

SPECIES DISTRIBUTION MODELLING OF TWO SPECIES ENDEMIC TO THE PHILIPPINES TO SHOW THE APPLICABILITY OF MAXENT

M. Z. G. Untalan¹, D. F. M. Burgos¹, K. P. Martinez¹

¹Department of Geodetic Engineering, University of the Philippines Diliman -
(mguntalan, dmburgos, kpmartinez1)@up.edu.ph

Commission IV

KEY WORDS: Maxent, Species Distribution Modelling, Endemic Species, Likelihood

ABSTRACT:

Maxent is a machine learning model used for species distribution modelling (SDM) that is rising in popularity. As with any species distribution model, it needs to be validated for certain species before being used to generate insights and trusted predictions. Using Maxent, SDM of two endemic species in the Philippines, *Varanus palawanensis* (Palawan monitor lizard) and *Caprimulgus manillensis* (Philippine nightjar), were created using presence-only data, with 14 *V. palawanensis* and 771 *C. manillensis* occurrences, and 19 bioclimatic variables from BIOCLIM. This study shows the consistency to historical facts of Maxent on two endemic species of the Philippines of varying nature. The applicability of Maxent on the two very different species show that Maxent has high likelihood to give good results for other species. Showing that Maxent is applicable to the species of the Philippines gives additional tools for ecologists and national administrators to lead the development of the Philippines in the direction that conserves the biodiversity of the Philippines and that increases the productivity and quality of life in the Philippines.

1. INTRODUCTION

1.1 Background

Species distribution models (SDM) is utilized for mapping and monitoring of animal and plant species thus posing as a resource inventory and conservation mapping tool. SDM maps habitat using species observation and environmental and/or spatial predictors. It is commonly used to map habitat suitability across a landscape, e.g. a low suitability means that the confidence that a species is not present in an area is high. SDMs also acts as a 'dynamic information architecture' because of its flexibility to revise and further improve a model as new data becomes available (Natureserve, n.d.).

The earliest studies of SDM used BIOCLIM - an algorithm that uses climate data to determine the environmental requirements of a species to map out suitable locations for a species habitat. BIOCLIM became widely used then as it only requires species occurrence data (latitude, longitude, and elevation). The climate interpolation methods used by BIOCLIM was also used to create the WorldClim database. It is a common source of climate data for SDM studies. (Nix et. al., 2014). Climate data that can be obtained from WorldClim are 19 bioclimatic variable derived from monthly temperature and precipitation, representing annual trends, seasonality, and extreme or limiting environmental factors.

Another popular algorithm for SDM studies is Maxent which uses the concept of maximum entropy from information theory to create a probability distribution on the environmental variables for the likelihood of presence. Maxent performs SDM from presence-only records (Elith, Phillips, Hastie, Dusik, Chee, Yates, 2011) moreover, utilizes sets of environmental (such as climatic, soil, etc.) grids and georeferenced occurrence so as to represent each grid cell having a predicted suitability of conditions for the species (GBIF, 2017). Maxent support four outputs whose default is cloglog. It provides estimates ranging from zero (0) to one (1) of probability presence that is primarily

dependent on sampling size (specifically tile/quadrant size and observation time). (Philips, n.d.).Maxent has been widely used particularly for areas with limited systematic biological survey data and species occurrence records are presence-only records (Elith et. al., 2011). In 2014, about 76% of published SDM studies uses Maxent analyses and WorldClim data.

SDM literature for species endemic to the Philippines are limited. In this research, the SDMs of two species endemic to the Philippines are investigated to determine applicability of Maxent, along with the bioclimatic variables from the Worldclim database, in the Philippines.

1.2 Research Problem

Species distribution models (SDMs) have allowed ecologists and other scientists to study the ecological niches and geographic distribution of species. SDMs provide great predicting capability that contains several key insights into many species, provided that these models are fit for the species in consideration.

Open ecology data like occurrence records of species in the Philippines are scarce which means there is a scarce number of SDM literature on Philippine species. In this research, the applicability of Maxent for species endemic in the Philippines and using bioclimatic variables only to create SDM are investigated. SDM of two species endemic to the Philippines are to be created and assessed to determine the algorithm's applicability.

