|  |  |
| --- | --- |
|  | Likely Fruit Bat Habitats Compared to Ebolavirus Clustering |
|  |  |
|  | Tyler Gates  GEOG:4150  12/17/19 |

Table of Contents

1. Abstract …………………………………………………………… 2
2. Introduction …………………………………………………………… 3
3. Data and Methods …………………………………………………………… 4
4. Results & Discussion …………………………………………………………… 6
5. Figures/Tables …………………………………………………………… 8
6. References …………………………………………………………… 22

Abstract

Investigating sources which cause the spread of Ebolavirus in west Africa plays a key role in predicting the spread of and expanding knowledge of the virus in local communities. It is fairly difficult to pinpoint alternative causes outside of human bodily fluids (*Center for Disease Control, 2018*), but there is a high likelihood that fruit bats in this region contribute to the spread of this disease. The purpose of this study is to explore the possible distribution of fruit bats among Guinea, Liberia and Sierra Leone where the highest amount of ebolavirus clustering is estimated to have occurred in the 2014 outbreak. SatScan software, paired with ArcMap’s hot spot detection, will be implemented in order to estimate areas at high risk for the disease. This will be compared to and against the results of a habitat suitability model created by the MaxEnt software, an open source software created for the modeling of species distribution. The comparison between the 2 results suggested that there was, in fact, a relationship between the *pteropodidae* fruit bat family and spatial clustering of ebolavirus disease risk in western Africa. These findings may be useful for the tracking of bats which may be able to carry and spread this virus, as well as help communities in the areas of interest to better prepare themselves for future events or prevent them from happening altogether.

Introduction

The Ebola virus is one which causes a viral hemorrhagic fever and is believed to have originated in the central African region (*Feldmann, 2011*). In 2014, this virus had been contracted by over an estimated 15,000 people in Liberia alone, and continues to affect thousands of people in the region every year (*Humanitarian Data Exchange, 2019*). Ebola has an extremely high mortality rate among both humans and nonhuman primates at about 70% in countries with little-to-no access to advanced healthcare options. This is in part due to the difficulty of diagnosing this disease; the symptoms which present themselves early on in the disease’s life cycle are very similar to those of other health issues which are common in these areas (*Feldmann, 2011*).

The Ebolavirus has been reported to be eradicated in Liberia numerous times, but seems to resurface at random times, causing more outbreaks in previously “clean” areas (*Doucleff, 2016*). According to multiple sources (*Doucleff 2016, Allela et al. 2005*), this resurgence in disease is largely due to the fact that the Ebolavirus can remain dormant in male sperm cells even after a cure has been administered to a carrier. The immune system does not reach this area of the body when searching for the virus, so even if a previous carrier has been cured, having unprotected sex can transmit the virus to other hosts causing the disease to spread. While this is thought to be a large player in the spread of Ebola, this disease can be spread through numerous other carriers. The most common being human bodily fluids and blood, followed by fruit bats and nonhuman primates which have been infected (*Center for Disease Control, 2018*). The carriers can be dead or alive, as the disease will still live on in mammals which have passed and continue to excrete bodily fluids. Fruit bats – the Greater Long-Fingered Bat in particular - are thought to be the primary carrier of the virus. However, due to the scarcity of these animals in the wild, it has been difficult for biologists to prove that this is true (*ColumbiaMSPH, 2019*).

This study aims to assist in bridging the gap between the unknown and proof of bats being a contributor to the spread of ebolavirus in western Africa. There have been findings that fruit bats are able to carry the virus, but biologists and ecologists have had trouble finding the epidemic species, responsible for the outbreak, carried by a single bat (*Hayman et. al, 2012*). The first such case in west Africa to be found in a fruit bat was done so earlier this year by Simon J. Anthony and his team, and was found in a region of Liberia (*Grady, 2019*). With this knowledge, we hypothesize that the ebolavirus should appear to cluster in the West African region. We also hypothesize that the clusters will of virus occurrences will overlap with estimated bat habitats.

Data and Methods

A close up of a map

Description automatically generated The focus area of this study includes the countries hit hardest by the 2014 Ebola outbreak in West Africa; Guinea, Liberia and Sierra Leone. The shapefiles, 2014 population data and ebolavirus occurrence data from the Humanitarian Data portal. This information has a resolution at the district level, totaling 64 across the 3 countries. The Ebola data is collected weekly, designating the number of newly-confirmed cases which occur that week. This data is then aggregated by month and year for 2014 using pivot tables and normalized by population. Fruit bat occurrence data was taken from the Global Biodiversity Information Facility, and includes a total of 256 point data representing human sightings of the *Pteropodidae* fruit bat family across the west African region. These sightings range temporally from 2000 to 2014. This time period was chosen due to the rise in global temperature in this time, which may have caused bat habitats to change in location (*Lindsey, 2019*). There were also multiple environmental variable rasters which were used throughout this study which cover the entirety of West Africa, including categorical land use (*Tappan, 2016*), elevation (*DataBasin, 1996*), average yearly precipitation (*Matsuura, 2018*) and average global temperature (*Fick, 2017*).

Figure 1: Relevant Study Area

We believed there to be grouping in the outbreak of the ebolavirus throughout the study area, and wanted to deduce where this clustering might be taking place. A Moran’s I test was used to examine our data for spatial autocorrelation. This test will assess whether or not the statistical assumption of randomly distributed data is true (*Detect Spatial Autocorrelation, 2019*). The closer the number is to 0, the more random the data appears to be. If the p-value of this test is <.005, the null hypothesis of perfect randomness can be rejected. This test was ran with the definition of neighbors as Inverse Distance weighting, which decreases the influence of a location’s value as distance grows (*Shepard, 1968*). This was chosen due to the fact that the most common forms of transportation in the study region are walking and bush taxis, which do not allow for very long travel distances among the population (*Njoh, 2009*). Row standardization was also implemented, which is simply a technique used to adjust spatial weights in a dataset. According to Figure 6, the data was not spatially random, and clustering is probable in the dataset.

