trends that are currently observed in the use of fossil fuels and energy, in the exploitation of natural resources, among others, not predicting major changes (decrease or increase) compared to the current trend (O'Neill et al., 2016).

Table 1 - Variables used in the development of the Iberian desman distribution model, processed for the study area.

Variable	Code	Period	Values	Source
Average temperature annual	BIO1	Historic	10.1-1.9 ± 1.4 ºC	(Fick and Hijmans, 2017)
		Future	17.3-10.9 ± 0.9 ºC	
Maximum Temperature of the hottest month	BIO5	Historic	28.7-23.8 ± 0.8 ºC	
		Future	32.5-27.7 ± 0.9 ºC	
Average temperature of the Warmest Quarter	BIO10	Historic	20.9-16.3 ± 0.6 ºC	
		Future	27.3-22.9 ± 0.6 ºC	
Precipitation for the quarter drier	BIO17	Historic	127-75 ± 8.47 mm	
		Future	116.91 - 75.67 ± 7.74 mm	
Annual Water Balance	BH	Historic	1378.5-711.17±144.07 mm	Calculated in this work using ArcGIS Pro from (Fick and Hijmans, 2017) and (Trabucco and Zomer, 2019)
		Future	1375.3-756.11± 145.46 mm	
Summer Water Balance	BHV	Historic	69-1.33 ± 12.81 mm	
		Future	53.29-(-0.22) ± 11.53 mm	
Human footprint	PH	-	90-10	("Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic)," 2005)
Standard deviation of altituc	DP_ALT	-	172.99 – 4.72m ± 24.9	(Fick and Hijmans, 2017)
Hydrological network buffe (Strahler's order)	RASTER_BUFFER	-	Scale (6-1)	Developed in this work through ArcGIS Pro

3.5 Species Distribution Models

Species distribution models (MDEs), also known as Ecological Niche Models, have been widely used by ecologists and researchers in a variety of uses and applications (Zurell et al., 2020)

Some of the most common studies, carried out within the scope of EAWs, consist of estimating the suitability of habitat for a given species taking into account its ecological preferences, with a view to designing conservation measures (Morueta-Holme et al., 2010) or in predicting the expansion of invasive species (Barbet-Massin et al., 2018). Other types of studies, equally frequent in the literature, focus on the simulation of future situations, using reference bioclimatic data, projected for the various climate change scenarios (Forester et al., 2013)

Among the various practical uses of these models, the selection of sites with optimal habitat conditions for the reintroduction of locally extinct species stands out (Treves et al., 2022) (Jhala et al., 2021).

Although most researchers have been dedicated to the development of EDMs for terrestrial animals, there are also many examples of this type of analysis applied to aquatic animals (Melo-Merino et al., 2020).

3.6 MaxEnt

The MaxEnt software (Phillips et al., 2024), has established itself as one of the most popular tools in the modelling of species distribution, mainly due to its predictive accuracy, even for small samples, and its ease of use (Merow et al., 2013). In order for the model to be created, data on the presence of the species(ies) under study and a set of environmental and anthropogenic covariables, potentially explanatory of their distribution(s), are required (Phillips, 2017).

MaxEnt, a program specifically designed to create MDEs, works from a *Machine Learning*, based on the principle of maximum entropy (Phillips, 2017). This principle postulates that, in the face of a set of possible solutions, the one with the highest entropy should be selected, i.e., the one with the highest degree of uncertainty, respecting the constraints imposed (Wu, 1997), which in the case of MaxEnt, are the environmental and anthropogenic covariables. In practice, MaxEnt trains the model, maximizing the differences in values that the covariates assume between a portion of the known points of presence, with background points randomly generated by the program, whose presence or absence has not been measured (Merow et al., 2013). In this way, the software learns to discriminate which traits are most suitable for the species.

Respecting the ranges of values of the co-variables in which the species occurs, which basically reflect its ecological preferences, MaxEnt generates several distribution models. It is at this stage that the software selects the one with the highest entropy, i.e., the one that translates the suitability of the habitat in a more uniform and uncertain way. Thus, the final model will present the widest possible distribution within the constraints imposed by the covariates.

An important technique present in MaxEnt is regularization. This function prevents the software from creating a model that is overfitting to the training data (species presence data), which would compromise the model's performance in generalizing to other presence data.

The model used 428 attendance records for training and generated 9998 background points.

4. Results

4.1Model performance

The omission curves (blue and black line of the Figure 3) are useful for assessing the predictive capacity of the model produced. The omission in training samples indicates the proportion of locations of known presence (those that were actually used to train the model) but were incorrectly classified as absences.

Ideally, the omission in training samples should be low, meaning that the model is correctly identifying the sites of presence of the species. In this sense, the closer the omission in training samples to the predictive omission (which represents the omission rate estimated by the model for all data), the better the generalization of the model to new data, whose locations were not used in the training of the model.

Given the proximity between the omission curve in training samples and the predictive omission obtained in the present study (Figure 3), it is considered that the model will have a good ability to generalize to other data.

The envisaged background fraction - which measures the fractional value as a function of the cumulative threshold (red line in the Figure 3) -, is a metric used to assess how restrictive the model produced is. The fractional value measures the proportion of the background environment that the model considers to be suitable for the species to occur. High values mean that the model identifies many sites as suitable, regardless of how likely the species are to be present in those locations. The cumulative threshold, on the other hand, is a cut-off point that indicates the degree of certainty about the presence of the species. Low cut-off values indicate that the model is not very restrictive and that it takes for granted the presence of the species even in inappropriate places. High cut-off values indicate that the model is too restrictive and that it only takes for granted the presence of the species in very particular conditions.

In the present study, the model obtained is quite restrictive, since for a cut-off threshold of 50, only 10% of the environmental conditions are adequate (Figure 3). However, obtaining a restrictive model, as is the case, is not necessarily negative. Indeed, such a model is consistent with two important characteristics of the species under study: its rarity and its demanding etho-ecological preferences.

