Anita Chandrasekaran

Professor Ergun Simsek

DATA 606

11 May 2020

**Predicting Elephant Poaching**

**Introduction**

*Project Overview*

The aim of this project is to create a machine learning model which will aid in conservation efforts for the African elephant by examining patterns in illegal poaching and related factors. By examining indicators previously studied such as poverty and ivory trade, as well as novel indicators such as land use allocation, a time series analysis can yield insights into primary factors associated with elephant poaching.

*Background*

From the mid-20th century to today, the free roaming African elephant (*Loxodonta africana*) population has drastically reduced from 1.2 million to approximately 400,000 members (Thouless, et al., 2016). Due to this extraordinary drop in population size across multiple sites, the International Union for Conservation of Nature and Natural Resources (IUCN) has designated the African elephant as a vulnerable species, one level above endangered (Blanc, 2008). The primary cause for this reduction in elephant population has been popularly attributed to illegal poaching and high demands in the ivory trade (Wittemeyer, et al., 2014).

The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) monitors illegal killing of African elephants in designated sites across the African savannah where poaching is known to occur despite active efforts of local rangers and conservationists. In a 2019 press release, CITES analyzed elephant carcasses that were found to have been illegally killed along with associated factors that were thought to have been related (CITES, 2019). Using data collected from the Great Elephant Census, last conducted in 2014, and collated surveys from sites monitored by CITES, the Proportion of Illegally Killed Elephants (PIKE) is the most cited metric for demonstrating this growing poaching problem (Allen, 2014).

Previous works in this domain have aimed to identify potential hotspots of poaching activity, determine areas and times that would benefit from increased monitoring, and discover other factors in elephant population decline. Burn, Underwood, and Blanc affirm that the factors identified by CITES, such as corruption and poverty, are indicative of increased probability of poaching (2011). Additional exploration using quantifiable metrics to measure corruption have been conducted using data from Transparency International (Pring and Vrushi, 2019). Other aspects of governance and overall country health that were cited heavily included human development (Burn, Underwood and Blanc, 2011). These metrics are gathered from the UN Development Programme, compiled annually from a variety of primary sources such as surveys and direct measurements (UNDP, 2019). The World Bank also supplies a number of these general metrics such as GDP (World Development Indicators). While CITES maintains a database of ivory seizures, this is not available to the public for open exploration; however, demand in ivory and its global price have also been shown to highly correlate to elephant endangerment (Hauenstein, 2019).

Of the many factors identified in literature to have large or minimal impact on the PIKE statistic, human-elephant conflicts are the least explored area. According to two studies conducted by Elephants Without Borders, human encroachment and human conflict has a large negative impact on the elephant populations of Angola and Botswana (Schlossberg, Chase and Griffin, 2018; Schlossberg, Chase and Sutcliffe, 2019). This project uses agricultural growth as an indicator of human encroachment, to measure how human-elephant conflict contributes to the declining elephant population.

**Methodology**

*Data Collection*

A variety of data sources were used in this analysis in order to gain a comprehensive view of the environment in which elephant poaching takes place. To gauge site-specific population sizes, the Great Elephant Census was scraped (Allen, 2014). In addition, CITES provided a number of resources for poaching statistics and detailed data points (CITES, 2019). This elephant-specific data is particularly remarkable due to its collection methodology: much of it is acquired using indirect observation such as aerial surveillance, through statistical inferences such as dung count, and by ranger-conducted surveys. As a result, the baseline dataset used for this project must be considered cautiously.

The indicators used to measure impact on poaching and elephant decline included: Human Development Index and the Global Corruption Barometer scraped from the UN’s Human Development Reports database (UNDP, 2019); and, forest and agricultural land as a percentage of total surface area, and GDP per capita scraped from the World Bank’s World Development Indicators database (The World Bank, 2019). Incomplete or missing data was filled with mean values for the given indicator.

*Data Analysis*

As stated previously, the PIKE statistical value was chosen to represent poaching activities of a given site. This was simply calculated using the number of illegally killed carcasses found and the total number of carcasses found during survey. These values were collated and published by CITES. Exploratory data analysis was conducted to determine which factors were adequately representative of a time series equivalent to the elephant population data.

