**NASA DEVELOP National Program**



NASA Ames Research Center

*Fall 2015*

Puerto Rico Health & Air Quality

A Geospatial Assessment of Environmental Variability in Puerto Rico and its Relation to Confirmed Dengue Cases

 **Technical Report**

Final Draft – November 19, 2015

Andrew Nguyen (Project Lead)

Alannah Johansen

Martha Sayre

Dr. Juan Torres-Pérez, Bay Area Environmental Research Institute (Science Advisor)

# 1. Abstract

Dengue fever is the fastest-growing vector-borne disease in the world. This deleterious illness is endemic in the Caribbean and can lead to hemorrhagic fever, shock, and death in severe cases; posing a major threat to the health of Caribbean communities. A high occurrence of the primary vector of the dengue virus *(Aedes aegypti*) has been detected in populated areas within Puerto Rico, contributing to several dengue outbreaks, including instances in 2010, 2012, and 2013. This study examined environmental conditions contributing to Confirmed Dengue Fever Cases (CDFC) for the island of Puerto Rico from January 2009 - December 2013 using monthly NASA Terra/ Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) 0.5° day land surface temperature (DLST) and night land surface temperature (NLST) products, 1 km Geostationary Operational Environmental Satellite system Puerto Rico Water Energy Balance (GOES-PRWEB) humidity products, 0.5° Climate Hazards Group InfraRed Precipitation and Satellite (CHIRPS) total precipitation (TP) modeled data, elevation, and land cover (LC). These data were incorporated into a Maximum Entropy Species Distribution Model to spatially delineate monthly potential dengue risk and to determine the permutation importance of variables based on CDFC. Land cover, elevation, and TP had the highest mean relative importance of environmental variables indicating that the disease spreads easily through low-elevation urban areas following periods of increased rainfall. Lastly, using Time Series Frequency Analysis, the correlation between MODIS 4km sea surface temperature (SST) products and CDFC were separately tested against environmental conditions to better understand the relationship between oceanic and land conditions contributing to dengue. SST and environmental conditions correlation coefficients indicate moderate to strong relationships.

**Keywords**

Remote Sensing, Maximum Entropy Species Distribution, MaxEnt, MODIS, *Aedes aegypti*

**2. Introduction**

**2.1 Background**

Dengue fever (DF) is a debilitating and potentially fatal mosquito-borne illness that is endemic in tropical regions. The Centers for Disease Control and Prevention (CDC) estimate that over 400 million people are infected globally each year by any one of four different serotypes, or variations, of the dengue virus (DENV) [CDC, 2014b]. Onset occurs within 3-7 days of the mosquito bite with symptoms that include fever, chills, rash, and vomiting, along with eye, muscle, and joint pain. The disease can further progress into dengue hemorrhagic fever (DHF) or dengue shock syndrome (DSS) in children, elderly individuals, and immunocompromised adults [Sharp et al., 2013]. Clinical signs at these advanced disease stages include severe bleeding, hypovolemic shock, disseminated intravascular coagulation (DIC), and death [CDC, 2014b; Sharp et al., 2013].

Globally, the virus is spread by several species of mosquito within the genus, *Aedes*, with the primary vector being *Aedes aegypti* [Cox et al., 2007]. *A. aegypti* is a domestic mosquito that lives and breeds near or within human-occupied structures [CDC, 2014a]. This species lays its eggs in a variety of water containers such as rain barrels, abandoned tires, plastic jugs, or pot holes on roads that are in close proximity to humans [CDC, 2014a; Barrera et al., 2011]. A secondary vector of DENV is *Aedes albopictus*, a rural mosquito that inhabits tall vegetation and tree holes or other containers that fill with water during the rainy season [Cox et al., 2007].

DF is endemic to the island of Puerto Rico, with anywhere from 3,000 to 9,000 cases occurring in non-epidemic years [CDC, 2014b]. However, DF epidemics on the island have been increasing in frequency and severity in recent years. The largest epidemic on record in Puerto Rico occurred in 2010, with over 26,700 suspected and 12,500 laboratory confirmed cases of the disease [Sharp et al., 2013]. In this year, the outbreak was unique because the majority of the cases were caused by serotypes 1 & 4, as opposed to 2 & 3, which had been circulating on the island from 2000-2009 [Barrera, Personal Communication, 2015]. The previous record outbreak occurred in 1998 and was similar to the 2010 outbreak, as both were caused by serotypes 1 & 4 and both occurred the year after major El Ni*ñ*o events [Barrera et al., 2011]. Also contributing to the 2010 outbreak were highly negative North Atlantic Oscillation index values that resulted in increased sea surface temperatures, and therefore increased rainfall in the region [Barrera et al., 2011]. Additionally, there have been two outbreaks since 2010. These occurred in 2012 and 2013, and both were primarily caused by serotypes 1 & 4 [Barrera, Personal Communication, 2015].

The primary vector known to contribute to DF epidemics in Puerto Rico is *A. aegypti*. This species exists in all regions of the island where humans are present with the highest numbers occurring in densely populated areas [Harrington et al., 2005]. Thus, the majority of these mosquitoes exist in developed urban regions, such as San Juan and Bayamon [Barrera et al., 2011]. Although other dengue vectors, such as *A. albopictus*,also exist on the island, the CDC Dengue Branch has never confirmed that any of these species are infected with DENV in Puerto Rico. Therefore, these species have not been designated as dengue vectors on the island [Barrera, Personal Communication, 2015].

There are multiple contributing environmental conditions that have been implicated in the spread of DENV, including increases in precipitation, sea surface temperature (SST), ambient temperature, and relative humidity (RH) [Johansson et al., 2009; Mendez-Larzaro et al., 2014; Patz et al., 1998; Barrera et al., 2011]. Increases in precipitation and temperature can result in the proliferation of *A. aegypti* by providing suitable breeding habitat and by hastening their development and reproductive cycles [Johansson et al., 2009]. Increases in temperature can also shorten the incubation period required for *A. aegypti* to become infectious after a DENV infected blood meal, which leads to an increase in the proportion of DENV vectors after a warming period [Johansson et al., 2009; Patz et al., 1998]. Also, increases in RH have been shown to result in increased ovipositon and hatching rates in *A. aegypti*, leading to increases in population density [Arruda Pedrosa de Almeida Costa et al., 2010] Additionally, overall SST has been correlated to dengue cases in Puerto Rico, Mexico, and New Caledonia [Mendez-Larzaro et. al., 2014; Ramos et. al., 2008; Diaz et. al., 2007]. Variations in TP are also influenced by SST anomalies during early dry months and late wet months of the year [Spence et al. 2004], potentially influencing DF incidence. In Puerto Rico, from 1982-2010, cooler surface waters are correlated to drier periods, while warmer surface waters are correlated to wet periods (R2=0.5) [Mendez-Lazaro, personal communication, October 27th 2015].

In order to study the relationship between favorable environmental conditions and increases in Confirmed Dengue Fever Cases (CDFC), this study utilized remotely-sensed products to quantify the contribution of environmental factors to DF epidemic and non-epidemic periods on the island of Puerto Rico from 2009 to 2013. To do this, the number of CDFC were analyzed from a continuous environmental geospatial modeling perspective.

**2.2 Project Objectives**

The primary objective of the first phase of this project was to utilize the Maximum Entropy (MaxEnt) Species Distribution Model within TerrSet’s Habitat and Biodiversity Modeler (HBM) as a predictive model for DF incidence in Puerto Rico based upon NASA Earth Observation (EO) environmental variables, as well as CDFC within the island of Puerto Rico. MaxEnt was used to model the suitability of dengue across Puerto Rico based upon the presence of CDFC in conjunction with LC, elevation, and contributing environmental factors. This method produced island-wide risk assessment maps of potential CDFC. The secondary objective was to perform a Time Series Frequency Analysis (TSFA) to explore the relationship between SST and CDFC, as well as the relationships between each of these and all other environmental variables. Project results provided a geospatial overview of both DF risk and the factors contributing to DF incidence in Puerto Rico from 2009-2013.

**2.3 Study Area**

Analyses were performed within the political boundaries of the Commonwealth of Puerto Rico. The island is an unincorporated United States territory located in the Caribbean Ocean with general coordinates of 18°15'N latitude and 66°30'W longitude (Figure 1). Puerto Rico is a small, densely-populated island with a total area of 9,104 km2 and a population of over 3.5 million. The climate is tropical with annual average temperatures between 21°C and 27°C, and a rainy season from April to November. Climate varies along the length of the island, with the drier regions occurring in the south. Land-based analyses were performed on the entire island, while correlations of SST to confirmed dengue cases extended past the study area to 256 km offshore (See 3.1 Data). *Aedes aegypti* have a known habitat preference for low-elevation urban areas [Lozano-Fuentes et al, 2012]. However, the entire spatial extent of the island was included in the MaxEnt model because the highest peak on the island is Cerro de Punta (1,338 m), which is well within the altitude range for *A. aegypti*. Studies conducted in Mexico showed common *A. aegypti* presence in altitudes up to 1,700 m above mean sea level, and occasional presence in altitudes from 1,700 m to 2,130 m [Lozano-Fuentes et al., 2014; 2012].

**2.4 Study Period**

The project examined the months from January 2009 through December 2013. This date range coincides with the most recent dengue outbreak years (2010, 2012, and 2013) according to CDFC *in-situ* data (See 3.1 Data).

**2.5 National Application(s) Addressed**

This project addressed the Health and Air Quality Application area within NASA’s Applied Science Program. By using NASA Earth Observing data, as well as modeled data products, this project focused on human welfare through the use of a Maximum Entropy Species Distribution Model to produce monthly DF risk assessment maps and to determine the contribution of environmental variables to DF epidemics in Puerto Rico. Model outputs will be used by public-health administrations to help predict and mitigate DF outbreaks in the future.

**2.6 Project Partners**

Currently, our end users, the Center for Disease Control (CDC)- Dengue Branch, University of Puerto Rico- Medical Sciences Department, and The Puerto Rico Department of Pubic Health use quantitative research on vector-borne diseases and outbreaks to inform public policy on vector control measures that can be taken to prevent the spread of diseases like DF.

The Puerto Rico Department of Health provides citizen services and public health announcements, and conducts health assessments pertaining to dengue awareness on the island. The agency reports on recent statistics and information regarding mosquito vector habitats, and publishes scientific literature related to various illnesses in Puerto Rico. The Dengue Branch of the CDC, located in San Juan, Puerto Rico, is dedicated to dengue research and health outreach. The agency employs public health practices such as education on the household spread of dengue and diagnostic testing. They also conduct molecular research and field investigations regarding dengue contraction and control. The results of this study will be able to inform both agencies on the contribution of environmental factors to DF outbreaks, allowing them to improve dengue prevention protocols.

**3. Methodology**

**3.1** **Data**

All data collected were obtained or downloaded in monthly time steps from January 2009 to December 2013. The data are divided into two categories: point *in-situ* data and raster environmental data.

**3.1.1 Point in- situ data**

***2000-2015 Confirmed Dengue Fever Cases (CDFC)***

A dataset of daily CDFC from January 2000 to August 2015 was obtained from the CDC, with a total of 44,338 CDFC. The CDC Dengue Branch tracks and monitors reported dengue cases and confirms these cases through laboratory testing. A vast majority of the data contain addresses, country name, zip code, and latitude/longitude coordinates. Cases without latitude/longitude coordinates or data that did not occur within the study period were excluded. This study used a total of 29,575 CDFC points within Puerto Rico. Increases, peak, and decrease of CDFC during this time period, generally, follow a seasonal time frame from late July to late January.

**3.1.2 Raster Environmental Data**

***2009-2013 Terra/ Aqua Moderate Resolution Imaging Spectroradiometer (MODIS)***

Two sets of *MODIS* data were downloaded to include (1) Sea Surface Temperature (SST) and (2) Land Surface Temperature (LST).

1. SST measurements (ºC) were obtained from NASA Earth Data Ocean Color website [NASA Ocean Biology, 2015]. Level 3 Aqua MODIS products of SST are at 4km resolution and products were derived using methods and algorithms produced by Brown & Minnett [1999].
2. MOD11C3v5, a MODIS level 3 data product, provided monthly DLST and NLST (Kelvin) at .05º resolution. LST data were downloaded for the h11v07 tile using Reverb, through the Land Process Distributed Active Archive Center (LP DAAC) website [LP DAAC, 2000].

***Climate Hazard Group InfraRed Precipitation with Stations (CHIRPS)***

TPdata (mm/month) at .05º resolution, from the Climate Hazard Group InfraRed Precipitation with Stations (CHIRPS) archive were downloaded from the University of California, Santa Barbara’s Climate Hazards Group website for each month of interest [CHG, 2015]. CHIRPS is a precipitation product from a combination of satellite and *in-situ* station data; it monitors drought and other environmental issues. The inputs to CHIRPS include modeled and Earth-observed precipitation data from the Climate Hazard Group’s Precipitation Climatology model, infrared satellite data from NOAA, Tropical Rainfall Measuring Mission (TRMM) data from NASA, NOAA Climate Forecast System data, and *in-situ* precipitation measurements [Funk et al., 2014].

***Geostationary Operational Environmental Satellite system Puerto Rico Water Energy Balance (GOES-PRWEB)***

Relative humidity (%) data were obtained from GOES-PRWEB which provides several island-scale estimated and modeled environmental datasets for Puerto Rico [GOES-PRWEB, 2009]. Incident radiation data are derived from Geostationary Operational Environmental Satellite system (GOES) and used to estimate net radiation, which is then used to further derive Photosynthetically Active Radiation. Additionally, solar radiation data from GOES are used to predict daily reference evapotranspiration along with an array of other important environmental conditions, including soil moisture. Precipitation data are sourced from NOAA’s Advanced Hydrologic Prediction Service to produce runoff and several other hydrological variables.

***2001 National Land Cover Database (NLCD)***

NLCD2001 for Puerto Rico was used as a land cover (LC) map and has a spatial resolution of 30 m. The map was a product of the Multi-Resolution Land Characteristics Consortium (MRLC) and was derived from Landsat 5 and 7 imagery. The Puerto Rico NLCD map was downloaded from the MRLC website [MRLC, 2001]. Since *A. aegypti* are known to occur in highest densities in developed and urban areas[Cox et al., 2007], the NLCD map was included as a static variable in the MaxEnt model in order to account for these habitat preferences.

