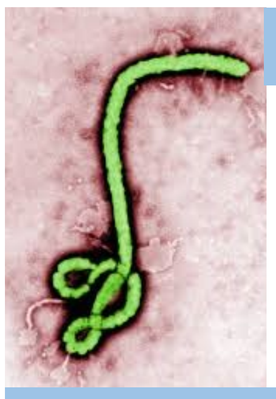
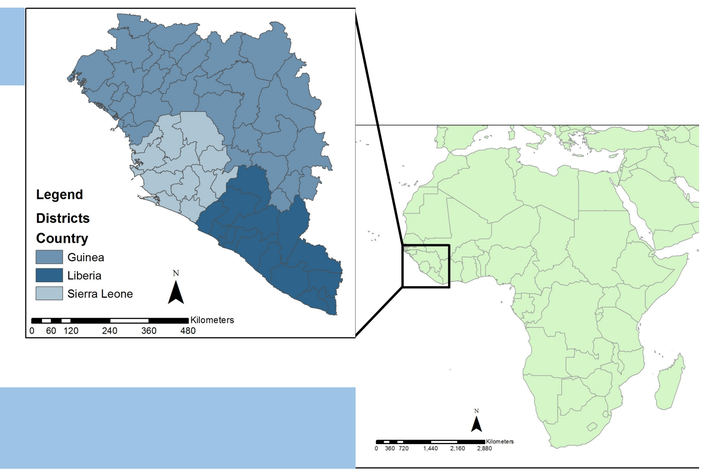
Understanding the Ebola Outbreak

Distribution and Predicting Factors of the 2013-2016 West Africa Outbreak



Introduction

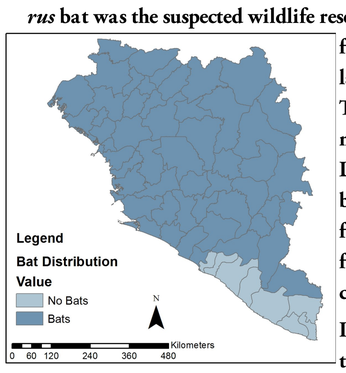
Ebola virus disease (EVD) is a communicable virus with a potentially high case-fatality rate. The largest and deadliest EVD outbreak occurred from late 2013 to mid-2016. Though the case-fatality rate was lower than previous EVD outbreaks at 39.5 percent, this outbreak infected and killed thousands more people than other outbreaks. The countries that were most impacted were Guinea, Liberia, and Sierra Leone, where approximately 28,616 people were infected with and 11,310 people died from EVD. The index case was a 2-year old child in a small Guinean village who came into contact with an infected *Mops condylurus* bat. Because this disease infected and killed so many more people than previous EVD outbreaks, it is important to understand why this EVD outbreak occurred where it did so that future outbreak locations can be predicted and the impact of future outbreaks can be mitigated. It is also important to look at which factors may be significant in predicting the initial stages of the outbreak as compared to the entire length of the outbreak, as these significant variables may change over the course of the outbreak. In this analysis, the first six months of 2014 were considered to be the initial stage of the outbreak, and the total outbreak was considered to be from early 2014 to mid-2016.



Methodology

The World Health Organization published data on weekly EVD cases within each district in Guinea, Liberia, and Sierra Leone. This data was cleaned in Excel and formatted to get the monthly EVD case distribution by district. The total number of cases for the first six months of 2014 and from the beginning of 2014 to mid-2016 were calculated. This table was then imported into ArcMap and joined to a layer with the Guinea, Liberia, and Sierra Leone districts from Global Administrative Areas. This layer, and the layers mentioned hereafter, were projected into Conakry 1905/UTM Zone 29N.