Vector Auto Regression (VAR) was chosen to analyze the multivariate time series data and determine the impact of each indicator (or group of indicators) on future prediction capability. Exploratory data analysis showed vastly different PIKE patterns and elephant population sizes; therefore, the datasets were separated by region. By using every possible combination of indicators (see Table 1) in tandem with the PIKE data values, training and validation datasets were split at Year=2015, leaving approximately 25% of the data for validation purposes. 

Training datasets for each region with the different combinations of indicators were fitted to a VAR model. Predictions were gathered and compared to the validation datasets using the root mean square error (RMSE), which measures the differences between the predicted PIKE values and the actual PIKE values. RMSE values can be compared within a dataset, for each region, but not across regions; therefore, the subsequent analysis will be broken down by region.

**Analysis**

*Exploratory Data Analysis*

Elephant population reports were restricted to specific sites used in the Monitoring of Illegal Killing of Elephants program, with regions defined therein (CITES, 2019). Approximately 200,000 elephants are currently in the South Africa region, 45,000 in Central and East Africa, and about 10,000 elephants in West Africa. This proportion has remained steady for the length of reporting recorded by the Great Elephant Census (approximately 20 years).

The chosen metric to quantify poaching was borrowed from CITES--this is called the Proportion of Illegally Killed Elephants (PIKE) and is calculated as a ratio of found carcasses with illegal causes of death to total carcasses. These values are reported by park rangers from the selected MIKE sites and are culled by CITES to add to its extensive database. This data is shared with all participant countries to encourage and facilitate exchange of intelligence related to poaching activities. Regionally, elephant population is highly concentrated in the south of the African subcontinent (Graph 1). 

Regionally, PIKE shows upward trends over time, except for Southern Africa (Graph 2). There are limited data available for South Africa MIKE sites due to inconsistent reporting and reporting bias. A larger scale PIKE graph over time shows the aggregated PIKE scores from all MIKE sites over time (Graph 3). This shows a slight downward trend in PIKE at the current time, but this may be preceding another upward trend as the line shows some seasonality.





*Forest Area.* Shown as a percentage of forested land area out of total land area, the forest area metric aids in determining the maximum capacity of elephants in a given region (UNDP). The upward trends in forested areas for Southern and Western Africa may be anomalous due to newly established protected zones, or because of inconsistent reporting. There is a positive correlation between forested areas and PIKE but may be influenced by these odd values (Graph 4).



*Corruption.* The Global Corruption Barometer is a quantifiable metric describing the influence of corruption on everyday lives. This indicator aids to determine how much political influence may be contributing to lax poaching efforts (Pring and Vrushi, 2019). Further exploration revealed no meaningful correlation between GCB and PIKE (Graph 5). This indicator was discarded from VAR analysis due to limited data.



*Human Development.* The Human Development Index is a composite score which factors in life expectancy, economic factors, and education to rank countries in their overall development (UNDP). HDI is gradually increasing for all countries, and there appears to be some kind of negative correlation between HDI and PIKE (Graph 6).



*Agricultural Land Use.* Agricultural land usage as a percentage of total land usage shows a negative correlation with PIKE over time, indicating that some kind of human encroachment in elephant territory may be influencing poaching activities (Graph 7).



*GDP per capita.* A similar negative trend in GDP per capita and PIKE is evident (Graph 8). Outstanding outliers in this indicator were discarded due to their very unique economic placement in the region.

*Vector Auto Regression*

The VAR models for each region were trained and predictions gathered for each of the 13 combinations of chosen indicators, as seen in Table 1. Models were evaluated using the Root Mean Square Error, the standard deviation of prediction errors (i.e., difference between predicted and actual values). Graph 9 shows the RMSE values for each VAR model of the indicator combinations and regions described previously.

VAR models with the highest RMSE were generally seen in the Central and Southern African datasets. The indicator combinations with the highest RMSE were: 4 (GDP only), 7 (HDI/GDP), 9 (HDI/Forest Area/GDP), 11 (Forest Area/GDP), 12 (Forest Area/Agri Land/GDP), and 13 (Agri Land/GDP). GDP appears to be the unifying indicator that consistently performs poorly when used to inform the VAR model for predicting PIKE values. 