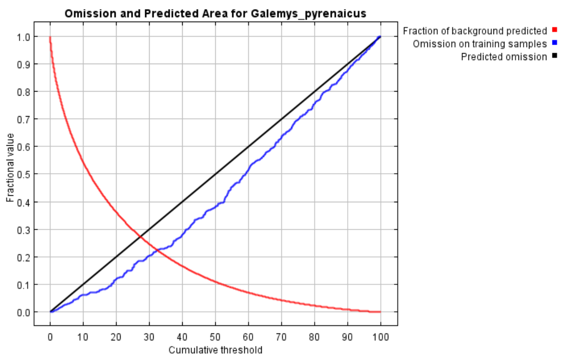


Figure 3 - Cumulative threshold plot of the iberian desman distribution model.

Depending on the limit of the classifier that is adopted, predictive machine learning models such as Maxent will be able to correctly classify some samples, but they will also produce false positives, which will inevitably create a bias in the results. An expeditious way to assess the performance of the model is, precisely, to evaluate the rate of true positives (or sensitivity) as a function of false positives (1- specificity), through the Receiver Operating Characteristic Curve (acronym ROC). In this sense, a perfect model would assign a higher value to all the places of presence of the species, compared to the background points (as mentioned above, they are not true absences, but "background" points, i.e., random samples created by the model in the environmental space where the species can potentially occur). So, ideally, the ROC would pass through the upper left corner of the graph, displaying a sensitivity of 100%, with no false positives. The area that this curve produces (acronym for AUC) is a good indicator, as it indicates how close to the perfect model (AUC=1) the model under analysis will be. In practice, the AUC indicates how good the model is at discriminating between the points where the species is present and the bottom points. However, it is important to note that among these background points generated by the software are true absences, but there will also be unsampled presences. For this reason, in addition to others, the use of AUC as an indicator of the performance of the models generated by MaxEnt has been criticized by several authors (Merow et al., 2013) (Shabani et al., 2018). Nevertheless, this metric remains one of the most used and popular in this type of study (Merow et al., 2013). The AUC value of the final model obtained was 0.853, which means that in a random selection of presence samples (true positives) and background samples, there is an 85.3% chance that the model will assign a higher score to true positives (Figure 4). In this sense, the performance of the model is considered very good.

4.2 Iberian desman Distribution Models

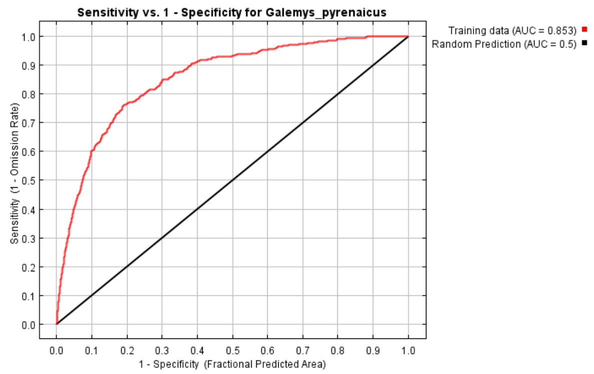


Figure 4 - Graphical representation of the Area over the Operator Curve (ROC) obtained for the Iberian Desman Distribution Model.

In the model produced for the historical period and for the projection of the future (Figure 5), only 13.55% of the study area represents humid habitats (watercourses and riparian gallery). This means that much of the region is not suitable for the iberian desman, as this species is a semi-aquatic mammal that lives exclusively in these environments. In this sense, 86.45% of the study area has a probability of less than 0.007 being suitable for the species. The model developed for the historical period revealed that for more than half (53.9%) of the available humid habitat, the probability of this habitat constituting the ecological niche of the species is very low (Table 2). The model for the future scenario, which includes bioclimatic variables affected by climate change, such as temperature increases, is, however, more serious. In fact, according to this projection, 74.4% of wetland habitats will have a very low probability of constituting the ecological niche of the species (Table 2). The same logic continues for the remaining ranges, with a general decrease in the probability of habitats being ideal for the species, with significant losses in areas with high or very high potential. This pattern is visible spatially, with few places benefiting from the climate change scenario (Figure 5) .