1.3 Significance

Varanus palawanensis and *Caprimulgus manillensis* are two species which are both endemic to the Philippines with very little literature. Nevertheless, showing that Maxent reveals information that agrees to these literature provides evidence that this model can be used for these species.

V. palawanensis is endemic to a specific portion of the Philippines, only in Palawan, while *C. manillensis* thrives all over the country, but only in the country. The former is very sensitive to its habitat while the latter can live almost anywhere on the country. The choice of *V. palawanensis* and *C. manillensis* was deliberate to show the flexibility of Maxent and therefore demonstrate high probability that it can be used for other species as well.

2. METHODOLOGY

2.1 Data Sources, Materials, and Software

The species occurrence data of *V. palawanensis* and *C. manillensis* were obtained from the Global Biodiversity Information Facility, an open-source website that provides species occurrence records. The occurrence records are georeferenced with WGS84 as the datum. The occurrences were limited to 1979 to 2019 and with human observation as the basis of record. Pertinent information about the occurrence records are shown in the table below.

Species	Number of Occurrences	DOI
<i>C. manillensis</i>	771	10.15468/dl.ncqdru
<i>V. palawanensis</i>	14	10.15468/dl.kvanyy

Table 1. Occurrence records

Code	Bioclimatic Variable	Remarks
BIO1	Annual Mean Temperature	
BIO2	Mean Diurnal Range	
BIO3	Isothermality	(BIO2 / BIO7) * 100
BIO4	Temperature of Seasonality	(Standard Deviation) * 100
BIO5	Maximum Temperature of Warmest Month	
BIO6	Maximum Temperature of Coldest Month	
BIO7	Temperature of Annual Range	(BIO5 - BIO6)
BIO8	Mean Temperature of Wettest Quarter	
BIO9	Mean Temperature of Driest Quarter	
BIO10	Mean Temperature of Coldest Quarter	
BIO11	Mean Temperature of Coldest Quarter	
BIO12	Annual Precipitation	
BIO13	Precipitation of Wettest Month	
BIO14	Precipitation of Driest Month	Coefficient of Variation
BIO15	Precipitation Seasonality	
BIO16	Precipitation of Wettest Quarter	
BIO17	Precipitation of Driest Quarter	
BIO18	Precipitation of Warmest Quarter	
BIO19	Precipitation of Coldest Quarter	

Table 2. Bioclimatic variables

The climate data used are bioclimatic variables downloaded from WorldClim. This contains 19 raster layers (Table 2) with a spatial

resolution of 2.5 minutes (4626.48 m by 4605.63 m). The bioclimatic variables are derived from the monthly temperature and rainfall values. Usually used in species distribution modeling and related ecological modeling techniques. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters) (WorldClim).

MaxEnt is an open source software which was used for the species distribution prediction of *V. palawanensis* and *C. manillensis*. The open source GIS software QGIS 3.4.7 was used to prepare the dataset, visualize and analyze the resulting predictions from Maxent.

2.2 Data Preparation and Cleaning

Bioclimatic variables are raster files covering global extents. For this study, we chose to analyze on the country level and the vicinity around the country.

2.3 Maxent

Steven Phillips created Maxent with the purpose of species distribution modelling. Together with this model, he had provided a Maxent software which is the standard for doing Maxent models for species. The software requires occurrence data and background or environmental data as primary input. Certain configurations were aligned to the purpose of this study. The output format chosen was cloglog which gives a rough estimate of a probability of presence. Nevertheless, the exact interpretation of cloglog, together with other output formats like logistic, is a topic of discussion. This is why the output must be interpreted when analysis has already been done, either through fieldwork or through research

2.4 Analysis

SDMs are great tools for modelling, but there is a danger of comfort and ignorance if left unvalidated through fieldwork and research. In this study, the outputs of Maxent are subjected to different analyses to show that the outputs of the model makes sense and are consistent with historical data and observations. The accuracies and variable importance should also be consistent with historical data.

3. RESULTS AND DISCUSSION

3.1 Scope: Country and Vicinity

Two scopes were chosen to validate the Maxent model. One looks solely at the Philippines and the other looks at the vicinity of the Philippines. These scopes are shown below. These two were chosen to explore the endemicity of the species within the Philippines and to the Philippines.



