Clark University
Graduate School of Geography

Module:
Advanced Raster GIS

Determining suitable protection areas for Guanaco (Lama guanicoe) in Patagonia

Benjamin Gaskill, Philipp Steinhilber, Aaron Richmond-Crosset

M.Sc. Geographic Information Science

Fall 2024
Lecturer: Prof. John Rogan

Submission date: 12/18/2024

Ben Gaskill (bgaskill@clarku.edu)
Philipp Steinhilber (psteinhilber@clarku.edu)
Aaron Richmond-Crosset (arichmondcrosset@clarku.edu)

Contents

List of Figures

List of Tables

Introduction

The South American camelids (SAC) (Artiodactyla, Camelidae) are a key component of the Andean biocultural heritage (Vilá & Arzamendia, 2022). The Guanaco (*Lama guanicoe)*, which is one of the two wild species of SAC, plays a fundamental role in the conservation of biodiversity and in the conservation of ecological balances (Vilá & Arzamendia, 2022). By spreading seeds across its habitat, it contributes to the process of vegetation regeneration and the diversification of species in the area (GONZÁLEZ et al., 2006). Moreover, the Guanaco is an important food source for many other species, such as pumas, foxes and condors (Wildlife Conservation Society, 2024). This endemic species adapted to extreme environments with a wide distribution in arid and semiarid ecosystems from Argentina, Bolivia, Chile, Ecuador and Peru, mainly from 3000 to 5000 m.a.s.l (Baldi et al., 2016). Identifying suitable habitats for the guanaco can help in the creation of protected areas, promote sustainable land use and wildlife-friendly management practices (Baldi et al., 2016).

For this project, suitable habitats for the guanaco were identified using observation data and environmental predictor variables. The results were then compared to existing protected areas to evaluate their effectiveness in conserving the species. Finally, the protected areas were classified based on the risk posed by anthropogenic factors such as roads and settlements.

Research Questions

1. Where are the suitable guanaco areas (fundamental niche)?
2. To what extent do current protected areas cover regions of suitable habitat for Guanaco, as shown in our species distribution model?
3. What is the level of risk to guanacos in each protected area?
4. Which areas should be prioritized for future protection to ensure the best conservation outcomes for the species?

Study Area

The study area of our research is Patagonia, South America. This region encompasses the countries of Chile and Argentina (Figure 1).

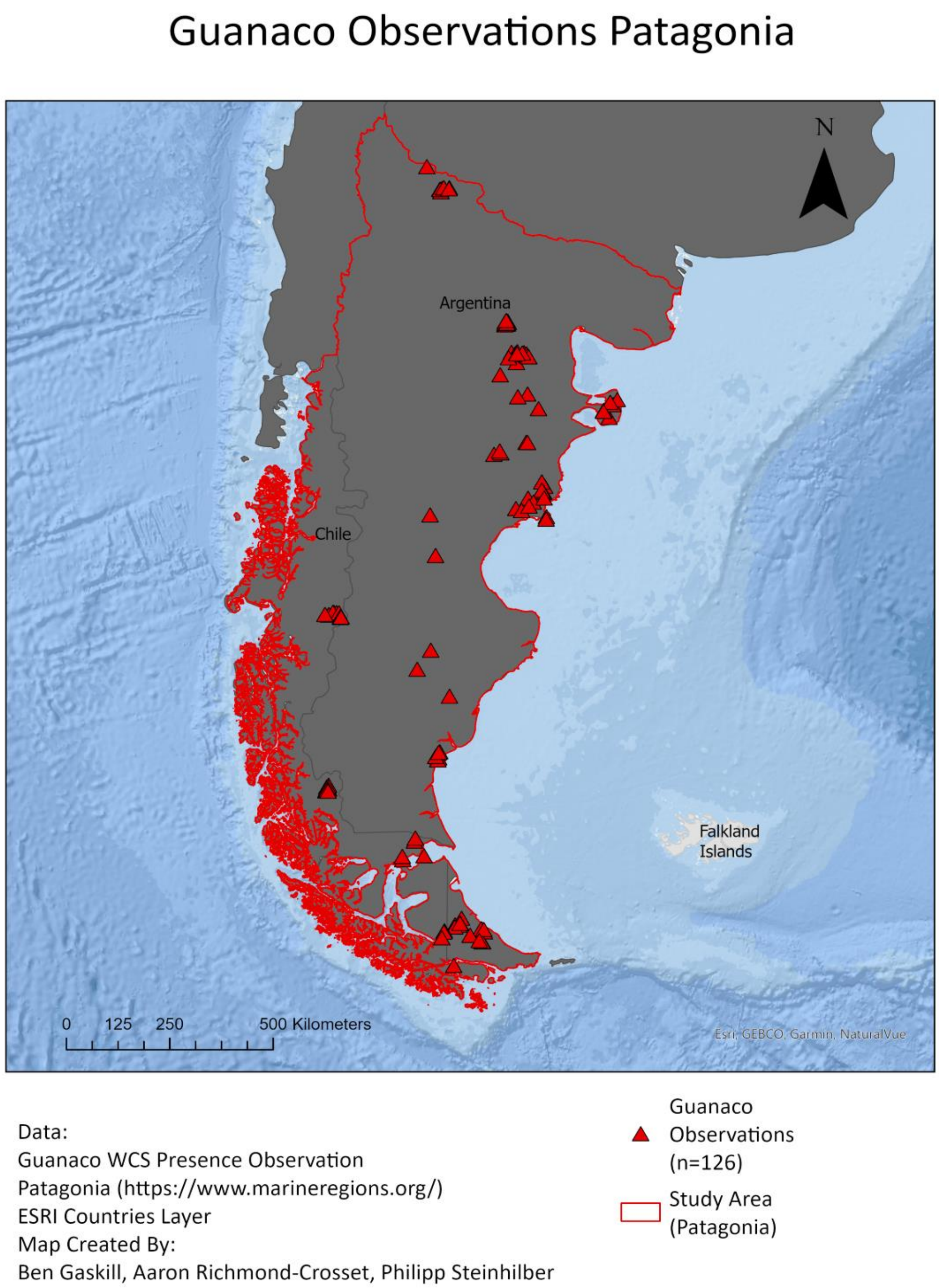


Figure 1: Study Area of Patagonia

Data Description and Preprocessing

Refer to Table 1 for a full list of our data sources. Our main data layer is Guanaco Presence Data, in the form of a point vector layer (VDC), courtesy of the Wildlife Conservation Society. Another essential data layer are polygons of protected and nonprotected areas (sourced from Protected Planet), including both state and non-state sponsored regions.

Our model requires multiple environmental variables to accurately model guanaco distribution. Bioclimatic variables, including monthly mean temperature and monthly mean precipitation, were sourced from the WorldClim organization. Both datasets have a spatial resolution of 2.5 minutes (~ 21 km^2 at the equator). These monthly data were downloaded and aggregated from 2000-2021 into multiple batches for efficient processing. The time intervals of these batches are 2000-2009, 2010-2019, and 2020-2021. A custom ArcPy Python script was developed and executed on the datasets to calculate mean cell statistics for each interval, then aggregated to a final mean temperature and mean precipitation raster for the entire time interval of 2000-2021.