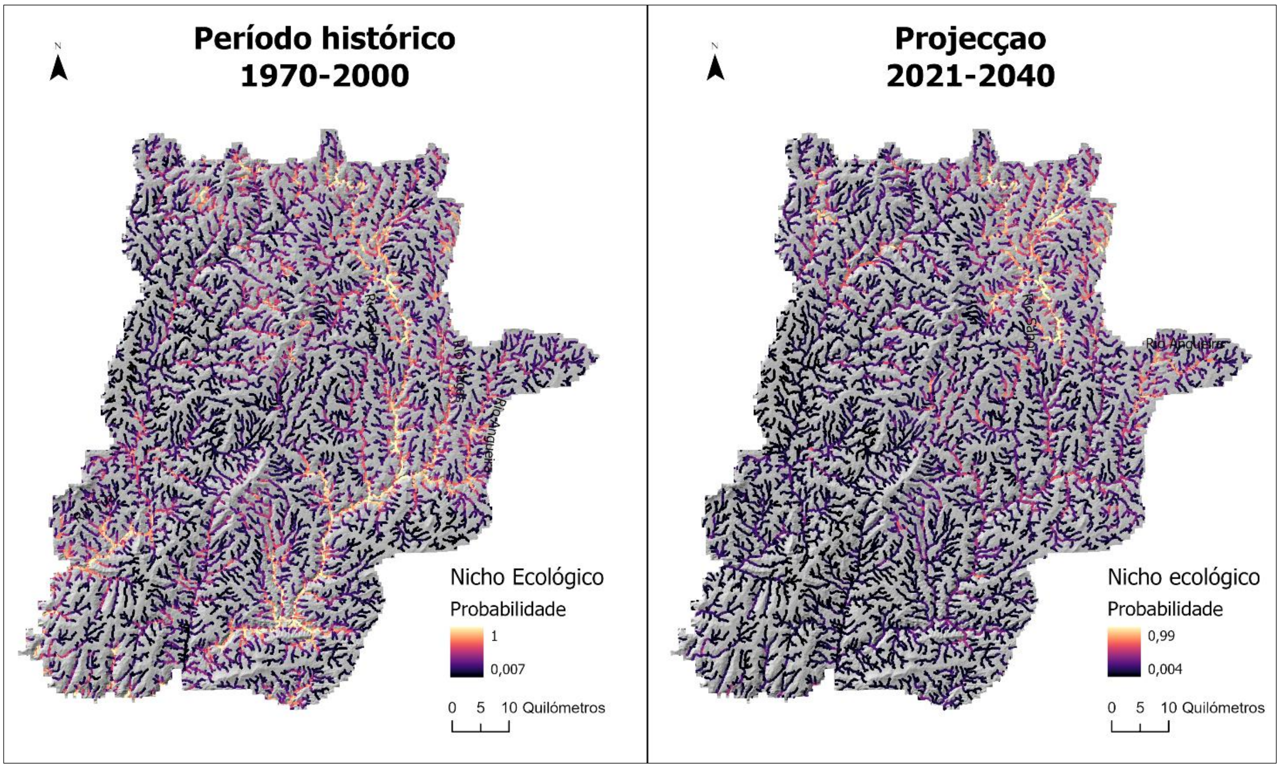


Figure 5 – Iberian Desman Distribution Models generated by MaxEnt software based on 9 variables.

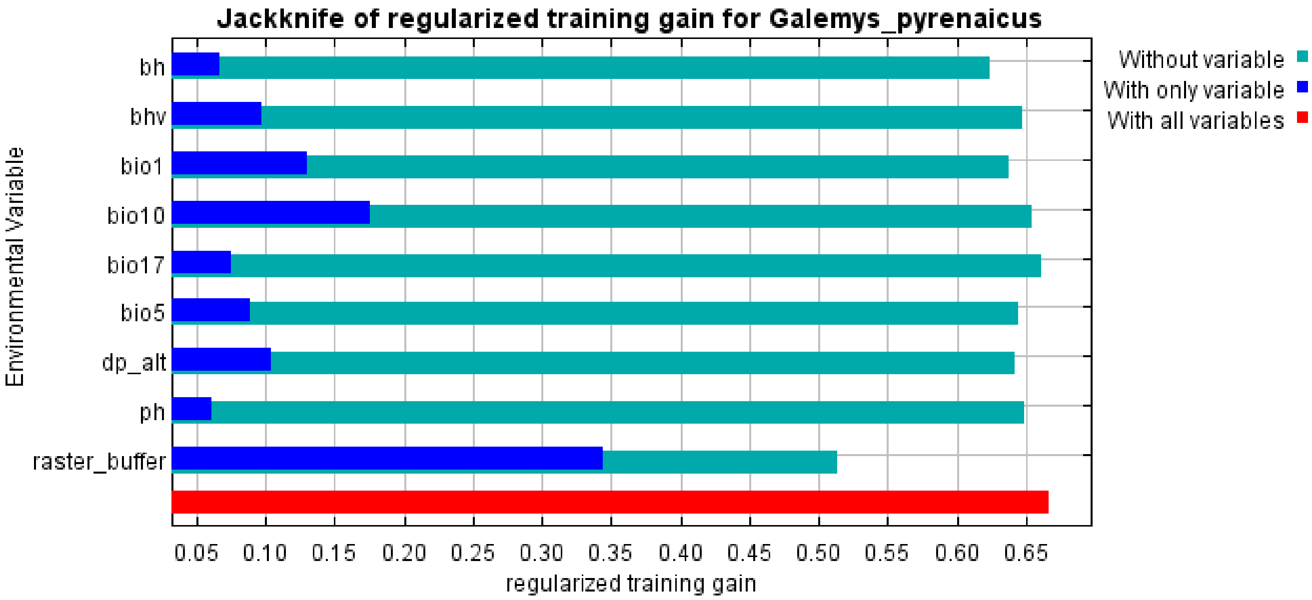
Table 2 – Probability of ecological niche, and consequently the probability of occurrence of the species, predicted for the historical period (1970-2000) and for the projection of a future scenario with climate change (2021-2040) CMIP6 SPSS2.

	Historical period		Projection		Difference
Ecological Niche Probability	**Pixels**	**%**	**Pixels**	**%**	**%**
Very low [0-0.2[	1185.159	53.9%	1636.962	74.4%	20.4%
Low [0.2-0.4[	514.259	23.4%	347.865	15.8%	-7.6%
Moderate [0.4-0.6[	250.478	11.4%	144.074	6.5%	-4.9%
High [0.6-0.8[	143.476	6.5%	48.824	2.2%	-4.3%
Very high [0.8-1[	104.064	4.7%	23.139	1.1%	-3.7%

4.3 Contribution of co-variables

One of the interesting results of MaxEnt is the graph that represents the cross-validation of the regularized training of the distribution model produced (Figure 6). This graph demonstrates the contribution of each of the variables to the construction of the model, showing what the model would gain from using only each of the variables, what it would gain if the variable were excluded, and the gain of the model when all variables are included.

As expected, the variable with the greatest predictive capacity of the developed model is raster_buffer, which represents humid habitats, since all points of presence of the species overlap these spaces (Figure 6). This variable was followed by the mean temperature of the hottest quarter (BIO10), mean annual temperature (BIO1), standard deviation of altitude (DP_ALT), water balance of the hottest quarter (BHV) and the maximum temperature of the hottest month (BIO5) (Figure 6). The variables that contributed the least to the construction of the model were the precipitation of the driest quarter (BIO17), the annual water balance (BH), and the human footprint (HP) (Figure 6).