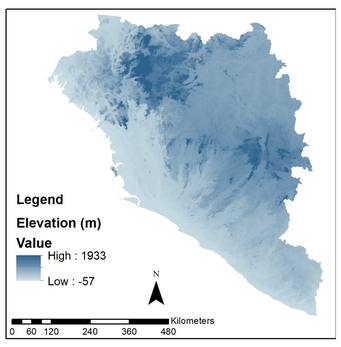
Several additional data layers were obtained to test for significance in predicting where the EVD cases occurred. Because the *Mops condylurus* bat was the suspected wildlife reservoir for this EVD outbreak, this species distribution was downloaded from the International Union for Conservation of Nature. Elevation (m) data was downloaded from NASA. The inland water source layers, one for rivers and canals and the other for lakes, were found from Digital Chart of the World. The Euclidian distance (m) from each of the water source layers was calculated. These rasters were then merged to get the minimum distance to a water source. Land cover data was obtained from the Global Land Cover Facility. Population density data (person/km^2) was found from the Oak Ridge National Laboratory. Average monthly temperature (°C) and total monthly precipitation (mm) data was obtained from WorldClim. These monthly average temperatures were averaged together to find a yearly average for temperature. These monthly total precipitations were summed together to find a yearly total for precipitation.

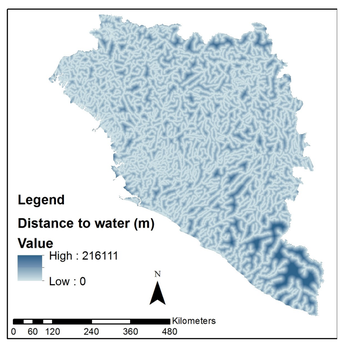


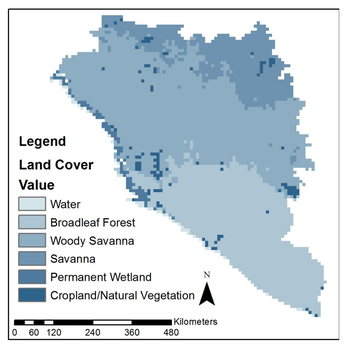
It was necessary to determine a single value for the districts for each layer mentioned above. For bat distribution a new field in the district layer was created, and districts with centroids within the bat *Mops condylurus* polygon, as well as one additional district that laid half within and half outside of the *Mops condylurus* polygon, were given a value of 1 to represent the presence of *Mops condylurus* bats, with the rest getting a value of 0 to represent absence of these bats. The total district area and land cover type areas were calculated as a new table. The proportion of evergreen broadleaf forests with the total district area was calculated, as EVD has been associated near areas of deforestation. This table was then joined with the district layer. Zonal statistics were used to create new tables with single values for each district for the remaining layers. The average values were used for the temperature, distance to water, and elevation layers. The median value was used for the population density layer. The sum was used for the precipitation layer. These tables were then joined to the district layer. The district layer was saved as a new shapefile and imported into GeoDa.

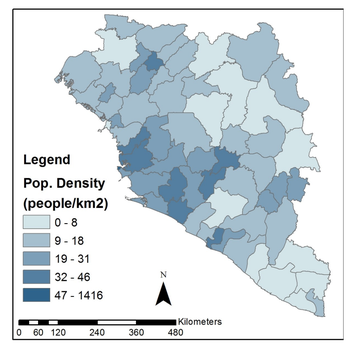
Four regressions were run in GeoDa to test the variables for significance. This included two regressions run for the initial stage of the outbreak and two run for the total outbreak to determine if different variables were significant in the two time periods. One of the regressions run for each of the time periods included the proportion of evergreen broadleaf forest as a variable and the other did not. This was done because the land cover data was less detailed than the rest of the data, so not including this variable in both regressions helps to test the significance of the more localized data. A variable with a probability value of less than 0.05 was considered significant.