Our next environmental variable is the MODIS Normalized Difference Vegetation Index (NDVI), a remote sensing-derived indicator of greenness and overall vegetation cover. Utilizing Google Earth Engine, we downloaded the mean NDVI for 2023 across Patagonia at a resolution of 250 meters.

Landcover data was sourced from the Global Land Analysis & Discovery (GLAD) team, as presented in the paper Global land cover and land use 2019. This dataset, developed by the University of Maryland within its Department of Geographical Sciences, provides global landcover information at 30-meter resolution. For our study, we clipped it to just our study area, Patagonia.

The Digital Elevation Model (DEM) was obtained from the USGS Digital Elevation Model (DEM) from the Hydrologic Derivatives for Modeling and Analysis (HDMA) database -- South America. The imagery provided elevation imagery at 30 m resolution and dated back to 2017. It consisted of 9 zip files each depicting a different region of the continent. For our study area, we clipped the data to just Patagonia.

Our final data layers are roads and settlements, sourced from OpenStreetMap.

Methods

Our study leverages the capabilities of the TerrSet Geospatial Monitoring and Modeling Software, namely the Habitat and Biodiversity Modeler. This tool enables modelers to accurately predict the distribution of a species, given point observation data and a set of environmental variables that define the species habitat (niche). Three different modeling approaches are offered for presence data as training input. The MAXENT (Maximum Entropy) modeling approach for presence only species distributions were developed by Steven Phillips, Miroslav Dudik and Robert Shapire (Phillips et al., 2017). We chose MAXENT due to the detailed and robust characteristics of the model's outputs, which will be discussed later in this paper.

A significant portion of our time was dedicated to data preprocessing, cleaning, and resampling. We faced numerous challenges related to differences in spatial resolution, inconsistencies with row and column size, and file formatting conventions between the multiple GIS programs we utilized (TerrSet, QGIS, ArcGIS Pro, MAXENT).

After downloading our data, the first step was to aggregate the mean monthly temperature and precipitation rasters into a single mean raster for all of South America. As discussed above, a custom Python script solved this issue. Our next step was bringing in our vegetation data. Due to the size of our study area, we hit significant roadblocks with the limitations of Google Earth Engine's export feature, frequently exceeding the export size limit. With an original goal to utilize 30-meter Landsat 9 NDVI, we decided to use the pre-processed MODIS data at 250 meters which solved our issues with data size. These datasets, in addition to landcover, DEM, presence observations, protected areas, roads, and settlements, were all projected to UTM zone 19 and clipped to the Patagonia study area polygon.

Before we could run our model, we had to solve numerous errors in TerrSet related to inconsistent resolutions, row and column size, and no-data errors. After weeks of testing and troubleshooting, our final solution was to resample the rows and columns to exactly match the resolution of our coarsest dataset, which was Mean Annual Temperature. A final cell size of 504 rows by 250 columns was applied to each of our datasets for MAXENT to run correctly.

Our parameters for the MAXENT model in TerrSett model include the default parameters, including a Logistic output format and options to include environmental variable response curves and a jackknife test. The environmental variables in our model include mean temperature, mean precipitation, NDVI, landcover, and the DEM.

The primary output of MAXENT is a suitability map for Guanaco in Patagonia (see Figure 2). The format of MAXENT's output is an ASCII file, which was converted to TerrSet's .RST as well as the traditional .TIF format using the GDAL Conversion Utility.

To create an anthropogenic risk map for Guanaco, we utilized the Distance Accumulation geoprocessing tool in ArcGIS Pro (see Figure 9). This tool was run using our roads and settlements layers to create Euclidian distance maps, which in turn were resampled to a scale of 0 to 1 to match the scale of the suitability map.

Our next steps were to overlay the protected area polygons with our suitability and risk rasters and run zonal statistics to extract the mean, median, minimum and maximum values. Using these rasters as a mask, we extracted 160 polygons that fell completely within the extent of our rasters. Issues with negative and no-data values were fixed using a combination of reclassification and edits to the included values in the analysis. See Figures 6, 7, 8 and Table 3 for maps, graphs and tables of suitability statistics. See Figures 10 and 11, and table 4 for maps, graphs, and tables of risk statistics. In addition, refer to the attached Excel file titled "Guanaco Zonal Statistics."

Results

Figure 2 is a representation of the Maxent model for the Guanaco in Patagonia. Redish colors show areas with better predicted conditions. Black triangles show the presence locations used for training. At first glance, a distinct spatial pattern in the species distribution emerges within the study area, with a clear preference for habitats in the east, while the west appears to be largely avoided.