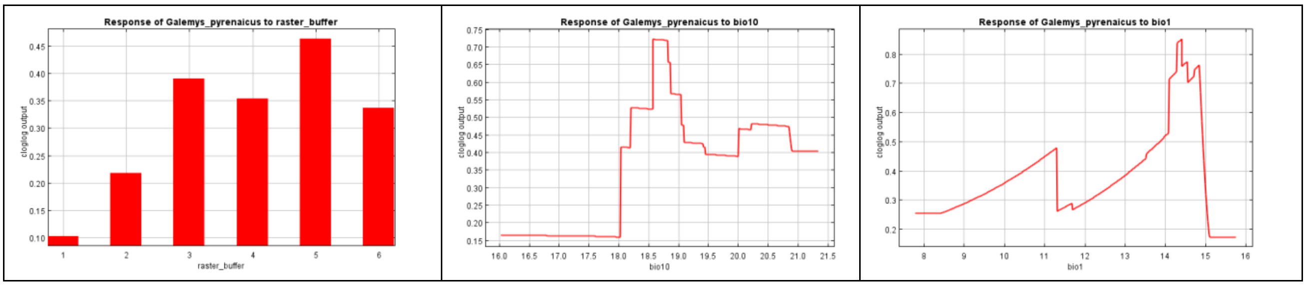


4.4 Covariate response curves

The analysis of the response curves of the model in relation to its co-variables can increase the knowledge about the ethological preferences of the species under study. Indeed, these curves show how the probability of a given area constituting an ecological niche varies over each variable.Figure 7).

In some cases the general trend of probability is one of increase or decrease; In other cases, there is a step pattern, which may be caused by errors, or by the species' preferences/avoidance for specific ranges of values. Generally speaking, the probability varies according to the conditions summarized in the Table 3.

Figure 6 - Contribution of variables to the construction of final models.



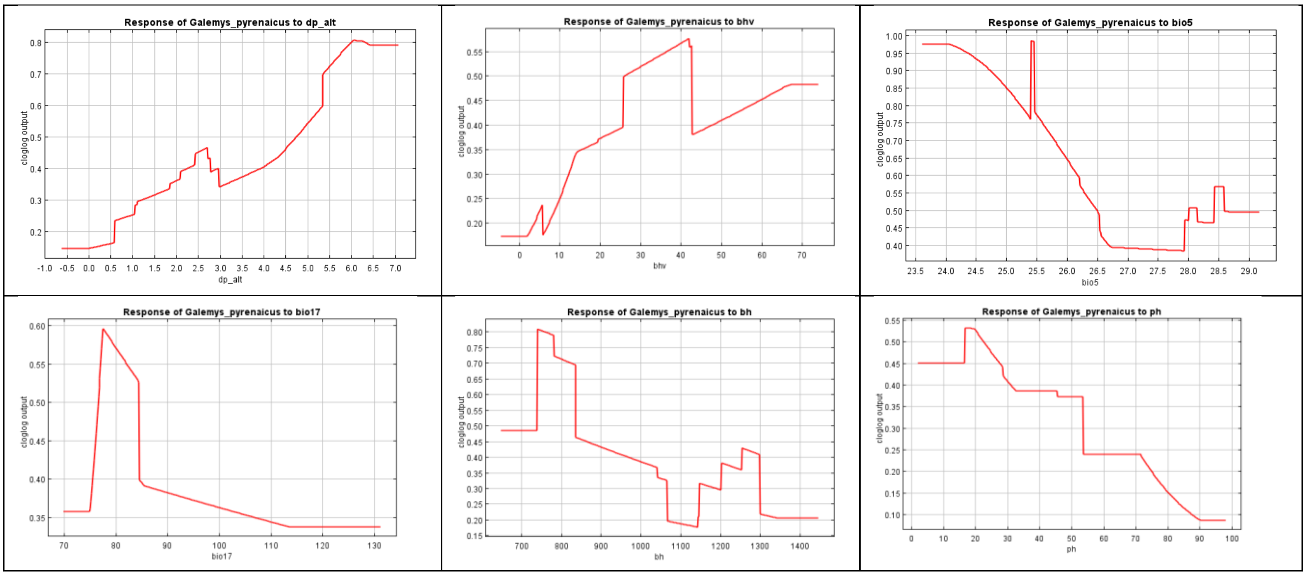


Figure 7 - Response curves of the Galemys pyrenaicus distribution model against the 9 co-variables used.

Table 3 - Variation of the probability of an ecological niche, as a function of the parameters of each variable.

Variable	Probability
RASTER_BUFFER	Highest for Strahler's order 3 and 5
BIO10	Very low up to 18°C; a sharp increase in level up to 18.5°C; sudden drop to 19ºC and stability with oscillations from that value
BIO1	Increase with oscillations up to 14ºC; a sharp increase of up to 14.5°C; sudden drop from 15ºC; very low from 15ºC
DP_ALT	Increases continuously with increasing altitude, with few oscillations
BHV	Very reduced from 0-10mm; increase with oscillations up to 40 mm; sharp drop from that value, followed by a continued slight increase.
BIO5	Very high at 23.5ºC followed by a progressive drop to 26.5ºC, with the exception of a peak at 25.5ºC. Slight increase in level from 28ºC to 29ºC.
BIO17	Very low up to 75mm followed by a sudden increase that culminates in a peak of approximately 78mm; from that value a sharp drop to 85mm, and a progressive drop to 112mm (approximately); greatly reduced from that value
BH	Moderate up to 720 mm, followed by a peak at 790 mm; progressive descent with oscillations up to 1090mm (approximately); slight increase in level up to 1300mm, followed by sharp decrease to 1400mm
PH	Highest between 15 and 20 (low values of human disturbance); progressive decrease with oscillations up to 100 (high human disturbance values)

5. Discussion

5.1 Limitations of the Model

Although a scientific and rigorous methodology was followed, it is important to mention that this work, being essentially a learning exercise, started from a diverse set of data, with very different dates, objectives and sampling designs. It is noteworthy that the sampling effort was not homogeneous for the entire study area, and that the methodological design was oriented to the location, rather than the species. In fact, a lot of data has been collected as part of environmental impact studies, so they cover very restricted geographical areas, such as the valley of the Sabor River. This concentration of points of presence of the species in the study area is even more evident in the kernel density analysis (Attachment C). This constraint has significant implications that will necessarily diminish the predictive capacity of the model, mainly because it biases the environmental conditions that the model can determine as being relevant to the species. A striking example of this bias can be seen in the variable raster_buffer of the Figure 7, the result of which indicates that for order 3 and 5 water lines, the probability of ecological niche increases. This result contradicts the study developed by Quaglietta et al., 2018, which shows that the headwaters of rivers constitute one of the last strongholds for the survival of the species. It is clear that the result obtained in the scope of this work is due to the use of data collected in an environmental impact study, whose sampling effort was concentrated only on the places where the infrastructure was built, so that the headwaters of the water lines are under-sampled. Other variables that are probably also affected by this problem are BH, BIO10 and BIO17. Contrary to what would be expected, the probability of an ecological niche decreases with the increase in the annual water balance, increases with the increase in temperature in the warmer quarter (although stabilizing from 20.5ºC), and decreases with the increase in precipitation in the drier quarter. On the contrary, PH, BHV and DP_ALT showed variations that reflect the current knowledge about the species, i.e., the probability of ecological niche increased with the increase in altitude, with the decrease in the human footprint, and with the increase in the summer water balance. To a certain extent, the BIO1 variable also corresponded to expectations, since it decreases sharply from 15ºC.

