Southern Bhutan Ecological Forecasting

Modeling Asian Elephant (*Elephas maximus*) Habitat Suitability along the Southern Bhutan Border with NASA Earth Observations

 **Technical Report**

Final Draft – August 6th, 2020

Palchen Wangchuk

Tashi Choden

Kuenley Pem Dem

Sonam Choden

Kelzang Jigme

**Advisors**

Joe Spruce (Science Systems & Applications, Inc., Research Consultant, Diamondhead, MS)

Dr. Kenton Ross (NASA Langley Research Center)

Sean McCartney (Science Systems & Applications, Inc., NASA Goddard Space Flight Center)

# 1. Abstract

Asian elephants (*Elephas maximus*) are a flagship species essential for the functioning of forest ecosystems, and they also have cultural significance in Bhutan. Elephants receive the highest legal protection as listed under Schedule I of the Bhutan Forests and Nature Conservation Act, 1995. Yet, they face threats of extinction due to poaching for ivory as well as the loss and fragmentation of their habitat. Due to the recent clearing of forests and the growing populations in these areas, there has been an increase in incidents of human-elephant conflict. These conflicts have been detrimental to farmer’s annual harvests and livelihoods and have led to retaliatory killing and injury of elephants in southern Bhutan. The DEVELOP Southern Bhutan Ecological Forecasting team partnered with the Bhutan Foundation and Bhutan Tiger Center to help address this problem. The team integrated NASA Earth observations, including Landsat 5 Thematic Mapper (TM), Landsat 8 Operational Land Imager (OLI), Terra Moderate Resolution Imaging Spectroradiometer (MODIS), and the Shuttle Radar Topography Mission (SRTM) to acquire information on land cover change and elephant habitat suitability along the southern border of Bhutan. The team utilized Esri ArcGIS Pro and Software for Assisted Habitat Modeling (SAHM) for data analysis, modeling, and visualization. The team used elephant occurrence data and environmental variables to model current habitat suitability for migrating elephant populations. This analysis provided partners with maps to inform decisions about the placement and conservation of elephant corridors and helped build their capacity to use satellite data for future studies and project planning.

**Key Terms**

remote sensing, habitat modeling, land cover change, elephant corridors, habitat loss, ArcGIS Pro, SAHM

# 2. Introduction

***2.1 Background Information***

The Asian elephant (*Elephas maximus*) has been listed as an endangered species by the International Union for Conservation of Nature Red List (IUCN, 2017) and included in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (Bisht, 2002). These elephants are usually found in a variety of habitats, including grasslands, secondary forests, scrublands, cultivated areas, tropical evergreen forests, moist deciduous forests, dry deciduous forests, and dry thorn forests (Sukumar, 2003). Although there are a variety of threats to their existence, the greatest is the increasing number of human-elephant conflicts (HECs) (Nature Conservation Division, 2018). In Asia, elephant habitat loss and degradation has occurred over extensive areas, which has led to increased conflicts with humans. In five countries (Bangladesh, Bhutan, China, Nepal, and Vietnam) of the 13 range states for the Asian elephant, the number of wild elephants was less than 200 in 2011 (Fernando and Pastorini, 2011).

In Bhutan, Asian elephants are a flagship species that are essential for the functioning of forest ecosystems. They also hold strong cultural significance. The elephants in Bhutan receive the highest legal protection as listed under Schedule I of the 1995 Forests and Nature Conservation Act. However, they still face threats of extinction due to the illegal poaching for ivory sales and the loss and fragmentation of their habitat. Recent clearings of forests, creating close proximity of human settlements to forested areas, has led to an increased number of HECs.

Currently, there are strategic efforts in southern Bhutan to help mitigate human retaliation to human-elephant conflicts by the Department of Forests and Park Services. These include the installation of solar electric fences, lighting fires/fire crackers, creating loud sounds, and physically chasing away the elephants together with support from locals (Bhutan Nature Conservation Division, 2018). Although many methods of harmless retaliation exist, the best method is identifying elephant movement corridors and potential elephant refuge areas that won't interfere with growing urban sprawl.

The area that we focused on was the southern Bhutan border, particularly the Gelephu area, and the study period we looked at was from 1999 to 2019. Gelephu serves as a trade post between Bhutan and India because of its location. The development of the local airport in Gelephu and the increase in urban settlements has brought humans ever closer to Asian elephant habitat, and this has led to an increase in HECs, which has caused unrest among the local population (Bhutan Nature Conservation Division, 2018). The study area is surrounded by the Royal Manas National Park, Jigme Singye Wangchuck National Park, and Phipsoo Wildlife Sanctuary. The area is a warm, fertile region with plenty of rainfall. Gelephu boasts a sub-tropical climate, having warm and dry winters and wet and dry summers. The winter months usually experience the colder temperatures (averaging between 17-22 degrees Celsius), and the summers are usually much warmer (averaging around 28+ degrees Celsius) (Meteobox, 2020).

For this DEVELOP project, we used NASA Earth observations and elephant sighting data to create an elephant habitat suitability model. We acquired specific variables that affect elephant occurrence in and around the study area. The variables were chosen through multiple conversations with the project partners and through a literature review. More information about these variables can be found in the in the Methodology section, below. We constructed Land Use Land Cover (LULC) change maps that gave us better insight into the geographical changes and developments in the Gelephu area in southern Bhutan. This allowed us to visualize the data to help supply proper information to the partners at the Bhutan Foundation and Bhutan Tiger Center and also supplemented information on LULC change provided by Yangchen et al. (2015). We referred to previous research regarding forest cover loss and fragmentation in the area to better understand how to analyze the variable-specific data on elephant habitat and habitat use (Padalia et al., 2019). We utilized the Software for Assisted Habitat Modeling (SAHM) method for incorporating the elephant occurrence data along with the variable-specific data that we acquired to compute an elephant habitat suitability model using methods from Morisette et al. (2013). Conducting a process similar to Sharma et al. (2019), we analyzed our model outputs using the Boosted Regression Tree (BRT) to examine the relationship between habitat suitability and variables like elevation, slope, precipitation, land surface temperature, distance to roads, distance to water sources and population density.



*Figure 1.* Study Area Map of Southern Bhutan Border.

***2.2 Project Partners & Objectives***

The DEVELOP Southern Bhutan Ecological Forecasting team partnered with the Bhutan Foundation and the Bhutan Tiger Center to address this issue of increasing human-elephant conflicts in Southern Bhutan. The Bhutan Foundation is a non-profit organization that specializes in funding and supporting different projects in, or pertaining to, Bhutan. The Bhutan Tiger Center is a center for research, education, outreach, and policy under the direction of the Ministry of Agriculture and Forests in Bhutan, Department of Forests and Park Services. The elephant research is also pertinent to the Tiger Center because it manages and are in charge of the collection of elephant data and informing the public regarding this subject. Our methodology and processes can be used as a reference for future surveys and projects that they will conduct on their own.

The objectives of the project were to use NASA Earth observation data and other resources to investigate land cover change and elephant habitat suitability along the southern Bhutan border. Then in ArcGIS Pro, we compiled LULC classification maps to serve as a reference for historical land use trends from 1999 to 2019. Finally, we utilized the Software for Assisted Habitat Modeling (SAHM) program to create elephant habitat suitability models. The goal of our analysis was to provide the partners with decision-supporting information about placement and conservation of elephant corridors with further engagement to integrate NASA Earth observations for future studies and projects.