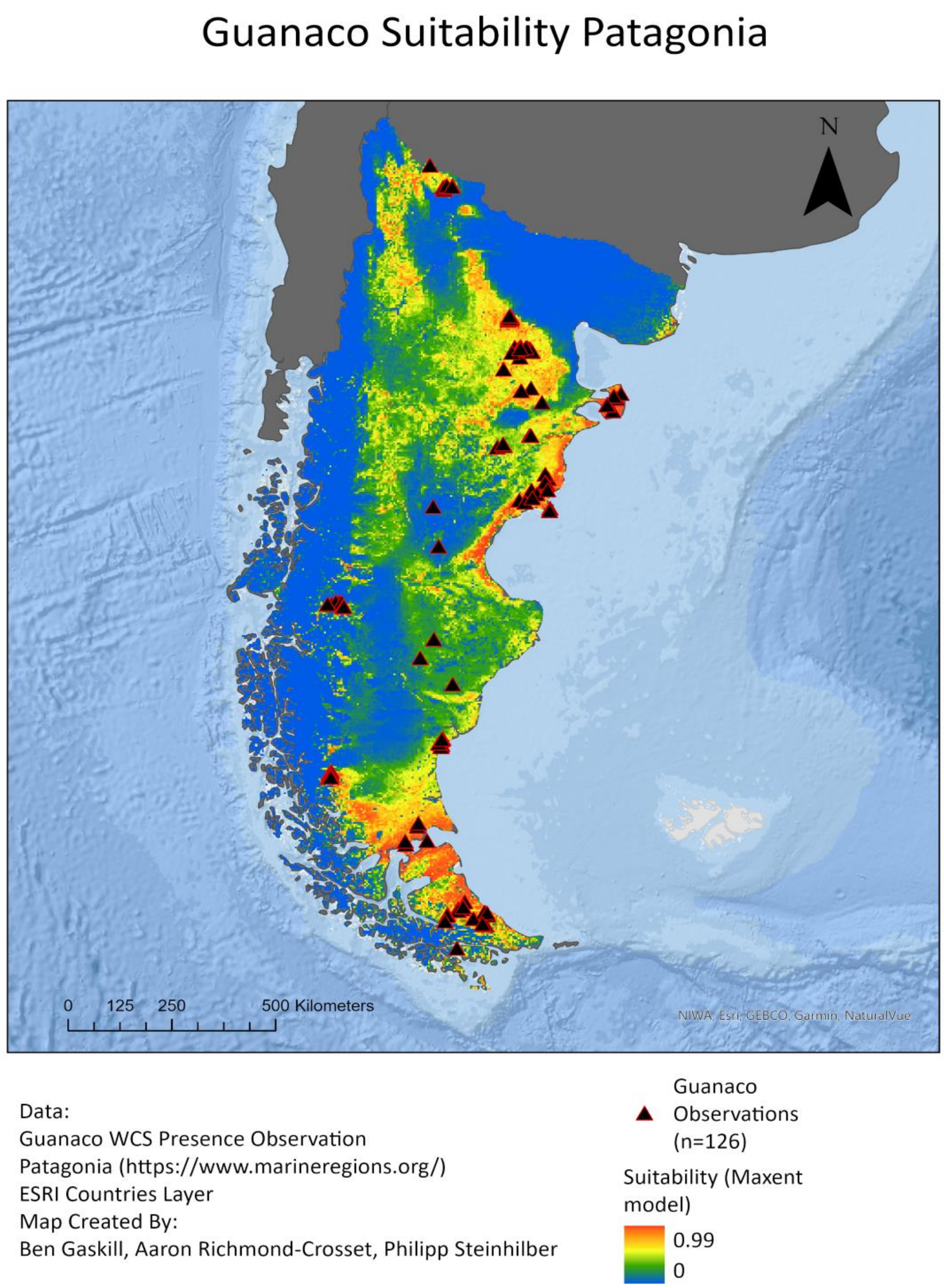


Figure 2: Guanaco Suitability Patagonia

Figure 3 illustrates the omission rate and predicted area as functions of the cumulative threshold for your Maximum Entropy species distribution model.

The blue line represents the omission rate, which shows the proportion of presence training samples that fall outside the predicted suitable area as the cumulative threshold increases. Ideally, this line should stay close to the black diagonal line, indicating the model's predictions align well with the training data. The red line shows the fraction of the background area that is predicted as suitable habitat. As the threshold increases, the predicted suitable area decreases, meaning fewer areas are classified as suitable.

The graph suggests that at low cumulative thresholds, the omission rate is very low, indicating the model captures most of the training presence records. However, as the threshold increases, the omission rate rises, meaning more suitable areas are excluded. Overall, the graph shows the robustness of the model, as the omission rate behaves as expected, and the predicted area decreases smoothly with increasing thresholds.

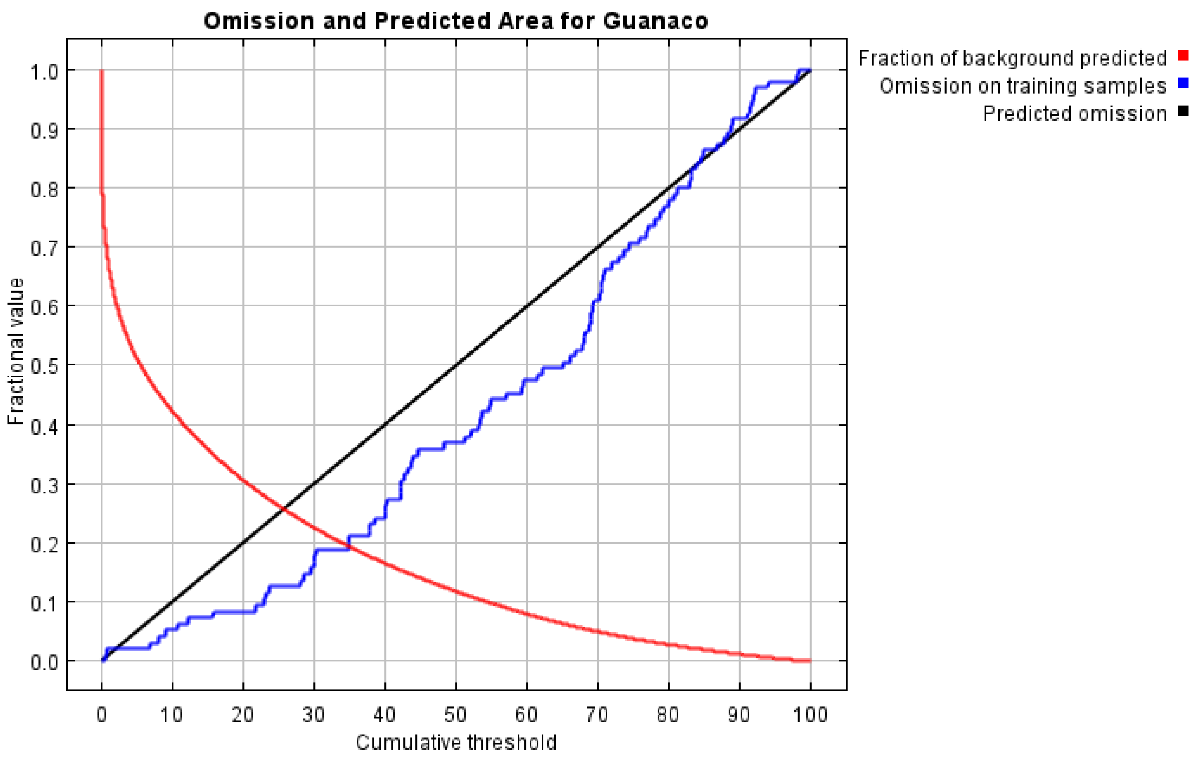


Figure 3: Omission and Predicted Area for the Guanaco

Figure 4 provides further insight into the performance of the model. The training area under the curve (AUC value is 0.875, which indicates that the model has high predictive accuracy. An AUC value close to 1 means that the model has a strong ability to distinguish between suitable and unsuitable habitats, whereas an AUC of 0.5 would indicate random prediction with no predictive power (black line). An AUC of 0.875 demonstrates that it performs significantly better than random chance.

The shape of the curve itself provides further insights. The steep initial rise shows that the model correctly identifies a large portion of suitable habitats early on, achieving high sensitivity. As the curve flattens near the top, it indicates that the model reaches highest sensitivity while predicting progressively larger areas. This behavior suggests that the model effectively identifies suitable habitats while also maintaining good specificity, meaning it minimizes the number of unsuitable areas falsely predicted as suitable.

The high AUC value also reflects the strength of the environmental predictor variables used in the model, such as NDVI, precipitation, or temperature, which appear to be strongly correlated with guanaco habitat suitability.