Figure 1. Scope of the study (top: Philippines, bottom: vicinity of the Philippines)

3.2 *Varanus Palawanensis*

There were 14 occurrence data points for *V. palawanensis*. No test samples were prepared because of the small sample size. The visualized prediction of the Maxent model is shown below.

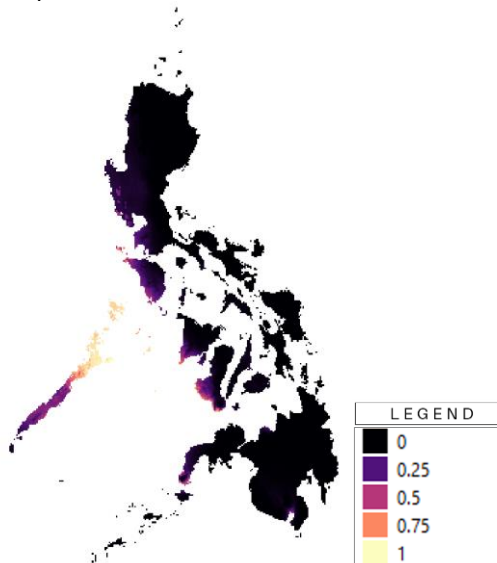


Figure 2. Visualized prediction of Maxent for *V. palawanensis*

The sensitivity VS specificity graph together with the area under the curve (AUC) is shown below. The high AUC shows that the model can predict the presence or occurrence with good accuracy

given the environmental variables which are bioclimatic variables. This suggests that the *V. palawanensis* needs strict climate conditions to thrive. This may be linked to the cold-blooded nature of this species - which means they rely on their environment to more or less regulate their temperature.

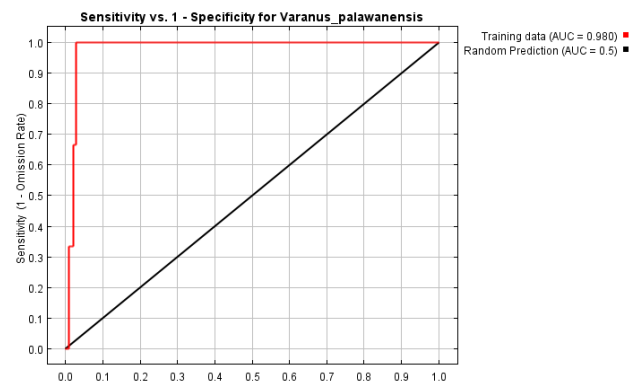


Figure 3. Sensitivity vs Specificity and AUC for *V. palawanensis*

Despite having no test samples, the model is validated through historic data and observation. It is known that *V. palawanensis* is endemic to Palawan which is consistent with the prediction of the model.

Endemicity has many factors, one factor is the availability of the right conditions for survival which includes both biotic like prey and abiotic factors like temperature. *V. palawanensis* is likely to be very sensitive to climate. The result of the prediction of Maxent for the presence of *V. palawanensis* in the vicinity of the Philippines shown below predicts that it is very likely to only be found in Palawan in the Philippines. This is consistent with the records that *V. palawanensis* is endemic to the Philippines.

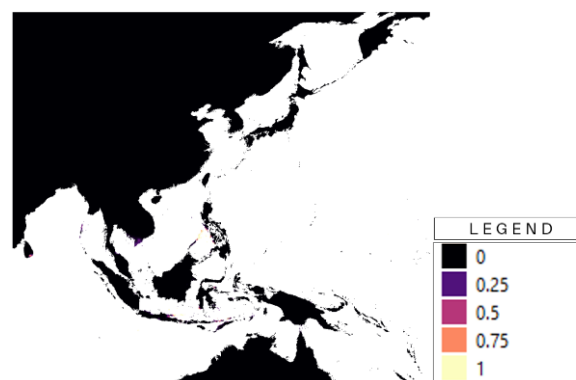


Figure 4. Visualized prediction within the Philippines' vicinity for *V. palawanensis*

3.3 *Caprimulgus manillensis*

Unlike *V. palawanensis*, *C. manillensis* is part of the Aves class commonly known as the birds class which are endothermic creatures, which means they can regulate their own body temperature. The ability to regulate body temperature enables animals to be more adaptive and more accepting of a variety of climatic conditions. *C. manillensis* is endemic to the Philippines and can likely thrive anywhere in the Philippines. The prediction of the Maxent model for this species on the Philippines is consistent with this information as seen below.