# 3. Methodology

***3.1 Data Acquisition***

Through literature review and discussion with NASA DEVELOP Science Advisors, the Goddard Space Flight Center Fellow, the Bhutan Foundation, and the Bhutan Tiger Center, the team considered nine variables to better understand elephant habitat suitability, which are listed in Table 1. The variables were chosen after multiple partner discussions and extensive literature review. Variable datasets and elephant occurrence datasets were collected for Southern Bhutan and the adjoining parts of North-Eastern India to increase the variety in training data for our models. The elephant occurrence data were found using the Global Biodiversity Information Facility (GBIF) data portal, providing open access to crowd source data about all types of life on Earth. Additional elephant occurrence data were generously supplied to us by the Bhutan Foundation and the Bhutan Tiger Center. The camera trap data we received were a byproduct of a different tiger research project the partners had ongoing and the additional elephant data they shared with us were collected via elephant radio collar. Both datasets were very helpful for our research.

Table 1.

*List of variables and their information.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Acquisition Method** | **NASA Earth Observations** | **Dates** |
| Reflectance data for Land Cover Land Use | Google Earth Engine (GEE) | Landsat 5 & Landsat 8 | 1999 & 2019 |
| Normalized Difference Vegetation Index (NDVI) Phenology | GEE | Landsat 8 | 2019 Spring, Summer, Fall, Winter |
| Distance to Roads & Urban Settlement | Socioeconomic Data and Application Center (SEDAC) | n/a | 2010 |
| Water Sources/  Distance to Water | HydroRIVERS | Shuttle Radar Topography Mission (SRTM) | 2000 |
| Elevation | DIVA-GIS | SRTM | 2020 |
| Slope | DIVA-GIS | SRTM | 2020 |
| Population Density | SEDAC | n/a | 2020 |
| Land Surface Temperature (LST) | GEE | Moderate Resolution Imaging Spectroradiometer  (MODIS) | 2019 |
| Precipitation | GEE; Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS) | n/a | 2019 |

The team formulated codes in Google Earth Engine (GEE) to obtain cloud-free Landsat data, for each season of the years 1999 and 2019, in order to compute vegetation greenness phenology data (using seasonal NDVI) for areas with Asian elephants. This method was then replicated, for a specific time period (e.g., annually for 1999 and 2019), to also download Landsat 5 and 8 scenes to produce LULC Classification Maps. The paths and rows for the Landsat scenes are given in Table 2. We also scripted codes in GEE to download Annual Average Land Surface Temperature (LST) and Annual Average Precipitation data products. For the LST code in GEE, we imported the Terra MODIS V6 LST product for image collection since it provided daily land surface temperature and emissivity values in a 1200 x 1200 km grid and used the ‘LST\_Day\_1km’ band to extract daytime land surface temperature. For precipitation, we imported Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS), which is a 30+ year quasi-global rainfall dataset. CHIRPS incorporates 0.05° spatial resolution satellite imagery with *in-situ* station data to create gridded rainfall time series.

Table 2.

*List of paths and rows for collected Landsat Scenes*

|  |  |
| --- | --- |
| **1999 & 2019** | |
| **Paths** | **Rows** |
| 136 | 41 |
| 137 | 41 |
| 138 | 41 |
| 136 | 42 |
| 137 | 42 |
| 138 | 42 |

Population density and Distance to Roads datasets were acquired from a Socioeconomic Data and Application Center (SEDAC) website. The population density dataset was structured for every five years from 2000 to 2020. River Networks data for the whole of Asia was downloaded from the HydroRIVERS website and the elevation data was downloaded from a DIVA-GIS website.

***3.2 Data Processing***

The team primarily used Esri ArcGIS Pro 2.6.0 for most of the data processing. Each of the datasets were imported into ArcGIS Pro in the form of GeoTIFFs and shapefiles. We mosaicked the Landsat scenes, and performed an unsupervised classification to produce LULC classification maps. We grouped the classes and reclassified them into six categories: Mature Forest, Immature Forest/Scrub/Herbaceous, Cultivated Land, Barren Land, Rivers and Snow/Ice which are the most significant in the context of elephant habitat. The 1999 and 2019 maps were recoded so that the classes are valued 10, 20, 30, 40, 50, 60 (1999) and 1, 2, 3, 4, 5, 6 (2019), then these added together using the Raster Calculator to generate a change map shown in *Appendix A*. We derived the slope data in ArcGIS Pro from the initial Elevation data we acquired from DIVA-GIS. The NDVI phenology data were then imported into ArcGIS Pro where we played with the color schemes, green represented most vegetation whereas red represented least vegetation.

We clipped the roads and rivers data to the study area polygon and calculated the Euclidean distance while projecting it to Indian UTM Zone 46N. Applying the Euclidean distance converted the vector data into raster data. We also clipped all the other variables to the polygon and resampled the cell size to 30 m, in order to maintain uniformity for inputing them as predictors to train the models in the Software for Assisted Habitat Modeling (SAHM).

***3.3 Data Analysis***

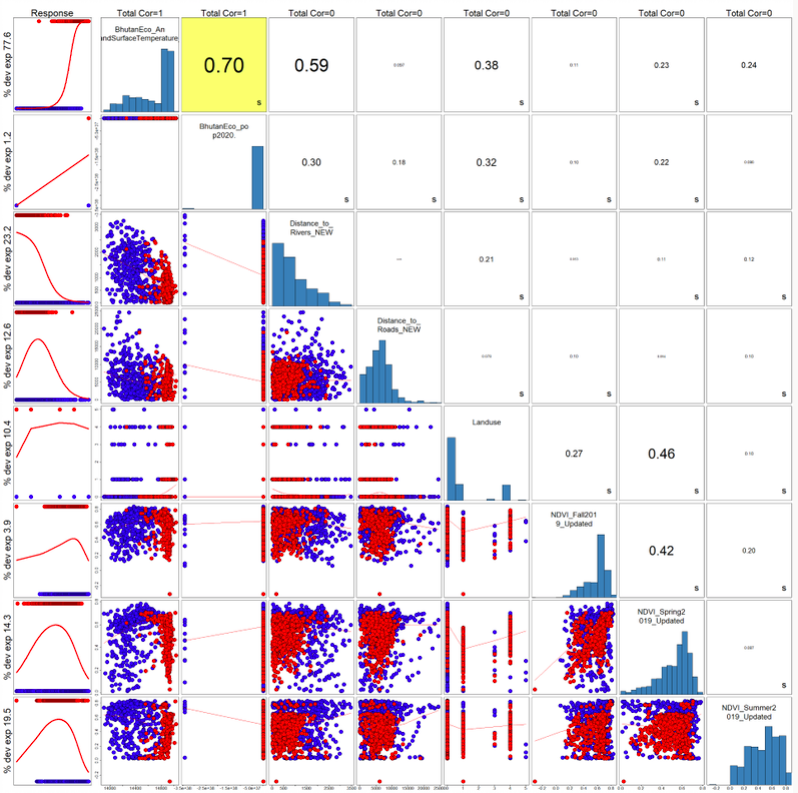
Using the VisTrails SAHM version 2.0.1, the team generated four models for studying elephant habitat suitability. We imported all nine variables as predictors and generated a template layer from the NDVI Summer raster. To create the template layer, all values of the raster were multiplied by 0 and then set to 1 in ArcGIS’ ‘Raster Calculator’ tool. Through SAHM, all other layers are resampled and aligned to the template layer projection. Since the elephant occurrence data were highly clustered, we had more presence than absence data. Some points existed within the same 30 m pixel, so a 50 m buffer was applied between each training point to prevent overfitting. Additionally, we balanced the presence/absence ratio by randomizing the selection of presence points and thereby excluding a large number of the presence points. These presence/absence data were used as the field data in building the SAHM models. While all the variables were used once as predictors, we used the four different seasons of NDVI, one NDVI for each season of the year, to capture seasonal variation.