Another limitation of this model is that it does not contemplate the profound changes that the landscape of the Northeast of Trás-os-Montes has undergone in recent decades, with the construction of the Baixo Sabor hydroelectric plant. It is likely that this infrastructure has caused significant impacts on the species' habitat. In this sense, future improvements to the model should consider this aspect.

5.2 Model Comparison: Historical Period and Projection

According to the model produced, a large part of the study area presents a very low probability of ecological niche of the species. There are some notable exceptions, namely the Sabor and Angueira basins, some water lines to the north, and the Tua river valley. It should be noted, however, that these areas benefited from the sampling effort concentration mentioned above. Regarding the projection, the loss of probability of ecological niche is remarkable. In fact, only a small section of the Sabor River and its adjacent watercourses, and a small part of the Mangueira River, will have suitable habitat for the iberian desman.

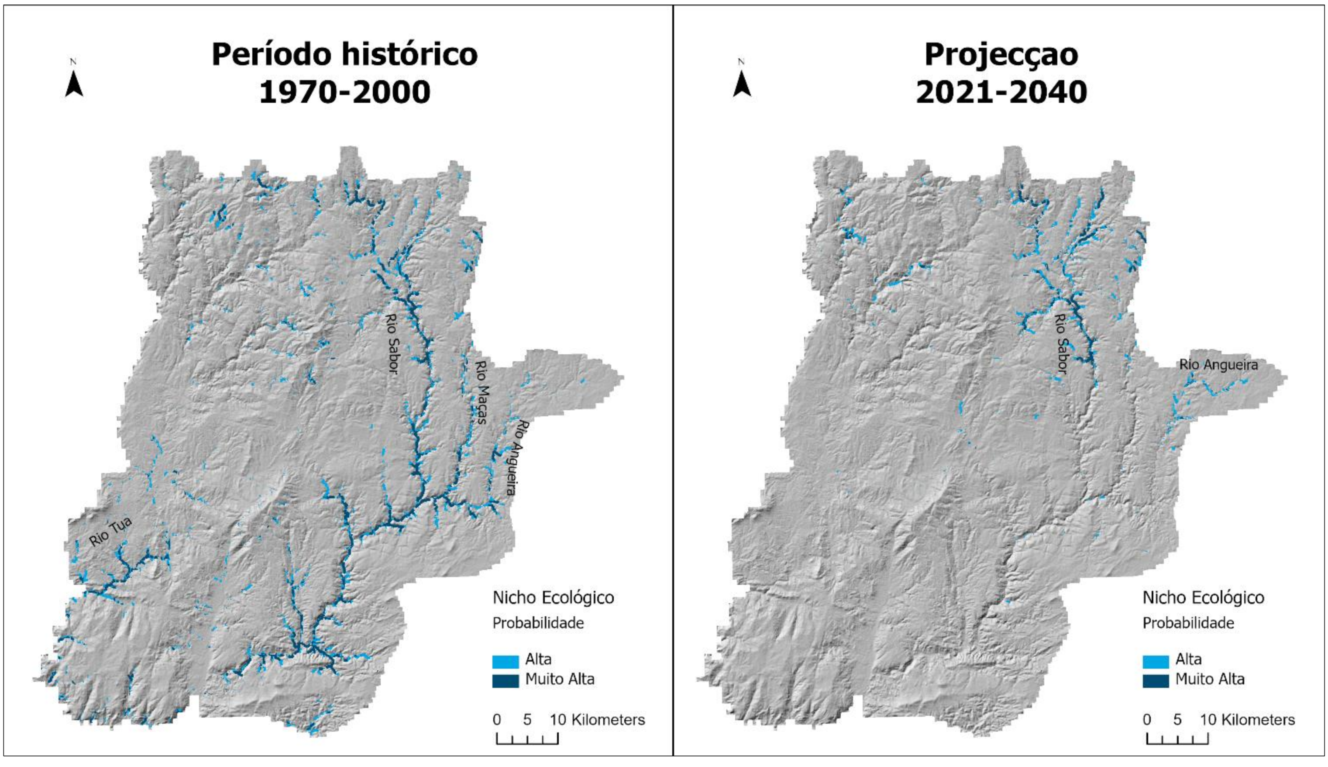


Figure 8 - Rivers and secondary watercourses that have a high probability, to a very high probability of constituting a habitat for the Iberian Desman.

This loss is corroborated by quantitative analysis, since there will be a difference of 12.9% for the moderate to very high classes in the probability of a given area constituting the ecological niche of the species.Table 3). The only class whose change was positive in the projection was the "very low" probability (0-0.2); For all other classes, the variation was negative. It is important to note that the model obtained, although restrictive, is not conservative, in fact, as discussed in the previous point, some variables did not behave as expected (such as water balance), which contributes to the model being less demanding. In this sense, it is quite likely that in a model with fewer limitations than the one developed in this work, the predicted ecological niche losses will be even higher.