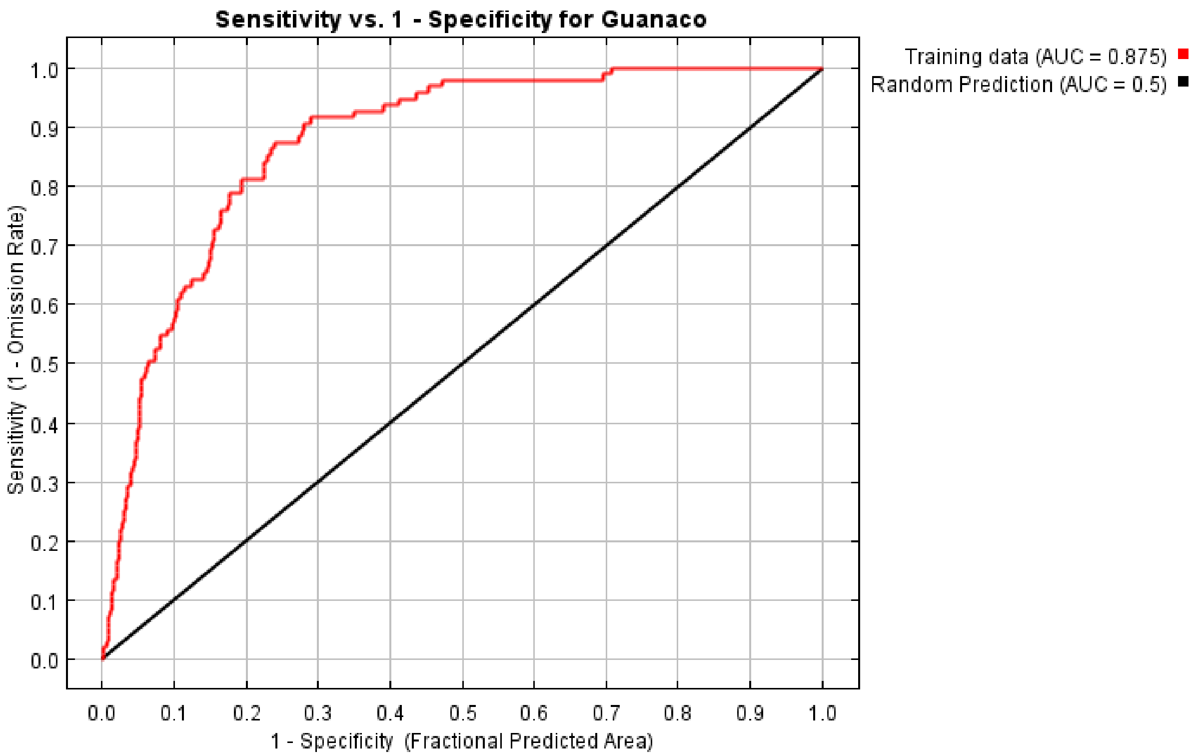


Figure 4: ROC for the Model

Analysis of variable contributions

The following Table 1: Relative Contribution of Environmental Variablegives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. The results indicate that NDVI had the highest importance, followed closely by precipitation.

Table 1: Relative Contribution of Environmental Variable

Variable	Percent contribution	Permutation importance
maxent_NDVI_msk_ASCII	32.3	29.1
maxent_precipitation_mask_ASCII	30.3	26.2
maxent_mean_temp_msk_ASCII	22.3	25.4
maxent_DEM_msk_ASCII	11.3	13.6
maxent_landcover_res_ASCII	3.8	5.8

Figure 5 shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is the NDVI, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is mean temperature, which therefore appears to have the most information that isn't present in the other variables.

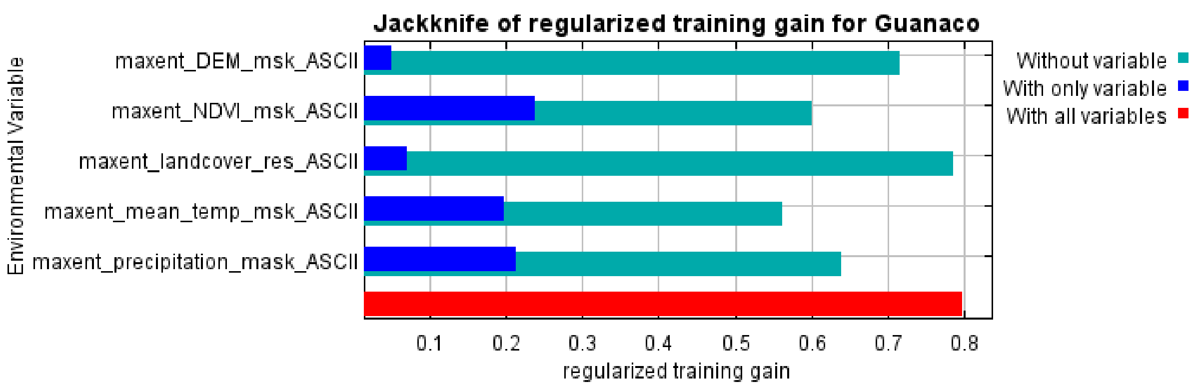


Figure 5: Jackknife Test of Variable Importance

Suitability Analysis

Based on the resulting suitability map (Figure 2), the mean suitability values for each protected area were extracted and plotted (Figure 6, Figure 7). The results clearly indicate that most protected areas with high suitability values are concentrated in the eastern part of Patagonia, while the western and northwestern regions are largely avoided.