***USGS Elevation***

The USGS 3D Elevation Program (3DEP) Digital Elevation Model (DEM) with a resolution of ⅓ arc second was used for elevation data. These LiDAR-based images were sourced from the USGS National Map Database [USGS, 2015]. Elevation can contribute to the available habitat for *Aedes aegypti* in a location due to the species’ climatic preferences [Lozano-Fuentes et al., 2012]. While elevation is not considered a weighted variable, geospatial results and model outputs that incorporate elevation data (See 3.2 Data Processing) will help link the role of elevation to occurrences of CDFC.

**3.2 Data Processing**

All monthly 2009-2013 CDFC, SST, DLST, NLST, RH and TP data were processed in order to conform with the data prerequisites of the Maximum Entropy Species Distribution Model within TerrSet; all of the monthly datasets were reprojected to the World Geodetic System (WGS) 1984 coordinate system.

***Point in- situ data***

Point shapefiles were created using known latitude/longitude coordinates for all laboratory CDFC using data supplied by project partners at the CDC Dengue Branch. Trends and separate point files were also created in monthly time steps for the study period (Figure 3 and 4).

***Environmental Variables***

Processing of LC, elevation, DLST, NLST, TP, RH, and SST was done using a combination of ArcGIS 10.3 Model Builder, ArcPy, and R-commander; this expedited the task of processing five years of monthly data for each variable.

All products were converted from their original forms into TIFFs and re-projected to the WGS 1984 coordinate system. The 3DEP DEM tiles containing parts of the island of Puerto Rico were then mosaicked together and clipped to the boundary shapefile of the island. All other products were subsequently resampled bilinearly to the ⅓ arc second DEM, which had the finest resolution among the variables, and clipped to the boundary shapefile of Puerto Rico. DLST and NLST were further processed to correct for a scale factor of .02, using the raster calculator tool in ArcGIS 10.3. All of the products’ *no data values* were set after processing back to their originals. The resultant TIFF files were then converted to RST format and further processed using TerrSet software. The values of the LC classes in Puerto Rico were reclassified into a numerically sequential order. Finally, all files were resampled to a uniform extent and several parameters within the metadata were updated. These include the addition of value units, as well as the classification of *no data values* as background.

SST data were converted from NetCDFs to TIFFs, re-projected, and clipped to a radius of 256 km off the coasts of the island based on park done by Mendez-Larzaro et al., 2014 .

**3.3 Analysis**

Analyses were conducted to geospatially predict CDFC and statically analyze environmental variables that are known to contribute to DF outbreaks.

**3.3.1 Maximum Entropy Species Distribution Model**

Maximum Entropy (MaxEnt) Species Distribution modeling was conducted for every month from January 2009 to December 2013 for the entire country of Puerto Rico using Clark Labs’ TerrSet 1.0. TerrSet is a software system that incorporates IDRISI GIS Analysis, image processing tools, and several modeling approaches for analysis of geospatial data [Clark University, 2015].

Due to the presence-only nature of the CDFC data, the MaxEnt portion of the Habitat Suitability / Distribution Module within TerrSet’s Habitat and Biodiversity Modeler (HBM) was chosen to model the geographic suitability of DF incidence in all of Puerto Rico. MaxEnt is a machine learning model that used the principle of maximum entropy to assign a predicted probability of suitable conditions for CDFC occurrences at each pixel location. The model accomplished this by starting with a uniform distribution, randomly assigning 75% of the presence points to be used for training, and then successively using each point to improve the fit of the model to the presence data [Phillips, 2006]. The remaining 25% of the data points were then used to test the accuracy of the predictions.

Model runs were performed in monthly time-steps from 2009-2013 to produce DF risk assessment maps based on the probability distribution of CDFC. In addition, MaxEnt model gave estimates of the relative contributions of each environmental variable to the final model output in the form of permutation importance values. MaxEnt estimates these values by tracking which variables most contribute to the overall fit of the model at each training step, then by measuring the decrease in fit when the values of each of the variables are randomly rearranged. The advantage to this estimation method is that it gives values of relative importance that are independent of the path used to calculate them [Phillips, 2006].

***Application of Time Lags to MaxEnt Model***

Once environmental conditions become favorable, time is required for mosquito populations to proliferate and for mosquitoes to become infectious after an infected blood meal. After an onset of symptoms begin to appear in infected people, cases can be diagnosed and confirmed through laboratory testing. Due to these delays between changes in the environment and the increase in DF incidence, it was necessary to apply a time lag for each environmental variable to the MaxEnt model runs. The lengths of the time lags were chosen for each variable based on previous studies. For DLST and NLST, a time lag of three months was chosen based on the work of Keating [2000] who showed that the increase in DF cases in Puerto Rico occurred 12 weeks following the peak in ambient temperature. A two-month time lag was chosen for TP based on work by Moore et al. [1978] whose results indicated that peak DF incidence in south-western Puerto Rico occurs 6-8 weeks after peak rainfall. A two-month time lag was also chosen for RH following the work of Gharbi et al. [2011] who used a Seasonal Autoregressive Integrated Moving Average model to determine that the strongest correlation between RH and dengue incidence in Guadeloupe, French West Indies occurred at a time lag of 7 weeks. The two month estimate of RH time lag was based on this study since no studies were found that determined the time lag between RH and DF incidence specifically in Puerto Rico.

***Analysis of Model Outputs***

Mean monthly permutation importance values for DLST, NLST, RH, and TP were calculated for all years, as well as both epidemic and non-epidemic years. These values were compared to each other and to the mean monthly number of training cases in order to determine trends in the factors that are contributing to dengue outbreaks.

MaxEnt also gives estimates of the predictive ability of the output for each model run. First, it gives values corresponding to the overall gain of the predictive ability of the model for both training and testing. Second, it calculates the area under the curve (AUC) for both training and testing curves on a sensitivity vs. 1-specificity graph [Phillips et al., 2006]. Third, it uses a binomial test of omission to calculate p-values testing the null hypothesis that there is no statistically significant difference between the ability of the model output or a random prediction to predict the location of test points. MaxEnt also performs a jackknife test that gives an additional measure of variable importance by measuring the training and testing gain for each variable by itself and for the model with that variable omitted [Phillips, 2006]. All of these measures were used to test for differences in model accuracy with respect to differing time lag scenarios (See IV. Results & Discussion – MaxEnt Time Lag Analysis)

***Habitat Assessment***

In addition to the MaxEnt model, a habitat assessment was performed using all CDFC from Jan 2009 - Dec 2013 along with the static variables of LC and elevation. The Habitat Assessment module within TerrSet’s HBM uses both a land cover and a habitat suitability map to assign regions of primary and secondary habitat, as well as potential corridors. Only developed areas were chosen to be included as potential habitat since *A. aegypti* is known to be dependent on humans for survival [Harrington, 2005]. All gap distances between and outside of the developed areas were estimated to be 400 m based on the work of Reiter et al. [1995] who estimated a 30 m - 440 m *A. aegypti* dispersal range in San Juan by using rubidium-marked eggs. Additionally, Harrington et al. [2005] used a release and recapture method to estimate a dispersal range of 12 m - 102 m in Puerto Rico with 87% of recaptures occurring within the same home they were released. This shows that the majority of *A. aegypti* would most likely not disperse beyond an estimated 400 m gap distance.

The habitat assessment model also uses a habitat suitability map as a weighted assessment of potential habitat locations. The habitat suitability map was generated for this study by using the MaxEnt model with inputs of all CDFC within the study period, as well as LC and elevation. Minimum weighted index values of 0.5 and 0.2 were chosen for primary and secondary habitat, respectively, since those values corresponded with the minimum values of known habitat preferences. Since both the presence of CDFC and the habitat preferences of *A. aegypti* were used to produce the habitat assessment map, the resultant habitat classifications can be designated as belonging to those *A. aegypti* individuals that are infected with any serotype of DENV.

**3.3.2 Time Series Frequency Analysis-CDFC and Sea Surface Temperature**

SST values for all pixels within the 256 km radius off the coast of the island were averaged to represent a given value per month and year (Fig. 2). These SST monthly averages were tested against monthly CDFC, DLST, NLST, TP, and RH using a Time Series Frequency Analysis (TSFA). Additionally, CDFC were also tested against DLST, NLST, TP, and RH. This results in two sets of TSFA Pearson correlations coefficients. These resulting coefficients would help us gain a better understanding of the relation between SST, dynamic environmental conditions, and CDFC.

Once values were standardized, the correlation coefficients were calculated. Correlation coefficients measure the strength of association between two variables over time [Helsel and Hirsch, 1993]. The following equation was used

Equation:

where r represents the correlation coefficient, N is the number of data points, x is series one, and y is series two [Box and Jenkins, 1976]. By comparing and correlating these data, we strengthened the association between SST and CDFC data.

**3.4 Time Lag Analysis**

***TSFA Analysis***

According to Focks and Barrera [2006], peak CDFC are detected six to eight weeks after the peak in seasonal rainfall in southwestern Puerto Rico and during the normal onset of rains in July and August. However, earlier than expected rainy seasons correspond to earlier CDFC, while later than expected start of the season correspond to later CDFC incidence.

Based on the wealth of literature concluding rainfall to be closely correlated to CDFC [Johansson et al., 2009; Mendez-Larzaro et al., 2014; Patz et al., 1998; Barrera et al., 2011],we adjusted the CDFC to TP TSFA results to account for a 5 month lag in TP. The relationship between CDFC and TP indicated a moderate correlation (r=0.38), a trend similar to related studies. However the 5 month lag and correlation value do not agree with a study conducted for Puerto Rico [Moore et al. 1978]. Moore et al. indicates CDFC detection approximately two months after the peak in seasonal rainfall within specific urban communities mainly based in the drier southwestern sections of the island when correlated with island-wide precipitation data. However, our study examines CDFC for the entire island 30 years later. Similar precipitation to CDFC correlation studies conclude 4-5 months lag in precipitation and are conducted in subtropical areas near our study time frame.

In order to account for the time frame between the occurrence time of SST to TP, a five month time lag was applied. To account for the time frame between the occurrence time of dynamic environmental conditions and CDFC, various lag time frames were applied for specific conditions in for our TSFA approach. A three month lag was applied to LST data while two months were applied to RH. Lastly, due to uncertainty of TP lag times to CDFC, two separate results were created based on 2 month and 5 month lag times. Therefore, this study presents modeled CDFC results to account for arguments regarding the proper lag times in the case of TP.

***MaxEnt Analysis***

To address TP 5 and 2 month lag time uncertainty, based on the results of the Time Series Frequency Analysis, a second set of model runs were performed which incorporated a 5 month time lag for TP while keeping all other time lags consistent with the previous model runs. In order to compare the results of the two time lags, an additional set of model runs was performed with no time lags applied. This set of model runs was used as a baseline for several parameters used to test model accuracy. These include training and testing gain and AUC, entropy, and testing omission evaluated at multiple thresholds. Values for each parameter were compared between the sets of time lagged model runs with the baseline values subtracted off from each. This was done in order to compare the relative improvement of each set of time lags over a set with no time lags applied. Two-sample t-tests were performed for all parameters and returned p-values that determined that there were no significant differences between the 2 month and 5 month TP time lags with respect to model accuracy or TP permutation importance. Additionally, no significant differences were found between either of the applied time lags and the no lag model runs. Multiple possibilities exist that could explain these results. First, since LC accounts for the majority of the relative importance, it may be that changes among the other variables would only slightly affect model results. Therefore, a change in the TP time lag would not significantly change the model. Second, it is possible that significant differences may have been seen if the study period were longer, which would have given a larger sample size.

**4. Results & Discussion**

**4.1 Maximum Entropy Species Distribution Model**

Results from all model runs showed that the overall greatest relative contribution to dengue incidence was due to land cover (60.3%±17.0%). This was followed by elevation (13.9%±15.6%), TP (11.8%±11.9%), DLST (5.9%±6.6%), RH (4.8%±5.3%), and NLST (3.3%±3.6%) (Table 1)(Figure 8, 10, & 11). These results indicate that the disease spreads easily through low-elevation urban areas following periods of increased rainfall. A similar pattern was seen on the risk assessment maps produced by the model, which showed the highest probability of suitable conditions for DF occurring in low elevation cities and coastal developed areas. In particular, the San Juan metropolitan area, Caguas, and Ponce showed the highest index values in the majority of model runs (Figure 6 & 7). It was also observed that, with the exception of high elevation regions, index values for the rest of the island increased dramatically during epidemic periods. A disparity was also noted for permutation importance values between epidemic and non-epidemic years. This was seen as an increase in the mean relative importance of LC and a decrease in the importance of elevation, TP, and DLST during epidemic years (LC=66.6%±13.5%, Elevation=10.4%±9.8%, DLST=4.5%±4.8%, NLST=3.2%±3.6%, RH=5.5%±5.5%, TP=9.4%±9.8%) when compared to non-epidemic years (LC=49.4%±17.1%, Elevation=19.9%±21.4%, DLST=8.3%±8.5%, NLST=3.4%±3.6%, RH=3.6%±5.0%, TP=15.3%±14.8%) (Figures 12 & 13). Examination of time series for each of the permutation importance values showed no indications of seasonal trends or patterns, with the exception of an increase in the relative importance of LC during epidemic periods. It was also grossly observed that the relative importance of TP was highest during 2011.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Environmental Variable** | **Percentage** |  |  |  |
| land cover | 60.3% ± 17.0% |  |  |  |
| elevation | 13.9% ± 15.6% |  |  |  |
| TP | 11.8% ± 11.9% |  |  |  |
| DLST | 5.9% ± 6.6% |  |  |  |
| RH | 4.8% ± 5.3% |  |  |  |
| NLST | 3.3%±3.6%) |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Table 1: Contributions of environmental variables to dengue incidence | | | | |

It should be noted that the geographic coordinates of the CDFC are geocoded based on addresses of those infected with DENV [CDC NCEH, 2005]. By using these data, the model accuracy therefore rests on the assumption that those with laboratory-confirmed dengue fever were infected at home. This assumption ignores the possibility that these individuals could have been infected at any location other than their current residence, however, there is no way of knowing the exact point of infection. Further, it was calculated from the US Census Bureau County to County Commuting Flows 2009-2013 dataset for Puerto Rico that only an estimated 48%±13% of people were commuting within the same county as their primary residence during the study period [USCB, 2015]. Therefore, this introduces a high level of uncertainty into the model, however, no alternative methods exist for geolocating CDFC at this time.