6. Conclusion

The present study developed a distribution model for the iberian desman for the historical period, and for a future scenario, affected by climate change, namely by the increase in temperature and by a lower volume of precipitation. Despite the limitations of the model in question (for more details, see the Discussion), this work confirmed the critical situation that this species currently faces, and that it will tend to worsen severely in the near future. According to the projection, there will be a significant reduction in the habitat suitability of the species, since the decrease of areas with high moisture levels will be an inevitability, if the bioclimatic forecasts are confirmed. The results underline the importance of adopting climate change mitigation strategies, otherwise relic species such as the iberian desman will become extinct due to lack of suitable habitats.

In this sense, it is suggested that this model be improved and that conservation and safeguarding efforts be developed in places that present conditions to ensure the survival of the species in the future.

7. References

Barbet-Massin, M., Rome, Q., Villemant, C., Courchamp, F., 2018. Can species distribution models really predict the expansion of invasive species? PLOS ONE 13, e0193085. https://doi.org/10.1371/journal.pone.0193085

Barbosa, A.M., Real, R., Mario Vargas, J., 2009. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (Galemys pyrenaicus) in Portugal and Spain. Ecol. Model. 220, 747–754. https://doi.org/10.1016/j.ecolmodel.2008.12.004

Dakhil, M.A., El-Keblawy, A., El-Sheikh, M.A., Halmy, M.W.A., Ksiksi, T., Hassan, W.A., 2021. Global Invasion Risk Assessment of Prosopis juliflora at Biome Level: Does Soil Matter? Biology 10, 203. https://doi.org/10.3390/biology10030203

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37(12): 4302-4315.

Forester, B.R., DeChaine, E.G., Bunn, A.G., 2013. Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions. Divers. Distrib. 19, 1480–1495. https://doi.org/10.1111/ddi.12098

Gunst, R.F., Webster, J.T., 1975. Regression analysis and problems of multicollinearity. Commun. Stat. 4, 277–292. https://doi.org/10.1080/03610927308827246

INE, 2021. Population and housing census - 2021 Census.

Jhala, H.Y., Qureshi, Q., Jhala, Y.V., Black, S.A., 2021. Feasibility of reintroducing grassland megaherbivores, the greater one-horned rhinoceros, and swamp buffalo within their historic global range. Sci. Rep. 11, 4469. https://doi.org/10.1038/s41598-021-83174-4

Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic), 2005.

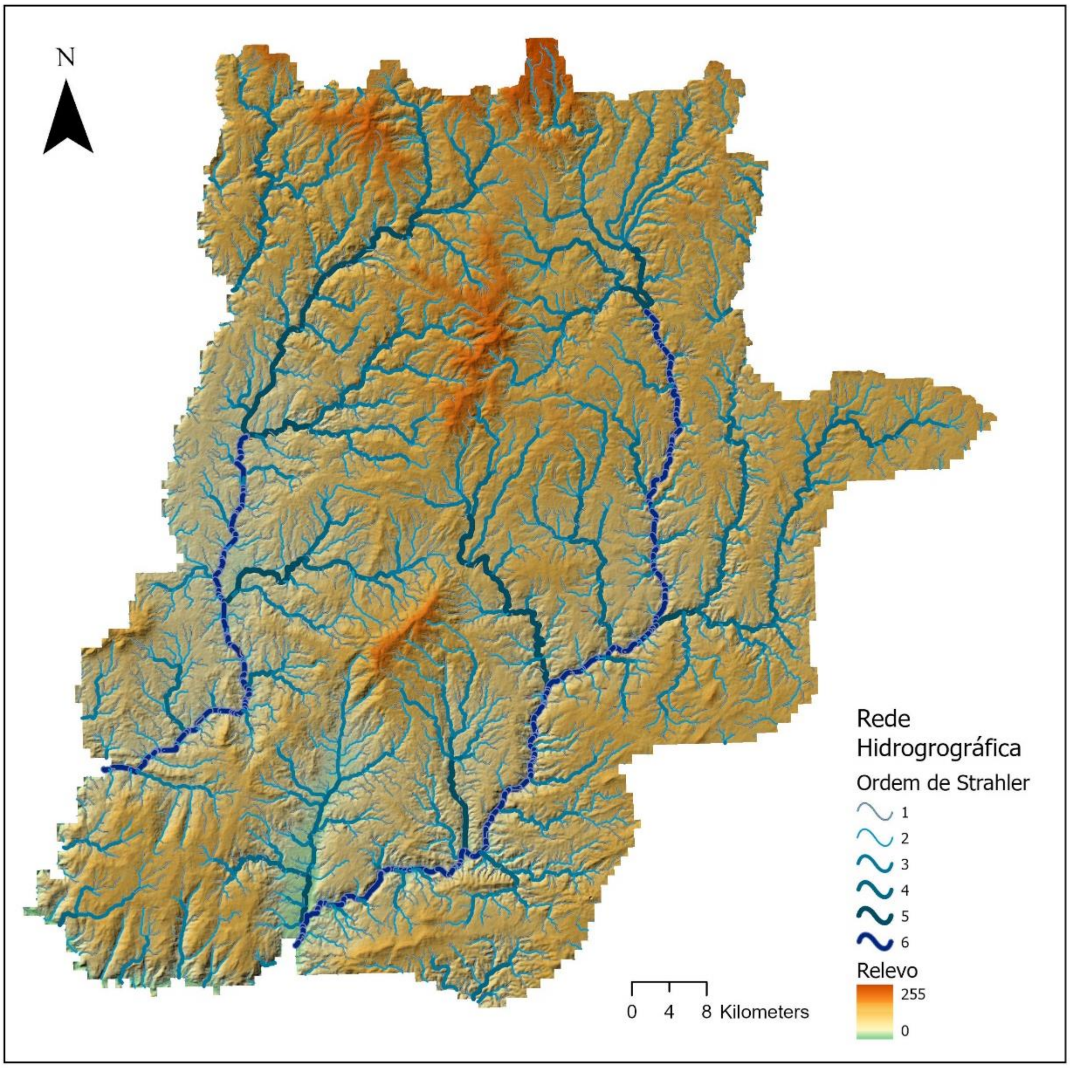
Liu, X., Blackburn, T.M., Song, T., Wang, X., Huang, C., Li, Y., 2020. Animal invaders threaten protected areas worldwide. Nat. Commun. 11, 2892. https://doi.org/10.1038/s41467-020-16719-2

Mathias, M.D.L., Fonseca, C., Rodrigues, L., Grilo, C., Lopes-Fernandes, M., M. Palmeirim, J., Santos-Reis, M., Alves, P.C., Cabral, J.A., Ferreira, M., Mira, A., Eira, C., Negrões, N., Paupério, J., Pita, R., Rainho, A., Rosalino, L.M., Tapisso, J.T., Vingada, J., 2023. Red Book of Mammals of Mainland Portugal. FCiências.ID / ICNF.

