Puerto Rico Ecological Conservation

Mapping Land Cover to Inform Endangered Frog and Bird Species Distribution and Conservation Planning in Puerto Rico

Spring 2025 | North Carolina – NCEI
April 4th, 2025

Authors: Emma Rubin (Analytical Mechanics Associates), Julia Mateski (Analytical Mechanics Associates), Nicholas Montalbano (Analytical Mechanics Associates), Joseph Gibbs (Analytical Mechanics Associates)

Abstract: Climate change, deforestation, and urban expansion are threatening endemic birds and amphibians of the Puerto Rican Archipelago, altering their habitat and distribution. WildMon, a nonprofit organization, uses eco-acoustic surveys and other technologies to create species distribution models that inform conservation of these species. This project assessed the feasibility of using NASA Earth observations to make land cover products to improve WildMon's models. We computed cloud-free, temporally composited multispectral reflectance images of Puerto Rico using Landsat 8 Operational Land Imager (OLI) and Landsat 9 OLI-2 imagery from 2013-2015 and 2023-2025, respectively. Due to insufficient older historical satellite data, the pre-existing 2001 National Land Cover Database (NLCD) map was leveraged as a historical baseline. To create the land use change maps, we used the NLCD deep learning classifier tool in ArcGIS to process the intermediate and current imagery into land cover maps. Given the tool was developed with continental U.S. data over temperate environments, this classifier did not appear to perform as well on a tropical landscape. We derived bioclimatic variables using Empirical Bayesian Kriging Regression Prediction tool in ArcGIS to interpolate DAYMET weather data into 30-m resolution grids, using Shuttle Radar Topography Mission (SRTM) topographical data as an explanatory variable. These variables and the 2023-2025 composited imagery were used as inputs to an unsupervised Self-Organizing Map (SOM) algorithm to obtain a more detailed current land cover classification. The SOM was effective at mapping detailed land cover classes when combined with manual editing using bioclimatic and elevation data. Due to a lack of in situ data, we could not validate the accuracies of maps created in this project. This analysis did provide WildMon with valuable data products they can integrate into their species distribution models, supporting wildlife and habitat conservation efforts in Puerto Rico.

Key Terms: remote sensing, land use land cover, Landsat, Puerto Rico, biodiversity hotspot, unsupervised classification, wildlife habitat

Advisor: Joseph Spruce (NASA Langley Research Center)

Lead: Tallis L. Monteiro (North Carolina – NCEI)

1. Introduction

Climate change and land use changes are significantly altering species distributions globally, contributing to shifts in ecosystems and biodiversity patterns. In Puerto Rico, these anthropogenic and other environmental changes have significantly reshaped avian and amphibian habitats. The island's diverse ecosystems—including tropical rainforests, dry forests, and mangroves—are increasingly threatened by habitat fragmentation from urbanization and agricultural expansion (Lugo & Helmer, 2004; Miller & Lugo, 2009). Inefficient land use planning has led to extensive urban sprawl, further reducing habitat connectivity and isolating many bird and frog populations (Martinuzzi et al., 2007). Additionally, changes in precipitation patterns and temperature seasonality have altered the availability of suitable habitats (Campos-Cerqueira et al., 2021). Understanding how changes in climate and land cover influence species distributions is essential for informing future conservation efforts in the face of continuing environmental change.

Advances in NASA Earth observations and geospatial analytics provide an opportunity to enhance traditional species distribution models, which rely on ecological survey data. Integrating recurring and historical land cover maps into species distribution models allows conservationists to monitor habitat changes over time. Using satellite imagery also allows scientists to analyze land cover on a broad synoptic scale, where relying solely on field data would be impractical. This approach has been successfully demonstrated in previous research, such as Montero et al. (2021), where researchers mapped the land cover of Costa Rica to inform species richness of plants and birds using Landsat imagery. The latter study used machine learning to detect significant land cover changes across three time points, allowing them to statistically correlate land use change with species distribution.

This project was conducted in partnership with WildMon, a non-profit organization that develops globally accessible technology to support species monitoring and conservation planning. WildMon integrates emerging technologies such as AI and eco-acoustics to create a holistic approach to ecological monitoring. The organization contributes to conservation projects worldwide, including in the Puerto Rican Archipelago, where it is working to further our understanding of bird and frog species distributions. Understanding these distribution changes is critical, as range shifts can cause a range of impacts on plant and animal species, such as habitat fragmentation and changes in disease transmission (Pecl et al., 2017). Additionally, the U.S. Fish & Wildlife Service's Caribbean Ecological Services Field Office, the Puerto Rico Department of Natural and Environmental Resources' Division of Land Resources, Para la Naturaleza, and the University of Puerto Rico leveraged our results to inform decisions on land acquisition, land management, and species conservation through the use of NASA Earth observations.

Marconi Campos-Cerqueira, the Chief Science Officer of WildMon, studies the distribution of bird and frog species across Puerto Rico. In 2021, he and his team investigated suitable habitats for 13 bird species and 8 frog species of greatest conservation importance (Campos-Cerqueira et al., 2021). A key outcome of this research was the identification of "Always Suitable Areas" (ASAs)—habitats expected to remain viable for these species into the future. Notably, 80% of ASAs for both birds and frogs fell outside current protected areas. Their analysis found that precipitation during the warm quarter of the year was the most significant factor in influencing the distribution of both birds and frogs, while temperature seasonality played a major role in determining solely bird distribution. Despite the enduring presence of ASAs, species distributions are not static. Previous research showed that at least six Puerto Rican species experienced a range shift to higher elevations between 1998 and 2015, with increasing regional temperatures identified as the primary explanatory factor (Grau et al., 2003). Research related to extreme climate and weather events also revealed response differences between frogs and birds. Birds were more affected by drought than by hurricanes, while frogs, in contrast, showed no significant changes in distribution (Campos-Cerqueira & Aide, 2021). These studies point toward the importance of understanding how land use and land cover change, climate, and extreme weather events impact the distribution of bird and frog species.

We evaluated the feasibility of using NASA Earth observation data to map land cover to ultimately enhance species distribution models and support conservation efforts across the Puerto Rican Archipelago (Figure 1). Specifically, we aimed to map and analyze land cover changes in Puerto Rico and its sister islands, from 2001 to 2025 using remote sensing data. This was done through the creation of three simple land cover maps separated by intervals of about 10 years. To create these maps, land cover classifications were performed on three maps, two of which were derived from temporally composited imagery. To support Wildmon's creation of current species distribution maps, we also performed a more detailed land cover classification on the map derived from current imagery. This map included several additional layers, such as bioclimatic variables and terrestrial indices like Normalized Difference Vegetation Index (NDVI), that can be aligned with WildMon's eco-acoustic survey data and provide a better understanding of the different causes of species range shifts.

Figure 1. True color synoptic image of the Puerto Rican Archipelago including the main island of Puerto Rico, plus Isla de Mona, Isla de Culebra, and Isla de Vieques.

2. Methodology

2.1 Data Acquisition

2.1.1 – Earth Observations

To obtain cloud-free imagery for our land cover classifications, we used two methods, one for aiding the land cover change analysis and another for compiling the detailed land cover map. Because of the contrasting parameter requirements of the deep learning classification tool used to create our land cover change maps and the LandTrender (Landsat-based Detection of Trends in Disturbance and Recovery) imagery, we were unable to use this imagery in its delivered form as an input for the NLCD deep learning tool in ArcGIS. Ultimately, we used the U.S. Geological Survey's (USGS) Earth Explorer Landsat Collection 2 Level 2 imagery for our land cover change analysis maps, and LandTrendr-composited Landsat reflectance imagery for our separate, more detailed land cover map. Landsat data from the LandTrendr tool were employed due to there being fewer amounts of no data pixels.