Clustering was assumed in our data after the autocorrelation test, and a hot spot analysis was performed. Two different tests were conducted; the first being the Getis-Ord Gi\* test through ArcMap, and the second being done through the SatScan Cluster Analysis. This was done to ensure that there were no errors in the datasets, as well as to compare the outputs of different algorithms. The Getis-Ord Gi\* algorithm will calculate the area in which clusters of high values or low values exist within the data (*Murack, 29*). The clusters only appear if they are significant, which is determined by the region’s values compared to those around it. The neighbor parameter of this test was also defined as Inverse Distance. The SatScan software was also used to search for hot spots, and returned similar results. In a purely-spatial analysis, which was specified for this test, the software uses a circular window which grows and moves around the area in order to scan for clusters (*Kulldorff,* *6*). The model is set to output only clusters with less than 50% of the population being at risk, preventing the entire study area from becoming one large, meaningless cluster. Clusters cannot appear at the center of other clusters, so that there will not be districts singled out when they should be part of a larger grouping. This model used Poisson statistics, which are typically used to find probabilities of events which are likely to be rare within a certain bound of area (*Routledge*). This model outputs all clusters which are found, along with their log-likelihood ratio, p-values, relative risk score and observed. These values define how likely a cluster is to occur if an event does happen versus if it does not, the statistical significance of the clusters, risk inside of the cluster versus outside of the cluster and the number of cases which occur inside of the cluster, respectively. Each cluster is created in a separate shapefile which contains this data, in order for it to be displayed cartographically. These were then overlaid on one another with different levels of transparency, colors and patterns in order to differentiate between them and indicate where the highest levels of overlap occur. These SatScan results were then compared with the Getis-Ord Gi\* results. The most accurate and complete results, which SatScan happened to give, were compared to a fruit bat habitat analysis,

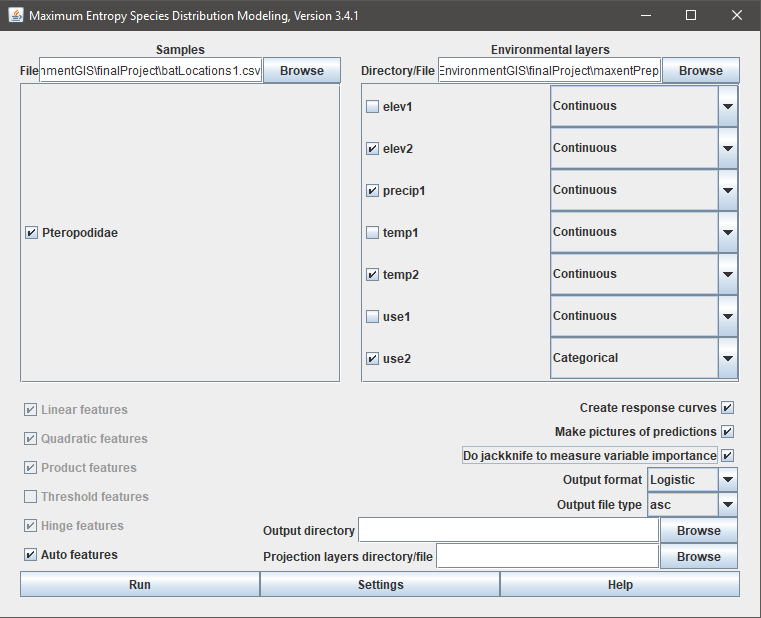
 Recorded fruit bat sightings and habitats are difficult to study, as they are few are far between in this region of Africa. According to the GBIF bat sighting data, there were only around 84 recorded sightings across the entire study area between the years of 2000 and 2014; this is why the bats were searched for by family, as it widens the pool of small bats which are able to spread the virus to humans. In order to discover if the clusters found previously are in fact coinciding with ideal bat habitats, a niche analysis needed to be performed on the available data. This model was created using the MaxEnt software, which is a program dedicated to modeling the distribution of species from presence data (*Elith, 2010*). This modeling software takes in known-presence data in the form of a .csv, which contains lat/lon coordinates for each sighting. It then takes in environmental raster files in ASCII format, continuous or categorically defined, and uses these data to estimate a species’ distribution across a space (*Philips & Dudik, 2008*).

Figure 2: MaxEnt Interface

In order for the raster files to be used in the software, they need to have identical geographic extents, projected to the same coordinate systems and have equal pixel resolutions. These steps were completed using ArcMap toolboxes. The raster images were first clipped to the area which bat sightings extended to, after which the smallest extent of the files were found; this happened to be the precipitation raster. All files were resampled to a resolution of (0.0083m, 0.0083m) using the nearest neighbor resampling technique (*Parker, 1983*). All files were reprojected to WGS 1983, and then clipped to the extent of the precipitation dataset in order to maintain an identical extent. These data are then exported as an ASCII filetype to be used in the software. Before running the model, some settings needed to be modified in order to present the least-biased results possible. 25% of the data was selected to be used as training data in order to prevent inflating the measure of our model performance, and used the Bootstrap replicated run type due to a smaller number of occurrence data. The maximum number of iterations was also increased to 1500 in order to allow for the model to reach the convergence threshold successfully. The selected outputs include a raster image of predicted suitable habitats, an analysis of omission and commission, environmental response curves and an analysis of variable contributions. After all relevant models were completed and analyzed for significance, the SatScan cluster model was overlaid atop the MaxEnt raster output to compare locations of importance.

Results

Results of both cluster and hot spot detection from SatScan and ArcMap Getis-Ord Gi\* tests show similar grouping from both algorithms. The SatScan clusters which have higher relative risk scores – 1, 3 and 5 – all match up with the output of the hotspot analysis from ArcMap (Figures 7,8 and 9). The SatScan results were used for comparison against the habitat suitability model due to its completeness and statistical outputs. According to Table 1, the most statistically significant clusters occur in the middle of the study region. This confirms the first hypothesis of the stud, being that there is in fact significant clustering of the ebolavirus throughout the 3 countries of interest. The most likely clusters to occur out of the given list, according to their LLR score, are clusters 1, 2 and 3. There is an elevated risk of the ebolavirus in Port Loko, Western Rural Area, Kailahun and Kerouane.

The outputs of the habitat suitability model concluded that there are estimated fruit bat species distributions in certain areas where significant clusters had occurred. The areas of Lofa, Macenta and Western Area Rural are a part of multiple significant clusters, and contain some of the most suitable habitats for the bats according to the MaxEnt model. This confirms our hypothesis that there are a higher number of likely cases of ebolavirus among the suitable bat habitats in western Africa. Figures 10, 15 and 16 all demonstrate this interpretation of the data in multiple visualizations.

Discussion

There were multiple districts in the northern region of Guinea which had a 0% infection rate, yet did not appear as a cold spot through the Getis-Ord Gi\* test. This was a fairly unexpected result, as the program usually detects areas with lower values such as this. The output may have been like this due to the fact that the normalized ebolavirus rates were extremely low values, which may not be well-supported by this algorithm. The difference between these 0 values and the others was fairly low, which may have also contributed to this finding.

I was also surprised to see that there was a region in southern Liberia in which there were no clusters or hot spots of the ebolavirus, but had a very high probability of appearing in the bat species distribution. This finding suggests that an area being an ideal region for bats to reside does not necessarily imply that this region will also see a higher rate of infection among the population. This does not go against our hypothesis, but does suggest that fruit bats are not a major contribution to the spread of this virus.

There were multiple limitations in this study, with the highest number being in the MaxEnt model. The variables chosen were indicative of what describes an ideal bat habitat, but not all of these rasters contributed a great amount to the calculation of the model. If more time was committed to the finding of higher quality or alternate environmental variables, the model may have ended up being more accurate and better represent the areas in which these bats may be residing. One dataset in particular, the precipitation raster, was a limitation in and of itself. It is an average rainfall amount over the course of nearly a century, which may not be truly representative of what the average rainfall of 2014 may have been. The resolution of multiple rasters also needed to be resampled using nearest neighbor interpolation, which may have caused some values to change.