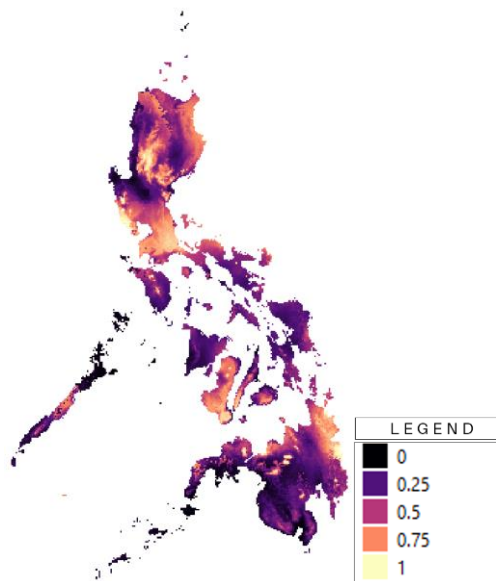


Figure 5. Visualized prediction of Maxent for *C. manillensis*

There were 771 occurrence data points for *C. manillensis*. This allowed the splitting of the data into a training set and test set with a 70:30 split. The sensitivity and specificity graph for this model together with its AUC is shown below. The graph shows good and acceptable performance for the model. The AUC for the train and test set differ by a considerable amount which is likely due to the fitness of *C. manillensis* to a wide range of climatic conditions in the Philippines.

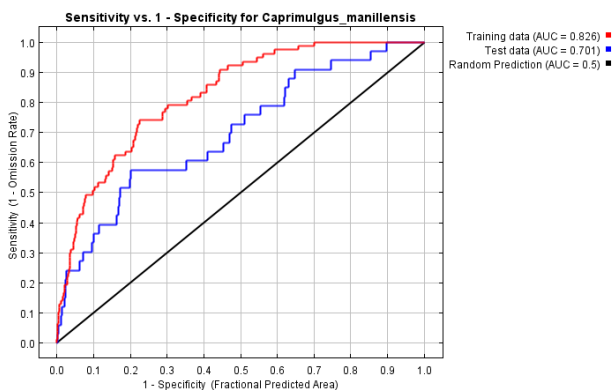


Figure 5. Sensitivity vs Specificity and AUC for *C. manillensis*

It is true that *C. manillensis* can live practically anywhere in the Philippines which may suggest that it has no problems adapting to a wide range of climate conditions, but it is endemic to the Philippines which suggests that it may be comfortable with the varying climate throughout the archipelago, but it may not be fit for just any climate conditions especially around the Philippines. The results of Maxent for the vicinity of the Philippines shown below predicts that *C. manillensis* can survive somewhere else, but places far from the Philippines which may explain the endemic status of the species. Aside from abiotic and biotic factors, dispersal or movement also plays a huge role in endemism. To illustrate this, consider the following scenario. Suppose a terrestrial species that live in an isolated island can thrive in another island, but because it cannot travel to that distant island, it remains endemic to that island.

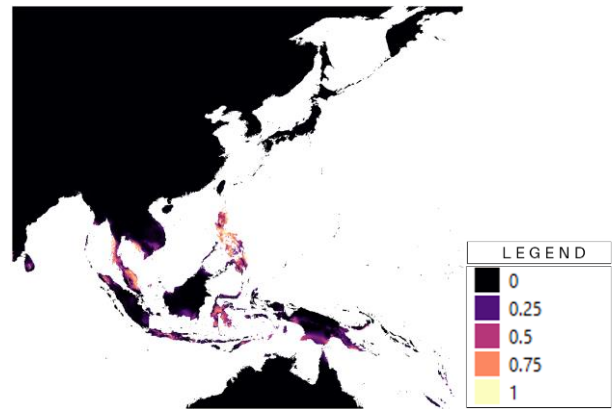


Figure 6. Visualized prediction within the Philippines' vicinity for *C. manillensis*

Different species in the same genus may not thrive in the exact environment conditions, but they tend to have similar and related environment conditions. To get a hint if *C. manillensis* can survive in the indicated environments, species of the same genus are explored. The cloglog prediction for the presence or occurrence of *C. manillensis* is turned into a binary map indicating where it is likely present. To create the binary map, the occurrence data points for the species is coupled with their cloglog Maxent values and the cloglog values within the raster which are greater than the first quartile of cloglog values coupled to the occurrence data points are labelled as where *C. manillensis* usually thrives. The binary map is shown below.