For the first method, we obtained Landsat 8 Operational Land Imager/Thermal Infrared Sensor (OLI/TIRS) imagery from 2013-2015 for the intermediate land cover map and Landsat 8-9 OLI/TIRS imagery from 2023-2025 for the present-day map. Imagery was acquired from the USGS Earth Explorer. We chose imagery with the lowest amount of cloud cover available throughout the dry period (November-April) but expanded our search by +/- two months from this time to obtain sufficient imagery for compositing. This period was chosen due to the relatively lower cloud cover and to maintain phenological consistency. Imagery was obtained spanning the study area.

The second method for acquiring imagery was through the U.S. Forest Service's online and open-access application called LandTrendr Data Explorer (Kennedy et al. 2010). This application creates a temporally composited and cloud-masked raster based on dates of interest and the Landsat platforms selected by the user. We selected Landsat 8-9 OLI/TIRS imagery from 2023-2025 for the present-day map (Table 1) with the same temporal parameters as the method previously described. We obtained median composited Landsat imagery for the present-day map using default settings on Landtrendr. This included two composited rasters from each year, one from September 1 – December 31 (excluding 2025) and the other from January 1 – April 30.

2.1.2 – Pre-existing Land Cover Maps

Due to data quality issues and limited availability of imagery from the 1990s to early 2000s, we acquired a pre-existing land cover map to serve as our historical baseline. While multiple land cover maps of Puerto Rico exist before 2005 (Gould et al., 2008; Kennaway & Helmer, 2023), we ultimately decided to utilize the 2001 National Land Cover Database (NLCD) Commonwealth of Puerto Rico Land Cover Layer offered by the U.S. Geological Survey (Homer et al., 2004; Table 2). A raster of this map was downloaded from the Multi-Resolution Land Characteristics (MRLC) Consortium website. We chose this map in coordination with project partners which allowed us to utilize the Land Cover Classification (Landsat 8) deep learning model trained on the 2016 NLCD database to classify our intermediate and current dates of land cover maps (ESRI, 2020). Doing so allowed us to apply a congruent NLCD classification system to each land cover map to calculate land cover change.

Additionally, we acquired the LANDFIRE 2022 Existing Vegetation Type (EVT) map of Puerto Rico from the U.S. Geological Survey site to cross-reference against our output and to extract some classes for our final detailed map (LANDFIRE, 2023). We chose this map because it is one of the few up-to-date land cover maps of Puerto Rico. The researchers who created this map were also able to integrate in situ data and other validation means to distinguish spectrally similar classes, particularly those related to development and other high-reflectance categories.

2.1.3 – Topographic Data

We obtained elevation data through the NASA Shuttle Radar Topography Mission (SRTM) Version 3 Digital Elevation Models (DEMs) collection on NASA Earth Data Search (Table 1). These DEMs contain 30-m resolution elevation data with a vertical accuracy of ±16-m (Mukul et al, 2017). These data were collected by the Space Shuttle Endeavour during its mission in February 2000.

2.1.4 – Climate Data

We acquired DAYMET weather station data from Earth Data Search. These data were collected from 24+ unique weather stations across the main island of Puerto Rico. We acquired data ranging from January 1st, 2023, until December 31st, 2023 (Table 2). DAYMET weather station data are collected daily and include datasets for minimum temperature, maximum temperature, and precipitation.

Table 1

Earth observations used in this study

Sensor and Platform	Spectral Bands and Features	Level	Spatial Resolution	Temporal Coverage	Project Use
Landsat 8 OLI	Landsat Bands 2 - 7	Collection 2 Level 2 Surface Reflectance	30-m	2013-2015	Imagery used in land cover classifications
Landsat 9 OLI	Landsat Bands 2- 7	Collection 2 Level 2	30-m	2023-2025	Imagery used in land cover

		Surface Reflectance			classifications
SRTM	Elevation (DEM)	Version 3 Vertical Accuracy ±16 m	30-m	February 2000	Extract topographic variables

Table 2
Ancillary data used in this project

Name	Temporal Coverage	Product Type	Data Type	Data Provider	Use Case
DAYMET Weather Station Data	2023	Point Data	Version 4.1	DAYMET	Interpolating climate data across Puerto Rico
2001 National Land Cover Database Map of Puerto Rico	2001	Land Cover Map	Raster, 30-m	MRLC	Historical baseline land cover map
LANDFIRE 2022 EVT Puerto Rico US Virgin Islands	2022	Land Cover Map	Raster, 30-m	USGS	Supplement detailed map & cross-reference for accuracy
Forest sites and land uses of Puerto Rico from the USFS Guide to the ecological systems of Puerto Rico	2009	Land cover maps and guides to the ecosystems of Puerto Rico.	Text and Image	USFS	Visual data validation and class creation

2.2 Data Processing

2.2.1 – Earth Observations

Before we could use the NLCD-based deep learning classifier, we first had to remove cloud-related pixels from each Landsat scene and composite the imagery to create cloud-free images of the Archipelago (Table A1). To remove cloud-related pixels, we used the Con Tool by selecting only cloud-free pixels and setting the rest of the values to no data. Next, we used the Clip raster function to clip the surface reflectance imagery for each scene to the corresponding cloud-masked QA bands. Following this step, we used the Cell Statistics tool to composite the imagery for each section of the study area by using the median surface reflectance value of each pixel. Once we completed this process for each section of the study area, the imagery was combined using the Mosaic to New Raster tool with the mean mosaic operator.

To obtain cloud-free median composited LandTrendr imagery, we used the Set Null tool in ArcGIS Pro to set all cloud-masked pixels (value of –32768) to no data. This process was performed on ~5–6 composited rasters for each study period. Once we set the cloud-masked values to no-data for all of the composited LandTrendr rasters, we performed a final median composite for each study period. To do this, we used the Cell Statistics tool in ArcGIS, processing the imagery as multiband rasters and ignoring no data values in the calculations.

2.2.2 – Pre-existing Land Cover Maps

To prepare the 2001 NLCD map as our historical baseline, we reprojected and reclassified the raster to align with our intermediate and current land cover maps. First, the raster was reprojected using the Project Raster tool in ArcGIS Pro to WGS84 to align with the rest of the data. Next, we reclassified the map using the Reclassify tool to a broader classification schema. All NLCD forest classes were combined into a single general "Forest" class, "Developed Open" and "Developed, Low Intensity" were combined into one "Developed, Low Intensity" class, "Developed, Medium Intensity" and "Developed, High Intensity" were combined into one "Developed, High Intensity" class, and finally, the NLCD "Herbaceous" and "Hay/Pasture" classes were combined into one "Herbaceous/Pasture." The remaining NLCD classes were left unchanged, resulting in a total of 10 classes.

Similarly, the 2022 LANDFIRE map was reprojected to WGS84 using the same means described above. Using the Extract by Attribute tool in ArcGIS Pro, open water, developed classes, quarries, mines, gravel pits, and cultivated crops were extracted and subsequently clipped out of the current imagery and other input layers to the detailed map classifier algorithm. We used the LANDFIRE map because, in addition to being one of the few relatively up-to-date land cover maps of Puerto Rico, this classification used in situ data and other means of validation to identify spectrally similar classes.

2.2.3 – Topographic Data

To obtain DEMs for the creation of topographic indices, we reprojected the raw SRTM data to the WGS84 coordinate system and clipped it to the study region. We then used the ArcGIS Pro tool Mosaic to New Raster to merge the 6 tiles to a 30-m resolution to create the final digital elevation model (DEM). To prepare for calculating hydrologic analysis, we filled the sinks in the raster to ensure accurate flow calculations using SAGA GIS.