The ‘Covariate Correlation and Selection’ tool in SAHM gives an output chart that shows each predictor and the covariate correlation between each of them. This enabled us to eliminate highly correlated variables. For example, if two variables have a covariate correlation value greater than 0.7, we eliminated the variable with less deviance explained and kept the other to run the models. If the highly correlated variables are not filtered, it can lead to unstable model fits (Morisette et al., 2013). After removing these variables, we ran the selected variables through four different modeling techniques (Boosted Regression Tree (BRT), Generalized Linear Model (GLM), Multivariate Adaptive Regression Splines (MARS) and Random Forest (RF)), which were run simultaneously in the same SAHM workflow in Vistrails. Each of the models produced 1) an elephant occurrence probability map, 2) evaluation statistics, including model prediction sensitivity and Area Under the ROC Curve (AUC), and 3) confusion matrice, which can be used to compare performance across all the models.

4. Results & Discussion

***4.1 Analysis of Results***

The ‘Covariate Correlation and Selecton’ tool showed that Land Surface Temperature (LST), Elevation, Slope and Precipitation were highly correlated. Out of these four predictors, we eliminated Elevation, Slope and Precipitation from the model runs and only kept LST because it had the highest deviance explained. This means that, for the given field data, LST captured more of the training data variation than the other three correlated predictors. After this down-selection process, the remaining predictors (LST, Distance to Rivers, Distance to Roads, Population Density, NDVI for each season, and Land Cover) were used to compute the four models.



*Figure 2.* ‘Covariate Correlation and Selection’ Output; showing only the variables selected for modeling. Percent deviance explained can be found on the left of the response curves for each predictor.

To choose the most accurate model, we looked at the evaluation graphs (Figure 3). The model fits well if the sensitivity and specificity curves for Receiver Operating Characteristic (ROC) plot for cross validation are well above the diagonal line which represents random (Figure 3 below; Morisette et al., 2013). All of our evaluation graphs had that criterion fulfilled. So, we looked at the Area Under the ROC Curve (AUC) and the threshold values to choose the apparent best performing model. The AUC and the threshold describe the model’s ability to discriminate between the presence (True Positive) and absence (False Positive) points. AUC gives an aggregate of the performance of a model across all possible thresholds. AUC ranges in value from 0-1 and the greater the AUC value the better the model performance (Toshniwal, 2020). The threshold also ranges in value from 0-1. For a high AUC of 0.9-1, a threshold of 0.5 is considered to be the most ideal measure of separability (Narkhede, 2018). This statistical metric can also be used to distinguish between True Positive and True Negative.

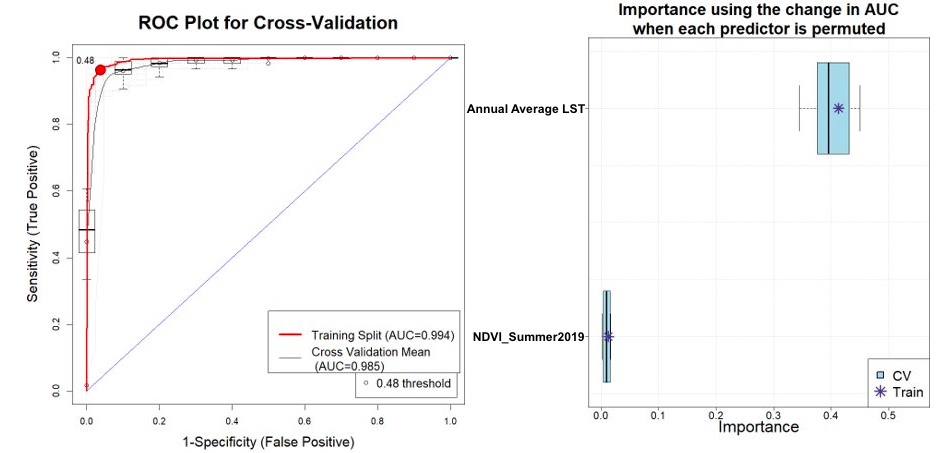
BRT had the most preferable AUC and threshold values of 0.994 and 0.48 respectively, making it the better performing model of the four we ran. The GLM, MARS, and RF also had high AUC values between 0.9-1 (Appendix A). However, the latter 3 models had thresholds of 0.89, 0.62, and 0.54 respectively (Table 3) which are not as good as BRT thresholds for classifying absence and presence points since they are far from the ideal threshold.

Table 3.

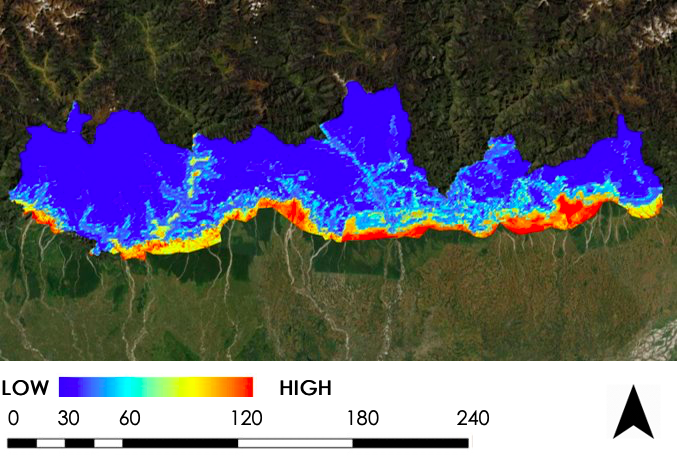
*AUC and threshold values for the Boosted Regression Tree, Generalized Linear Model, Multivariate Adaptive Regression Splines, and Random Forrest models*

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **Training Split AUC** | **Cross Validation AUC** | **Threshold** |
| BRT | 0.994 | 0.993 | 0.48 |
| GLM | 0.994 | 0.993 | 0.89 |
| MARS | 0.99 | 0.985 | 0.62 |
| RF | 0.988 | 0.987 | 0.54 |