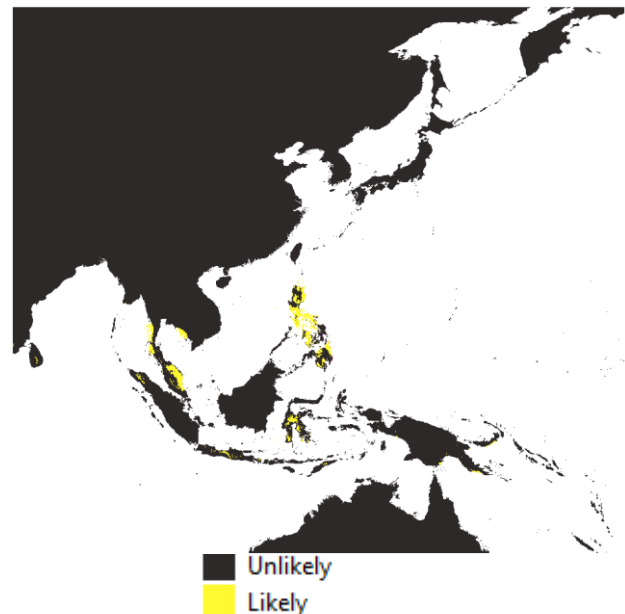


Figure 7. Binary map of the likelihood of occurrence of *C. manillensis*

The presence of *Caprimulgus* in the places marked where *C. manillensis* can thrive confirms the likelihood of *C. manillensis* being able to survive in these marked places. This is illustrated below.

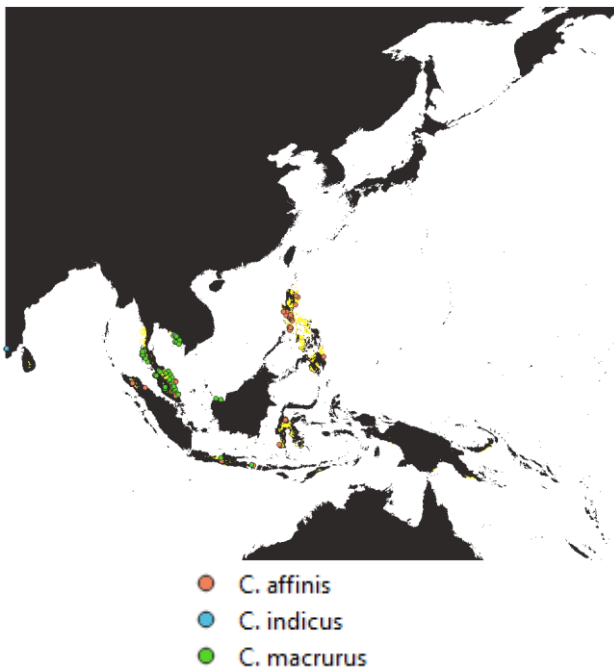


Figure 8. Occurrence of Caprimulgus within the Philippines' vicinity

Three species of the Caprimulgus genus which are affinis, indicus, and macrurus can be found in the places where manillensis is likely to be in. Although this is not definite evidence that *C. manillensis* can thrive in places outside the Philippines, it does give an idea of likelihood.

3.4 Maxent in the Philippines

The results and findings above show the possibility of use of Maxent for two species in the Philippines. Only a few literature can be found on these two species and none discusses the key environmental attributes that they need in order to survive. The limited information on their ecological niches based on abiotic factors has lead the modelling to rely on the default parameters usually used for terrestrial species which are the bioclimatic variables. Further studies on these species can help improve modelling by being able to identify environmental attributes that are crucial to the survival of these species. Despite using only climate variables, the model was able to be consistent historical facts. This suggests that these species are sensitive to some aspect of climate. Further discussion on the effect of certain aspects of climate to these species may provide starting points for further studies in ascertaining important environmental variables in modelling their distribution both in the geographical space and in the environmental variables space.