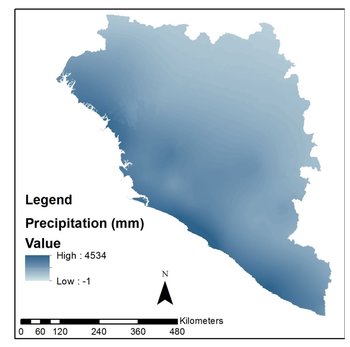
Exploratory spatial data analysis was also performed in GeoDa. Boxmaps were created to show the outlier districts that had significantly more or fewer EVD cases than other districts. Spatial autocorrelation analysis was performed using Univariate Local Moran's I using a first order Queen's contiguity weights matrix. This weights matrix was chosen because it accounts for each of the district's neighbors regardless of if they are horizontally, vertically, or diagonally next to the district in question and disregards districts that are not immediately next to the district in question. This analysis outputted the global Moran's I value and cluster and significance maps.

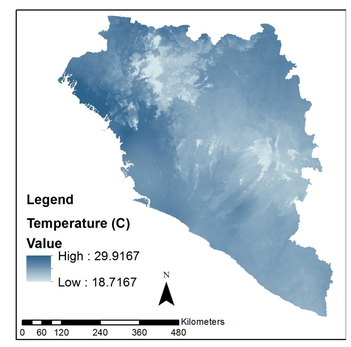












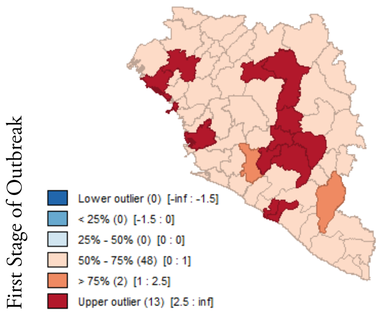
Results

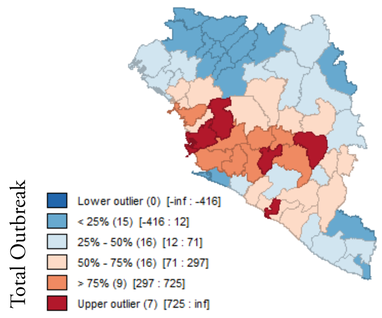
None of the variables in either of the regressions for the initial stage of the outbreak were significant. The R-squared values were also close to 0, indicating that a very little amount of the case distribution was explained by the variables tested. For the total outbreak regression including the evergreen broadleaf forest, temperature, population density, and elevation were significant. For the total outbreak regression not including the evergreen broadleaf forest, population density, elevation, and distance to water were significant. Both regressions had R-squared values close to 0.5, indicating that approximately half of the case distribution was explained by the variables tested. The findings that lower temperature, increased population density, and lower elevation were significant in predicting EVD case distribution are consistent with the results of other studies (Pigott et al., Ng et al., Kramer et al.). The significance of distance to water has not been found in other studies.

With Evergreen Broadleaf Forest		
Variable	Coefficient	Probability
Temperature	-268.418	0.02107
Population Density	2.30836	0.0000
Elevation	-2.03269	0.00388

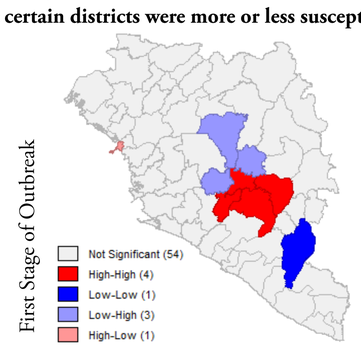
Without Evergreen Broadleaf Forest		
Variable	Coefficient	Probability
Population Density	2.23843	0.0000
Elevation	-1.40881	0.00588
Distance to Water	5.34601	0.03578

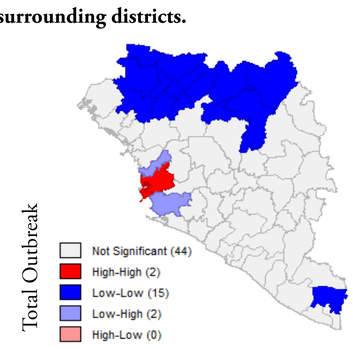
Boxmaps were created to show the case distribution among the districts both in the initial stage of the outbreak and the total outbreak. In the first stage of the outbreak, there were no low areas in general and no lower outliers because the majority of the areas at this point had no more than one EVD case. In the total outbreak, however, there were low areas, though still no lower outliers, because by this point most districts had had EVD cases, even if just a few. In both the initial stage of the outbreak and the total outbreak there were upper outliers, meaning there were several districts that had significantly more EVD cases than other districts. An upper outlier in the initial outbreak was a district with more than 2.5 EVD cases, and for the total outbreak it was a district with more than 725 EVD cases, indicating that most districts in these three countries experienced high numbers of EVD cases at some point during the outbreak.