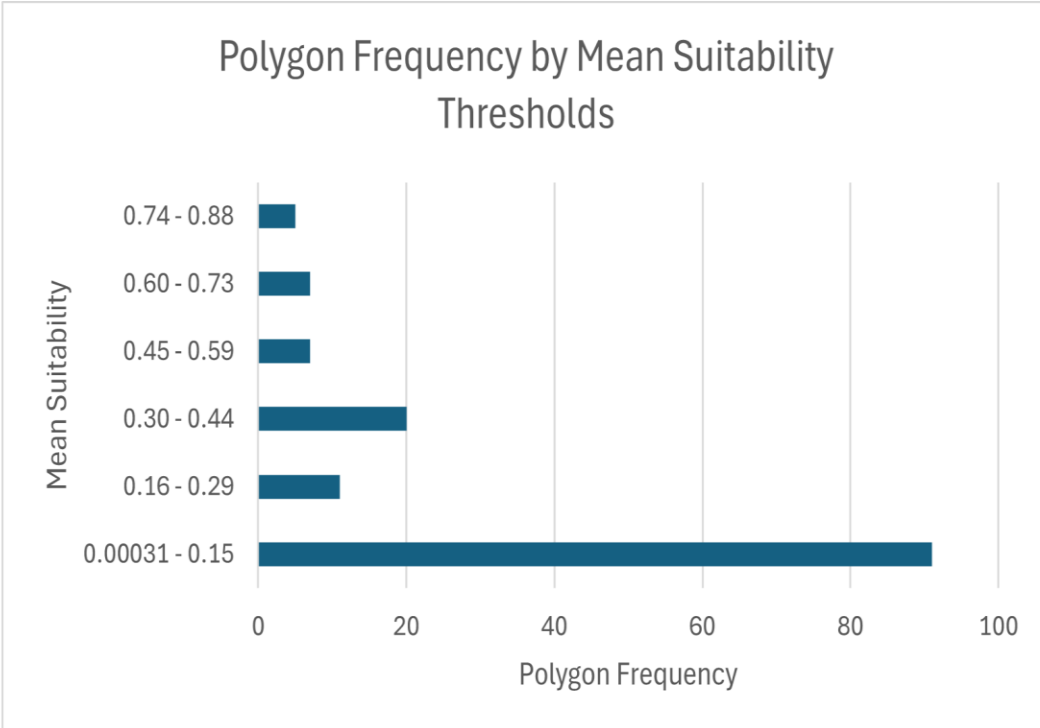
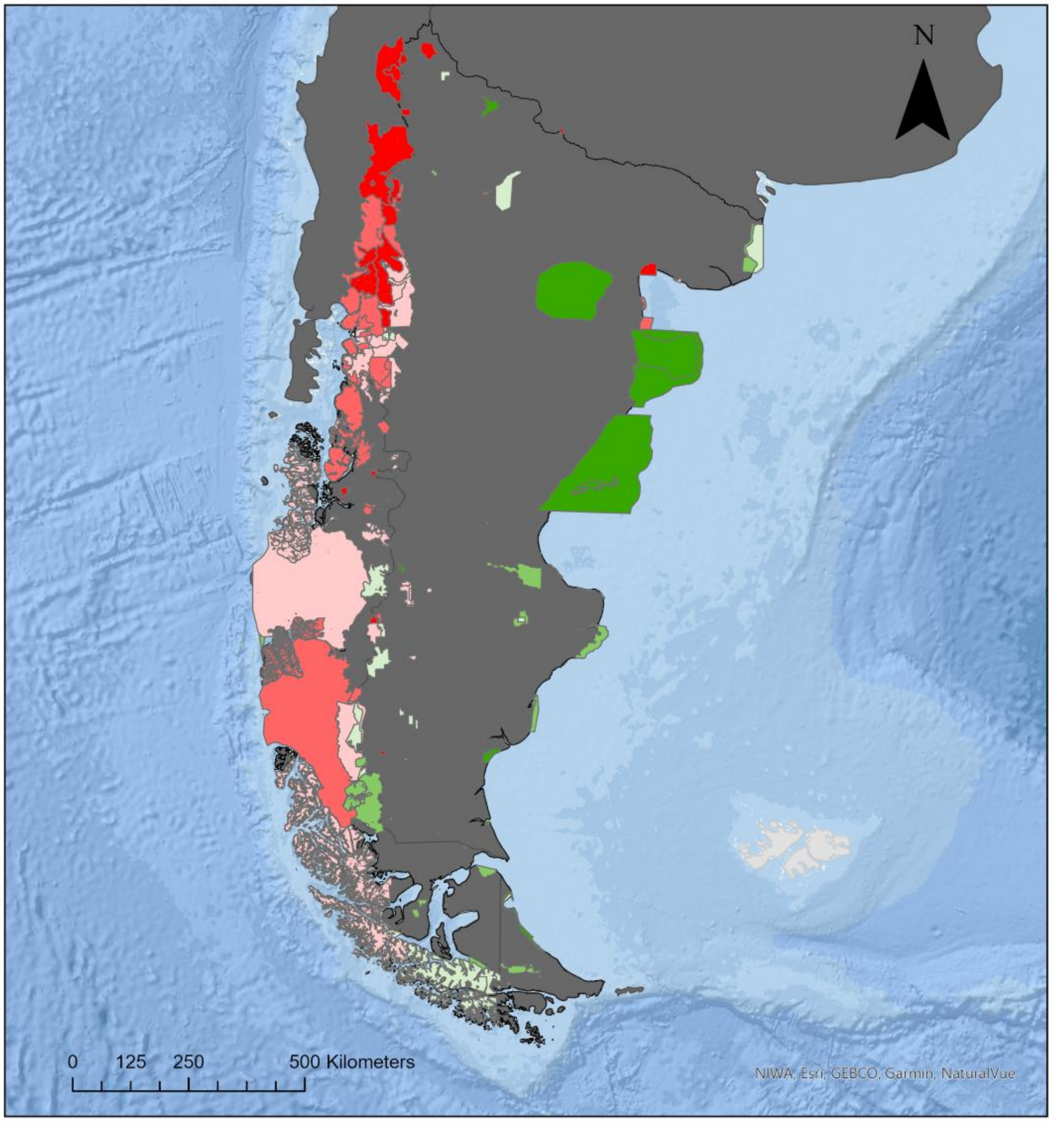


Figure 6: Polygon Frequency by Mean Suitability Thresholds

Mean Suitability per Protected Area



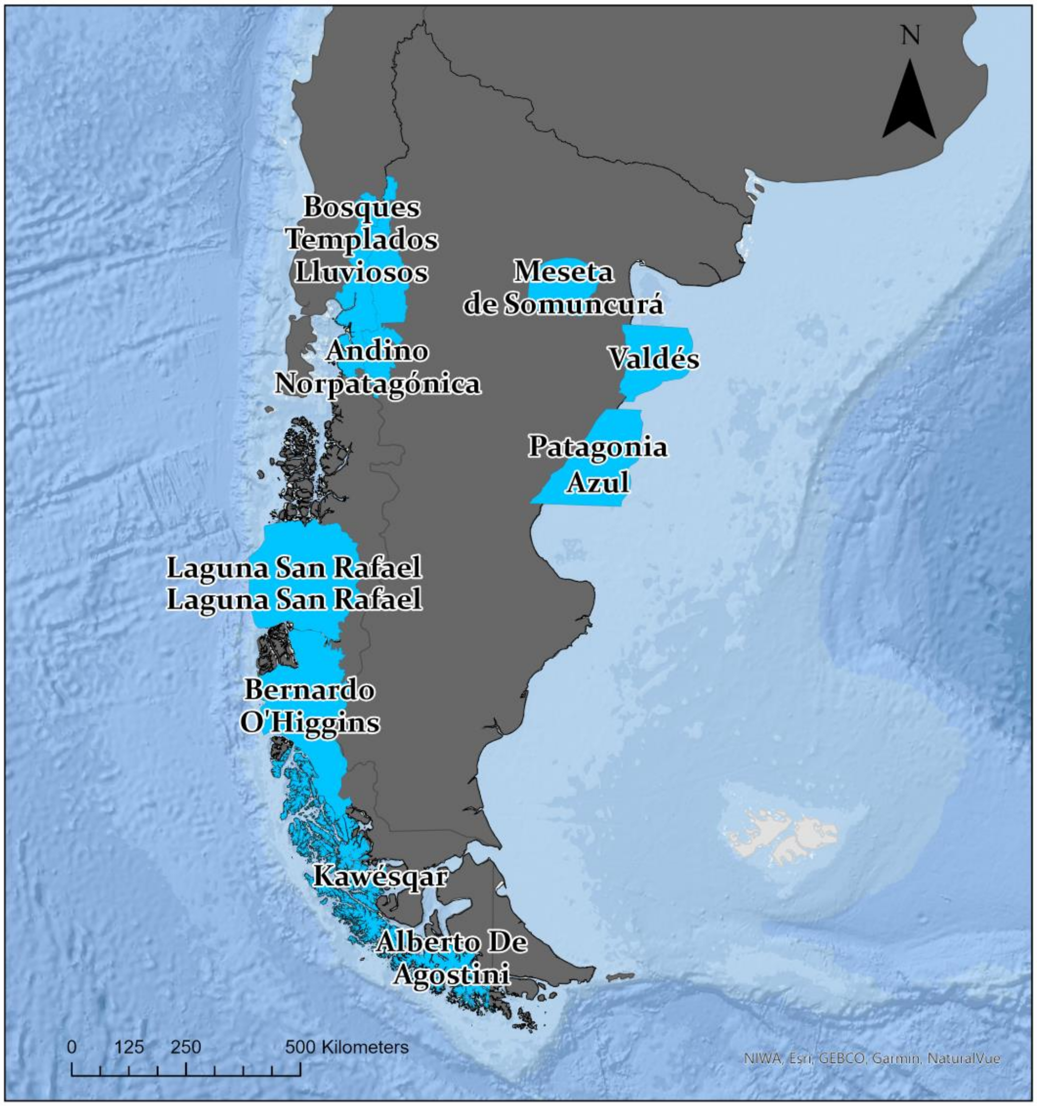
Data:
Protected Areas Boundaries
(https://www.protectedplanet.net/en)
Patagonia (https://www.marineregions.org/)
ESRI Countries Layer
Map Created By:
Ben Gaskill, Aaron Richmond-Crosset, Philipp Steinhilber