Variations in the number of training CDFC for each model run may also have resulted in variations in the model results. In order to show this, mean permutation importance values for each variable were calculated for model runs with the number of training samples: TS<100 (LC=48.9%±19.8%, Elevation=23.0%±22.7%, DLST=5.7%±7.3%, NLST=1.5%±2.3%, RH=1.7%±4.3%, TP=19.2%±17.8%), 100⩽TS<500 (LC=60.6%±11.8%, Elevation=12.4%±11.8%, DLST=6.9%±6.9%, NLST=4.7%±4.2%, RH=5.9%±5.1%, TP=9.5%±7.8%), and TS⩽500 (LC=73.6%±12.5%, Elevation=6.8%±3.9%, DLST=4.3%±5.4%, NLST=2.5%±2.0%, RH=6.3%±5.8%, TP=6.4%±4.3%). As with the analysis between epidemic and non-epidemic years, mean LC permutation importance increased, and standard deviations decreased for all variables, with increases in the number of training samples (Figure 14). Therefore, it is possible that the disparity seen between epidemic and non-epidemic years could be due to an artifact of the model, with the changes in permutation importance values simply being due to an increase in model accuracy when the number of training samples is large.

***Habitat Assessment***

A habitat assessment map for DENV infected *A. aegypti* was produced by combining *A. aegypti* habitat preferences with CDFC locations, as well as LC and elevation data. The resultant map located habitat of DENV infected *A. aegypti* to be along the majority of Puerto Rico coastline as well as inland urban areas, such as Caguas (Figure 9). Analysis of the habitat assessment map showed that 96.6% of all CDFC were contained within the habitat and corridor boundaries, with 86.8% being in the primary or secondary habitat and 9.8% in the primary or secondary corridors. Additionally, the total percentage increased to 97.3% when CDFC, within a distance of 100 m of the range boundaries, were included in the analysis.

**4.2 Time Series Frequency Analysis-CDFC and Sea Surface Temperature**

Two sets of TSFA correlations were conducted to better understand the relationship between SST and CDFC to various environmental variables that include NLST, DLST, RH, and TP. The strong relationship between SST values and NLST (r=0.81) indicates a strong relationship (Figure 16), while DLST resulted in a value of 0.40 with a moderate relationship (Figure 17). SST to RH indicated a value of 0.45 (Figure 18). Lastly, SST to TP showed a value of 0.60 with a 5 month time lag applied (Figure 19). For the CDFC set, NLST resulted in a 0.37 value (Figure 20), while DLST resulted in 0.41(Figure 21). Both correlations had a 3 month time lag applied. CDFC to RH resulted in a -0.13 value indicating little to no correlation with a 2 month time lag (Figure 22). A 5 month time lag was applied to CDFC to TP resulting in a 0.38 value indicating a moderate relationship (Figure 23). However, a 2 month time lag resulted in a value of -0.36, indicating little to no relationship. Lastly, CDFC to SST were values resulted in an r value of 0.41 implying a moderate to strong relationship (Figure 24).

**4.3 Future Work**

Due to the high relative importance assigned by the MaxEnt model to the static variables of LC and elevation, it was difficult to determine trends in CDFC with respect to environmental conditions. Therefore, now that it has been established that CDFC occur most often in low-elevation urban areas, the next phase of the project will narrow the study area to the extent of the habitat and corridor boundaries of the habitat assessment map. This will allow the analysis to focus solely on the relative importance of each of the environmental conditions contributing to DF outbreaks. In addition, other environmental variables will be added to the model including ambient temperature, wind speed, evapotranspiration, and soil moisture. This will allow for a more complete understanding of the effect of environmental changes on DF epidemics.

# 5. Conclusions

# MaxEnt results indicated a strong relationship between dengue risk and LC. This is in agreeance with much of the literature, as *Aedes aegypti* breeding conditions are based in areas with high population densities and areas where water can accumulate in small to medium sized bodies. Additionally, elevation is most likely a factor of population density; the higher in elevation, the less populated areas exist in Puerto Rico. These lower-risk areas in higher elevation can be seen in the geospatial model results. Lastly, MaxEnt and TSFA results indicate TP as a moderate to strong driver for dengue risk, most likely due to accumulated bodies of water. DLST, NLST, and RH are also know drivers of *Aedes aegypti* breeding. Although not as significant compared to LC, elevation, and TP, they indeed contribute to dengue risk and should be considered as drivers. Due to environmental variability, unpredictable occurrences of outliers, and a large spatial study area, the TSFA results are considered moderate to strong relationships. Thus, further studies are recommended to examine and better understand other conditions as contributors to dengue risk from a geospatial island-wide perspective.

Supplementing infectious disease control efforts with analyses of Earth observations will help to increase our understanding of how these diseases are spread, and what can be done to prevent them. By applying remote sensing and geospatial modeling techniques to the field of disease ecology, the scientific community and policy makers are able to learn about how changes in the environment can contribute to outbreaks, and predict where epidemics are most likely to occur. This will allow public health authorities to allocate resources, educational efforts, and medical support to the communities that are most at risk.

# 6. Acknowledgments

We would like to acknowledge Dr. Pablo M*é*ndez-L*á*zaro from Medical Sciences Campus of the University of Puerto Rico, and Dr. Roberto Barrera from the U.S. Centers for Disease Control and Prevention (CDC) Dengue Branch for project advisement and data sources. We would also like to thank Jessica Cabrera from the Puerto Rico Department of Health for her overall support of the project. Lastly, we thank Chippie Kislik and Victoria Ly for their support and supervision for the project as the management team at DEVELOP NASA Ames Research Center.

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 NNL11AA00B and cooperative agreement NNX14AB60A.

# 7. References

Arruda Pedrosa de Almeida Costa, E., E. Maria de Mendonça Santos, J. Cavalcanti Correia, and C. Maria Ribeiro de Albuquerque (2010), Impact of small variations in temperature and humidity on the reproductive activity and survival of *Aedes aegypti* (Diptera, Culicidae), *Rev. Bras. Entomol., 54*(3), 488-493.

Barrera, R., M. Amador, and J. MacKay (2011), Population Dynamics of *Aedes aegypti* and Dengue as Influenced by Weather and Human Behavior in San Juan, Puerto Rico, *PLOS Neglected Tropical Diseases*, 5(12), e1378.

Box, G. E. P., and G. M. Jenkins (1976), *Time series analysis: forecasting and control*, Holden-Day, Oakland, California.

Brown, O. B., P. J. Minnett, R. Evans, E. Kearns, K. Kilpatrick, A. Kumar, R. Sikorski, and A. Závody (1999), MODIS Infrared Sea Surface Temperature Algorithm Algorithm Theoretical Basis Document Version 2.0, *University of Miami*, 33149–1098.

Climate Hazard Group (CHG) (2015), Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) Version 2, January 2009 to December 2013, University of California, Santa Barbara. Available from: ftp://[chg-ftpout.geog.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global\_monthly/tifs/](http://chg-ftpout.geog.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_monthly/tifs/) (last acessed 5 October 2015).

Centers for Disease Control and Prevention (2014), Dengue and the *Aedes aegypt*i mosquito,  <http://www.cdc.gov/dengue/resources/30Jan2012/aegyptifactsheet.pdf> (last acessed 11 November 2015).

Centers for Disease Control and Prevention (2014), Dengue Epidemiology, <http://www.cdc.gov/Dengue/epidemiology/index.html> (last accessed 5 October 2015).

Centers for Disease Control and Prevention (CDC) National Environmental Public Health (NCEH) Tracking Program (2005), A geoprimer: Environmental public health tracking, Version 1.0, <http://www.cdc.gov/nceh/tracking/pdfs/geoprimer.pdf> (last acessed 11 November 2015).

Cox, J., M. E. Grillet, O. M. Ramos, M. Amador, and R. Barrera (2007), Habitat segregation of Dengue vectors along an urban environmental gradient, *Am. J. Trop. Med. Hyg.*, 76(5), 820–826.

Clark Labs (2015), Clark University. [https://clarklabs.org/products/.](https://clarklabs.org/products/)

Eisen, L., and S. Lozano-Fuentes (2009), Use of Mapping and Spatial and Space-Time Modeling Approaches in Operational Control of Aedes aegypti and Dengue, edited by A. J. Tatem, *PLoS Neglected Tropical Diseases*, *3*(4), e411, doi:10.1371/journal.pntd.0000411.

Focks, D. A., and R. Barrera (2006), Dengue transmission dynamics: Assessment and Implications for control. *WHO Report of the Scientific Working Group meeting on Dengue, Geneva, 01/2006*.

Funk, C. et al. (2014), A Quasi-Gloabal Precipitation Time Series for Drought Monitoring, US Geological Survey, 1-5.

Gubler, DJ (2002), Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends Microbiol 10: 100–103.

GOES-PRWEB (2009), Relative Humidity,  <http://academic.uprm.edu/hdc/GOES-PRWEB_RESULTS/> (Updated Daily).

Harrington, L. C., T. W. Scott, K Lerdthusnee, T. C. Coleman, A. Costero, G. G. Clark, J. J. Jones, S. Kitthawee, P. Kittayapong, R. Sithiprasasna, and J. D. Edman (2005), Dispersal of the Dengue vector Aedes aegypti within and between rural communities, *Am. J. Trop. Med. Hyg.*, 72(2), 209–220.

Helsel, D. R., and R. M. Hirsch (1993), *Statistical Methods in Water Resources*, Elsevier.

Hurtado-Díaz, M., H. Riojas-Rodríguez, S. J. Rothenberg, H. Gomez-Dantés, and E. Cifuentes (2007), Short communication: Impact of climate variability on the incidence of dengue in Mexico: Impact of climate variability on the incidence of dengue in Mexico, *Tropical Medicine & International Health*, *12*(11), 1327–1337, doi:10.1111/j.1365-3156.2007.01930.x.

Johansson, M. A., F. Dominici, and G. E. Glass (2009), Local and Global Effects of Climate on Dengue Transmission in Puerto Rico, edited by E. Massad, *PLoS Neglected Tropical Diseases*, *3*(2), e382, doi:10.1371/journal.pntd.0000382.

Land Processes Distributed Active Archive Center (LP DAAC) (2000), MODIS 11C3

Land Surface Temperature and Emissivity Monthly L3 Global 0.05Deg CMG, Version 5, January 2009 to December 2013, https://lpdaac.usgs.gov/dataset\_discovery/modis/modis\_products\_table/mod11c3, NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.

Little, E., R. Barrera, K. C. Seto, and M. Diuk-Wasser (2011), Co-occurrence Patterns of the Dengue Vector Aedes aegypti and Aedes mediovitattus, a Dengue Competent Mosquito in Puerto Rico, *EcoHealth*, *8*(3), 365–375, doi:10.1007/s10393-011-0708-8.

Lozano-Fuentes, S. C., Welsh-Rodriguez, A. J. Monaghan, D. F. Steinhoff, C. Ochoa-Martinez, B. Tapia-Santos, M. H. Hayden, and L Eisen (2014), Intra-annual changes in abundance of *Aedes (Stegomyia) aegypti* and *Aedes (Ochlerotatus) epactius* (Diptera: Culicidae) in high-elevation communities in Mexico, *J. Med. Entomol*., 51(4), 742-751.

Lozano-Fuentes, S. C., M. H. Hayden, C. Welsh-Rodriguez, C. Ochoa-Martinez, B. Tapia-Santos, K. C. Kobylinski, C. K. Uejio, E. Zielinski-Gutierrez, L. Delle Monache, A. J. Monaghan, D. F. Steinhoff, and L Eisen (2014), The dengue virus mosquito vector *Aedes aegypti* at high elevation in Mexico, *Am. J. Trop. Med. Hyg*., 87(5), 902-909.

Méndez-Lázaro, P., F. E. Muller-Karger, D. Otis. M. J. McCarthy, and M. Peña-Orellana (2014), Assessing climate variability effects on Dengue incidence in San Juan, Puerto Rico, *Int. J. Environ. Res. Public Health*, 11, 9409-9428.

Moore, C. G., B. L. Cline, E. Ruiz-Tiben, D. Lee, H. Romney-Joseph, and E. Rivera-Correa (1978), *Aedes aegypti* in Puerto Rico: Environmental determinants of larval abundance and relation to dengue virus transmission, *Am J Trop Med Hyg*, 27, 1225-1231.

Multi-Resolution Land Characteristics Consortium (MRLC) (2001), National Land Cover Database Puerto Rico Version 1,<http://www.mrlc.gov/nlcd01_data.php>, U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, MRLC Project, Sioux Falls, SD.

Patz, J. A., W. J. Martens, D. A. Focks, and T. H. Jetten (1998), Dengue fever epidemic potential as projected by general circulation models of global climate change., *Environmental Health Perspectives*, *106*(3), 147.

Phillips, S. (2006), A brief tutorial on Maxent. *AT & T Research*, http://www.cs.princeton.edu/ ~schapire/maxent/tutorial/tutorial.doc.

Phillips, S., R.P. Anderson, and R.E Schapire (2006), Maximum entropy modelling of species geographic distributions, *Ecological Modelling*, *190*, 231–259.

Ramos, M. M. et al. (2008), Epidemic dengue and dengue hemorrhagic fever at the Texas–Mexico border: results of a household-based seroepidemiologic survey, December 2005, *The American journal of tropical medicine and hygiene*, *78*(3), 364–369.

Reiter, P, M. A. Amador, R. A. Anderson, and G. G. Clark (1995), Short report: Dispersal of Aedes aegypti in an urban area after blood feeding as demonstrated by rubidium-marked eggs, *Am. J. Trop. Med. Hyg., 52*(2), 177-179.

Spence, J. M., M. A. Taylor, and A. A. Chen (2004), The effect of concurrent sea-surface temperature anomalies in the tropical Pacific and Atlantic on Caribbean rainfall, *International Journal of Climatology*, *24*(12), 1531–1541, doi:10.1002/joc.1068.

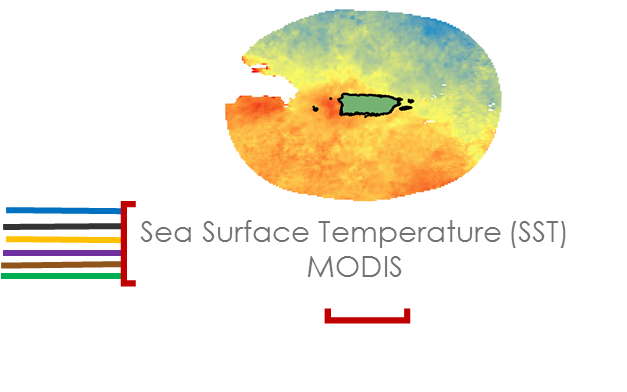
Sharp, T. M., E. Hunsperger, G. A. Santiago, J. L. Muñoz-Jordan, L. M. Santiago, A. Rivera, R. L. Rodriguez-Acosta, L. Gonzalez Feliciano, H. S. Margolis, and K. M. Tomashek (2013), Virus-Specific Differences in Rates of Disease during the 2010 Dengue Epidemic in Puerto Rico, *PLOS Neglected Tropical Diseases*, 7(4), e2159.