2.2.4 – Climate Data

First, we converted DAYMET daily weather station data from the. netCDF format to a format usable by ArcGIS. Leveraging Python, we combined the observed climate variable, latitude, longitude, and date into a .csv for analysis in ArcGIS Pro. To fit the format needed to calculate several of the bioclimatic variables, we aggregated the data from daily observations to monthly averages for temperature data and monthly sums for precipitation data. Using the latitude and longitude, we extracted the elevation, slope, and aspect for each weather station point. This produces .csv files containing the following values for each weather station point: latitude, longitude, month, climate variable value, elevation, aspect, and slope. We repeated this process for each of the three DAYMET climatic variable datasets: precipitation, maximum temperature, and minimum temperature.

Monthly mean temperature was a necessary input to some of the bioclimate variable equations, but the weather stations that we used in this project only measures minimum and maximum temperature values. To ameliorate this, we averaged the temperature maximum and minimum by month to compute mean monthly temperature by month.

2.2.5 – Terrain and Spectral Index Calculations

We calculated several indices to potentially include in the algorithm to produce the current, detailed map. Not every index and terrain variable derived was used for this purpose as further analysis narrowed down which were most relevant to include. We calculated a topographic wetness index (TWI) to better understand how water accumulates across the topographically diverse islands (Equation 1). Prior to calculating flow accumulation, we derived the flow direction from the DEM. Flow direction indicates the steepest downhill path for each pixel in the raster, showing the direction in which water will flow. We performed this calculation using the D8 Flow Direction algorithm in SAGA GIS, where flow direction is assigned to one of the eight adjacent neighboring cells. After flow direction was derived, we calculated flow accumulation (A) using SAGA GIS. Flow accumulation refers to the number of cells that flow into a pixel. Flow direction is used to identify which cells add water to a particular downstream cell. We performed the final TWI calculated

using the TWI function in SAGA GIS using flow accumulation (A) and slope (B; Kopecky et al., 2021). Higher TWI values indicate areas with water accumulation potential.

$$TWI = \ln\left(\frac{A}{\tan(B)}\right) \quad (1)$$

We derived NDVI from the final composited imagery from section 2.2.1. Using ArcGIS Pro, we clipped the final composited imagery to only include the study area. Using the Raster Calculator tool, we input the red and near-infrared bands to obtain the final NDVI raster (Equation 2; Rouse et al., 1974). NDVI quantifies vegetation greenness, density, and health. The index ranges from –1 to 1, with values closer to 1 indicating dense vegetation. Values near 0 usually represent barren land. Negative values typically represent water or clouds.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (2)$$

We derived Normalized Difference Moisture Index (NDMI), which is used to detect moisture levels in vegetation, from the final composited satellite imagery (Equation 3). To do this, we clipped the final composited imagery from section 2.2.1 to only include the study area, and input Equation 3 into Raster Calculator to make the final raster (Wilson & Sader, 2002).

$$NDMI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \quad (3)$$

We derived the Modified Normalized Difference Water Index (MNDWI) from the final composited satellite imagery (Equation 4). We used MNDWI instead of the Normalized Difference Water Index (NDWI) because the classification algorithm misclassified urban development as water when NDWI was applied. MNDWI uses SWIR instead of NIR bands in the formula, which helps to differentiate open water features from built-up areas (Xu, 2006).

$$MNDWI = \frac{(Green - SWIR)}{(Green + SWIR)} \quad (4)$$

2.3 Data Analysis

2.3.1 – Climate Data

For this project, we calculated eight bioclimatic variables (Table B1). Using Python, we first applied the necessary computations to the inputs associated with each bioclimatic variable using the weather station data (O'Donnell & Ignizio, 2012). For example, for bioclimatic variable 1, or BIO 1, the temperatures recorded at each weather station were averaged by month, then that product was averaged across the year. Given the need for bioclimatic data that span the entire archipelago rather than a limited number of points, we leveraged these calculated weather station observations to interpolate prediction surfaces for each variable chosen. Specifically, we used the Empirical Bayesian Kriging Regression Prediction (EBK) geoprocessing tool in ArcGIS Pro. This tool combines two other interpolation methods, regression analysis and kriging, allowing for heightened accuracy than either of these methods alone (Krivoruchko & Gribov, 2019).

To begin this process of rasterizing the bioclimatic data, we imported the .csv files created for each bioclimatic variable into ArcGIS Pro using the XY Table to Point data upload option, creating point feature classes of the weather stations and their associated calculated climate values. These point layers served as inputs to the EBK Regression Prediction tool with the climate variable field selected as the dependent variable. For interpolations related to temperature, only elevation was selected as the input explanatory raster as initial analysis revealed a strong correlation between these variables and weak relationships with other

variables, like slope or aspect. For the interpolation related to precipitation, we selected elevation, NDVI, and NDMI as input explanatory variables for the same reasons, albeit these correlations overall were not as strong. Regarding the output layers options, while naming only the geostatistical layer is required to run the tool, we also provided names for the other two output options: prediction raster and diagnostic feature class. The prediction raster creates an interpolation surface covering the whole study area and the diagnostic feature class provides information about model accuracy such as the root-mean-square-error and mean absolute error. All other parameters were left as default. In the tool environment, the processing extent was set to the extent of all data in layers, the mask was set to our study area shapefile, and the default option for the coincident points option was removed, leaving it blank. With these set, we ran the tool for each variable described (Table B1).

Due to the computational intensity of this tool, we were unable to calculate all 19 bioclimatic variables detailed by the U.S. Geological Survey in our timeframe (O'Donnell & Ignizio, 2012). We chose the 8 specific bioclimatic variables we calculated due to their importance to our partner's research and their apparent ability to be interpolated accurately. We chose to interpolate bioclimatic variables that are based on mean values to avoid inaccuracies associated with interpolating totals. Since this rule eliminated estimating any precipitation-related bioclimatic variables, we chose to create our own, listed as Average Daily Precipitation (Table B1).

2.3.2 - Simplified Land Cover Maps

To obtain simplified land cover maps for use in the land cover change analysis, we used the Classify Pixels Using Deep Learning tool in ArcGIS with the ESRI premade Land Cover Classification (Landsat 8) deep learning model. This classification model used Landsat 8 and 9 imagery and was trained using randomly sampled locations of specific land cover types within the continental United States (CONUS) 2016 NLCD database. We utilized this model on our cloud-free manually composited imagery to obtain land cover maps with the NLCD classification schema for our intermediate and current periods. Finally, we reclassified the maps to align with our historical baseline schema previously described in 2.2.2. This reclassification was done because the CONUS-trained deep learning model misclassified specific forest types, wetland types, and herbaceous and hay/pasture land cover classes.

2.3.3 - Land Cover Change Maps

We obtained land cover change maps using the Change Detection Wizard in ArcGIS Pro. The Change Detection Wizard creates a streamlined process for change detection analysis by producing layers that represent the categorical change in land cover over time. First, we configured the two rasters and set the tool to the categorical change method. Next, we set up the class configurations, changing the filter method to only output pixels where land cover change had occurred, and including all the classes in the change detection workflow. Finally, we set smoothing to a 5 x 5 neighborhood to quiet the noise for the change detection visualization but did not do so for the change calculations. We carried out this process three times to create our change detection maps, including our historic to intermediate maps, intermediate to current, and finally, historic to current.

2.3.4 - Detailed Current Land Cover Map

Given the lack of in situ data needed for supervised classification, we used an unsupervised algorithm to map land cover for our detailed present-day map. We employed Self-Organizing Maps (SOMs) over other unsupervised algorithms like K-means because SOMs do not assume linearity and are effective at preserving topological relationships for high dimensional data. SOMs also do not require a predefined set number of clusters, making them more adaptable than the K-means algorithm. To perform our SOM classification, we used the MiniSom package in Python (Table D1).