Mean Suitability per Polygon

- 0.000307 - 0.007698
- 0.007699 - 0.031605
- 0.031606 - 0.095186
- 0.095187 - 0.213728
- 0.213729 - 0.439913
- 0.439914 - 0.879138

Figure 7: Mean Suitability per Protected Area

Largest Protected Area (sq km)



Data:
Protected Areas Boundaries
(https://www.protectedplanet.net/en)
Patagonia (https://www.marineregions.org/)
ESRI Countries Layer
Map Created By:
Ben Gaskill, Aaron Richmond-Crosset, Philipp Steinhilber

Figure 8: Largest protected areas by km²

Risk Analysis

In the next step, an anthropogenic risk map was generated using the distance to settlements and roads within the study area. Euclidean distances to both settlements and roads were calculated, and the resulting rasters were combined and rescaled to range from 0 to 1, where 0 represents the lowest risk and 1 the highest risk (Figure 9).

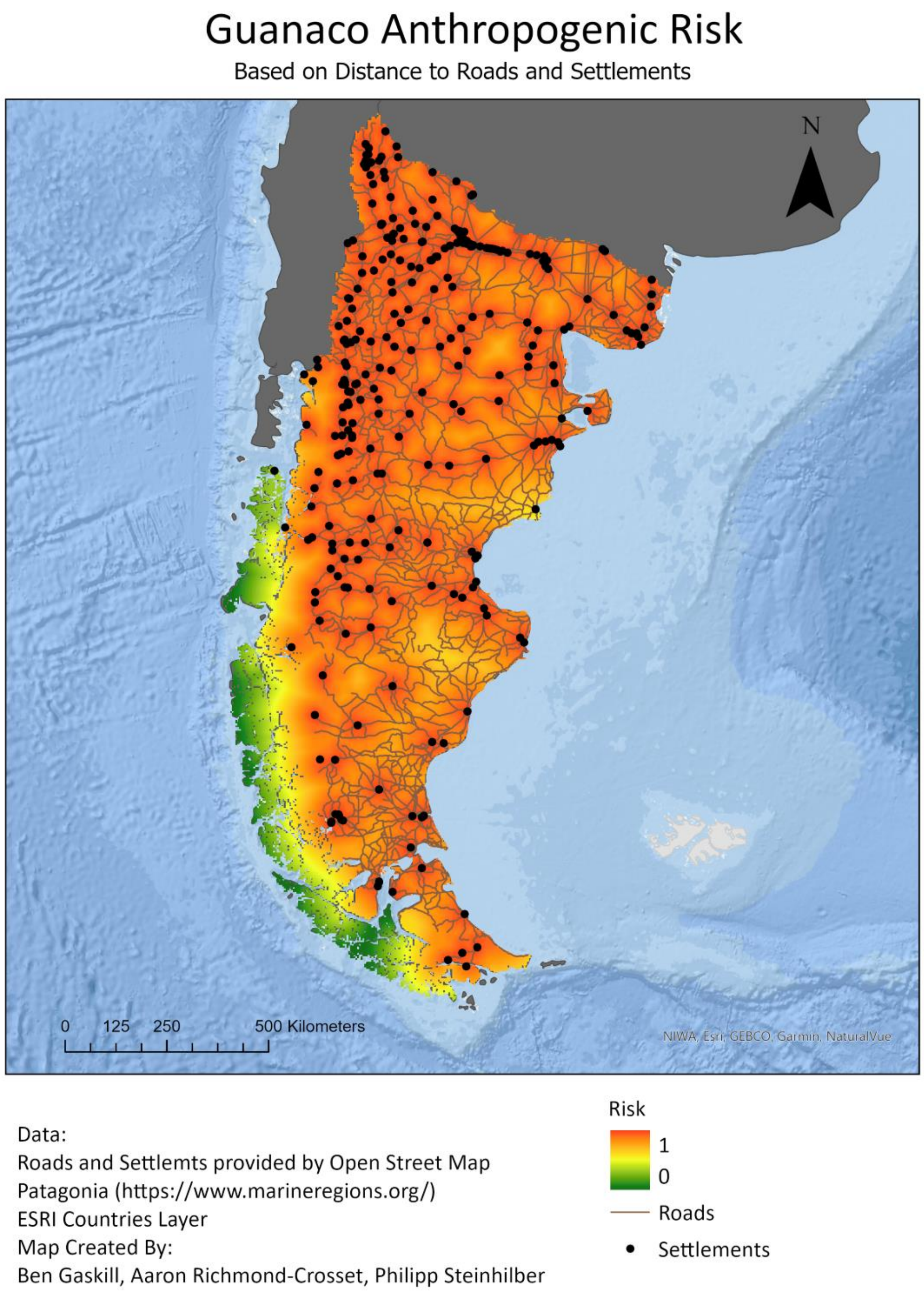


Figure 9: Guanaco Anthropogenic Risks

From the risk map, the mean risk factor for each protected area was extracted (Figure 10). The spatial pattern contrasts with that seen in Figure 7. In this case, protected areas with the lowest mean risk factor are concentrated in the northwestern part of the region.