US Census Bureau (USCB) (2015), Commuting (Journey to Work) Main, http://www.census.gov/hhes/commuting.

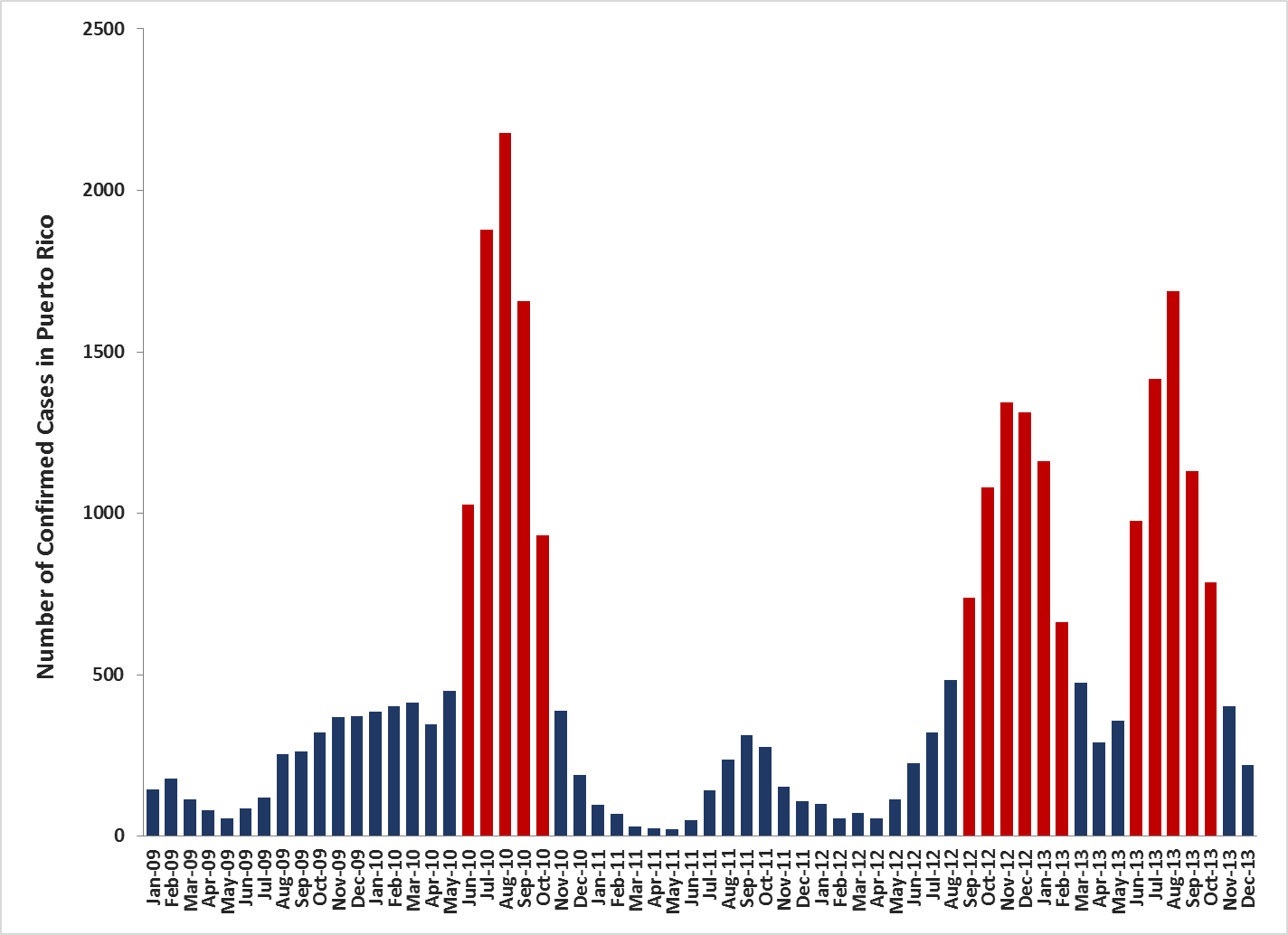
US Geological Survey (USGS) The National Map (TNM) (2015), 3D Elevation Program (3DEP), <http://nationalmap.gov/3DEP/index.html>.

**Appendix A: Figures**

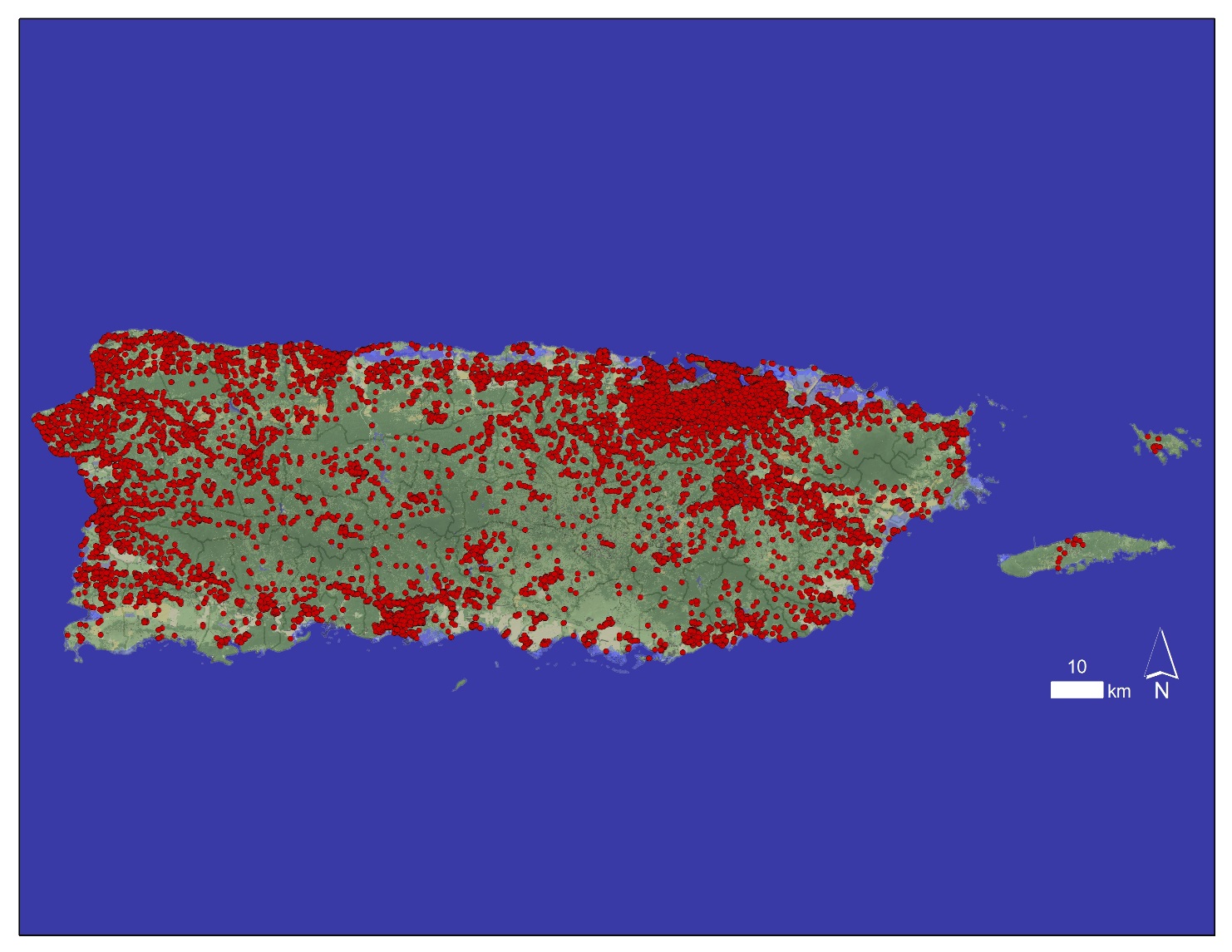
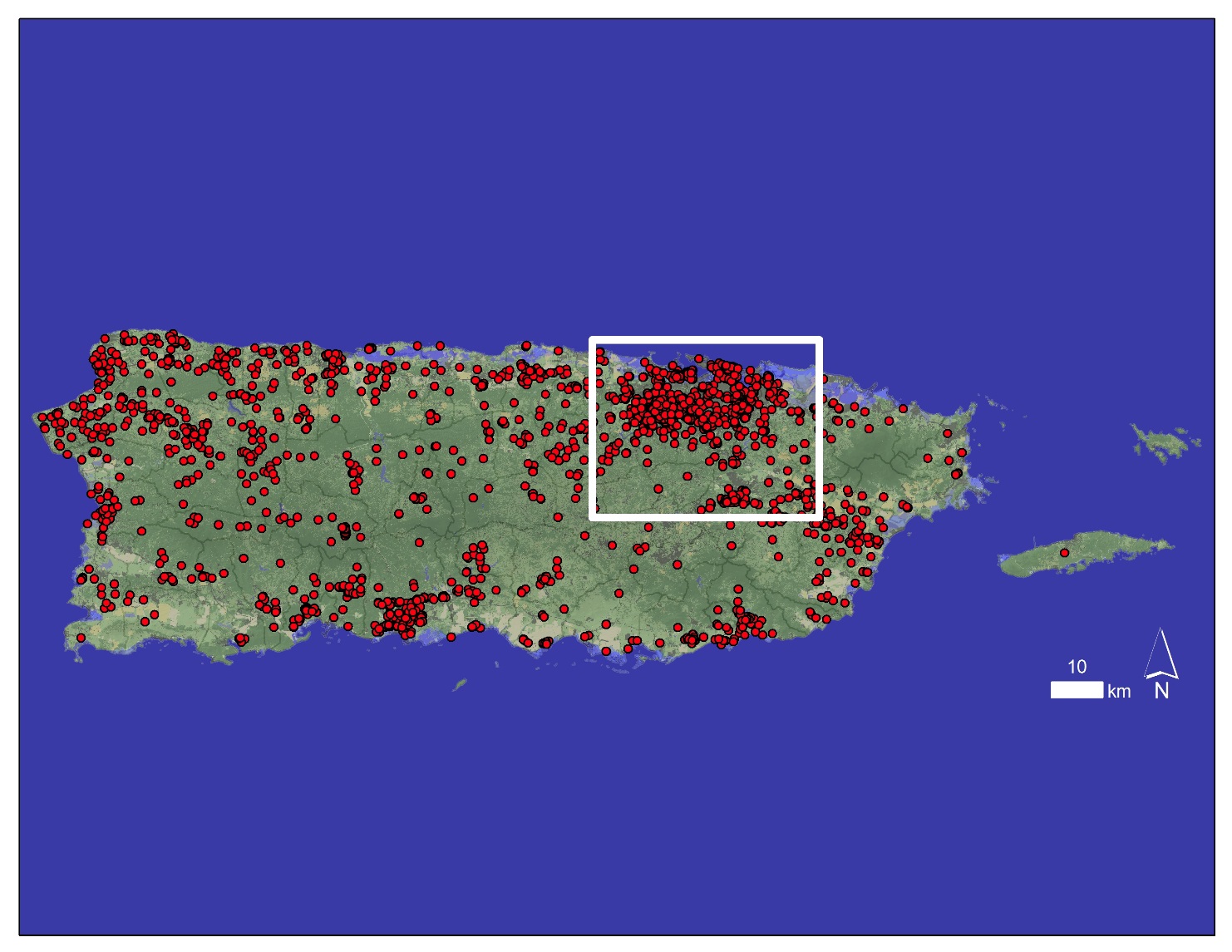
*Figure 1*: Study area 1: Analyses using the MaxEnt model were conducted within the political boundaries of the Commonwealth of Puerto Rico.



*Figure 2*: Study area 2: The study area was extended 256 m off the coast of Puerto Rico for all TSFA analyses using SST.