For this process, we first input multispectral imagery, NDVI, annual mean temperature (BIO 1), and MNDWI into one stack (Table D1). Next, we scaled the data to ensure that the values stayed in a consistent range. We specifically utilized Standard Scaler, which centers the data around zero and makes all data points have a standard deviation of one, before running Principal Component Analysis (PCA) for dimensionality

reduction. PCA reduces the number of variables in the dataset while still preserving the most information possible. During PCA, the algorithm aims to find features that maximize variance. If one feature in the data stack varies more than others due to its scale, it would ineffectively dominate the principal components. After PCA, we obtained Scree and Cumulative Explained Variance Plots to identify the ideal number of Principal Components (PCs) to add to our final model. We found that three PCs were the ideal number to minimize computational power while preserving variance.

We then input the three PCs into the SOM algorithm from the MiniSom package in Python. Throughout the trials of this process, we tested numerous combinations of hyperparameters. Our final map output used a 7 by 7 grid size, a sigma value of 1, 5,000 training iterations, and a learning rate of 0.1 (Appendix B). We used random weights to initialize the neuron weights. We first tried using the random training method from the MiniSom package. However, this method was too computationally intensive. Instead, we leveraged the batch training method. After training, we set the algorithm to assign the best matching neuron to each pixel in the dataset and visualize the classification output.

In our first few iterations of this model, we classified all pixels in Puerto Rico. However, after visual evaluation, we realized that the SOM was not separating cultivated land from urban development well and was having difficulties identifying wetlands. To circumvent this, we opted to use the classes that were extracted from the 2022 USGS LANDFIRE map described in section 2.2.2 to mask out high-reflectance cluster classes that the algorithm was having difficulty distinguishing. Finally, we input the masked layers into the SOM to classify specific and biologically relevant vegetation types.

2.3.5 - Detailed Current Land Cover Map Refinement & Manual Editing

Next, the output raster from the SOM model was imported into ArcGIS for refinement and manual editing. The original map produced by the model had 20 classes. The outputs of SOMs are not defined to specific, labeled classes and therefore require manual work to distinguish, combine, and label classes. To do so, we first grouped the pixels in the SOM output classes into general categories, including forest, open space, and unclassified. To further distinguish specific classes like dry forest or beach, we then used conditional statements of elevation (DEM) and precipitation (Average Daily Precipitation) ranges related to each land cover type (Table D2; Murphy and Stallard, 2012, Chapter C). Coastal buffers were created to distinguish classes by filtering for pixels that only appear by the coast. A 500-m buffer was created and applied to the raw SOM output classes that were found to contain wetland, ensuring that only pixels within 500-m of the coast were included in the potential wetland class. A similar buffer of 50-m was applied to classes containing high reflectance pixels or open space to extract a beach class.

Finally, once all classes were established and reclassified, the LANDFIRE urban, cultivated, and water pixels were added to the map using the Raster Calculator tool in QGIS. The resulting layer was a final, detailed land cover map. We used the Zonal Statistics tool in ArcGIS to obtain pixel count values for statistical analysis and data visualization.

3. Results

3.1 Analysis of Results

3.1.1 Land Cover Change Maps

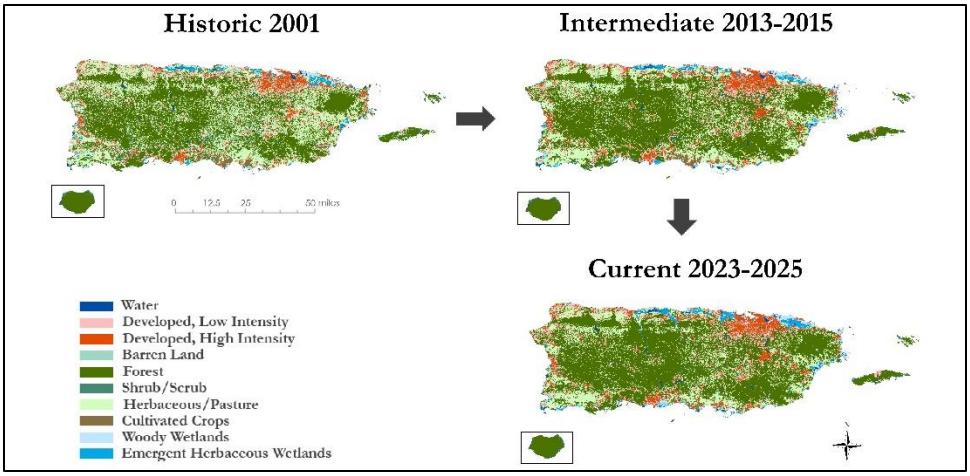

Figure 2. Land Cover Change Maps. Intermediate and current maps derived from Landsat 8 and 9 OLI cloud free imagery. Historic map is a reclassified version of the NLCD 2001 map.

This methodology produced three simplified land cover maps across three time periods, historic, intermediate, and current or present-day (Figure 2). We analyzed the number of pixels in each land cover class to better understand how they have changed between our three maps (Figure C1; Figure C2). There was a noticeable decrease in pasture/herbaceous and shrub/scrub land cover both from historic to intermediate and from intermediate to current. There was also a steady increase in forest area, about 20% from historic to intermediate and 6% from intermediate to current.

3.1.2 Climate Data

In all, we were able to produce eight bioclimatic variable maps at the 30-m resolution (Figure B1). The annual mean temperature and the mean temperature of the wettest quarter maps have similar spatial patterns, but different temperature values. This depicts the strong correlation between elevation and temperature, as well as the relationship of temperature to seasonal precipitation effects. The annual mean precipitation map shows a similar spatial pattern to previous precipitation maps, with the central and upper right portions of Puerto Rico being wetter and the southwest and northwest portions of the island being drier (National Weather Service & National Oceanic and Atmospheric Association, n.d.).

3.1.3 Detailed Land Cover Map

We computed a detailed current land cover map that included 14 terrestrial land cover classes, as well as open water and unclassified areas (Figure 3). We quantified and analyzed the number of pixels in each land cover class in the detailed land cover map to better understand the abundance of different classes (Figure D1). Overall, we found that developed and urban land cover classes took up 18%, open space occupying 23%, cultivated occupying 2%, and forest classes occupying about 57% of the island's area. Out of this 57% of forest cover, about 38% of this was lowland moist forest, 10% was lowland dry forest, and 8% was submontane wet forest. We also distinguished montane rain forest (also described as cloud forest) and mangroves, two land cover types of critical conservation importance due to their value to endemic species (Colón-Adorno, 2020). About 61 km^2 of mangroves and other coastal wetlands and 54 km^2 of lower montane rain forest were identified.