Melo-Merino, S.M., Reyes-Bonilla, H., Lira-Noriega, A., 2020. Ecological niche models and species distribution models in marine environments: A literature review and

spatial analysis of evidence. Ecol. Model. 415, 108837. https://doi.org/10.1016/j.ecolmodel.2019.108837
Merow, C., Smith, M.J., Silander Jr, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x
Morueta-Holme, N., Fløjgaard, C., Svenning, J.-C., 2010. Climate Change Risks and Conservation Implications for a Threatened Small-Range Mammal Species. PLoS ONE 5, e10360. https://doi.org/10.1371/journal.pone.0010360
O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K., Sanderson, B.M., 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geosci. Model Dev. 9, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016
Phillips, S., Dudík, M., Schapire, R., n.d. Maxent software for modeling species niches and distributions (Version 3.4.1).
Phillips, S.J., 2017. A Brief Tutorial on Maxent.
Presence of the species: GBIF.org (21 February 2024) GBIF Occurrence Download, 2024. https://doi.org/10.15468/dl.xd5xgf
Quaglietta, L., Paupério, J., Martins, F.M.S., Alves, P.C., Beja, P., 2018. Recent range contractions in the globally threatened Pyrenean desman highlight the importance of stream headwater refugia. Anim. Conserv. 21, 515–525. https://doi.org/10.1111/acv.12422
Shabani, F., Kumar, L., Ahmadi, M., 2018. Assessing Accuracy Methods of Species Distribution Models: AUC, Specificity, Sensitivity and the True Skill Statistic.
Sillero, N., Barbosa, A.M., 2021. Common mistakes in ecological niche models. Int. J. Geogr. Inf. Sci. 35, 213–226. https://doi.org/10.1080/13658816.2020.1798968
Strahler, A., 1957. Quantitative analysis of watershed geomorphology. Eos Trans. Am. Geophys. Union 38, 913–920. https://doi.org/10.1029/TR038i006p00913
Trabucco, A., Zomer, R.J., 2019. Global Soil Water Balance Geospatial Database. CGIAR Consortium for Spatial Information.
Treves, A., Terenziani, A., Angst, C., Comino, E., 2022. Predicting habitat suitability for *Castor fiber* reintroduction: MaxEnt vs SWOT-Spatial multicriteria approach. Ecol. Inform. 72, 101895. https://doi.org/10.1016/j.ecoinf.2022.101895
Wu, N., 1997. The Maximum Entropy Method. Springer, Berlin.
Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., Lillard, J.W., 2014. A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int. J. Appl. Sci. Technol. 4, 9–19.
Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A standard protocol for reporting species distribution models. Ecography 43, 1261–1277. https://doi.org/10.1111/ecog.04960

Attachment A - Hydrographic Network generated through the Hydrology tools of ArcGIS Pro software



Attachment B – Analysis and Selection of Covariates

At an early stage, the variables used for the development of the model (Table B1) were pre-selected based on the literature on the species.

Table B1 – Variables analyzed within the CP and PCA matrix.

CODE	VARIABLE
BIO1	Average annual temperature
BIO5	Maximum temperature of the hottest month
BIO10	Average temperature of the warmest quarter
BIO11	Average temperature of the coldest quarter
BIO12	Annual rainfall
BIO17	Precipitation from the driest quarter
BIO18	Precipitation of the wettest quarter
BH	Annual Water Balance
BHV	Summer Water Balance
DP_ALT	Standard deviation of altitude

In order to reduce the dimensionality of the model and avoid multicollinearity, a Pearson correlation matrix (PC) was performed (Figure B1). The variables that exhibit the greatest correlation with each other are:

- BIO10 and BIO1 - 0.97 correlation;
- BIO17 and BIO18 - correlation of 0.94;
- BIO18 and BHV - correlation of 0.92;
- BIO17 and BHV - correlation of 0.79;
- BHV and BH: 0.77 correlation.

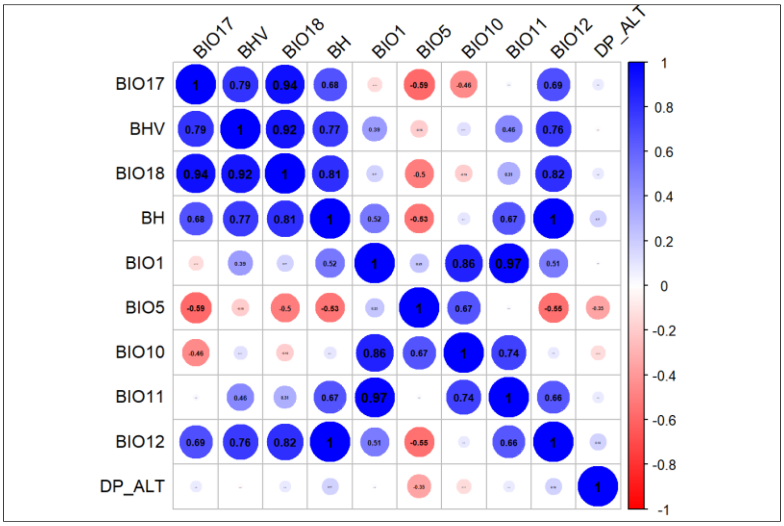


Figure B1 – Correlation matrix of the bioclimatic variables (Pearson's coefficient) for the study area obtained through the R software from a multi-band raster created in the ArcGis Pro software.

In order to deepen the understanding of the correlation between the variables, and to identify which ones contributed more significantly to the variability of the data, a Principal Component Analysis (PCA) was performed. The screeplot revealed that the first 3 principal components explain most of the variability in the data (Figure B2).