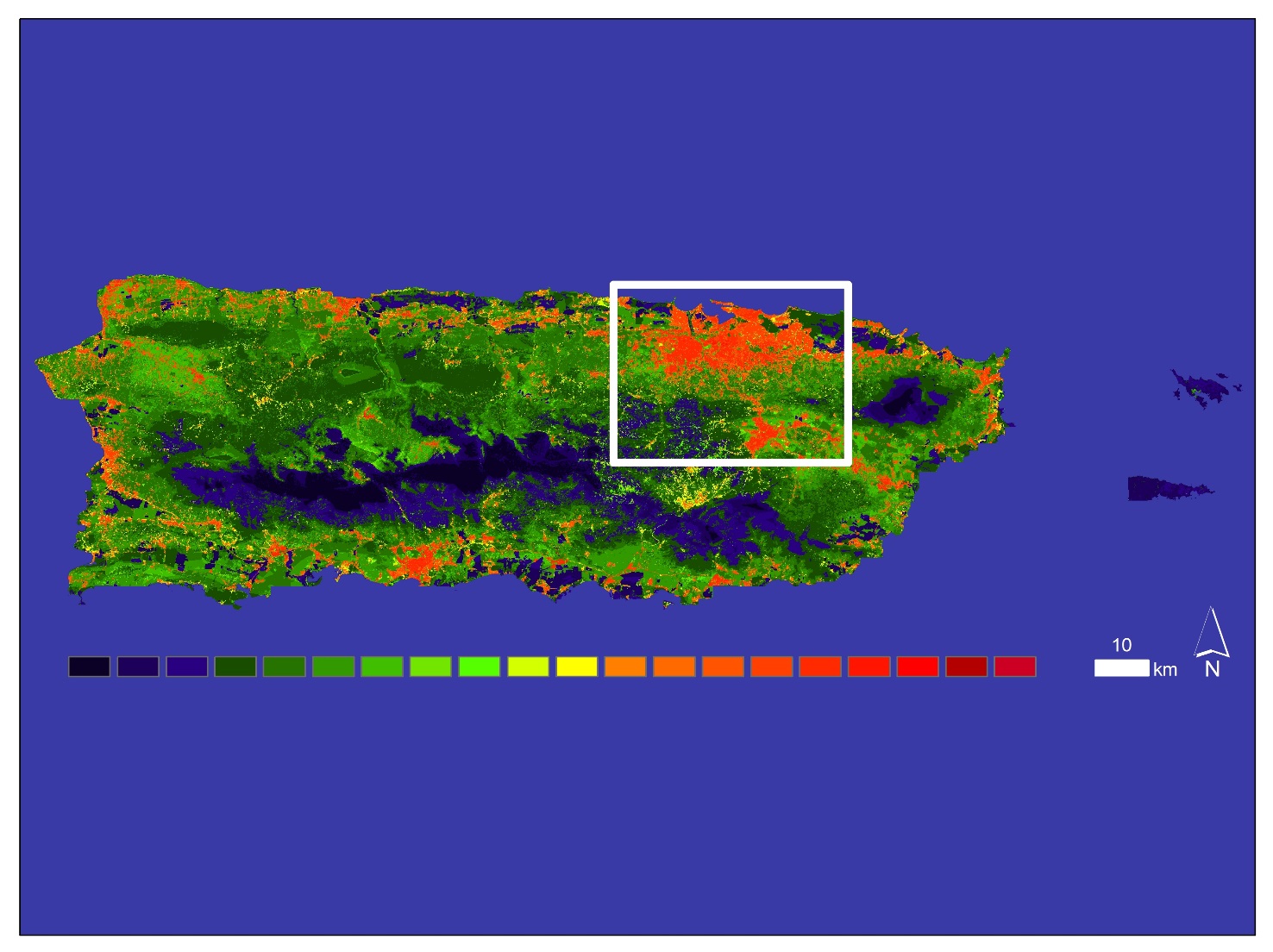
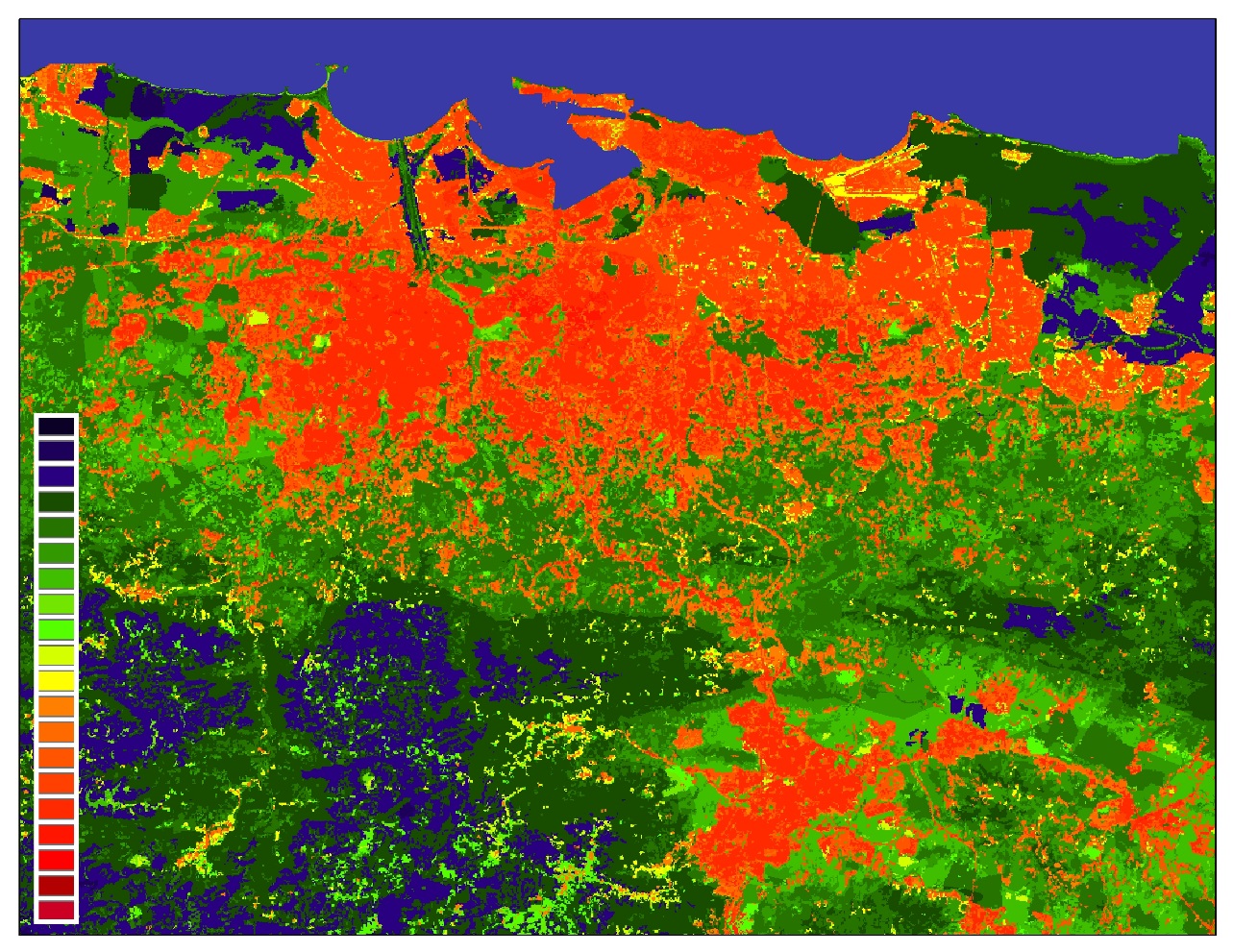


*Figure 3*: Timeline of CDFC in Puerto Rico from January 2009 – December 2013. Epidemics occurred in 2010, 2012, and 2013.



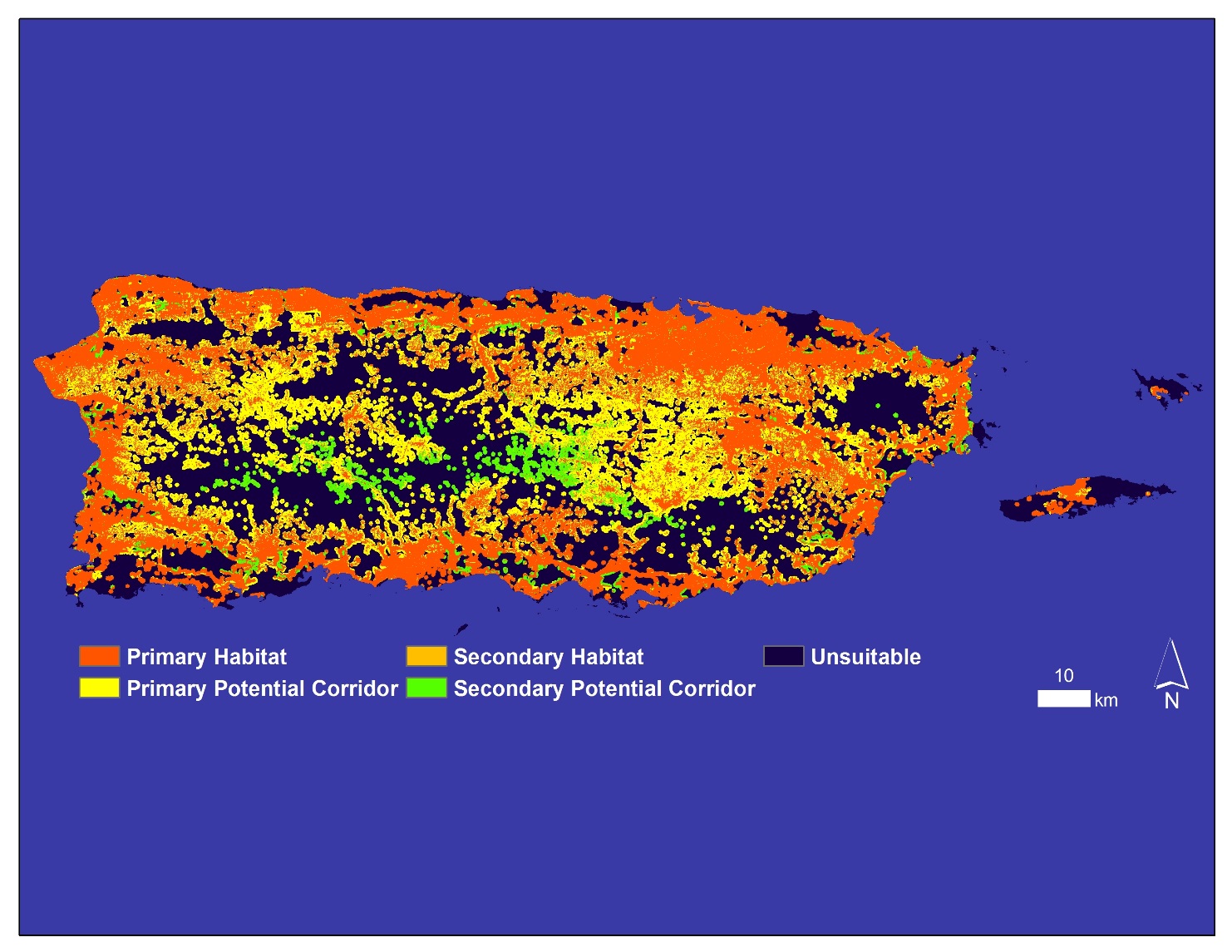
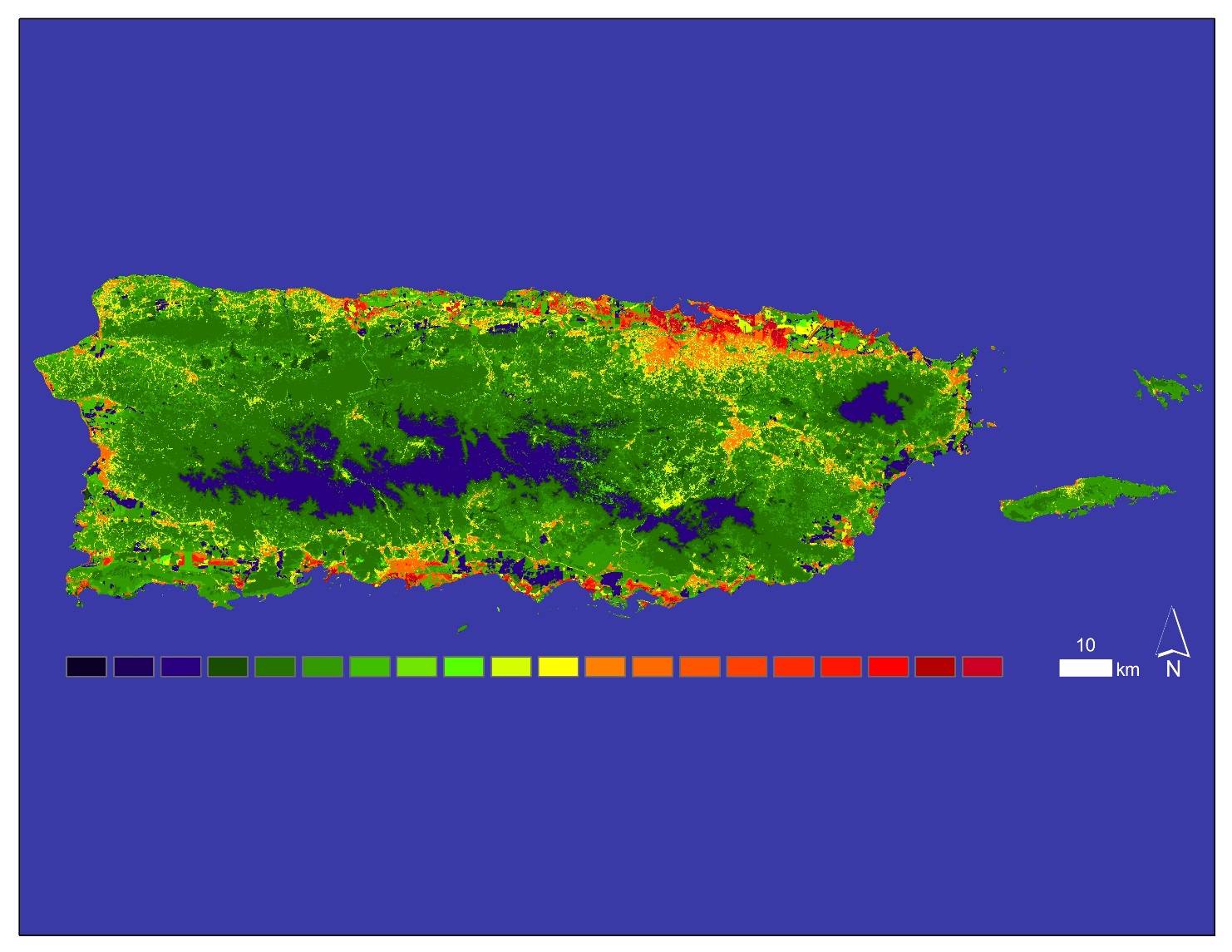
*Figure 5*: Map showing the locations all CDFC that occurred during August 2010, the month with the highest number of CDFC during the study period. The San Juan metropolitan area is highlighted.

*Figure 4*: Map showing the locations all CDFC that occurred in Puerto Rico from January 2009 – December 2013.