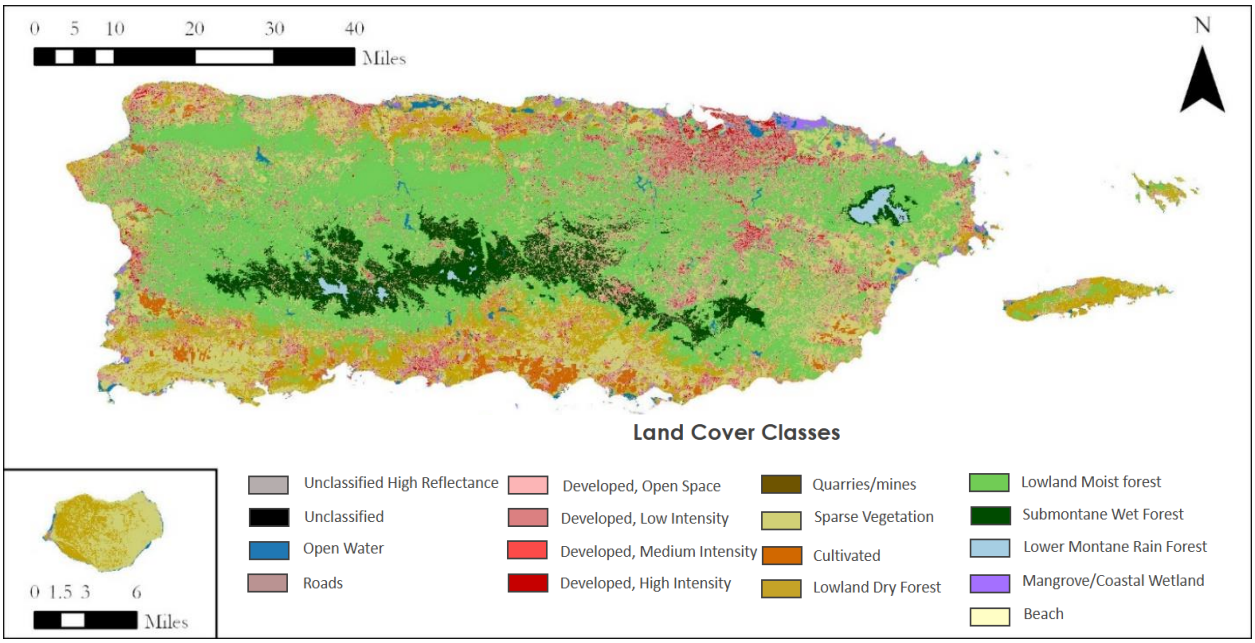

Figure 3. 2023-2025 detailed map of Puerto Rico with land cover classification legend, shown in the left corner is the Isla of Mona

3.2 Errors and Uncertainties

Throughout this analysis, we found that the most classification errors or uncertainties arose from the lack of in situ data. This prevented us from leveraging a potentially more robust supervised classification algorithm to classify land cover types. Using a supervised classification algorithm such as a Support Vector Machine algorithm or an XGBoost algorithm would have likely yielded more accurate results, assuming sufficient training data. Using an unsupervised classification algorithm as tried made it difficult to delineate between spectrally similar land cover types, such as cultivated crops versus urban development. Additionally, this hindered our ability to validate the results of our classifications. Specifically, we could not run typical classification validation quantifications such as confusion matrices. Instead, we had to rely on visual inspection of imagery, which can lead to subjective classification errors, particularly when 30-m resolution imagery does not show enough detail to clearly and consistently distinguish certain land cover types by eye.

For the land cover change analysis, our inability to access the imagery used to create the 2001 NLCD map created another source of uncertainty. Through this, we had limited ability to assess the spectral signatures of the classes to validate our simplified spectral scheme. The classification methodology used to create the NLCD 2001 map may not align well with the NLCD deep learning classifier tool we ran on the intermediate and current imagery, as sensors and preprocessing techniques have changed since 2001. Additionally, upon further inspection there appeared to be distortion or misalignment between some of the coastal areas of the 2001 NLCD compared with our intermediate and current maps and other current pre-existing landcover maps such as the 2022 LANDFIRE map, introducing additional potential error and uncertainty to our change maps including this historical baseline map. The NLCD deep learning classifier tool presented uncertainty because it was trained on continental US data collected over temperate landscapes and not tropical landscapes like Puerto Rico. Tropical environments can exhibit distinct seasonal vegetation patterns and therefore can have significantly different spectral properties. This complexity can lead to potential misclassifications, such as classification of forest types based on wetness and/or deciduousness. Given these environmental gradients, we resorted to aggregating all forest types into one category for maps using the NLCD classification schema.

In terms of the bioclimatic variables, some of the weather stations in the DAYMET data sources had missing data, which may have caused missed patterns in precipitation and temperature means. Additionally, weather

stations in Puerto Rico may disproportionally be in areas of lower elevation, which could cause bias in the interpolation.

4. Conclusions

4.1 Interpretation of Results

In the bioclimatic variable analysis, we were able to demonstrate the effectiveness of using Empirical Bayesian Kriging Regression Prediction for interpolating weather station data. The ability to generate 30-m resolution bioclimate data maps will be extremely beneficial to the partners as they develop species distribution models. Avian and amphibian species have different climate preferences, and having access to moderate spatial resolution data can enhance the accuracy of habitat suitability assessments and distribution models. After creating the detailed land cover classification map for 2023-2025, we cannot fully conclude that the use of SOMs for land cover classification is effective. To make this conclusion, this methodology must be tested when there is accurate and extensive ground validation data, which we did not have access to. While the SOM algorithm proved effective at identifying certain land cover patterns, it struggled in differentiating spectrally similar land cover types like agriculture and urban development. In the future, adding more variable inputs to the SOM model could be beneficial in improving accuracy. Being able to track urban sprawl and anthropogenic development is critical for monitoring bird and frog habitat loss. However, through the process of creating our final detailed map through masking and leveraging the LANDFIRE maps, we conclude that using multiple data sources can enhance classification accuracy. When comparing the final output of the SOM algorithm, it visually shows increased accuracy in comparison to the classification based on only the multispectral data from 1 data source. While this study could not produce concrete conclusions about the use and accuracy of these classification tools and outputs, our research provides ample avenues to carry out further analysis.

4.2 Feasibility & Partner Implementation

Through our analysis, we found that NASA Earth observations showed potential for aiding WildMon in their wildlife species distribution modeling. This methodology to create the bioclimatic map layers will enable the partners to generate bioclimatic variables for subsequent years. In the context of climate change, having up to date and high-resolution climate variable map layers will be useful inputs for species modeling. One goal of WildMon's next project is to use bioclimatic variables to consistently identify suitable areas for future habitat monitoring and restoration. To add more data to this model, the partners could interpolate DAYMET weather station data for past years using the interpolation method outlined in this study. However, this may be easier for temperature-related bioclimatic variables rather than precipitation-related bioclimatic variables, as precipitation related variables require interpolating totals, which are very sensitive to missing data. Also, temperature variables only used the elevation model as an explanatory variable, whereas precipitation variables relied on NDVI derived from Landsat imagery.

The land cover change maps and detailed map generated with our methodology may not be reliable for decision-making or analysis as further validation of these products would be necessary to determine the accuracy and level of error present. Barring in situ data, a random sample of points can be taken from the land cover maps we created and compared with high-resolution imagery to manually test their accuracy. With in situ data, more robust validation efforts may be pursued, such as the computation of confusion matrices for estimating overall and individual land cover class accuracies. In their current state, these products may provide value for general directional changes in certain land cover classes that appear to be more accurately represented such as forest. Moving forward, the methodology and cloud-free composited imagery for the intermediate and current periods could be leveraged to improve the land cover change analysis and calculations between these periods. One potential approach would be to train the NLCD deep learning model on training data created across the study area to enhance the accuracy of the generated NLCD maps. Another option would be to test supervised or unsupervised classification on the intermediate and current data, respectively, to try to produce more refined land cover and subsequent land cover change maps. Overall, our products—the bioclimatic layers, topographic variables & indices, and land cover maps—provide several

potential avenues to be useful for our partners in developing their wildlife species distribution models and planning the next natural resource conservation steps in Puerto Rico.

5. Acknowledgements

The Puerto Rico Ecological Conservation Team would like to thank Dr. Marconi Campos-Cerqueira (WildMon), Kris Harmon (WildMon), and the rest of the WildMon for their partnership and insight, Joe Spruce and Dr. Kenton Ross for their science advising, our Lead Tallis Monteiro (North Carolina – NCEI) for her support throughout the term, and Brent Bowler (Virginia – Langley) for Project Coordination advice and feedback.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract 80LARC23FA024.