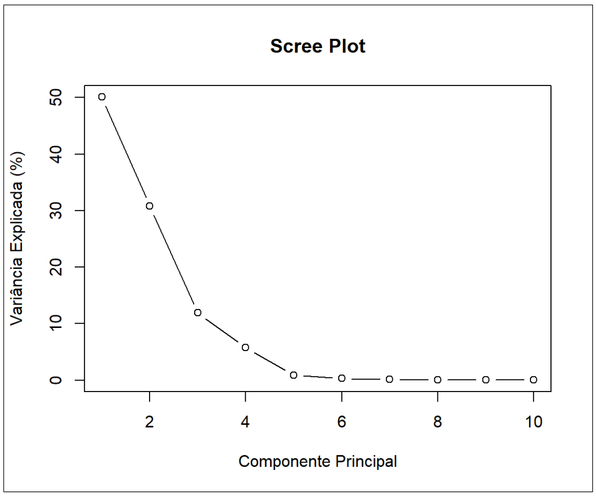


Figure B2 – Screeplot, which represents the percentage of variance for each principal component, generated from the R software.

Positive correlations between variables occur when two vectors point in the same direction; negative correlations are identified when the vectors point in opposite directions; Perpendicular vectors show little or no correlation. The PCA analysis of the first two principal components (Figure B3) reveals that there is a significant correlation between:

- BIO12, BH and BHV;
- BIO 18 and BIO17;
- BIO17 and BIO5;
- BIO1 and BIO11.

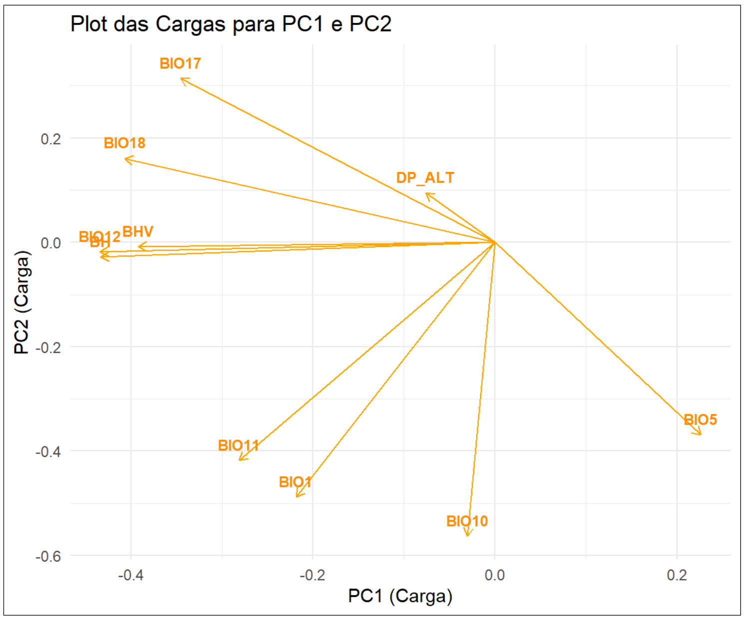


Figure B3 – Biplot of the PCA generated from the R software.

Finally, an analysis of the factor loadings of the variables (Figure B4) revealed that the variables that contributed the most to Principal Component 1 were:

- BIO5;
- BIO17;
- BIO1;
- BIO10 and BHV.

The most important variables for Principal Component 2 were:

- BH;
- BIO10;
- BHV;
- BIO17.
-

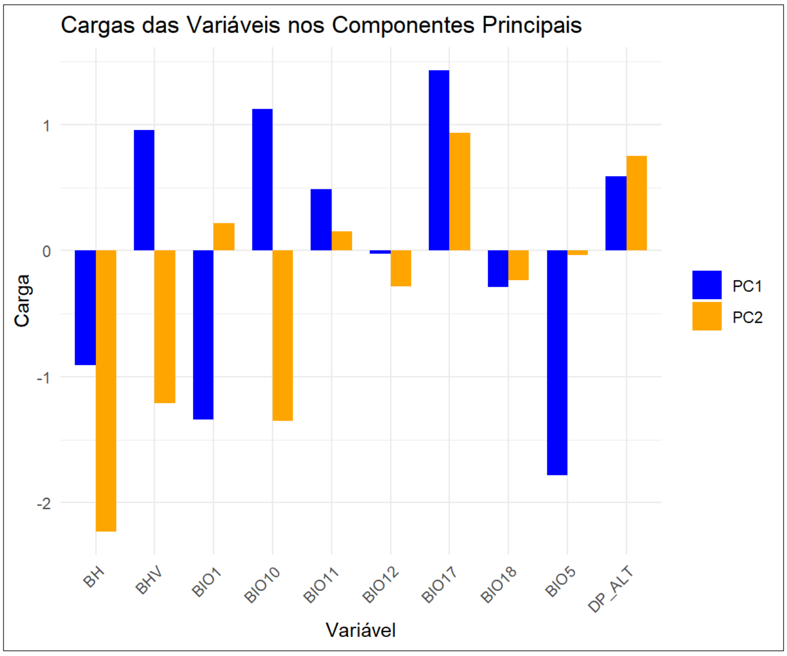


Figure B4 – Contribution of the variables to PC1 and PC2.

The cross-referencing of these three analyses made it possible to rank and eliminate the variables that exhibited strong correlation in the PCA and in the Pearson correlation matrix, and to select those that contributed the greatest variability to the model (Figure B4). However, it should be noted that knowledge about the etho-ecological requirements of the species was also incorporated into the selection process. For this reason, we chose to maintain two correlated variables – BH and BHV. Although BHV presents a lower variability, and consequently less predictive capacity compared to BH, this parameter is extremely important for the species, since the iberian desman needs continuous water flows in the period when this resource is scarcer (summer). The remaining variables selected were BIO1, BIO5, BIO10, BIO17 and DP_ALT. Although this last variable contributes little, it is a variable that has a low correlation with the others, so it will have the potential to provide different and useful information to the model.

Attachment C – Kernel density of points of presence of Galemys pyrenaicus