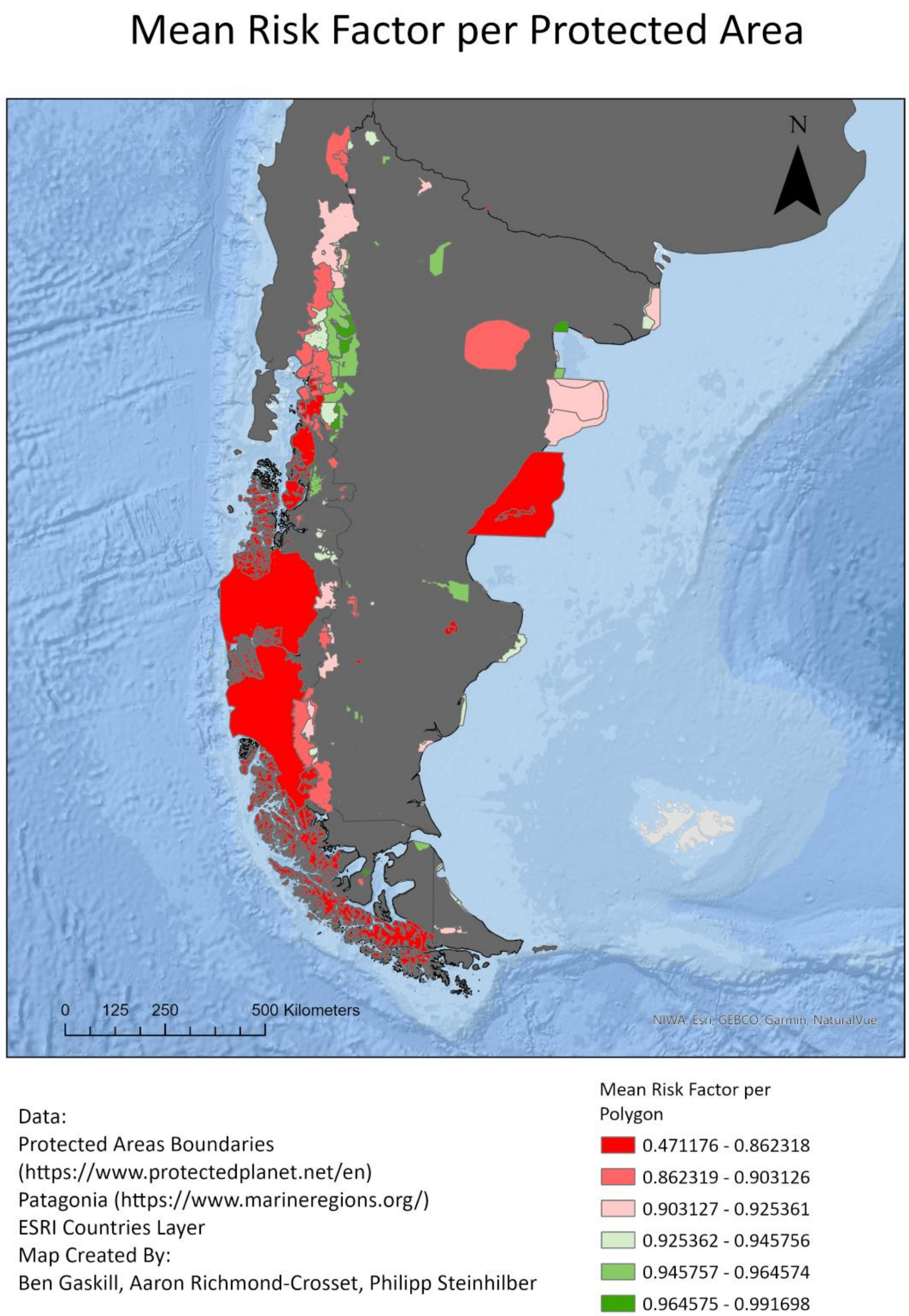


Figure 10: Mean Risk Factor per Protected Area

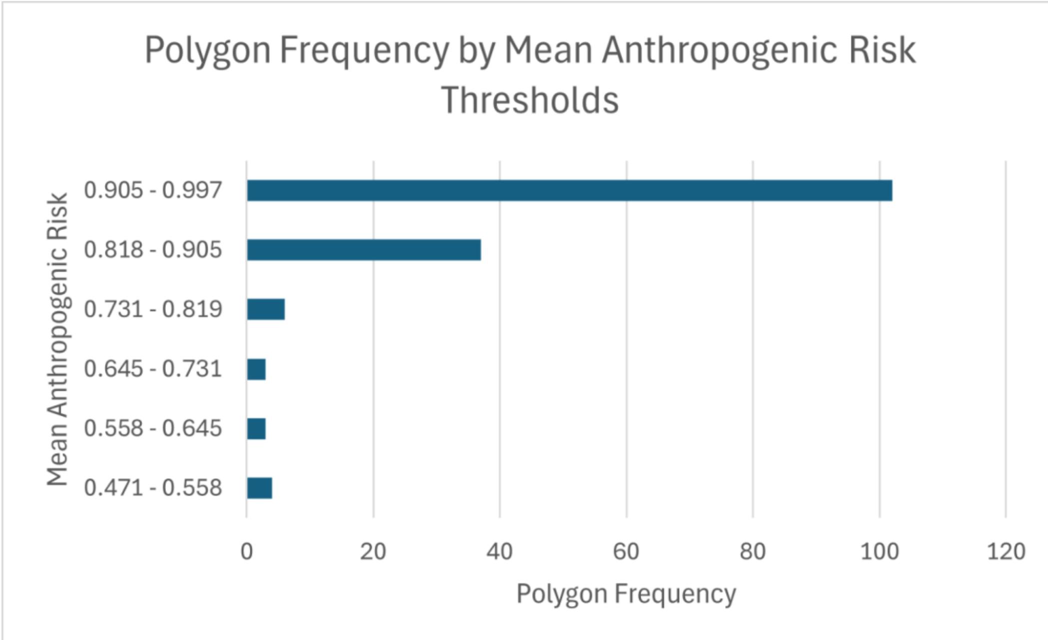


Figure 11: Polygon Frequency by Mean Suitability Thresholds

With this information extracted, the protected areas in Patagonia can now be analyzed based on their mean habitat suitability and vulnerability (indicated by the risk factor). Figure 12 reveals a clear spatial pattern, in which the largest cluster of protected areas exhibits high vulnerability but low suitability values. The vulnerability factor remains high across nearly all protected areas. Existing protected areas that encompass the most suitable habitat for the guanaco, while also facing the highest risk from anthropogenic disturbance, are in the upper right corner of the graph. These areas are particularly crucial for the conservation of the species and should be the primary focus for future expansion.