Figures/Tables

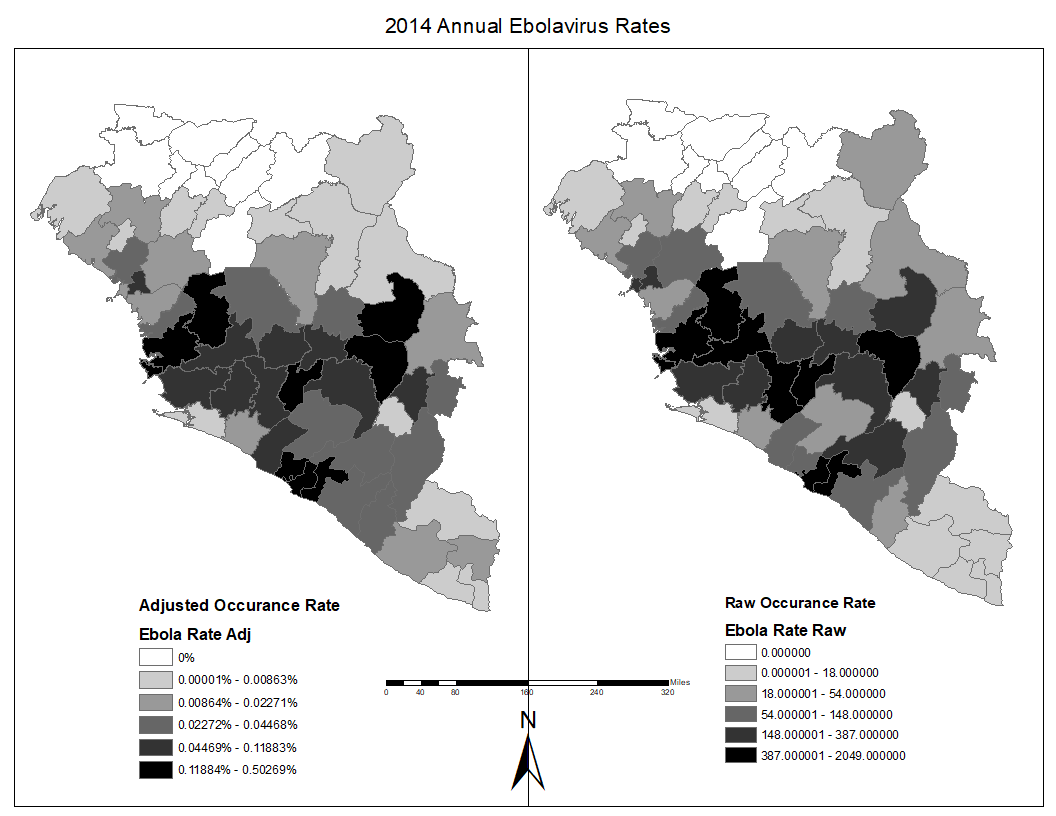


Figure 3: Raw and Adjusted Ebolavirus Occurance Data

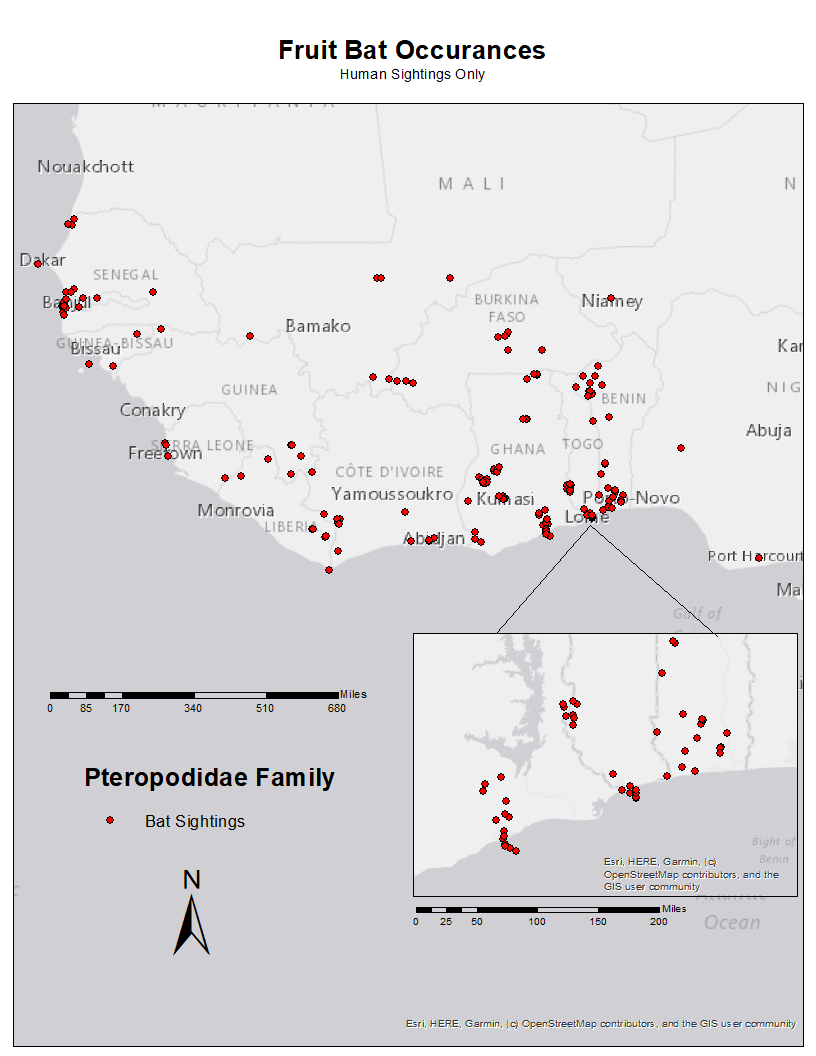


Figure 4: Fruit Bat Occurance Data

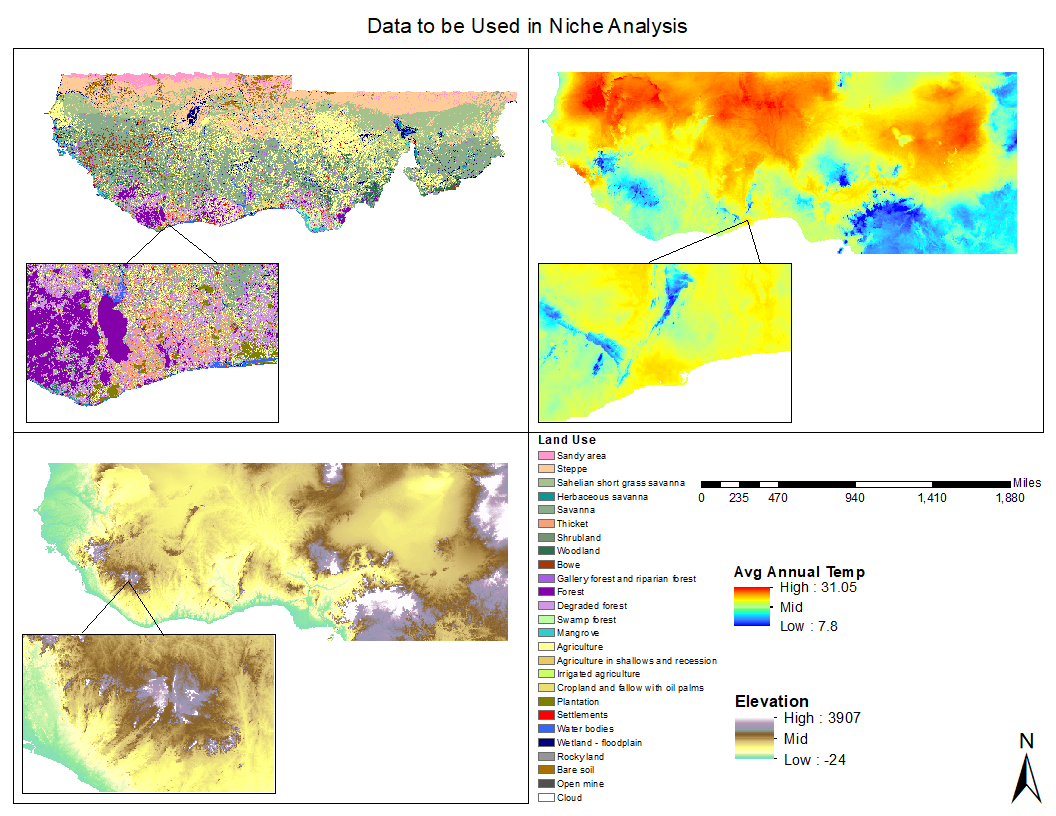


Figure 5: Raster Image Data to be Used in MaxEnt Software