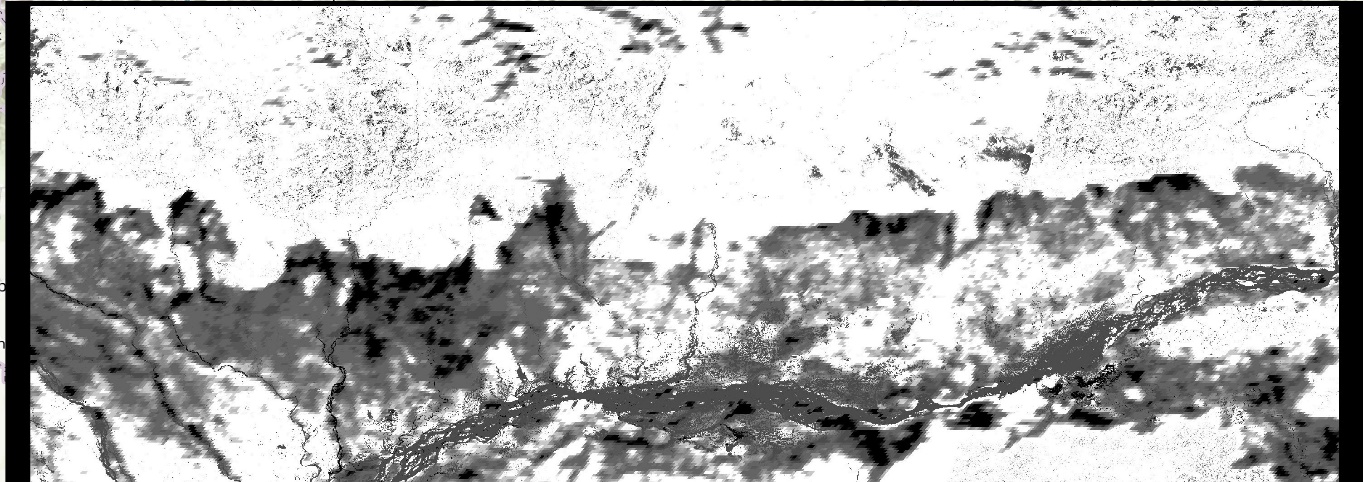
Upon further analysis, we determined the most significant predictor in each model by looking at the SAHM generated graph titled “Importance using the change in AUC when each predictor is permuted.” In all four models, LST is a significantly important predictive factor in that most of the model’s prediction was based on this one variable. This importance is likely related to the high correlation LST showed with elevation and slope, leading to a situation in which the LST variable is also indicative of the mentioned terrain variables. This makes sense that temperatures in the study area for a given time of year tend to decrease with increasing elevation.

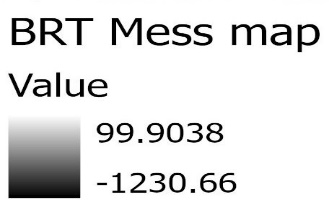
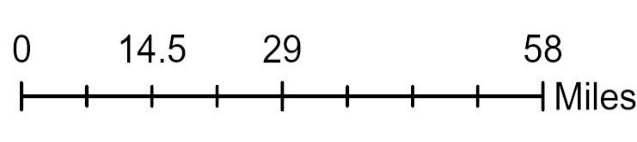


*Figure 3.* ROC Plot for Cross-Validation and Predictor Importance using the change in AUC when each predictor is permuted for BRT

The following map (*Figure 4*) is the probability of elephant occurrence map produced by the BRT model. The areas in blue represent the lowest probability of elephant occurrence and red represents the highest. There are areas of medium to medium-high probability of occurrence (light blue to yellow) the mountains, mostly concentrated along water sources, as well as some variation along the southern border of Bhutan. The border area contains the highest probability of occurrence overall. *Figure 4.* BRT probability map for elephant habitat suitability.

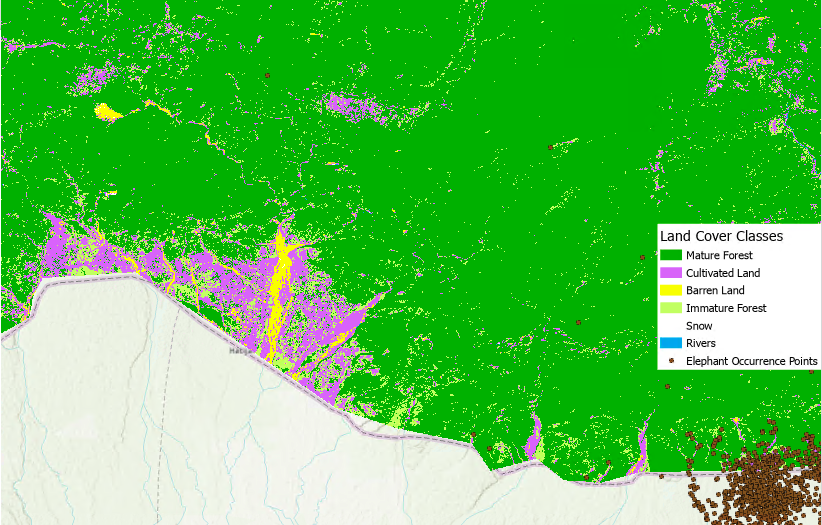
SAHM created Multivariate Environmental Similarity Surface (MESS) map to evaluate the interpolation and extrapolation of the model (Elith et al., 2010). Negative values in the MESS map indicate the areas where predictor values of at least one predictor were outside of the range observed in the training data. For example, if elevation values in the training data range from 100-600 m, predicted areas that have an elevation of 700 m would have a negative MESS value. Similarly, MESS values of zero indicate that locations aren't outside the range of values but are on the edge. This map will help the partners identify areas where adding camera traps would provide valuable data for model refinement.





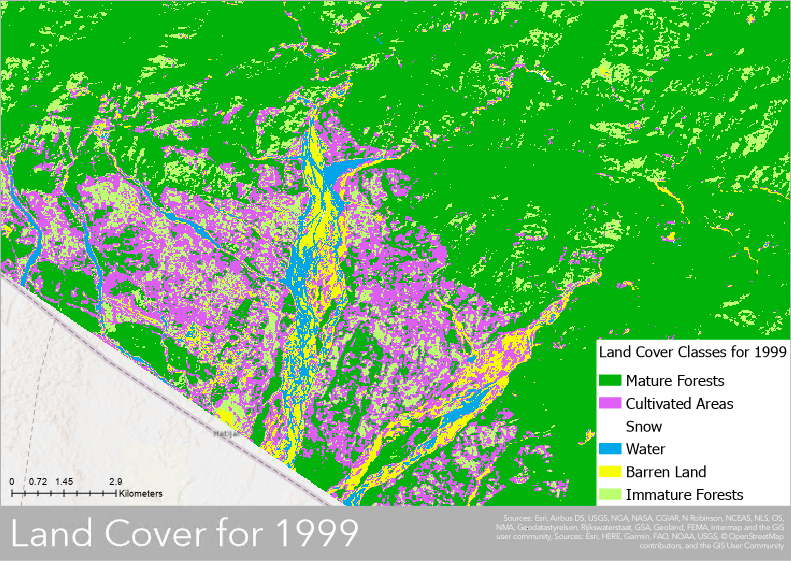
*Figure* *5.* SAHM-derived Multivariate Environmental Similarity Surface map