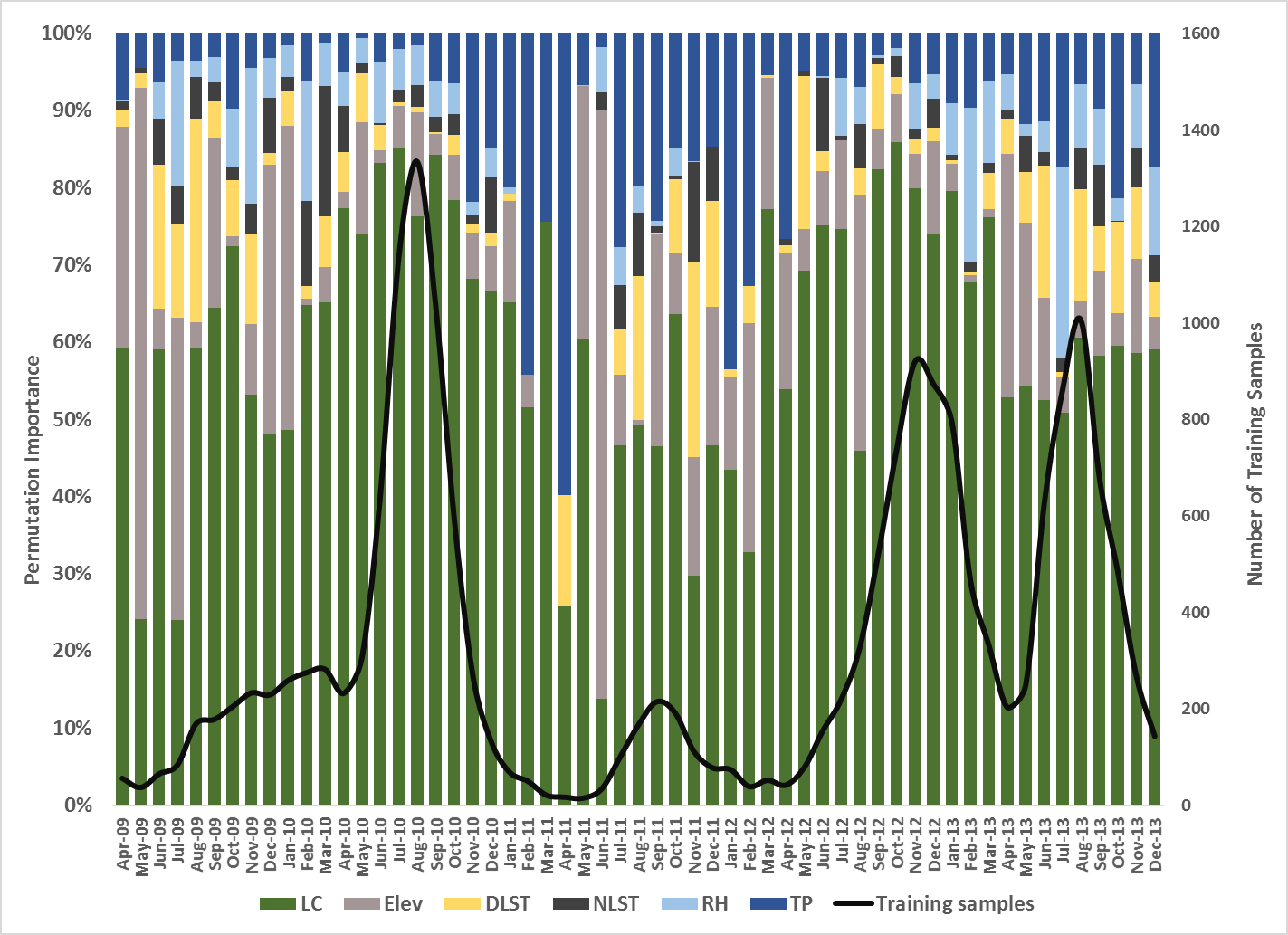
*Figure 6*: DF risk assessment map produced by the MaxEnt model for August 2010, the month with the highest number of CDFC during the study period. Warmer colors represent regions of higher DF risk. The San Juan metropolitan area is highlighted.

*Figure 7*: Close up view of the San Juan metropolitan area from the August 2010 DF risk assessment map produced by the MaxEnt model. Warmer colors represent regions of higher DF risk.

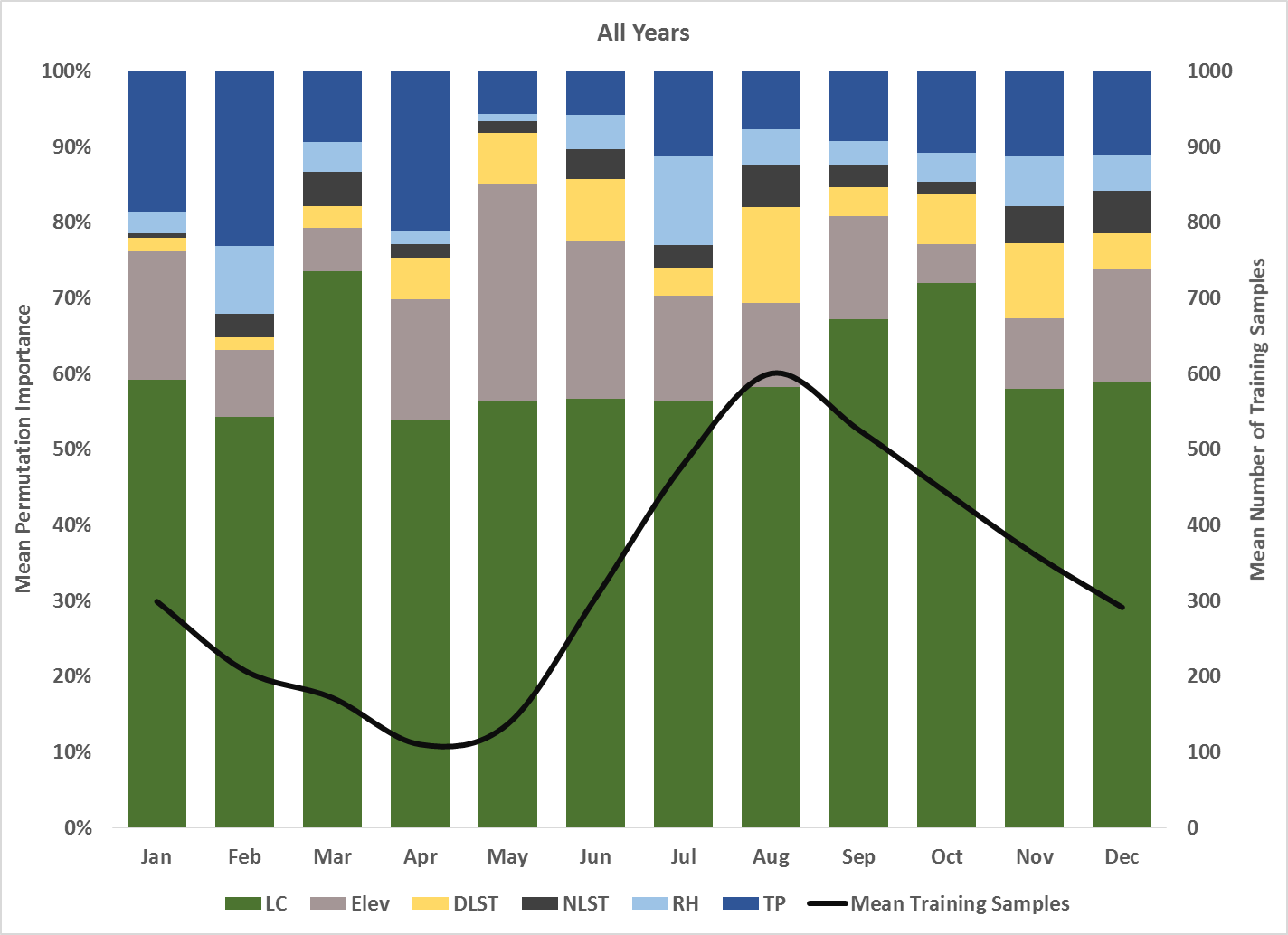


*Figure 9*: Habitat assessment map for DENV infected *A. aegypti* produced by TerrSet’s Habitat Suitability Module within HBM using map inputs of LC and habitat suitability.

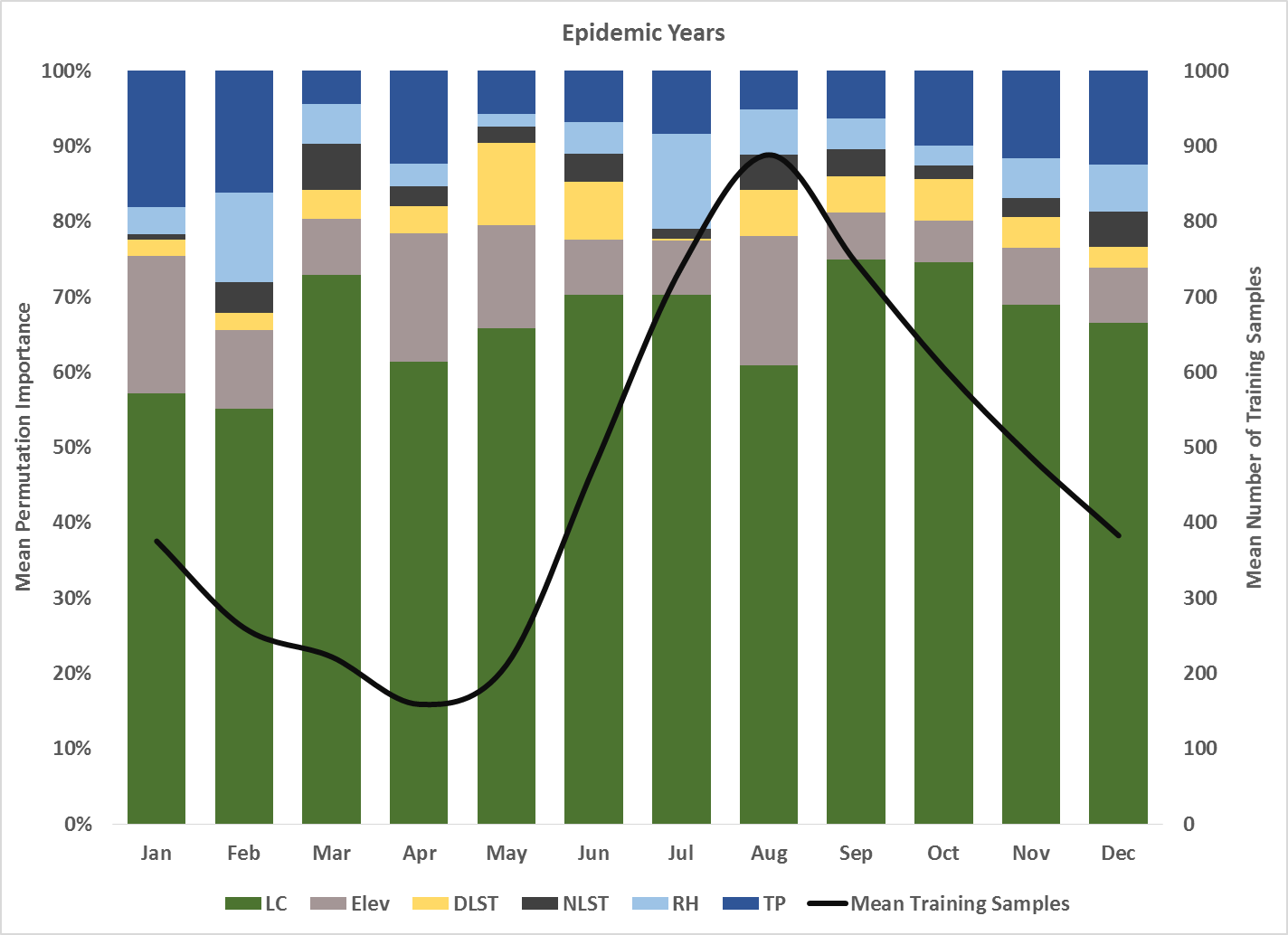
*Figure 8*: Habitat suitability map for DENV infected *A. aegypti* produced by the MaxEnt model using inputs of all CDFC from January 2009 – December 2013, as well as LC and elevation. Warmer colors represent regions of higher habitat suitability.



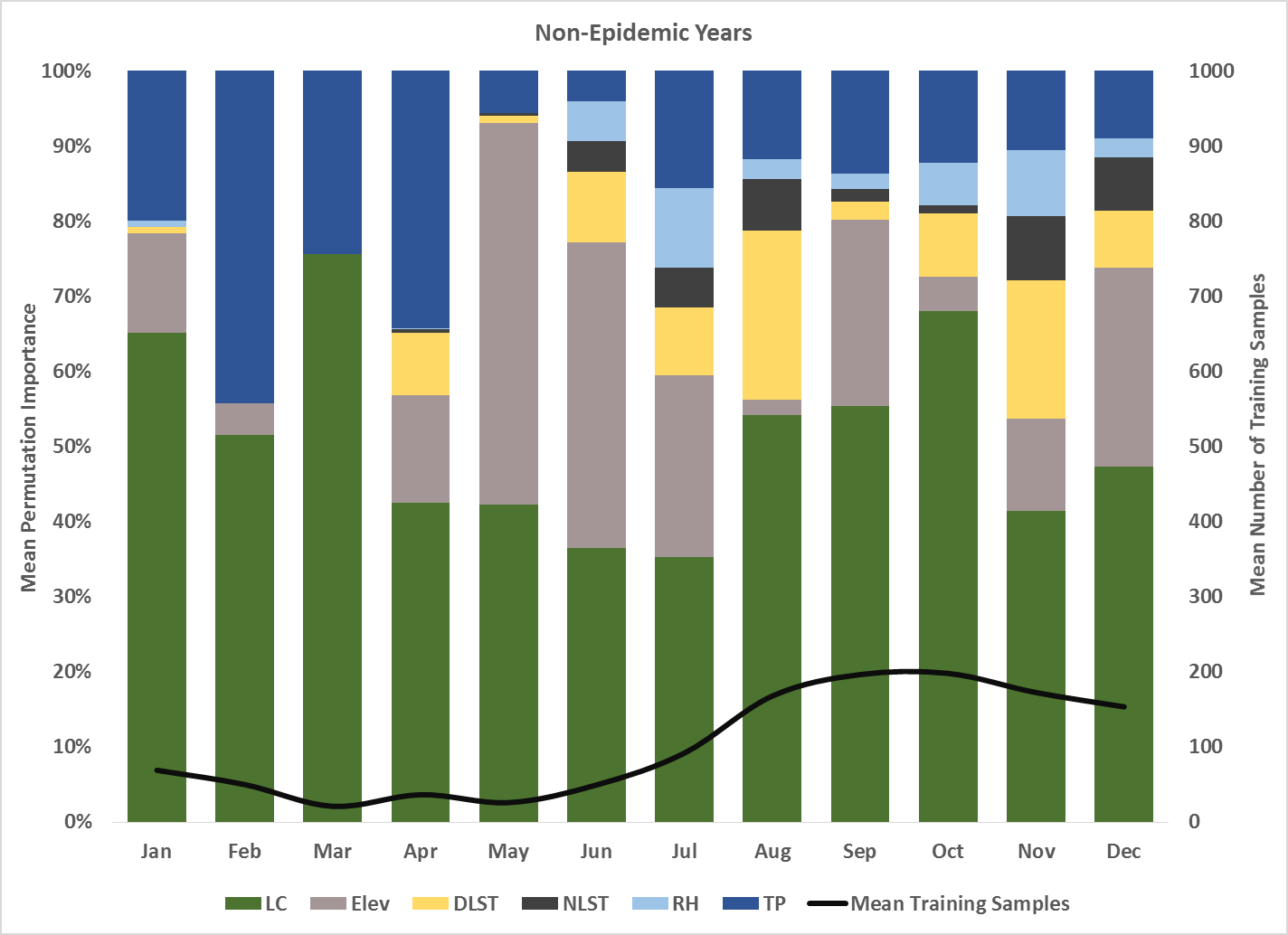
*Figure 10*: Timeline of MaxEnt output of permutation importance values for all environmental variables in Puerto Rico from January 2009 – December 2013.