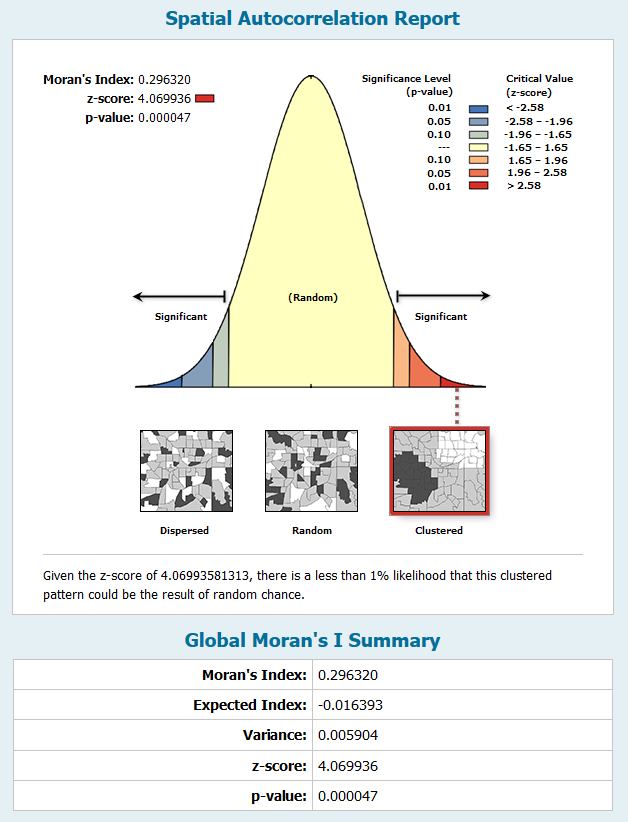


Figure 6: Results of ArcMap Global Moran's I Test. This shows that there is significant non-randomness in the dataset.

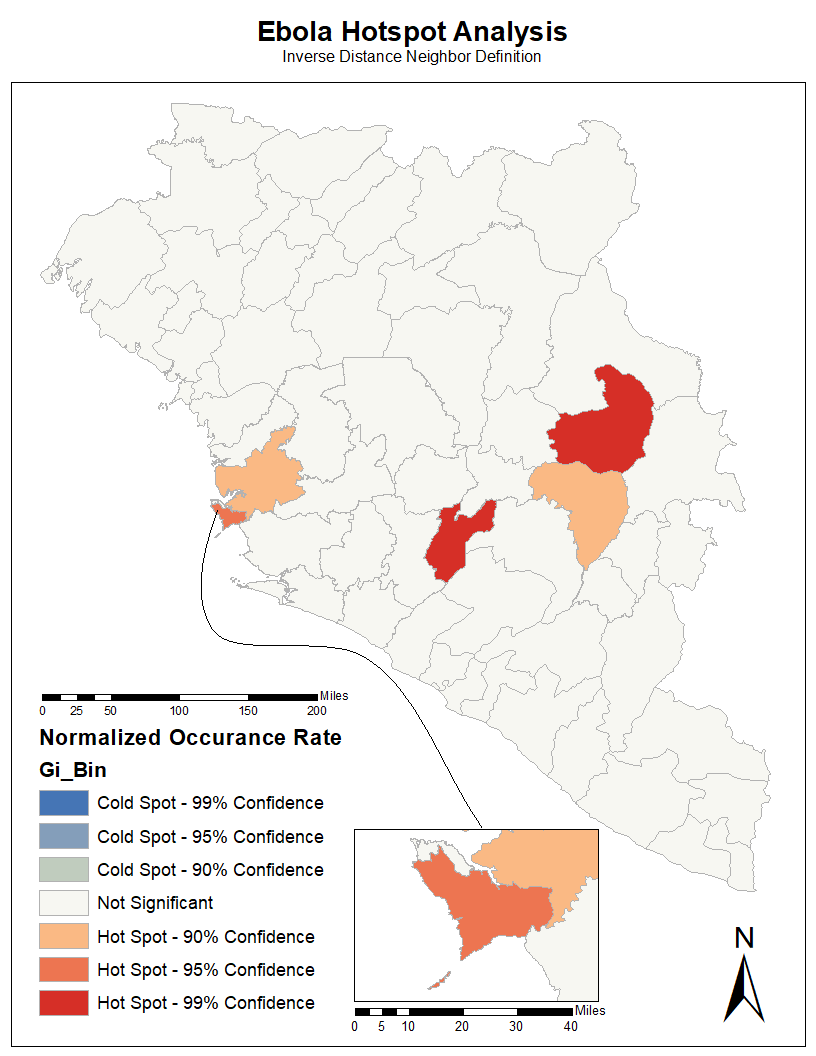


Figure 7: Getis-Ord Gi\* Test from ArcMap software.