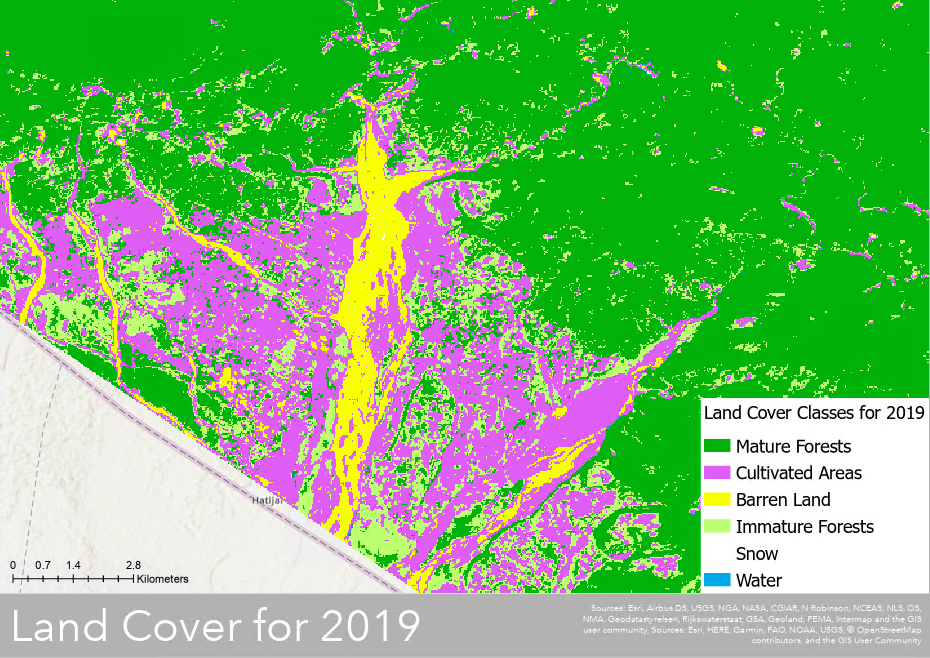
Figures 6 and 7 pertain to the project’s Land Cover maps zoomed in to show the Gelephu region. The Land Cover maps for 1999 and 2019 for the whole study area are given in Appendix B. The elephant occurrence data have been plotted over *Figure 6*. The elephants tend to occur more in the cultivated areas or human settlements, mature and immature forests, and near water bodies. Some uncertainty may have been introduced from sampling bias, as most of the presence/absence data we used were collected from that patch of area. If we had data from other locations, the result might have varied. Another observation is that the Gelephu region has large areas of cultivated land aside from the forests. Cultivated areas also include human settlements and roads. It makes sense to see large areas of purple (cultivated areas) because of the recent developments that have been taking place in Gelephu.



*Figure* *6*. Land Use and Land Cover Classification Map zoomed in onto the Gelephu Region with the elephant occurrence data points for 2019 overlaid.

*Figure 7* shows the comparison between the Land Use and Land Cover classification maps for the years 1999 and 2019. Most of Southern Bhutan consisted of mature forests and immature forests with few barren lands and cultivated areas in 1999. However, with development, the southern region changed into containing more cultivated areas as compared to mature and immature forests. As shown in the 2019 map below, mature forests have formed into immature forests due to settlements. On the other hand, immature forest further away from settlements have changed into mature forest likely because those areas have not been disturbed.





*Figure 7.* Comparisons of Land Use and Land Cover Classification for the Gelephu region for the years 1999 (top) and 2019 (bottom).

Some of the errors and uncertainties in the project are discussed here. The Roads data were obtained from SEDAC and SEDAC collected the data from VMAP0 which is a crowd sourced data set. The year for the Roads data was also not specified in the metadata which means it is not necessarily for the year 2019. Additionally, some parts of the elephant occurrence data were obtained from Global Biodiversity Information Facility (GBIF) which is crowd sourced data. We gathered data from GBIF and combined it with the data sent to us by the Bhutan Tiger Center and the Bhutan Foundation. The elephant occurrence data in Bhutan was not entirely collected from the elephant camera traps. Some of the data obtained from the Bhutan Tiger Center was a byproduct of tiger research surveys and not specifically designed for surveying elephant occurrence. So, this may have caused some sampling bias of elephant occurrence data.

The *in-situ* elephant survey data that we used spans from the year 2014-2019. This could have also affected our model given this disparity in the data collection study period. Also, the high amount of clustering within presence points probably led to overfitting of the models. Field data that captures more of the variation of elephant presence/absence across the landscape would help refine models and provide further insight into suitable habitat beyond the lower elevation areas identified by these preliminary maps. As mentioned previously, the land surface temperature was highly correlated to other variables (e.g., temperature and slope) in the tested models. More model evaluation and model input product validation are also needed, including accuracy assessment of the land cover maps generated for the project.

***4.2 Future Work***

A follow-on project is recommended to build upon the results, products, and capabilities resulting from our initial study. Such work could consider additional relevant *in-situ* data to help increase the accuracy of the elephant suitability models by reducing sampling biases to a minimum. More work is needed to further develop certain modeling inputs and outputs. Future participants could perform more extensive analyses in regards to LULC maps and LULC change trends. We only compiled LULC maps for our project at two timesteps, one at the beginning (1999) and one at the end (2019) of our 20-year study period. To gather more information on historical land use trends, future teams could also collect and compute LULC maps for intervals within the 20-year period and also for broader periods than 20 years. The LULC data we used to run the habitat suitability model pertains to the year 2019. To improve the accuracy of the project results, LULC datasets could be gathered from multiple years to compute results on SAHM and then compared to assess trends and to perhaps forecast future change, not only with respect to LULC, but also in terms of elephant habitat suitability. Regarding SAHM modeling, more habitat suitability predictor variables could also be tested such as moisture indices, tree cover layer, and elephant disturbance to vegetation in known elephant occurrence zones. The future term could also look at the more spatially resolute Landsat LST data, broken up into seasonal layers and taking into account both day time and night time LST instead of the Annual Average MODIS LST used here. Landsat LST could possibly outperform MODIS LST (Parastidis et al., 2017). Landsat data could be used to further assess if LST is the most significant driver of the models or how using MODIS LST compared to Landsat LST affected our results. Improved models, methods, and data products can be provided to project end users, leveraging for the established workflow from this initial study. With the foundations we have laid down, a future project will be able to continue building upon our foundational work and improve upon our processes and end product accuracy in their results.

# 5. Conclusions

In conclusion from our 10-week project, results suggest that viable models for elephant habitat suitability have been compiied using a combination of NASA Earth observations and *in-situ* data on elephant occurrence. We found that Land Surface Temperature (LST), Slope, and Elevation, highly correlated predictors in the SAHM models, are drivers of elephant habitat suitability. LST was shown to be a strong driver for all of four tested models within the SAHM framework. A study by Parastidis et al.(2017) found that Landsat LST accuracy is heavily influenced by land cover specific emissivity. Our study used coarser resolution MODIS LST products at ~1km resolution that were generated using a different method than that used in the mentioned Landsat LST study. Something to consider for future replications of our work is to try using Landsat LST instead of MODIS LST**.**

Another key component that could have affected these predictors is the disparity between absence and occurrence points in our elephant occurrence data set, resulting in alterations to limit any sampling bias in our research. From the four models we generated using SAHM, the Boosted Regression Tree model yielded the strongest results in habitat suitability for our project, however, this was generated using only two predictor variables. The four models generated four habitat suitability maps that suggest and display where future research can be focused to expand understanding of Asian elephant habitats in Bhutan. The results indicated the importance of roads, waterways, forest cover, and land surface temperature in regards to elephant occurrence. Roads also indicate connections to cultivated areas and settlements where elephant occurrences have been found. There was also more frequent elephant occurrence occurring along the southern border with India, which may indicate that some elephants may be coming into Bhutan from India.

From the landcover change map that was calculated, there was a significant change in cultivated areas due to an increase in human settlement. Mature forests surrounding cultivated areas have also been harvested for logging or agricultural purposes and then reverted into immature forests. In general, forests were frequent for both dates of LULC maps, especially for more rural areas that are adjacent to where human settlements are most prevalent.