3.5 Contribution of Aspects of Climate to the Prediction of Occurrence

Since Maxent is a machine learning model which performs iterations to improve performance, it can track by how much the predicting capability of the model improves by changing the weight to specific environmental variables. Through this, certain variables are seen as more important than others.

The prediction for the occurrence of *V. palawanensis* was largely governed by precipitation of driest month, precipitation of the warmest quarter, and temperature annual range which contributes 48.8%, 26.2%, and 19.4% to the prediction, respectively. This suggests that precipitation plays a huge role in the survival of this

species. Additionally, temperature range or the difference between the hottest temperature and coldest temperature also affects this cold blooded species.

The prediction of the occurrence of *C. manillensis* was largely governed by the mean diurnal range, annual precipitation, mean temperature of warmest quarter, and precipitation of warmest month which contributes 24.4%, 13.4%, 12.6%, and 12.4%, respectively.

3.6 Jackknife Analysis on the Importance of each Environmental Variable

The discussion has so far focused on the importance of environmental variables framed on how much they contribute to the prediction while using all the variables. The next discussion will frame importance in a different way. The next discussion will look at the performance of the model when certain variables are removed.

The jackknife test consists of two processes per variable. The first process looks at the prediction ability of the model with just a specific variable. The second process looks at the prediction ability of the model if that specific variable is ignored and only the rest are taken as inputs.

The jackknife results for *V. palawanensis* as shown below show that the model was able to do relatively well when only BIO18: Precipitation of Warmest Quarter was used for prediction. This means that it has the most information by itself concerning the habitat of *V. palawanensis*. The model performed worst without BIO14: Precipitation of Driest Month which means that there is key information contained in this variable which is not present in the other variables.

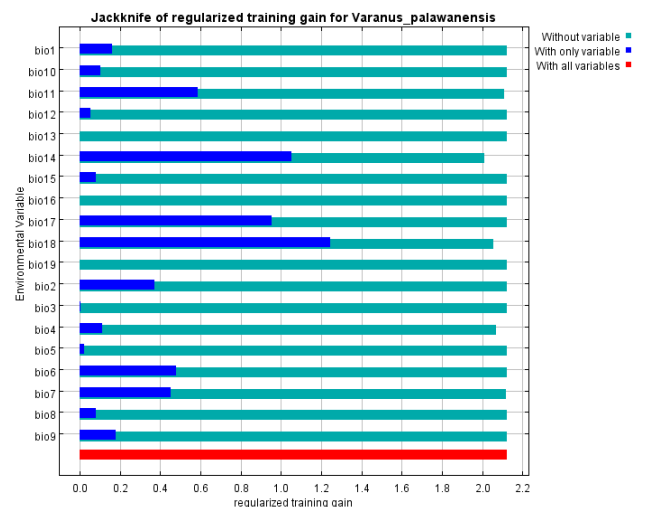


Figure 9. Jackknife analysis for *V. palawanensis*

The jackknife results for *C. manillensis* for AUC is shown below. The AUC is a metric that related sensitivity and specificity - a higher AUC means better predicting performance for the model. The results below shows that BIO5: Max Temperature of Warmest Month alone will yield the greatest performance for the model and that BIO8: Mean Temperature of Wettest Quarter alone will yield the poorest performance for the model which suggests that is has crucial information for predicting the likely habitats of *C. manillensis*.