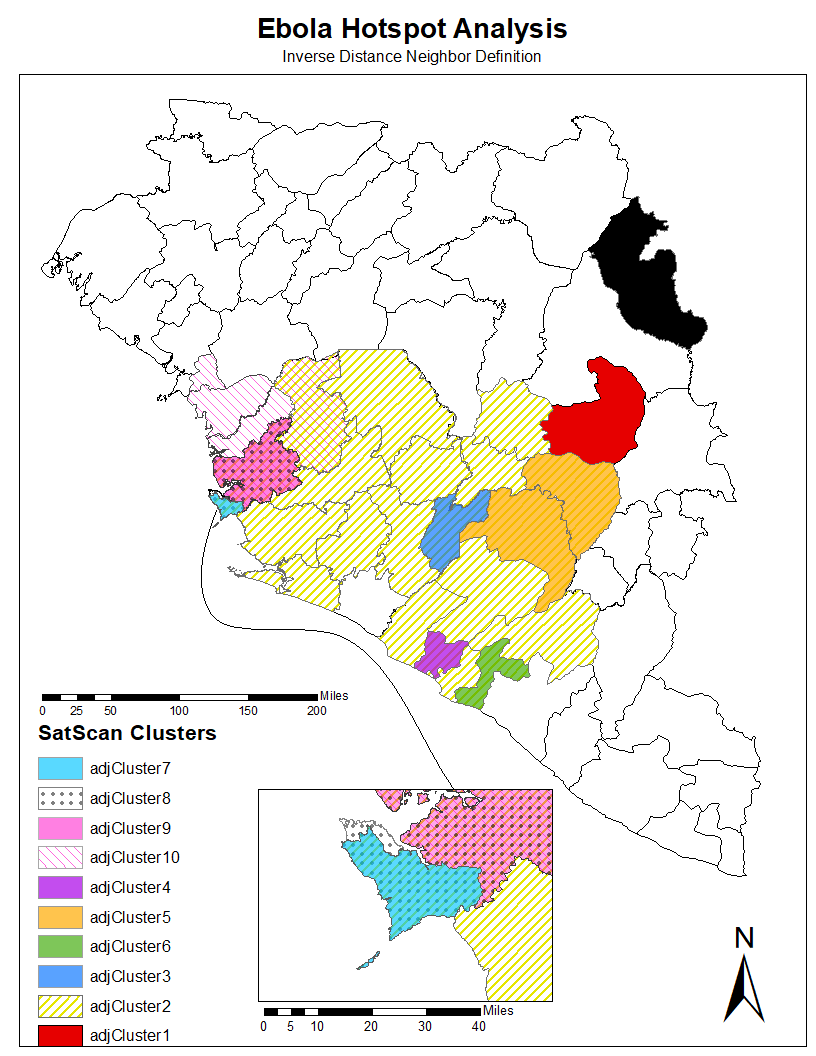


Figure 8: Geographic results of SatScan hotspot detection software.

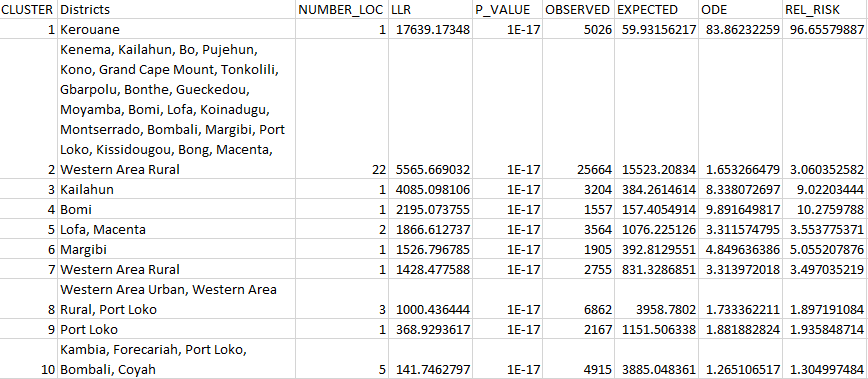


Table 1: SatScan cluster detection statistical output

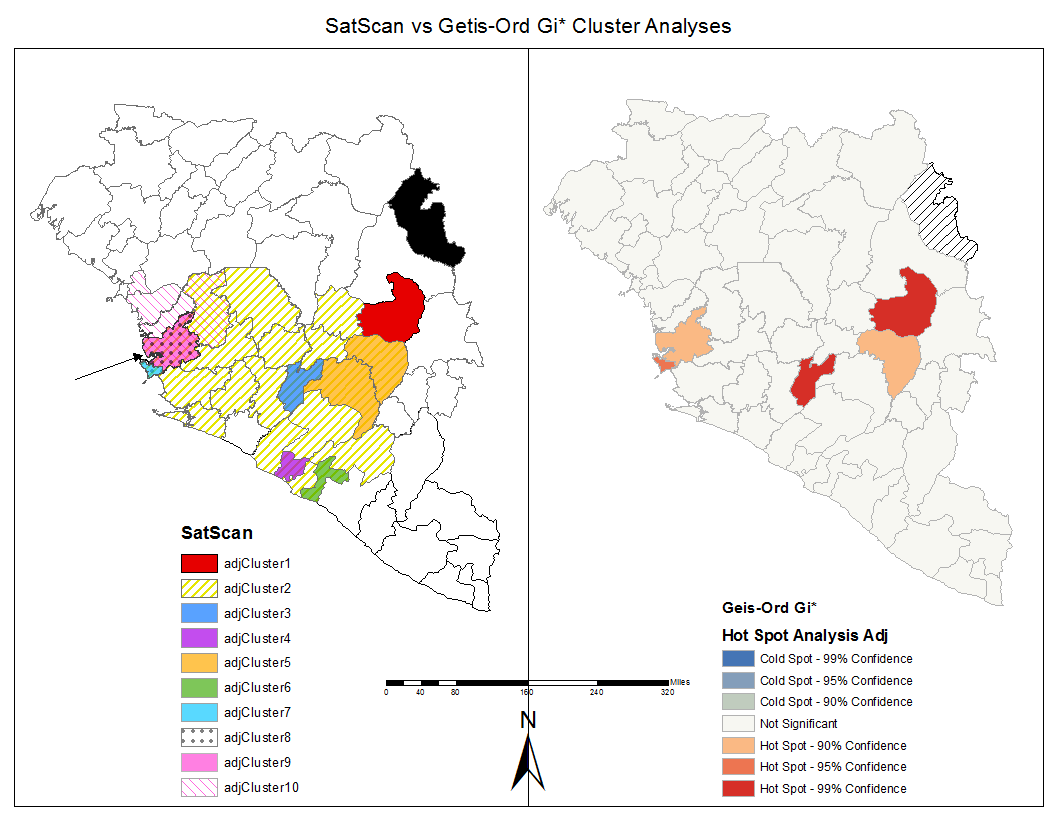


Figure 9: Comparison of Getis-Ord Gi\* and SatScan cluster detection results.