The project yielded methods, maps, data products, and documentation that will provide the Bhutan Foundation and Bhutan Tiger Center information to use in reducing Human Elephant Conflicts and aiding wildlife management. To continue building upon our research, a follow-up project is recommended to strengthen the accuracy of the currently obtained results, products, and capabilities resulting from our initial study. In the future, more *in-situ* data should be gathered to avoid any sampling bias in the training data for the models. We created Land Use and Land Change maps for 1999 and 2019 only, so another recommendation would be to create more maps from different years to increase the accuracy of habitat suitability results. After building upon our project parameters, future research should also take into consideration additional environmental variables (e.g.., tree canopy cover and elephant disturbance zones) to use as predictors for the habitat suitability models. We hope that our work will serve as a strong foundation for future projects and will improve decision-making capabilities of the Bhutan Foundation and Bhutan Tiger Center.

# 6. Acknowledgments

The Southern Bhutan Ecological Forecasting team would like to thank our science advisors, mentors, and our partners at the Bhutan Foundation and Bhutan Tiger Center for their support and guidance:

* Joe Spruce, SSAI Research Consultant, Diamondhead, MS, Science Advisor
* Dr. Kenton Ross, NASA Langley Research Center, Science Advisor
* Sean McCartney, SSAI, NASA Goddard Space Flight Center
* Tshewang Wangchuk, Bhutan Foundation, Director
* Dr. Tshering Tempa, Bhutan Tiger Center
* Darcy Gray, DEVELOP Fellow at Goddard Space Flight Center
* Peder Engelstad, Natural Resource Ecology Laboratory, Colorado State University
* Kristen Dennis, CO Lead, Geoinformatics Fellow

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors 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 NNL16AA05C.

# 

# 7. Glossary

**AUC** - Area Under the ROC Curve

**BRT** - Boosted Regression Tree

**CHIRPS** - Climate Hazards Group InfraRed Precipitation with Station data - Incorporates 0.05o satellite resolution with *in-situ* data to create rainfall time series

**EO** - Earth observations; Satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time

**GBIF** - Global Biodiversity Information Facility

**GEE** - Google Earth Engine

**GLM** - Generalized Linear Model

**HEC** - Human Elephant Conflict

**IUCN** - International Union for Conservation of Nature

**LULC** - Land Use Land Cover

**LST** - Land Surface Temperature

**MARS** - Multivariate Adaptive Regression Splines

**MODIS** - MODerate resolution Imaging Spectroradiometer - NASA imaging sensor gathering 36 spectral bands of entire Earth

**NDVI** - Normalized Difference Vegetation Index - An index of land cover measuring vegetation greenness

**RF** - Random Forest

**ROC** - Receiver Operating Characteristic

**SAHM** - Software for Assisted Habitat Modeling

**SEDAC** - Socioeconomic Data and Applications Center

**SRTM** - Shuttle Radar Topography Mission

# 8. References

Bisht, S. S. (2002). *An overview of Elephant Conservation in India*. Indian Forester. http://www.indianforester.co.in/index.php/indianforester/article/view/2469/0.

Center for International Earth Science Information Network - CIESIN - Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia. (2013). Global Roads Open Access Data Set, Version 1 (gROADSv1). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved July 10, 2020, from https://doi.org/10.7927/H4VD6WCT.

Center for International Earth Science Information Network - CIESIN - Columbia University. (2018). Gridded population of the world, version 4 (GPWv4): Population count, revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved June 20, 2020, from <https://doi.org/10.7927/H4JW8BX5>

CENTER. (2013). USGS Landsat 8 Surface Reflectance Tier 1. Google Earth Engine, accessed July 2020. doi://10.5066/F71835S6

# Data Basin. (2006). World Wildlife Fund (WWF) HydroSHEDS (BAS) - Asia HydroRIVERS Version 1.0 at 30s resolution [Data File]. Retrieved from <https://databasin.org/datasets/b38b6e6dfbea4e39b92a31e43bf0344f>

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution, 1*, 330–342. doi: 10.1111/j.2041-210X.2010.00036.x

EROS CENTER. (2013). USGS Landsat 8 Top of Atmosphere Tier 1. Google Earth Engine, accessed July 2020. doi://10.5066/F71835S6

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland J., Harrison, L., Hoell, A., & Michaelsen, J.. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data, 2*, 150066. doi:10.1038/sdata.2015.66 2015.

IUCN (2017). The IUCN Red List of Threatened Species. Version 2017-1. Retrieved from htps://www.iucnredlist.org

Jensen, J. R. (2014). *Remote sensing of the environment: an earth resource perspective*. Pearson.

Meteobox. (2020). *Weather statistics for Gelephu, Bhutan.* Retrieved from https://meteobox.com/bhutan/gelephu/statistics/

Morisette, J. T., Jarnevich, C. S., Holcombe, T. R., Talbert, C. B., Ignizio, D., Talbert, M. K., … Young, N. E. (2013, January 25). *VisTrails SAHM: visualization and workflow management for species habitat modeling*. Wiley Online Library. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0587.2012.07815.x>.

NASA Jet Propulsion Laboratory (JPL). (2013-2017). *NASA Shuttle Radar Topography Mission Global 1 arc second* [Data set]. doi:10.5067/MEaSUREs/SRTM/SRTMGL1.003

Narkhede, S. (2018, June). Understanding AUC – ROC Curve. *Towards data science.* Retrieved from https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Nature Conservation Division. (2018). *Elephant conservation action plan for Bhutan 2018-2028.* Nature Conservation Division, Department of Forests & Park Services, Ministry of Agriculture & Forests, Thimphu, Bhutan.

Padalia, H., Ghosh, S., Reddy, C.S., Nandy, S., Singh, S., & Kumar, A. S. (2019) Assessment of historical forest cover loss and fragmentation in Asian elephant ranges in India. *Environmental Monitoring and Assessment*, *191*, 802. <https://doi.org/10.1007/s10661-019-7696-5>

Parastatidis, D., Mitraka, Z., Chrysoulakis, N., & Abrams, M. (2017, November 23). *Online Global Land Surface Temperature Estimation from Landsat*. MDPI. https://www.mdpi.com/2072-4292/9/12/1208.

Running, S., Mu, Q., Zhao, M. (2017). MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1Km SIN Grid V006. NASA EODIS LP DAAC. DOI: 10.5067/MODIS/MOD11A1.006

Sharma, P., Adhikari, H., Tripathi,S., Ram, A.K., & Bhatarai, R. (2019). Habitat suitability modeling of Asian Elephant Elephas maximus (Mammalia: Proboscidea: Elephantidae) in Parsa National Park, Nepal and its buffer zone*. Journal of Threatened Taxa*, *11*(13).14643–14654. <https://helda.helsinki.fi/bitstream/handle/10138/310445/4467_Full_article_text_with_author_details_26032_1_10_20191025.pdf?sequence=1&isAllowed=y>