6. Glossary

Bioclimatic Variable – biologically meaningful indicators derived from climate data, typically temperature and precipitation

Composite – in reference to imagery, composites are created by stacking and melding several satellite scenes to create a single image, usually to remove cloud cover

Digital Elevation Model (DEM) – A raster data set that captures elevation values of individual pixels. Widely used for topographic analyses and to derive geospatial indices

Earth observations – Satellites and sensors that collect information about the Earth's physical, chemical, and biological systems over space and time

Eco-acoustics – the collection of audio data to investigate natural and anthropogenic sounds and their relationship to the environment and ecosystem over a wide range of study scales

Empirical Bayesian Kriging (EBK) Regression Prediction – A geostatistical interpolation method that combines kriging with regression analysis. Explanatory variables are used to enhance the interpolation accuracy

Environmental DNA (eDNA) – genetic material left behind by organisms in an environment, which can be collected and used to identify and monitor species presence, ecological community dynamics, and biodiversity

Interpolation – in the context of climate analysis, interpolation is the process of using existing data at particular geographic locations to predict values in unsampled areas that surround these points

LandTrendr – An algorithm that identifies changes in pixel values over time

Modified Normalized Difference Water Index (MNDWI) – An index similar to NDWI except it uses shortwave infrared and green bands instead of near-infrared. This index better distinguishes between water features and other areas of high reflectance, like urban areas

Mosaic – in the context of imagery, mosaics are created by stitching several adjacent satellite scenes together to create one whole image

National Land Cover Database (NLCD) – A dataset that contains Landsat data to provide information on land cover change

Normalized Difference Vegetation Index (NDVI) – A geospatial index that measures vegetation health and density based on the difference between near-infrared and red-light reflectance. NDVI values range from -1 to +1, with higher values indicating healthier and denser vegetation.

Normalized Difference Water Index (NDWI) – An index focused on identifying water by measuring the difference in reflectance between the near-infrared and green bands of light. NDWI values range from -1 to +1, with values closer to +1 indicating open water.

Normalized Difference Moisture Index (NDMI) – An index that measures vegetation water content or moisture levels in vegetation

OLI – Operational Land Imager, a sensor on Landsat that measures different wavelengths of light

Principal Component Analysis (PCA) – A dimensionality reduction algorithm that reduces the number of variables in a dataset while maintaining as much variance as possible. This is done through transformation of the original dataset into variables called "Principal Components."

Python – a versatile computer programming language used in various applications, including data analysis

Self-Organizing Maps (SOMs) – An unsupervised neural network algorithm that takes in high dimensional data and projects it to a two dimensional scale while preserving the topology of the data.

Species Distribution Models (SDMs) – Models that predict where a species is likely to be present, typically based off environmental condition data, land cover, and presence/absence data.

Topographic Wetness Index (TWI) – An index that quantifies topographic-driven variation in soil moisture using slope and flow accumulation, which are derived from DEM data. The higher the value of TWI, the more water accumulates in this area.

Thermal Infrared Sensor (TIRS) – A sensor that measures land surface temperature

7. References

Campos-Cerqueira, M., Terando, A. J., Murray, B. A., Collazo, J. A., & Aide, T. M. (2021). Climate change is creating a mismatch between protected areas and suitable habitats for frogs and birds in Puerto Rico. *Biodiversity and Conservation*, *30*(12), 3509–3528. https://doi.org/10.1007/s10531-021-02258-9

Campos-Cerqueira, M., & Aide, T. M. (2021). Impacts of a drought and hurricane on tropical bird and frog distributions. *Ecosphere*, *12*(1), e03352. https://doi.org/10.1002/ecs2.3352

Colón-Adorno, A. (2020). Analyzing changes in Puerto Rican mangroves and local conditions to identify restoration opportunities. *Tropical Resources*, *39*, 00–00. https://tri.yale.edu/tropical-resources/tropical-resources-vol-39/analyzing-changes-puerto-rican-mangroves-and-local

ESRI. (2020). Land Cover Classification (Landsat 8) [Map]. ArcGIS Online. Retrieved April 2, 2025, from https://www.arcgis.com/home/item.html?id=e732ee81a9c14c238a14df554a8e3225

Gould, W. A., Alarcon, C., Fevold, B., Jimenez, M.E., Martinuzzi, S., Potts, G., Quinones, M., Solórzano, M. & Ventosa, E. (2008). The Puerto Rico Gap Analysis Project Volume 1: land cover, vertebrate species distributions, and land stewardship. Gen. Tech. Rep. IITF-39. *Río Piedras, Puerto Rico: U.S. Department of Agriculture, Forest Service, International Institute of Tropical Forestry*, 165 p. https://doi.org/10.2737/IITF-GTR-39

Grau, H. R., Aide, T. M., Zimmerman, J. K., Thomlinson, J. R., Helmer, E., & Zou, X. (2003). The ecological consequences of socioeconomic and land-use changes in postagriculture Puerto Rico. *BioScience*, *53*(12), 1159–1168. https://doi.org/10.1641/0006-3568(2003)053[1159:TECOSA]2.0.CO;2

Homer, C. G., Huang, C., Yang, L., Wylie, B. K. & Coan, M. J. (2004). Development of a 2001 National Land Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, *70*(7), 829–840, https://doi.org/10.14358/PERS.70.7.829

Kennaway, T. A., & Helmer, E. H. (2023). Puerto Rico, Vieques and Culebra land cover and forest formations circa 2000. *Fort Collins, CO: Forest Service Research Data Archive*. https://doi.org/10.2737/RDS-2023-0022

Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment*, *114*(12), 2897–2910. https://doi.org/10.1016/j.rse.2010.07.008

Kopecký, M., Macek M., Wild, J. (2021). Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of the Total Environment*, *757*. https://doi.org/10.1016/j.scitotenv.2020.143785

Krivoruchko, K. & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, *32*, 2211-6753. https://doi.org/10.1016/j.spasta.2019.100368

Landscape Fire and Resource Management Planning Tools (LANDFIRE), Earth Resources Observation and Science Center (EROS), U.S. Geological Survey (USGS), and Hatten, T.D., 2023, LANDFIRE 2022 (230) Update: U.S. Geological Survey data release, https://doi.org/10.5066/P974JF8W

Lugo, A. E., & Helmer, E. (2004). Emerging forests on abandoned land: Puerto Rico's new forests. *Forest Ecology and Management*, *190*(1), 145–161. https://doi.org/10.1016/j.foreco.2003.09.012

Martinuzzi, S., Gould, W. A., & Ramos González, O. M. (2007). Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population census data. *Landscape and Urban Planning*, *79*(3–4), 288–297. https://doi.org/10.1016/j.landurbplan.2006.02.014

Miller, G., & Lugo, A. E. (2009). Guide to the ecological systems of Puerto Rico (No. IITF-GTR-35; p. IITF-GTR-35). U.S. Department of Agriculture, Forest Service, *International Institute of Tropical Forestry*, 437 p. https://doi.org/10.2737/IITF-GTR-35

Montero, A., Marull, J., Tello, E., Cattaneo, C., Coll, F., Pons, M., Infante-Amate, J., Urrego-Mesa, A., Fernández-Landa, A., & Vargas, M. (2021). The impacts of agricultural and urban land-use changes on plant and bird biodiversity in Costa Rica (1986-2014). *Regional Environmental Change*, *21*(48). https://doi.org/10.1007/s10113-021-01767-1

Mukul, M., Srivastava, V., Jade, S., & Mukul, M. (2017). Uncertainties in the Shuttle Radar Topography Mission (SRTM) Heights: Insights from the Indian Himalaya and Peninsula. *Scientific Reports*, *7*, 41672. https://doi.org/10.1038/srep41672

Murphy, S. F., & Stallard, R. F. (2012). Water quality and landscape processes of four watersheds in eastern Puerto Rico. *U.S. Geological Survey*. https://pubs.usgs.gov/pp/1789/