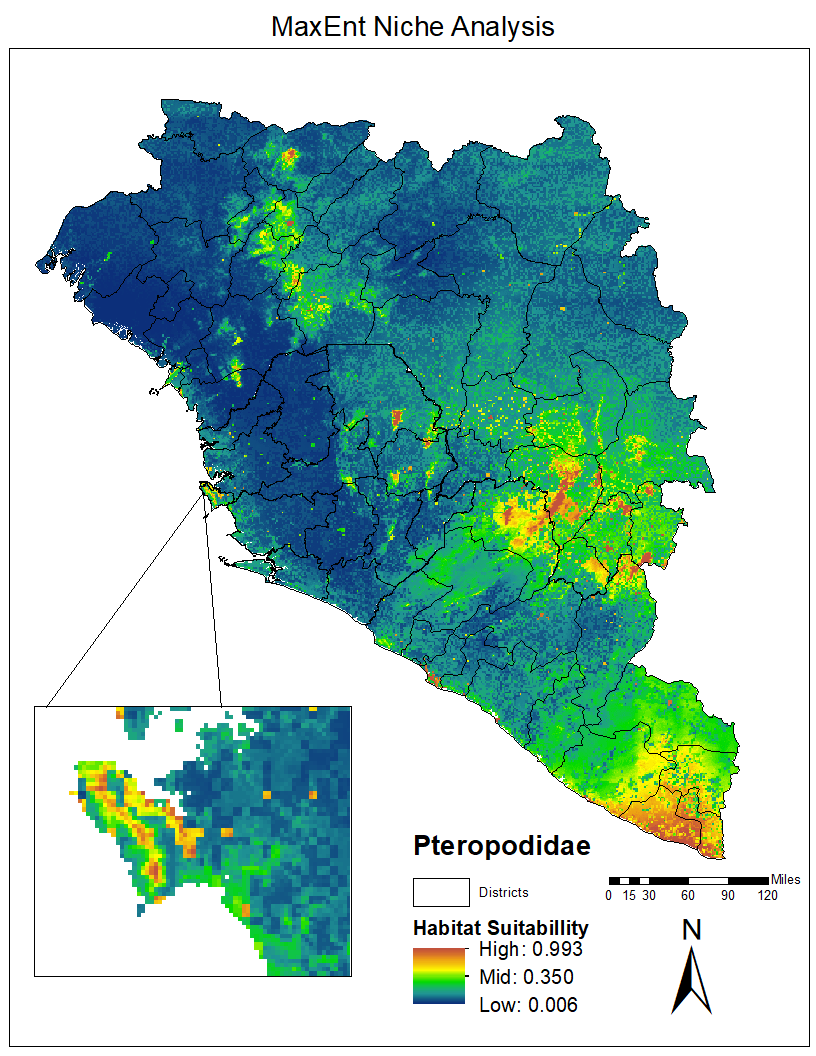


Figure 10: Geographic results of MaxEnt software for probable Pteropodidae bat family habitat.

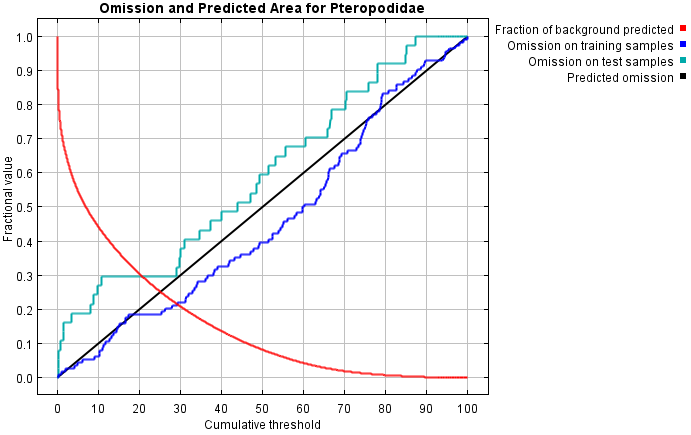


Figure 11: Omission rate and predicted area as a function of the cumulative threshold. The closer the blue and teal lines are to the 1:1 relationship (black line), the better the model fits.

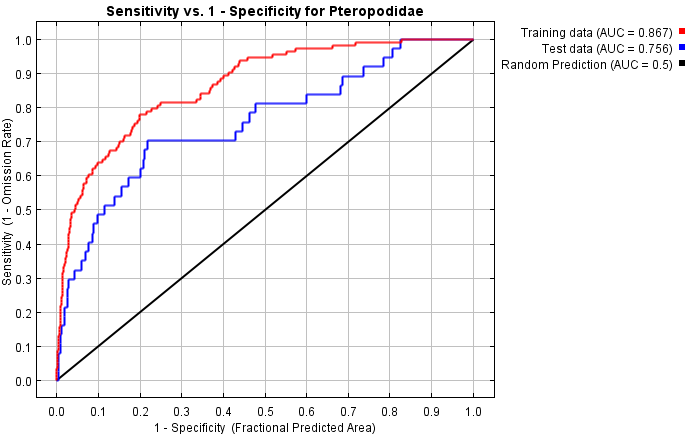


Figure 12: This graph depicts how well the model performs in predicting species occurances compared to a random selection of points. Sensitivity is how often actual occurances are predicted, and specificity is how often absences are predicted. A higher AUC indicates a more reliable model.

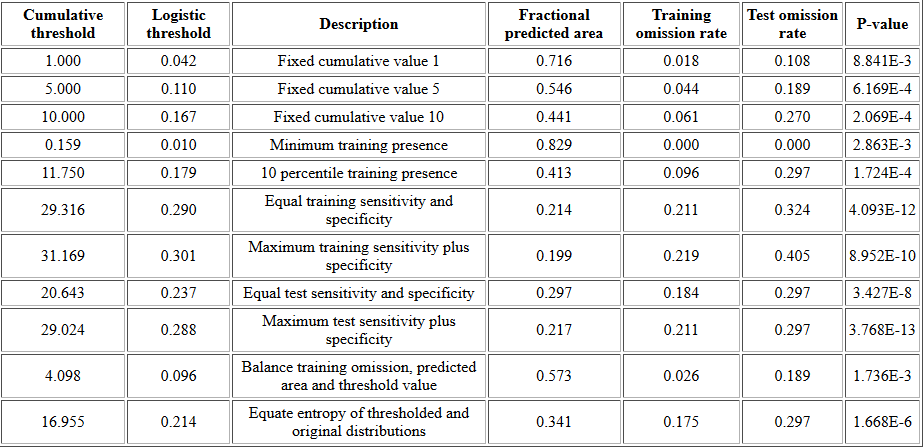


Table 2: Various common thresholds and ommission rates for selecting cutoff values to represent habitat and non-habitat. The p-values all being significant indicate that the model fits the data much better than randomly assigning points in space