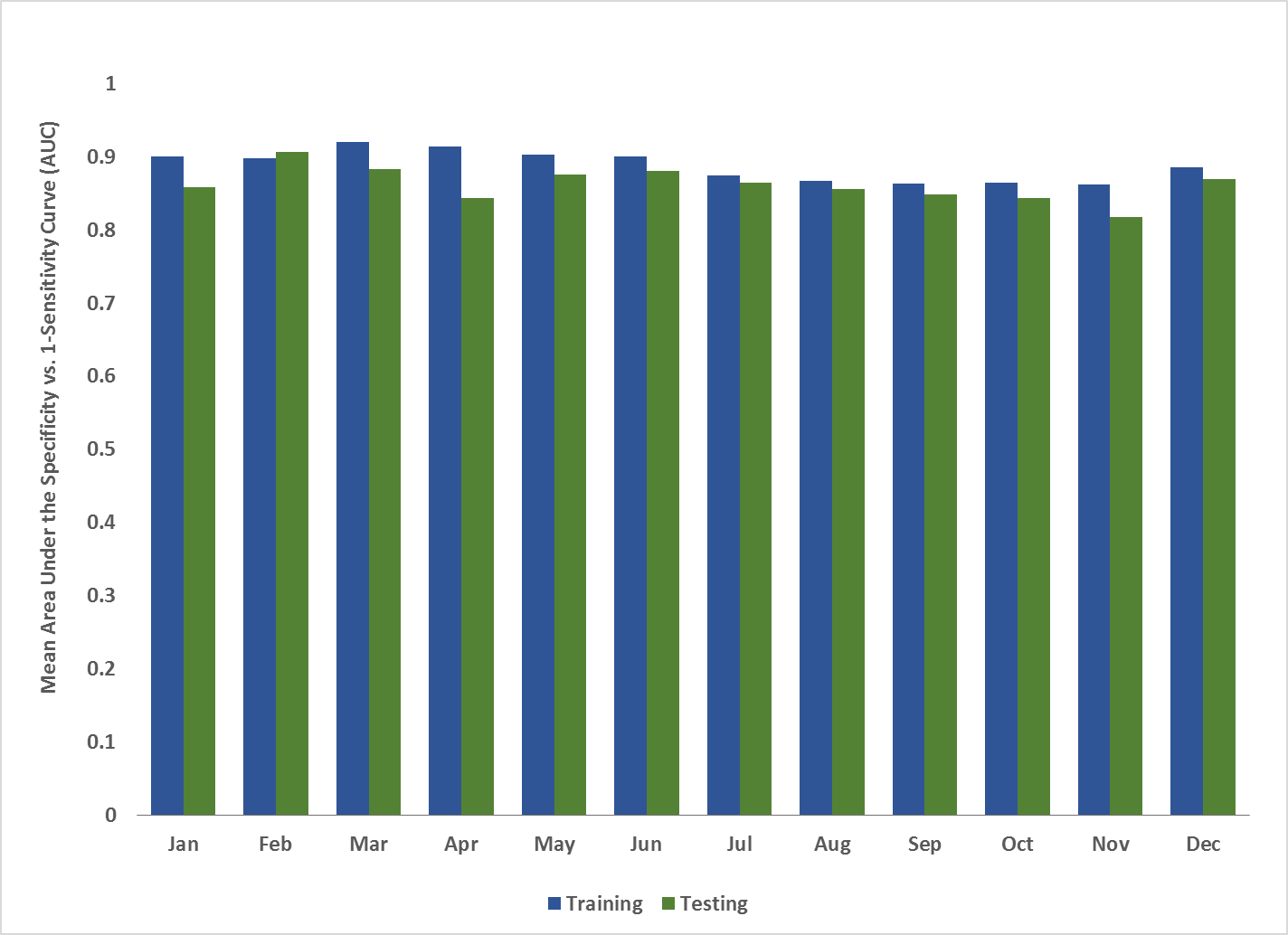
*Figure 11*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico from January 2009 – December 2013.



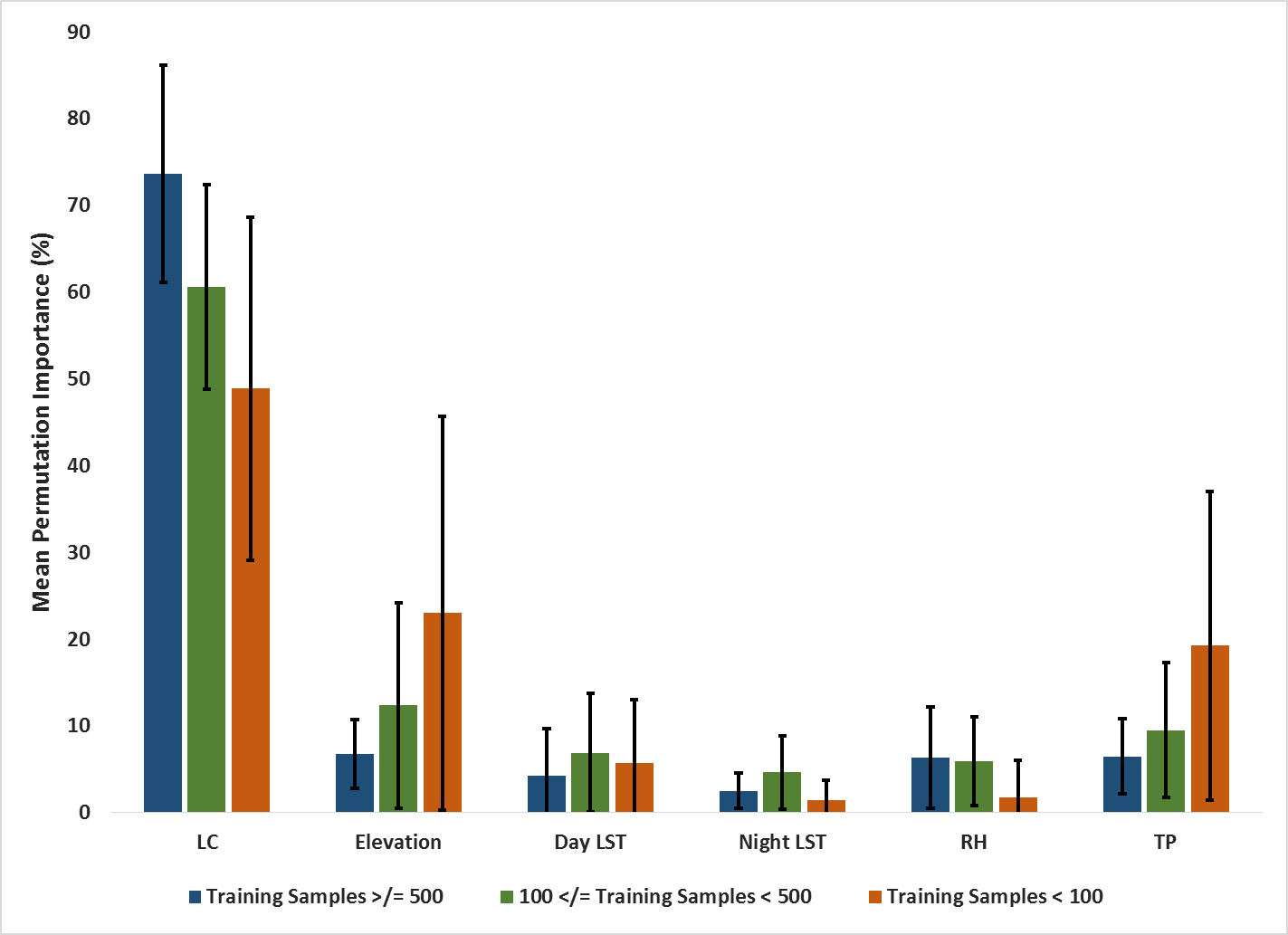
*Figure 12*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico during the epidemic years of 2010, 2012, and 2013.



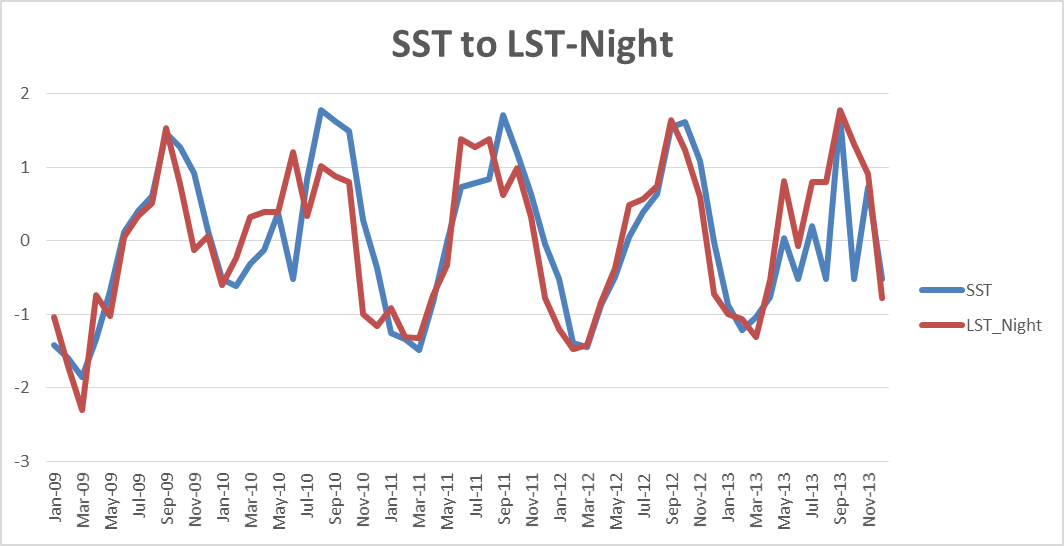
*Figure 13*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico during the non-epidemic years of 2009 and 2011.



*Figure 14*: Comparison of mean monthly training vs. testing AUC values determined from the Specificity vs. 1-Sensitivity Curve.



*Figure 15*: Comparison of mean permutation importance values for MaxEnt model runs with the number of training samples < 100, 100≤TS<500, and ≥ 500.



*Figure 10*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico from January 2009 – December 2013.

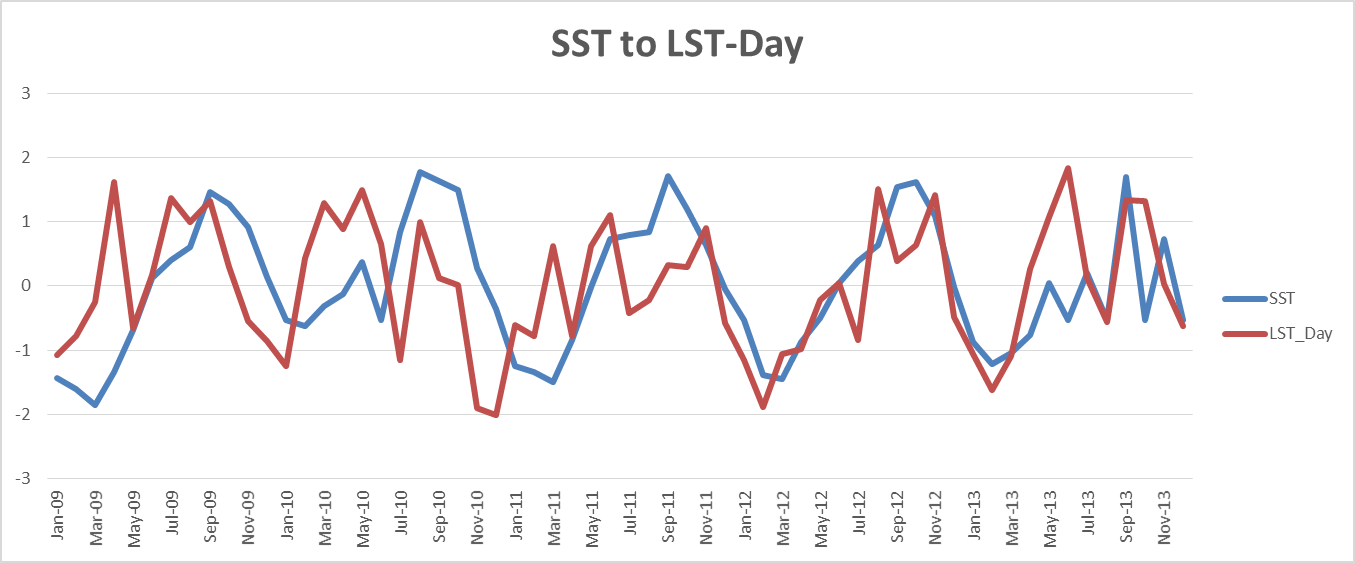
*Figure 10*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico from January 2009 – December 2013.

*Figure 12*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico during the epidemic years of 2010, 2012, and 2013.