National Weather Service & National Oceanic and Atmospheric Administration. (n.d.). Rainfall and Temperature Maps for Puerto Rico and U.S. Virgin Islands, 1991-2020 Monthly Normals. *National Weather Service & National Oceanic and Atmospheric Administration*. https://www.weather.gov/sju/climo_pr_usvi_normals

O'Donnell, M. S., & Ignizio, D. A. (2012). Bioclimatic predictors for supporting ecological applications in the conterminous United States (Data Series 691). *U.S. Geological Survey*. https://doi.org/10.3133/ds691

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., … & Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, *355*(6332). https://doi.org/10.1126/science.aai9214

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication*, *351*, 309–317. https://ntrs.nasa.gov/api/citations/19740022614/downloads/19740022614.pdf

Wilson, E.H. & Sader, S.A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment, 80*, 385-396. https://doi.org/10.1016/S0034-4257(01)00318-2

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

8. Appendices

Appendix A: *Imagery Pre-processing*

Table A1
Selected and Masked Landsat QA Bands for Cloud Removal Process in ArcGIS Pro

Pixel Value	Cloud-Free Selected Value	Cloud-Associated or other Masked Value
1		X
21762		X
21824	X	
21890		X
21952	X	
22018		X
22080	X	
22146		X
22208		X
22280		X
23826		X
23888		X
24082		X
24144		X
29986		X
30048		X
54724		X
55052		X

Table A1. Depicts the Quality Assessment (QA) values selected to create a cloud-free image of the study area. Numbers in the Cloud-Free Selected Value column represent pixels that are clear of clouds or other distortion.

Appendix B: *Variables & Indices*

Table B1

Bioclimatic Variables Calculated in this Study

Bioclimatic Variable Name	Bioclimatic Variable Number	Units	Inputs
Annual Mean Temperature	1	Degrees Celsius	Ave. monthly temperature
Max. Temperature of Warmest Month	5	Degrees Celsius	Monthly max. temperatures
Min. Temperature of Coldest Month	6	Degrees Celsius	Monthly min. temperatures
Mean Temperature of Wettest Quarter	8	Degrees Celsius	Ave. monthly temperature, total monthly precipitation
Mean Temperature of Driest Quarter	9	Degrees Celsius	Ave. monthly temperature, total monthly precipitation
Mean Temperature of Warmest Quarter	10	Degrees Celsius	Ave. monthly temperature
Mean Temperature of Coldest Quarter	11	Degrees Celsius	Ave. monthly temperature
Average Daily Precipitation	N/A	Millimeters	Ave. daily precipitation

Table B1. Shows the eight bioclimatic variables calculated for this study. These include seven formal USGS defined variables as well as one team-derived precipitation-related variable.

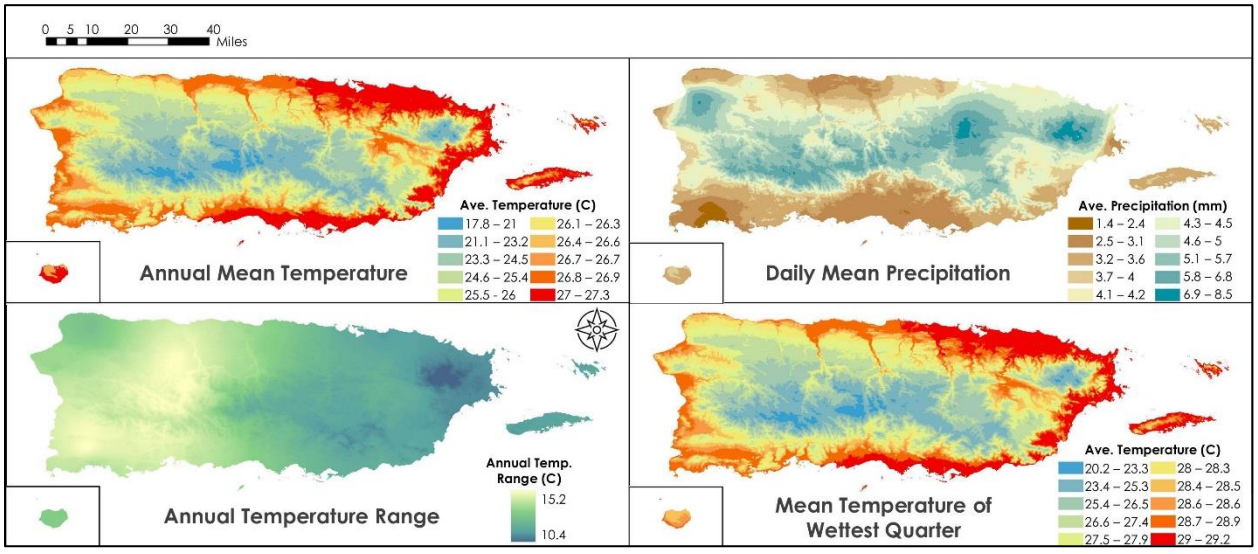

Figure B1. Four Bioclimatic Variable Visualizations this shows four of the eight bioclimatic variables calculated for this project including annual mean temperature, daily mean precipitation, annual temperature range, and mean temperature of the wettest quarter.

Appendix C: *Land Cover Change*

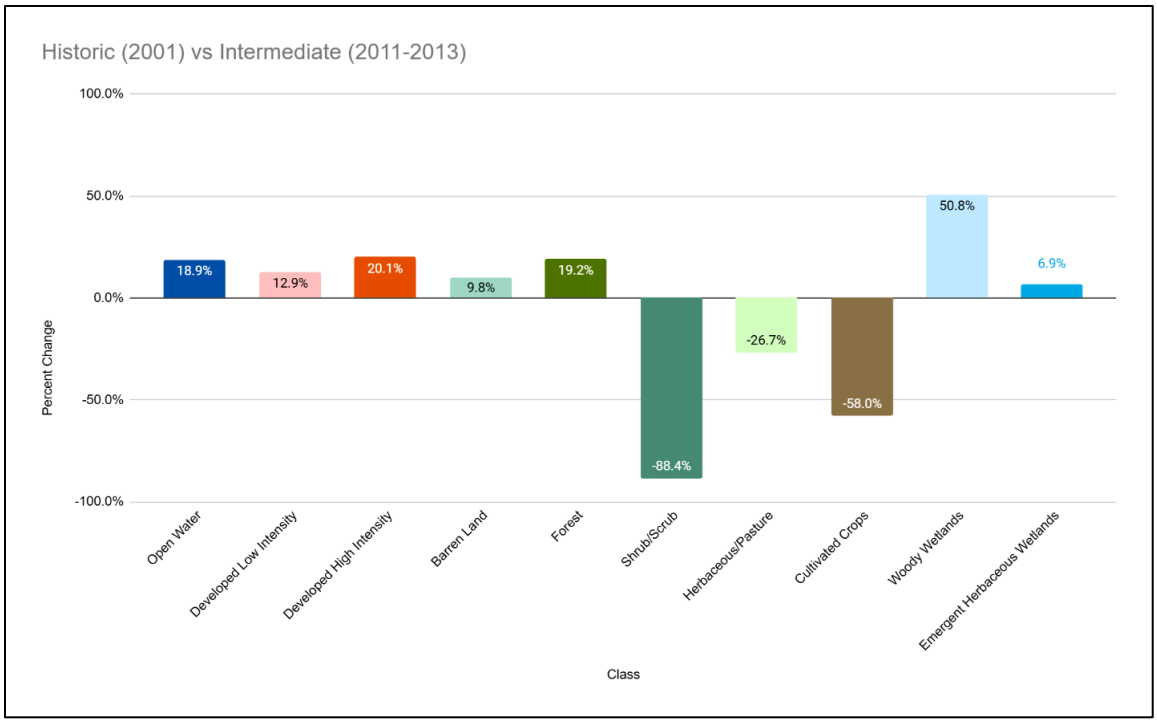

Figure C1. Historic vs. Intermediate Land Cover this depicts the percentage of change for each land cover class between the historic 2001 NLCD land cover map and the intermediate land cover map we created.