Spatial autocorrelation analysis was performed using Univariate Local Moran's I. In the first stage of the outbreak, the Global Moran's I was 0.160941 meaning there was a small amount of global clustering. In the total outbreak, the Global Moran's I was 0.38064 meaning there was some global clustering. The cluster maps show that there was local clustering at both time periods. There were a few areas of high-high clustering and low-low clustering, indicating there were areas with many EVD cases surrounded by other areas with many EVD cases and areas with few EVD cases surrounded by other areas with few EVD cases. These high-high and low-low areas could be useful in determining if there are commonalities among these districts that had low or high numbers of EVD cases. There was one high-low district in the initial stage of the outbreak, indicating there was one area with many EVD cases surround by areas with few EVD cases. In both time periods there were low-high clusters, indicating there areas with few EVD cases surround by areas with many EVD cases. These clusters could be useful to determine what factors make certain districts were more or less susceptible to EVD cases than the surrounding districts.





Conclusion

It was found here that population density and elevation were significant in predicting where EVD cases occurred over the total outbreak, and temperature and distance to water may also be significant as they were each significant in one of the total outbreak regressions. None of the variables tested were significant for predicting the cases in the initial stage of the outbreak, thus further analysis should be performed to determine which variables may be significant in predicting initial EVD cases. Knowing these variables would help to predict where future outbreaks may occur so that countries and their districts can be better prepared to limit or stop EVD transmission.

There were several outlier districts in both the initial stage of the outbreak and the total outbreak. There were also several clusters of districts with few cases being surrounded by other areas with few cases and districts with many cases being surrounded by other areas with many cases. These clusters should be studied further to determine what their commonalities are that may make them more or less susceptible to the spread of EVD. There was also clustering of districts with few cases being surrounded by areas with many cases and districts with many cases being surround by areas with few cases. These clusters should be studied further to determine what their differences are that may make them more or less susceptible than the surrounding districts to the spread of EVD. Understanding what makes certain districts more susceptible to EVD transmission can help to mitigate the impact of future outbreaks. If the factors associated with the areas that have few EVD cases can be understood and applied to the areas that had high numbers of EVD cases, future outbreaks may be smaller and effect fewer people.

There are several limitations to this study. As this outbreak was an emergency and likely stressful and chaotic for health workers and often occurring in low-resource settings, the case data may be inconsistent or inaccurate. For this analysis, the confirmed and probable cases were used which could be an overestimate if probable cases were in fact not EVD. Additionally, the land cover data was at a lower resolution than the other data used here so significant details may have been missed in this layer. Similarly, in assigning a single value to each district, which is a relatively large area, some of the more localized details are missed that could significantly effect results. Lastly, not all of the variables that could be significant in predicting where EVD would spread were tested here. Future studies could include additional variables to perform a more detailed analysis of which variables are significant in predicting the spread of EVD.

All maps use UTM Conakry 1905/UTM Zone 29N projection.

Data sources:

World Health Organization: Weekly cases by district (2016).

Global Administrative Areas: District boundaries (2015).

Environmental Systems Research Institute: Country boundaries (2017).

Digital Chart of the World: Inland water (1992).

WorldClim: Average temperature and total precipitation (2016).

International Union for Conservation of Nature: *Mops condylurus* bat distribution (2009).

National Aeronautics and Space Administration: Elevation (2000).

Oak Ridge National Laboratory: Population density (2005).

Global Land Cover Facility: Land cover (2012).

Marisa Zellmer

UEP 294: Advanced GIS

December 2017

Tufts