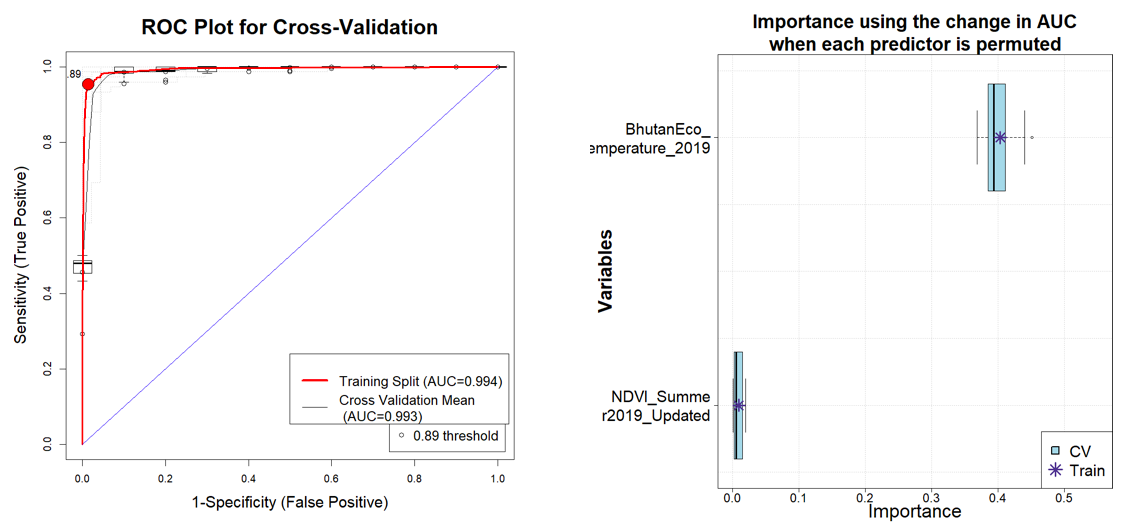
Sukumar, R. (2003). *The living elephants: Evolutionary ecology, behavior, and conservation*. Oxford University Press, New York, 478pp.

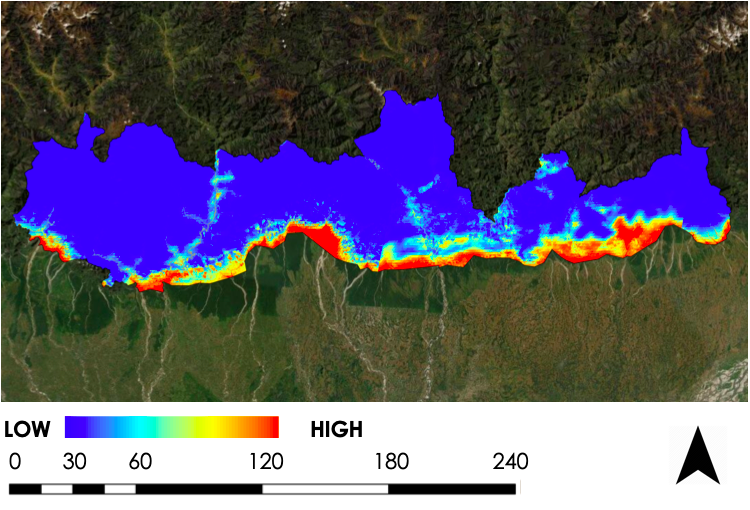
Toshniwal, R. (2020, January). Demystifying ROC Curves. *Towards data science.* Retrieved from https://towardsdatascience.com/demystifying-roc-curves-df809474529a

Yangchen, U., Thinley. U., & Wallentin, G. (2015). *Land use land cover changes in Bhutan 2000-2013.* College of Natural Resources, Royal University of Bhutan, Lobesa, Bhutan. Retrieved from <https://www.researchgate.net/profile/Gudrun_Wallentin/publication/288181926_Land_Use_Land_Cover_Changes_in_Bhutan_2000-2013/links/567ed87108ae051f9ae66aea.pdf>

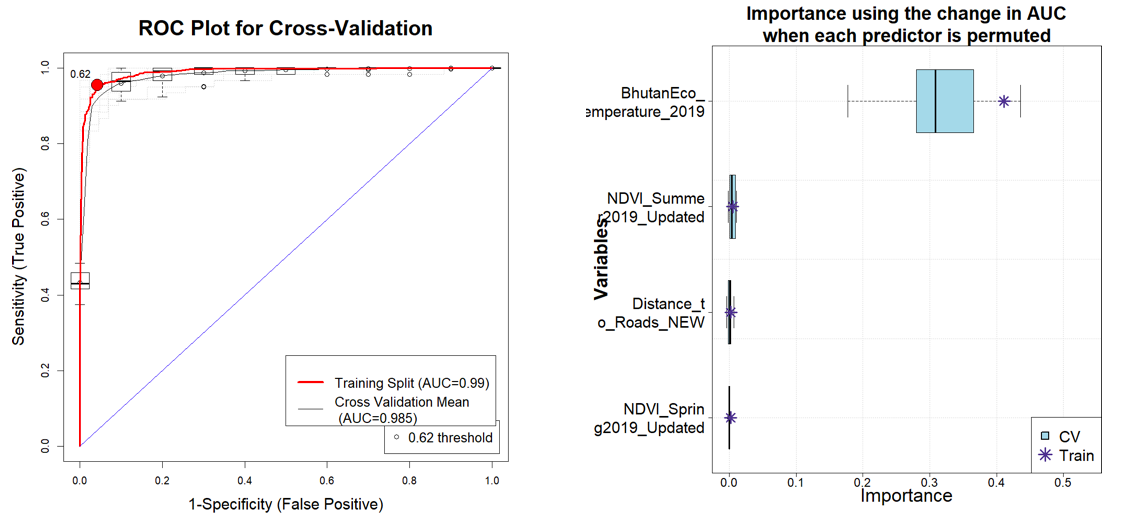
U.S. Geological Survey (2012). U.S. Geological Survey Earth Resources Observation and Science Center. Provisional Landsat TM Surface Reflectance. https://doi.org/10.5066/f7kd1vz9

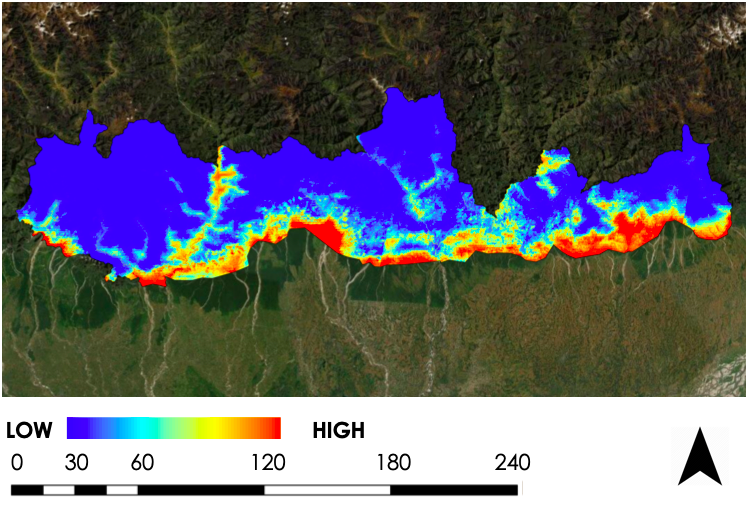
# 9. Appendix A

*Figure A1. ROC plot for Cross-Validation and Importance using the change in AUC when each predictor is permuted for GLM.*