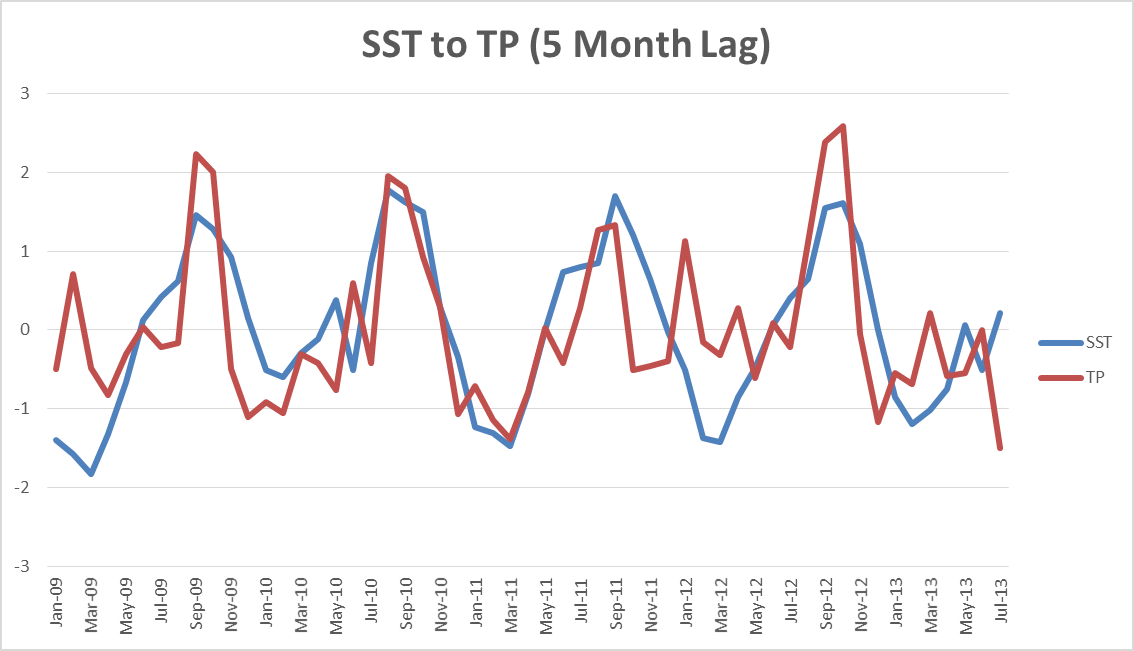
*Figure 10*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico from January 2009 – December 2013.

*Figure 10*: Timeline of MaxEnt output of mean monthly permutation importance values for all environmental variables in Puerto Rico from January 2009 – December 2013.

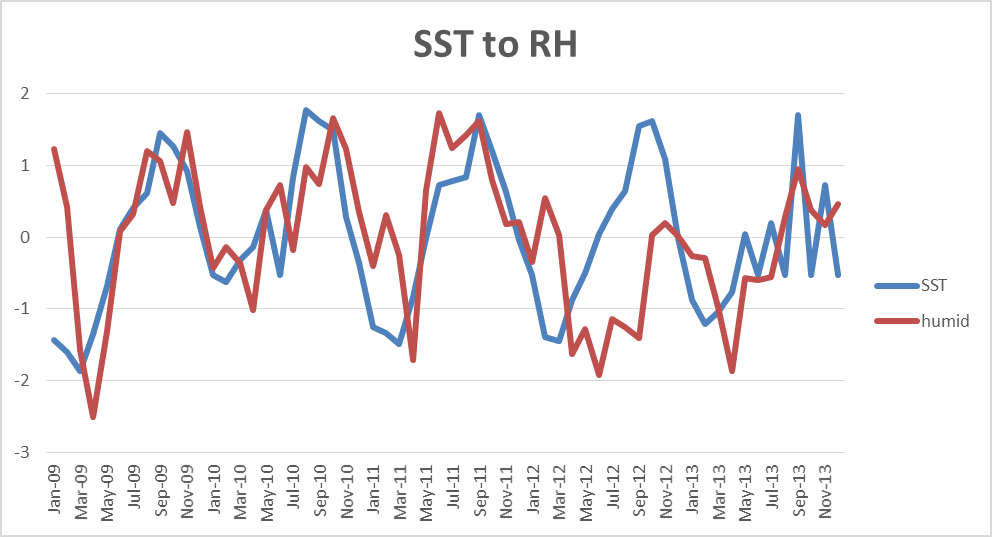
*Figure 16*: SST to NLST correlation resulted in an R value of 0.82



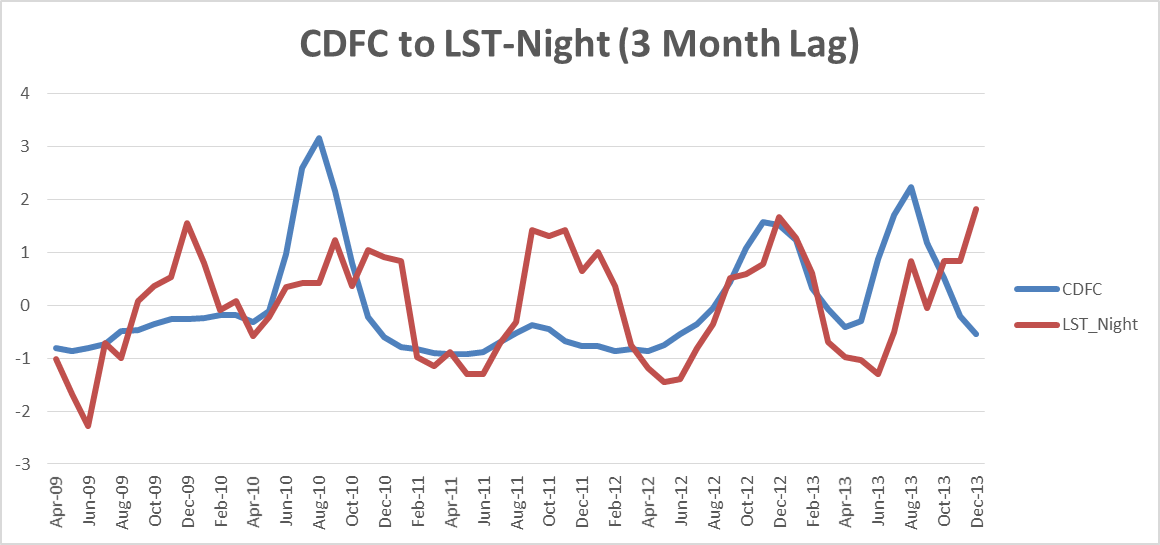
*Figure 17:* SST to DLST correlation resulted in an R value of 0.40



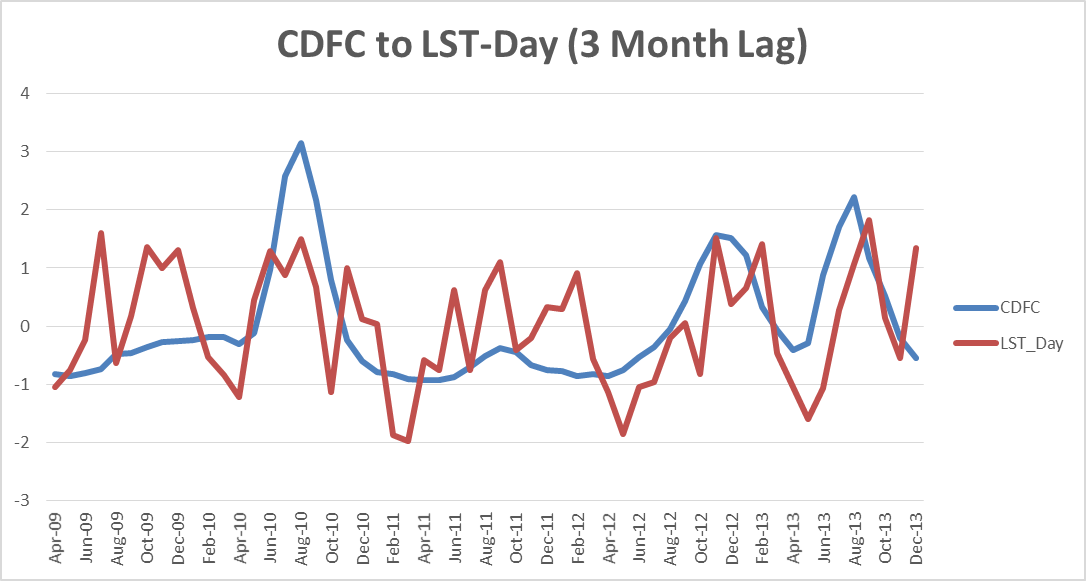
*Figure 18:* SST to TP correlation resulted in an R value of 0.60 with a 5 month lag applied.



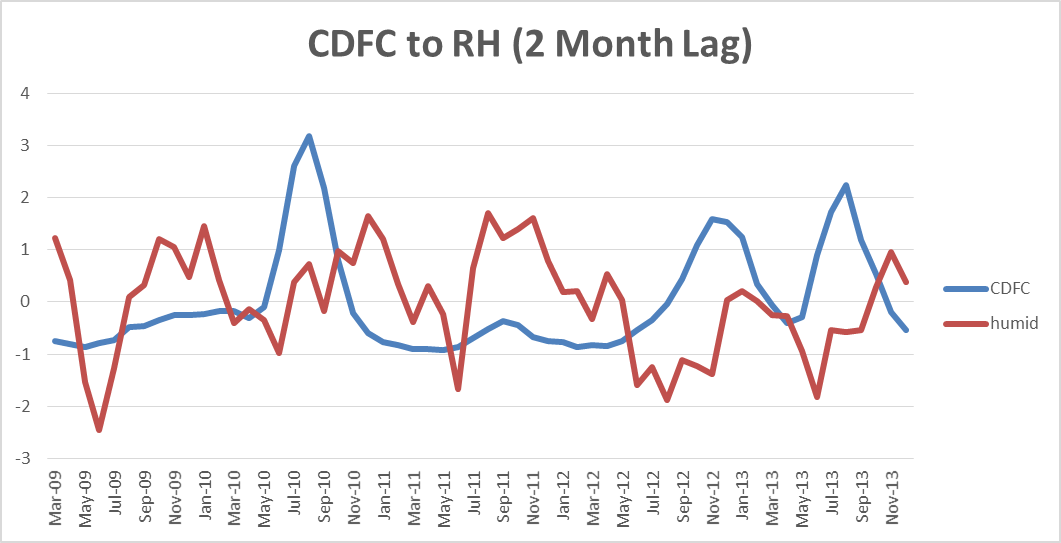
*Figure 19*: SST to RH correlation resulted in an R value of 0.45



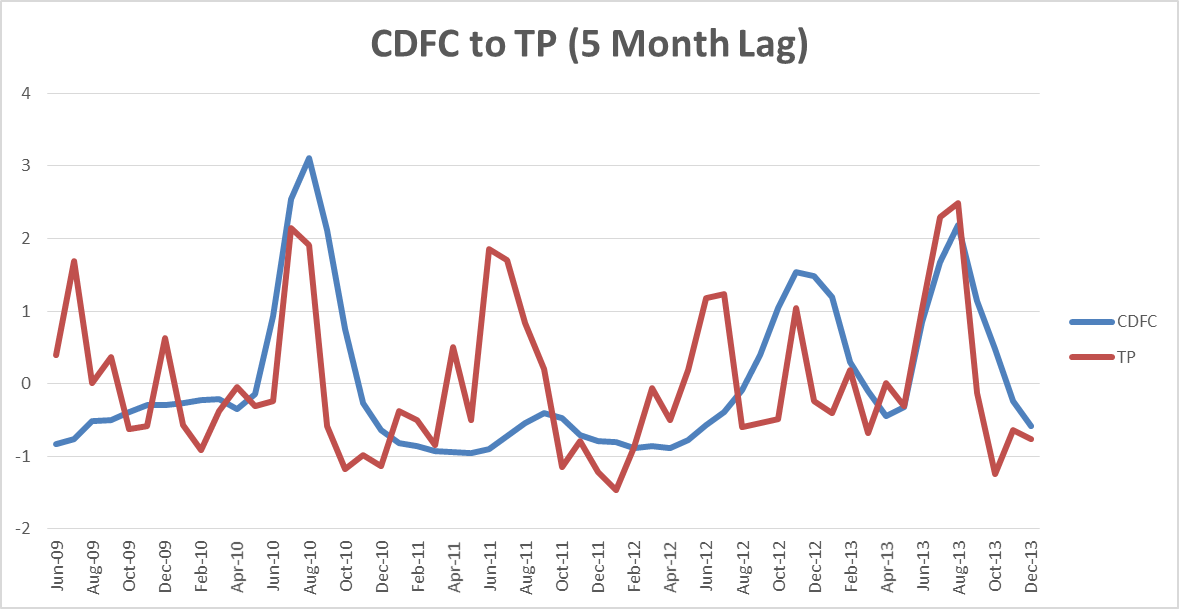
*Figure 20:* CDFC to NLST correlation resulted in an R value of 0.39 with a 3 month lag applied.



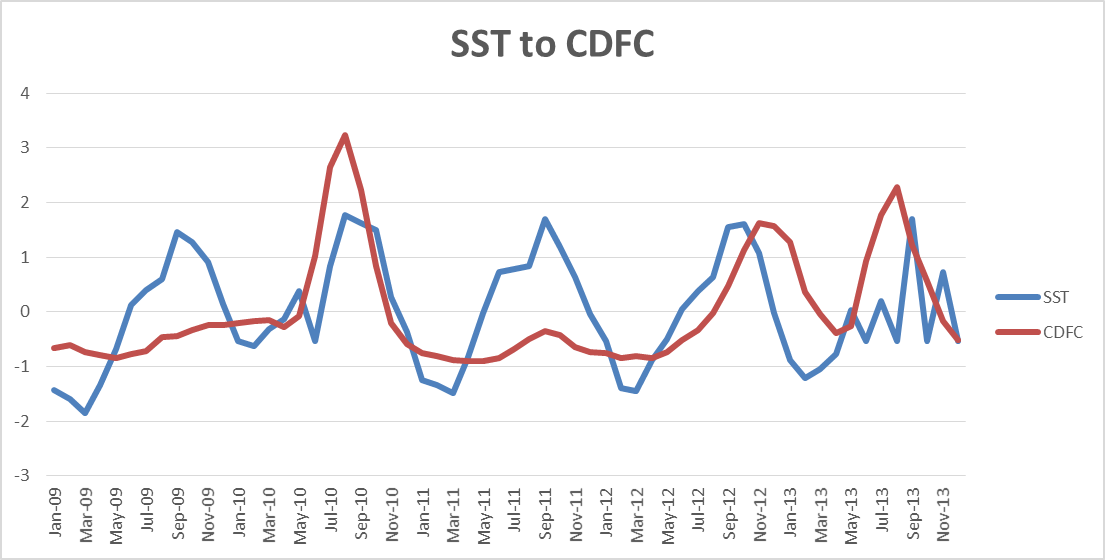
*Figure 21:* CDFC to DLST correlation resulted in an R value of 0.41 with a 3 month lag applied.



*Figure 22:* CDFC to RH correlation resulted in an R value of -0.13 with a 2 month lag applied.



*Figure 23:* SST to TP correlation resulted in an R value of 0.38 with a 5 month lag applied.



*Figure 24:* SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.

Figure 24: SST to CDFC correlation resulted in an R value of 0.41.