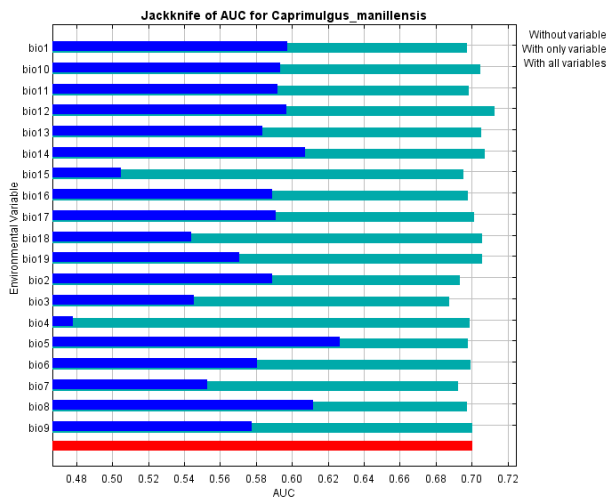


Figure 10. Jackknife analysis for *V. palawanensis*

3.7 Impact to Conservation

The likely places where species could thrive is not information that only ecologists should know. It is crucial that national administrators concerned with the development of the country know the distribution of the species in order to protect the ecological balance within the country. Ignorance to these could lead to a great reduction in the population of certain species. species distribution models (SDMs), together with ecological niche models (ENM), which works under the same conceptual framework as SDMs but is interpreted differently, allow concerned people to create responsible decisions that minimize the negative effects of human intervention with that natural course of landscape transformation - For example, deforestation to erect human settlements.

In this study, Maxent was examined for two species which are both endemic to the Philippines. Biodiversity of the Philippines is one of the richest in the world, containing 70% to 80% of the world’s plant and animal species. Using the insights in SDMs and ENMs such as Maxent can help conserve the rich biodiversity in the country. The table below shows the important aspects of climate to the two species. *V. palawanensis* seems to be affected by precipitation the most and *C. manillensis* seems to be affected by temperature the most.

Species	Most Important by Contribution	Allows greatest performance by itself	Seems to have the most key information
<i>V. palawanensis</i>	Precipitation of Driest Month	Precipitation of Warmest Quarter	Precipitation of Driest Month
<i>C. manillensis</i>	Mean Diurnal Range	Max Temperature of Warmest Month	Mean Temperature of Wettest Quarter

Table 3. Important aspects of climate that affects *V. palawanensis* and *C. manillensis*

Many aspects of urbanization and modernization threatens many species. To these two species, deforestation would likely have large adverse effects on the population, especially to *V. palawanensis* which appears to be more dependent on climate than *C. manillensis*. Massive deforestation in palawan especially in their habitat may disturb the local climate enough to affect the

population of *V. palawanensis*. Multiple high-rise buildings in a concentrated area promote higher temperatures because sun rays have to bounce to the long spaces in between the buildings with each bounce making the concrete they are incident on hotter. *C. manillensis* have been seen in urban areas - poor urban planning can make temperatures unfit for *C. manillensis*.

4. CONCLUSION

MaxEnt is rising in popularity in species distribution modelling and ecological niche modelling. In this study, Maxent has produced predictions for two endemic species in the Philippines which are consistent with historical observations. The considerable disparity in their nature and response to climate provides evidence that Maxent can likely be used for other species in the Philippines as well. The insights from the model can then be used to provide information that can be used for proper conservation of the two species which national administrators can use to lead the development of the Philippines in the direction that does not involve sacrificing or damaging the biodiversity of the Philippines.

5. REFERENCES

Elith, J., Phillips, Hastie, T., Dusik, M., Chee, Y., Yates, C., 2011. A statistical explanation of Maxent for ecologists. *Diversity and Distributions (Diversity Distrib.)* 17, 43-57. doi.org/10.1111/j.1472-4642.2010.00725.x.

Global Biodiversity Information Facility, 2019. GBIF Occurrence Download. doi.org/10.15468/dl.kvanyy.

Global Biodiversity Information Facility, 2019. GBIF Occurrence Download. doi.org/10.15468/dl.kvanyy.

Global Biodiversity Information Facility, 2017. Maxent. .gbif.org/en/tool/81279/maxent. (May 2019)

NatureServe. Advancing Regulatory Process and Conservation Outcomes with Improved Distribution Data for At-risk Species. www.natureserve.org (May 2019).

Nix, H., Booth, T., Busby, J., Hutchinson, M., 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current Maxent studies. *Diversity and Distributions (Diversity Distrib.)* 20(1), 1-9. doi/epdf/10.1111/ddi.12144.

Phillips, S., Dudik, M., Schapire, R. Maxent software for modeling species niches and distributions (Version 3.4.1). /biodiversityinformatics.amnh.org (May 2019).

Phillips, S. A Brief Tutorial on MaxEnt. biodiversityinformatics.amnh.org (May 2019).

QGIS Development Team, 2019. QGIS Geographic Information System. *Open Source Geospatial Foundation Project*. qgis.osgeo.org (May 2019).

WorldClim, 2019. World Climate Data. worldclim.org. (May 2019)