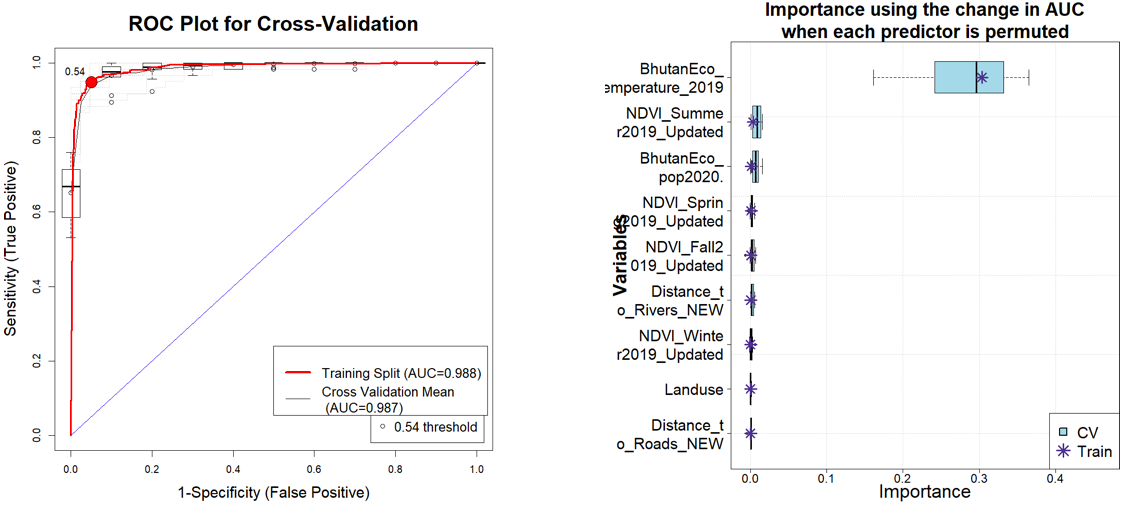
**

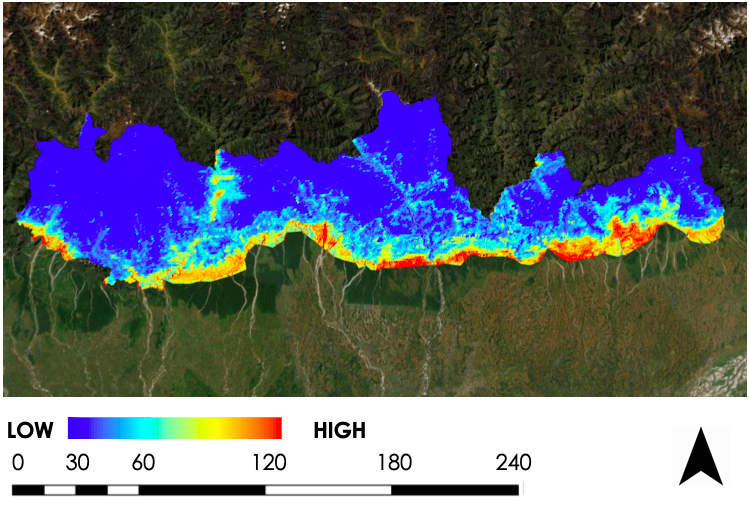
*Figure A2. Probability map produced by GLM.*

*Figure A3. ROC Plot for Cross-Validation and Importance using the change in AUC when each predictor is permuted given by MARS, 2019.*

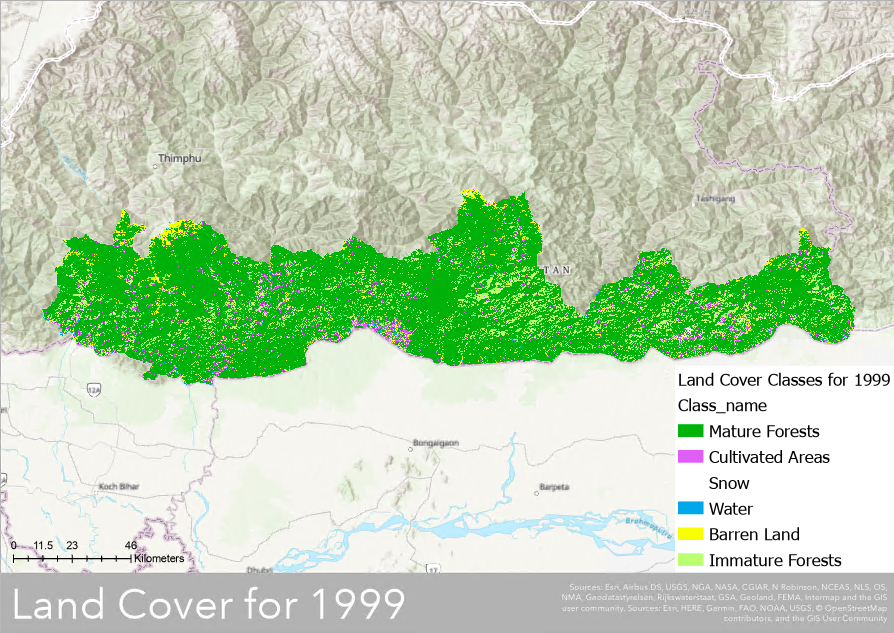
**

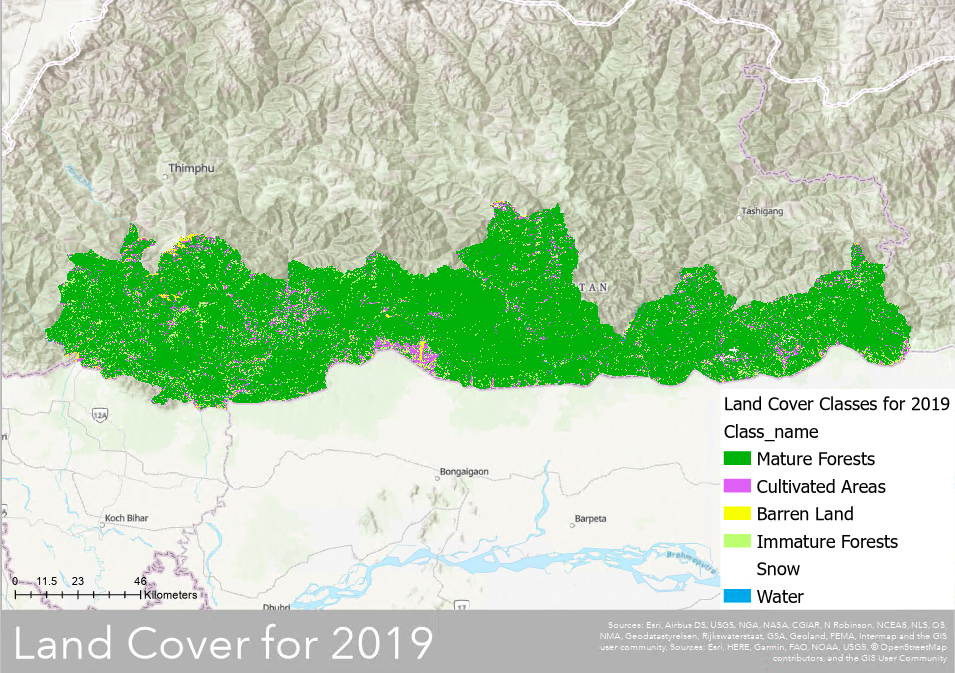
*Figure A4. Probability map produced by MARS.*

*Figure A5 ROC Plod for Cross-Validation and Importance using the change in AUC when each predictor is permuted given by RF.*

**

*Figure A6 Probability Map produced by RF.*

**

**

*Figure A7. Land Use and Land Cover Classification for 1999 (top) and 2019 (bottom).*