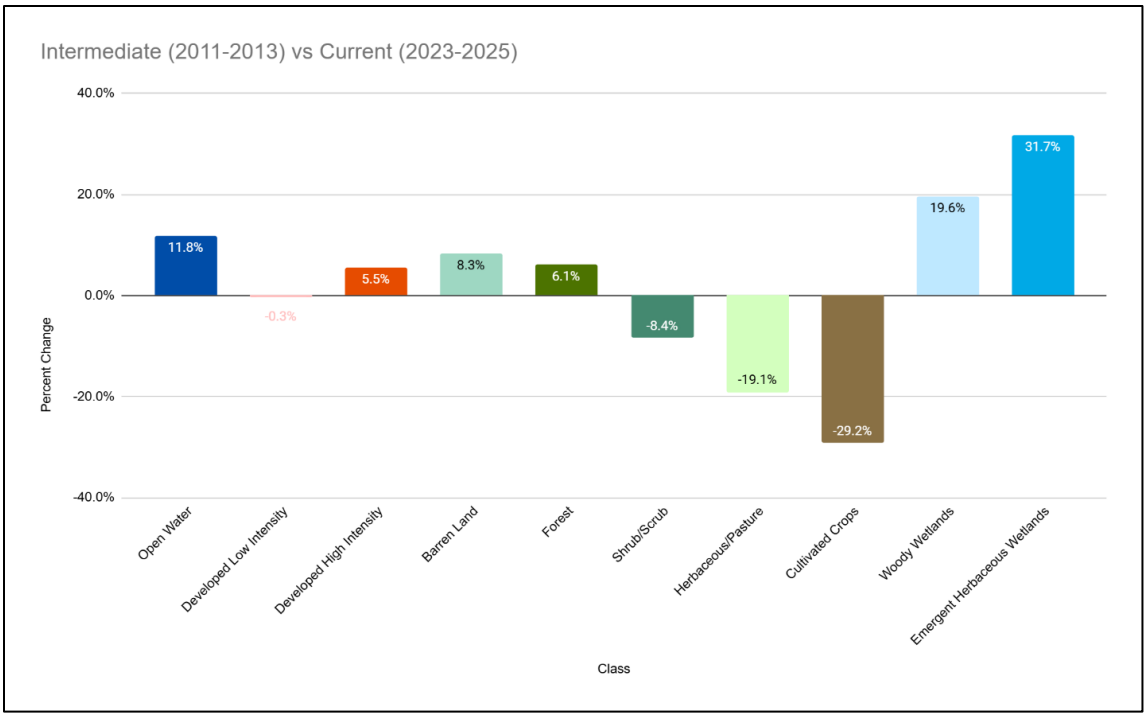

Figure C2. Intermediate vs. Current, or Present-day, Land Cover this depicts the percentage of change for each land cover class between the intermediate and current intermediate land cover maps we created.

Appendix D: *Detailed Land Cover*

Table D1

Parameters and Fields for the Self Organizing Map Algorithm (MiniSOM, Python 3.13.2)

Field	Parameter Used
Input Rasters	Multispectral Imagery, NDVI, annual mean temperature (BIO 1), MNDWI
Number of Principal Components	3
Sigma Value	1
Grid Size	7 x 7
Number of Iterations	5000
Learning Rate	0.1
Training Method	Batch training

Table D2

Conditional Statements used to Determine Classes

Class	Conditional Statement
Lowland Dry Forest	Elevation < 400 m AND Precipitation < 4 mm/day
Lowland Moist Forest	Elevation < 550 m AND Precipitation > 4 mm/day
Submontane Wet Forest	(Elevation > 600 m) OR (Elevation > 550 m AND Precipitation > 5 mm/day)
Lower Montane Rain Forest	Elevation > 700 m AND Precipitation > 7 mm/day
Mangrove/Coastal Wetland	Elevation > 30 m
Beach	Elevation > 8 m

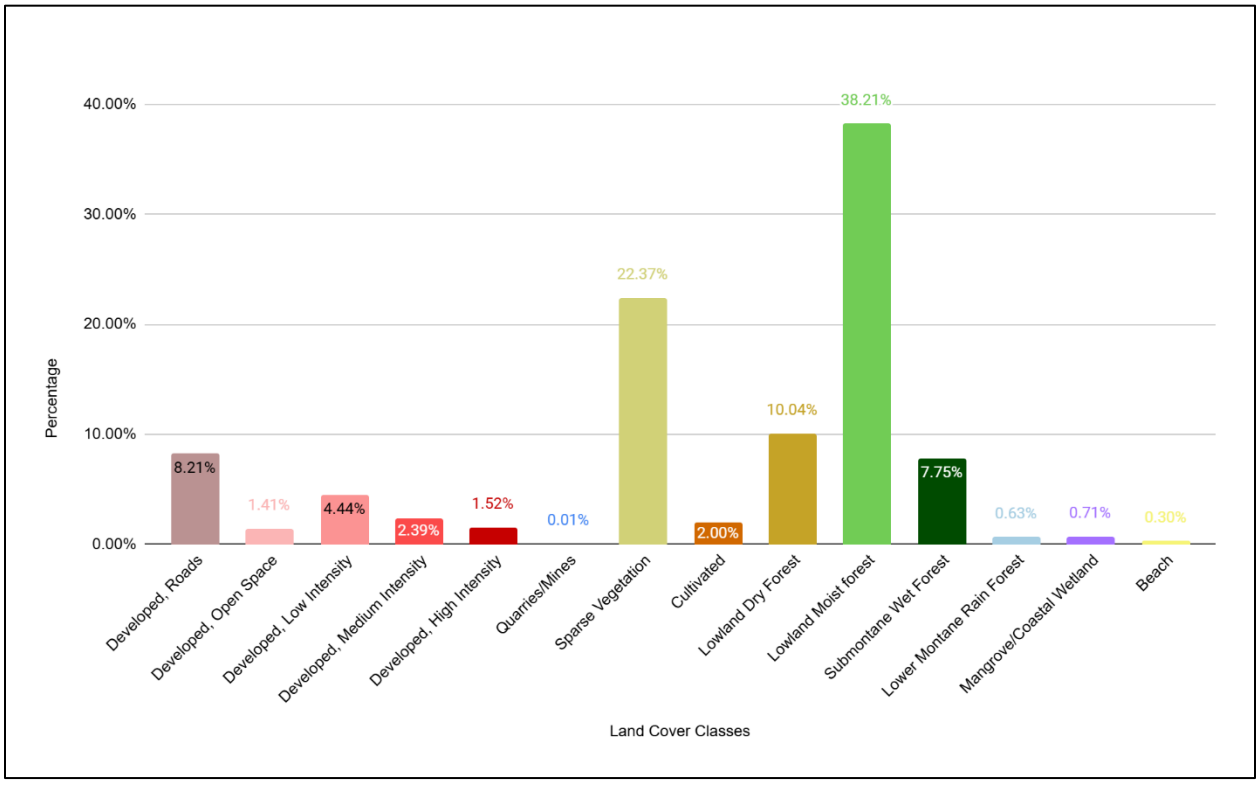

Figure D1. Proportion of each land cover type in the current, detailed map that Shows the proportion of each land cover type in the current, detailed map for 14 of the 17 classes. The remaining classes were related to no data or unclassified pixels.