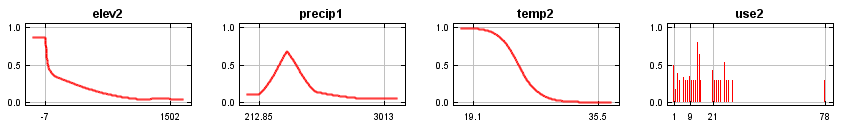
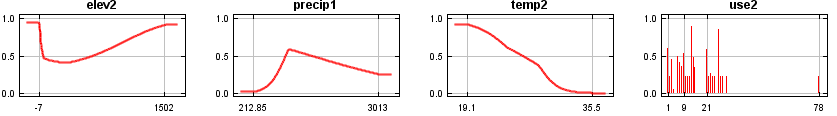
 

Figure 14: These response curves display the probability of species occurrence given values of each predictor variable. The more similar the graphs are to each other, the less correlated the variables are. The top graphs include average values of all other variables, whereas the bottom graphs are looking at the single variable’s impact on the predicted niche.

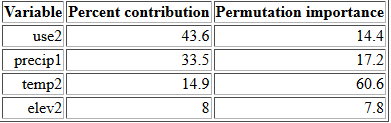
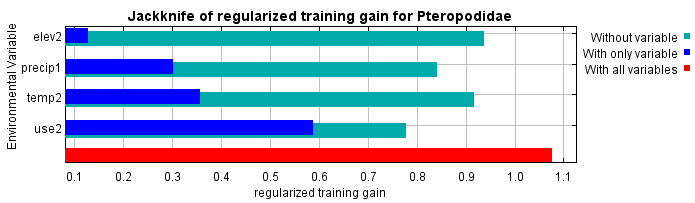
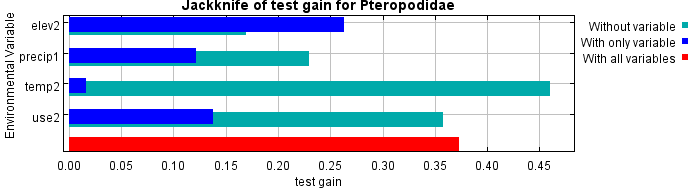


Table 3: Represents estimates of the relative contribution of each variable to the niche model.

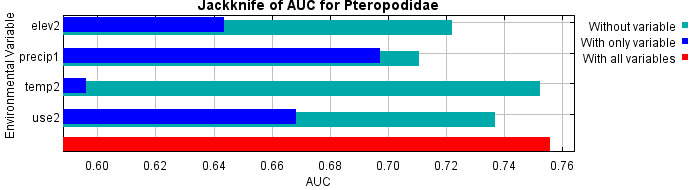


Table 4: Jacknife plots indicating the importance and contribution of each environmental variable to the overall suitability model.

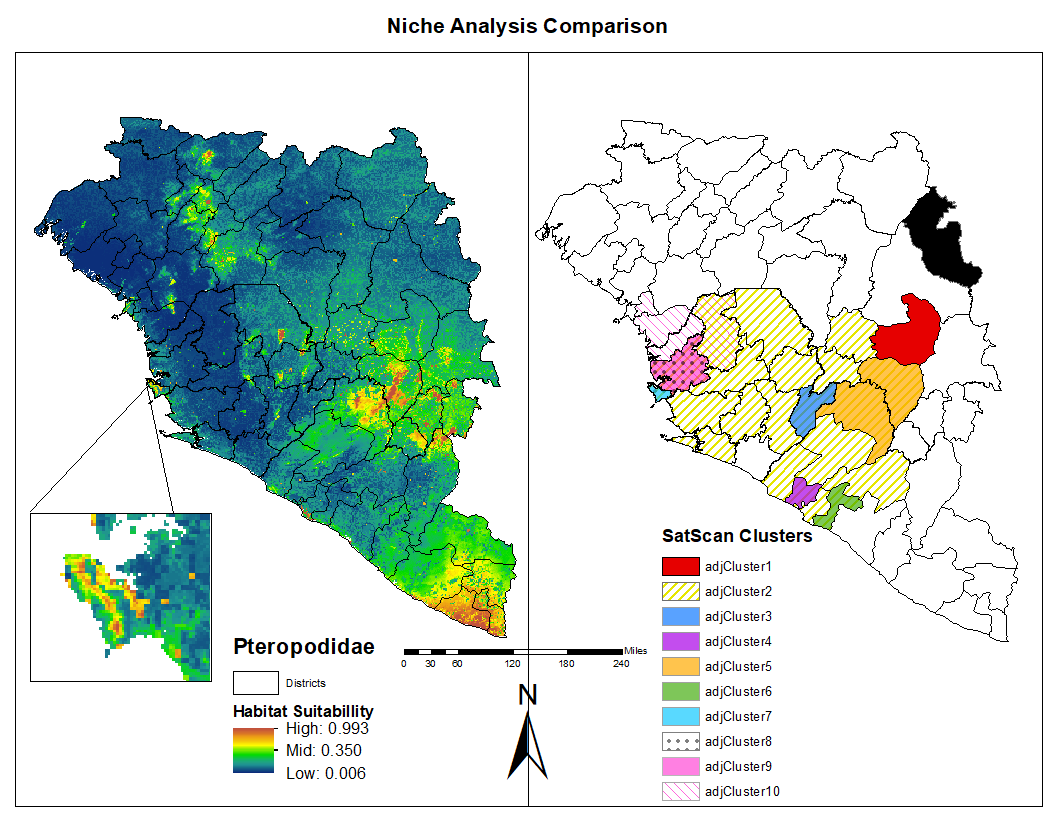


Figure 15: Comparison of MaxEnt habitat suitability and SatScan cluster detection results