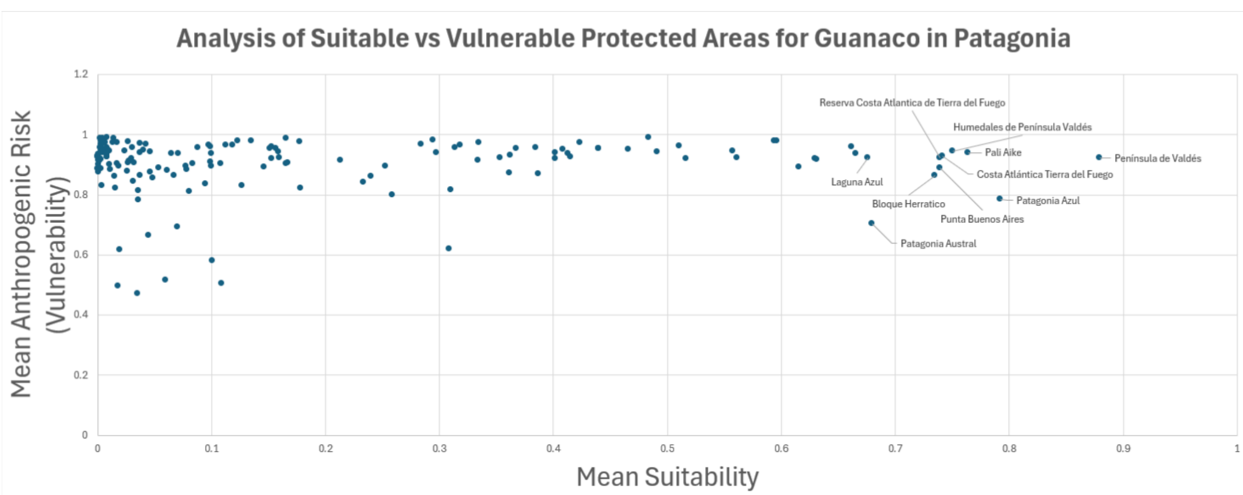


Figure 12: Analysis of Suitable vs. Vulnerable Protected Areas

Conclusion

The Analysis of suitable protection areas for Guanaco in Patagonia included determining the fundamental niche of the species based on observation data and relevant environmental variables. Comparing the result to already existing protected areas, a clear discrepancy between the species' habitat and the protected areas were exposed. While areas with highest suitability values occur predominantly in east/southeastern Patagonia, most protected areas are concentrated in the western part of the region.

Moreover, a risk parameter was calculated for every protected area based on anthropogenic disturbance. It is recommended resources be allocated first to the areas where there is high suitability and high anthropogenic risk, as depicted in Figure 12.

A complete list of protected areas, as well as their associated suitability and risk value, can be downloaded and should be viewed in Microsoft Excel (see Attachments).

Acknowledgements

We would like to thank the Wildlife Conservation Society (WCS) for providing Guanaco Presence data. Additional thanks to John Rogan, Florencia Sangermano and Adlai Nelson for their unwavering support and dedication to our project, Clark University and its students.

Tables

Table 2: Data Sources

Name	Source	Spatial Resolution	Temporal Resolution	Type
Guanaco Presence Data	Wildlife Conservation Society (WCS)	N/A	N/A	Points (VDC)
Conservation / Non-Conservation Land	Protected Planet	N/A	N/A	Polygons (.shp)
Normalized Difference Vegetation Index	Landsat 9 (derived and calculated using Google Earth Engine code)	250 meters	2023 Mean	.tif
Mean Monthly Temperature	WorldClim 2000-2021 (aggregated / calculated mean)	21 kilometers	2000-2021 Mean	.geotiff
Mean Monthly Precipitation	WorldClim 2000-2021 (aggregated / calculated mean)	21 kilometers	2000-2021 Mean	.geotiff
Landcover	Global Land Analysis and Discovery (Paper)	30 meters	2019	.tif
DEM	USGS, Verdin, K.L. Study area: zip9 Continent: all sa_dem.zip	30 meters	2017	.tif
Roads	OpenStreetMap	N/A	N/A	.shp
Settlements	OpenStreetMap	N/A	N/A	.shp

Table 3: Guanaco Habitat Suitability and Risk of the Top Ten Largest Protected Regions by Area

NAME	**Area (SQ Km)**	Mean Suitability Per Polygon	Mean Anthropogenic Risk Per Polygon
Laguna San Rafael	**51330.34435**	0.070491	0.693003
Bernardo O'Higgins	**44776.49542**	0.019981	0.618236
Patagonia Azul	**30595.69375**	0.792155	0.786045
Kawésqar	**28000.66399**	0.059861	0.515454
Andino Norpatagónica	**23574.31925**	0.040552	0.947678
Bosques Templados Lluviosos	**21619.27763**	0.026291	0.878134
Valdés	**19482.06621**	0.630151	0.919648
Laguna San Rafael	**17096.74667**	0.045266	0.665336
Meseta de Somuncurá	**16158.07156**	0.615532	0.8919
Alberto De Agostini	**11911.32559**	0.108777	0.505029

Table 4: Top Ten Protected Regions Most Suitable for Guanaco

NAME	Area (SQ Km)	**Mean Suitability Per Polygon**	Mean Anthropogenic Risk Per Polygon
Península de Valdés	3875.884185	**0.879138**	0.921621
Patagonia Azul	30595.69375	**0.792155**	0.786045
Pali Aike	50.941706	**0.763832**	0.940993
Humedales de Península Valdés	427.552529	**0.750784**	0.945664
Costa Atlántica Tierra del Fuego	761.640191	**0.741639**	0.928167
Punta Buenos Aires	82.398287	**0.73957**	0.888044
Reserva Costa Atlantica de Tierra del Fuego	794.261047	**0.73931**	0.923192
Bloque Herratico	26.282966	**0.734668**	0.86512
Patagonia Austral	1035.699518	**0.680027**	0.704061
Laguna Azul	13.034445	**0.675943**	0.921718

Table 5: Top Ten Protected Regions Most At-Risk for Guanaco

NAME	Area (SQ Km)	Mean Suitability Per Polygon	**Mean Anthropogenic Risk Per Polygon**
Bosque Petrificado Valcheta	1.641179	0.008326	**0.991698**
Laguna de los Cisnes	17.36832	0.483917	**0.991602**
Cañada Molina	1.701668	0.165591	**0.988507**
Playa Larga	0.321573	0.0024	**0.988259**
Capilla de Mármol	0.507264	0.014291	**0.988259**
Coyhaique	26.695029	0.005111	**0.987043**
Península San Julián	74.398124	0.294642	**0.981723**
Los Arrayanes	17.866741	0.003353	**0.980221**
Nant y Fall (Arroyo Las Caídas)	6.070165	0.135404	**0.979966**
Cuchillo-Cura	3.860666	0.594061	**0.979323**

Attachments

- Guanaco_Zonal_Statistics.exl

References

Baldi, R. B., Acebes, P., Cuéllar, E., Funes, M., Hoces, D., Puig, S., & Franklin, W. L. (2016). *Lama guanicoe*. The IUCN Red List of Threatened Species 2016: e.T11186A18540211. https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T11186A18540211.en

González, B. A., Palma, R. E., Zapata, B., & Marín, J. C. (2006). Taxonomic and biogeographical status of guanaco *Lama guanicoe* (Artiodactyla, Camelidae). *Mammal Review, 36*(2), 157–178. https://doi.org/10.1111/j.1365-2907.2006.00084.x

Phillips, S. J., Dudík, M., & Schapire, R. E. (2024). *Maxent software for modeling species niches and distributions* (Version 3.4.1). Available from http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed 2024.

UNEP-WCMC & IUCN. (2024). Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM) [Online], November 2024, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

Vilá, B., & Arzamendia, Y. (2022). South American Camelids: Their values and contributions to people. *Sustainability Science, 17*, 707–724. https://doi.org/10.1007/s11625-020-00874-y

Wildlife Conservation Society. (2024). Guanacos. *Wildlife Conservation Society*. Retrieved from https://www.wcs.org/our-work/species/guanacos