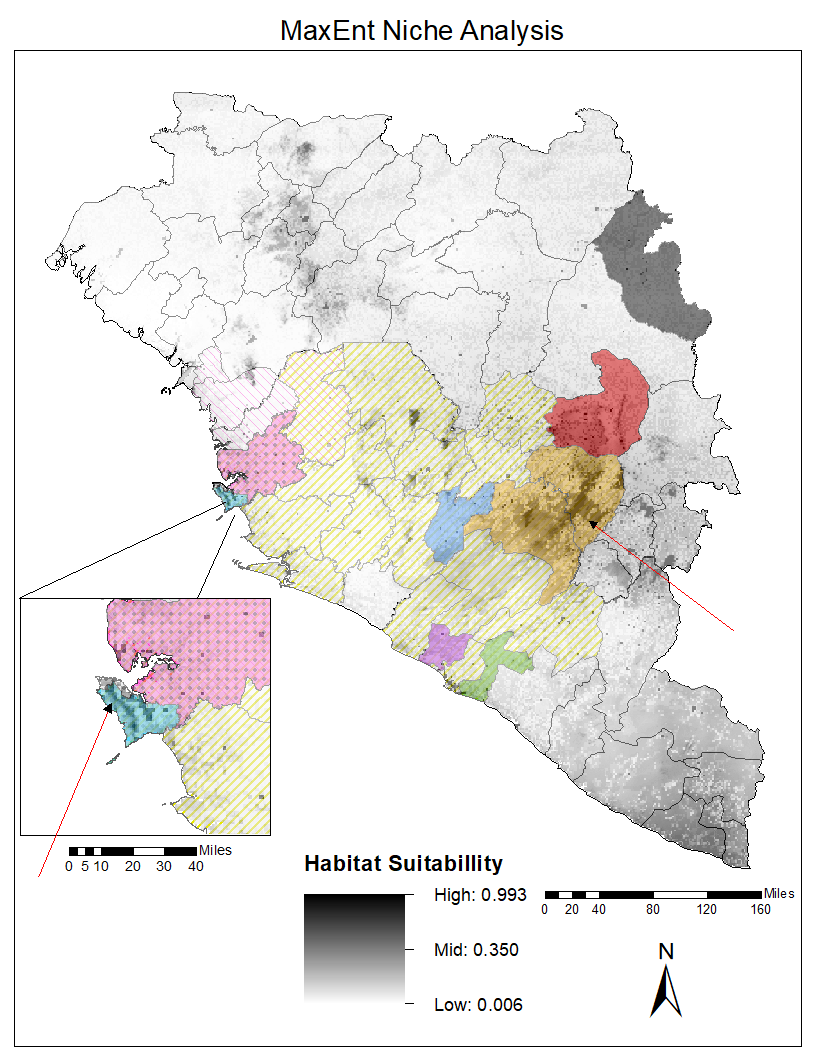


Figure 16: Overlay of SatScan clusters on top of the habitat suitability map in order to designate areas of overlap in the data

References

* Allela, Loïs, et al. “Ebola Virus Antibody Prevalence in Dogs and Human Risk.” *Emerging Infectious Diseases*, vol. 11, no. 3, 2005, pp. 385–390., doi:10.3201/eid1103.040981.
* ColumbiaMSPH. “Scientists Discover Ebola Virus in West African Bat.” *EurekAlert!*, Columbia University, 24 Jan. 2019, https://www.eurekalert.org/pub\_releases/2019-01/cums-sde012319.php.
* Doucleff, Michaeleen. “Ebola Carriers? Why The Virus Keeps Coming Back.” *NPR*, NPR, 29 Apr. 2016, https://www.npr.org/sections/goatsandsoda/2016/04/29/475801199/ebola-carriers-why-the-virus-keeps-coming-back.
* Elith, Jane, et al. “A Statistical Explanation of MaxEnt for Ecologists.” *Diversity and Distributions*, vol. 17, no. 1, 2010, pp. 43–57., doi:10.1111/j.1472-4642.2010.00725.x.
* “EVD Cases by District.” *Humanitarian Data Exchange*, Sept. 2019, <https://data.humdata.org/dataset/evd-cases-by-district>.
* Feldmann, Heinz, and Thomas W Geisbert. “Ebola Haemorrhagic Fever.” *The Lancet*, vol. 377, no. 9768, 2011, pp. 849–862., doi:10.1016/s0140-6736(10)60667-8.
* Fick, S.E. and R.J. Hijmans, 2017. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology.
* GBIF.org (02 December 2019) GBIF Occurrence Download <https://doi.org/10.15468/dl.nrxe1x>
* Grady, Denise. “Deadly Ebola Virus Is Found in Liberian Bat, Researchers Say.” *The New York Times*, The New York Times, 24 Jan. 2019, https://www.nytimes.com/2019/01/24/health/ebola-bat-liberia-epidemic.html.
* Hayman, David T.s., et al. “Ebola Virus Antibodies in Fruit Bats, Ghana, West Africa.” *Emerging Infectious Diseases*, vol. 18, no. 7, 2012, pp. 1207–1209., doi:10.3201/eid1807.111654.
* “History of Ebola Virus Disease Error Processing SSI File.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 18 Sept. 2018, <https://www.cdc.gov/vhf/ebola/history/summaries.html>
* “How Can I Detect/Address Spatial Autocorrelation in My Data?” *IDRE Stats*, UCLA, <https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-can-i-detectaddress-spatial-autocorrelation-in-my-data/>.
* Kulldorff, M, Information Management Services Inc. “SatScan v8.0: Software for the spatial and space-time scan statistics.” SatScan Software, 2009, <http://www.satscan.org/>.
* Lindsey, Rebecca, and LuAnn Dahlman. “Climate Change: Global Temperature.” *Climate Change: Global Temperature | NOAA Climate.gov*, NOAA, 19 Sept. 2019, https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature.
* Matsuura, Kanji, and Cort J Wilmott. “Terrestrial Precipitation: 1900-2017 Gridded Monthly Time Series.” *Global Precipitation Archive*, University of Delaware, Aug. 2018, <http://climate.geog.udel.edu/~climate/html_pages/Global2017/README.GlobalTsP2017.html>.
* Murack, Jennie. “Spatial Autocorrelation using GIS” [PowerPoint Presentation]. MIT Libraries. Available from: [https://www.libraries.mit.edu](https://www.libraries.mit.edu/).
* Njoh, Ambe J. “The Development Theory of Transportation Infrastructure Examined in the Context of Central and West Africa.” *The Review of Black Political Economy*, vol. 36, no. 3-4, 2009, pp. 227–243., doi:10.1007/s12114-009-9044-4.
* Parker, J. Anthony, et al. “Comparison of Interpolating Methods for Image Resampling.” *IEEE Transactions on Medical Imaging*, vol. 2, no. 1, 1983, pp. 31–39., doi:10.1109/tmi.1983.4307610.
* Phillips, Steven J., and Miroslav Dudík. “Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation.” *Ecography*, vol. 0, no. 0, 2008, doi:10.1111/j.0906-7590.2007.5203.x.
* Routledge, Richard. “Poisson Distribution.” *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 27 July 2017, https://www.britannica.com/topic/Poisson-distribution.
* Shepard, Donald. “A Two-Dimensional Interpolation Function for Irregularly-Spaced Data.” *Proceedings of the 1968 23rd ACM National Conference on -*, 1968, doi:10.1145/800186.810616.
* Tappan, G. G., Cushing, W.M., Cotillon, S.E., Mathis, M.L., Hutchinson, J.A., and Dalsted, K.J., 2016, West Africa Land Use Land Cover Time Series: U.S. Geological Survey data release, <http://dx.doi.org/10.5066/F73N21JF>
* “30 Arc-Second DEM of Africa.” *30 Arc-Second DEM of Africa | Data Basin*, NASA, 1996, https://databasin.org/datasets/2965da954b114ff3b47621e99e3b29ba.