Well-performing models shown in Graph 10 show more consistent RMSE values across regions. Indicator combination 1 (HDI only) shows the smallest error, meaning that this indicator by itself informs the VAR model the closest to the actual PIKE values observed.

Combinations 3 (Agri Land only), 6 (HDI/Agri Land), 8 (HDI/Agri Land/Forest Area), and 10 (Forest Area/Agri Land) performed consistently well across all regions and showed smaller RMSE values.

**Discussion/Conclusion**

Through the initial exploratory data analysis, several potential poaching indicators were identified as possible factors contributing to higher poaching rates represented by large PIKE values. Interestingly, the corruption indicator (GCB) has been identified as a key factor in previously conducted poaching prediction studies. Despite its potential utility, the limited availability of usable data for corruption made GCB an impractical indicator to use for model training. Through this, however, there were a number of indicators which stood out as potentially useful in predicting PIKE values and, therefore, poaching probability. Using combinations of four of these indicators (HDI, forest area, agricultural land area, GDP), regional datasets were used to train VAR models, ultimately to determine which indicators aided in closely predicting PIKE values.

The VAR models that were trained on regional datasets generally showed higher error for Central and Southern Africa. Limited available data for the Southern African region can account for these high error rates; error for Central Africa prediction values is likely due to discarding outliers, many of which belonged to this region. Additionally, higher RMSE was observed with indicator combinations using GDP in the model training. The correlation between GDP and PIKE for the African continent is generally negative; however, this becomes less obvious when viewed regionally. The unsteady nature of GDP in the African subcontinent due to fluctuating economies and populations likely contribute to the poor performance for this indicator.

The VAR models that showed low RMSE error typically used HDI and Agricultural land use during training. For these models, Central and Eastern Africa were the best performing datasets. This is particularly curious due to the high RMSE values for Central Africa models using GDP. This seems to indicate that poaching is not highly correlated to GDP, and that agricultural land use is a valid metric to use for predicting PIKE values across various African subregions.

It is evident that minimal elephant carcass data presents a major challenge in determining the validity of poaching indicators. Smaller datasets and inconsistent/subjective reporting contribute to the tenuous nature of any conclusions drawn from this analysis. However, a general conclusion can be made that quality of life and agricultural sprawl are highly influential on the rate of poaching of African elephants. Logically, it follows that poaching is not a necessity when governments place restrictions on human encroachment in elephant territory, thereby leading to smaller PIKE values for those sites. By reducing the possibility of human-elephant interactions in the wild, the probability of elephant poaching decreases.

**References**

Allen, P. "The Great Elephant Census." *The Great Elephant Census* (2014).

Blanc, J. "Loxodonta africana." *The IUCN Red List of Threatened Species* 2008 (2008): T12392A3339343.

Boult, Victoria L., et al. "Human‐driven habitat conversion is a more immediate threat to Amboseli elephants than climate change." *Conservation Science and Practice* 1.9 (2019): e87.

Burn, Robert W., Fiona M. Underwood, and Julian Blanc. "Global trends and factors associated with the illegal killing of elephants: a hierarchical Bayesian analysis of carcass encounter data." *PLoS one* 6.9 (2011).

Convention on International Trade in Endangered Species of Wild Fauna and Flora. “New report highlights continued threat to African elephants from poaching.” *CITES Press Release.* 10 May 2019.

Hauenstein, Severin, et al. "African elephant poaching rates correlate with local poverty, national corruption and global ivory price." *Nature communications* 10.1 (2019): 1-9.

Pring, Coralie, and Jon Vrushi. "Global Corruption Barometer: Africa 2019." *Transparency International* (2019).

Schlossberg, Scott, Michael J. Chase, and Curtice R. Griffin. "Poaching and human encroachment reverse recovery of African savannah elephants in south-east Angola despite 14 years of peace." *PloS one* 13.3 (2018).

Schlossberg, Scott, Michael J. Chase, and Robert Sutcliffe. "Evidence of a growing elephant poaching problem in Botswana." *Current Biology* 29.13 (2019): 2222-2228.

*World Development Indicators*: *Agricultural land (% of land area), Forest area (% of land area), GDP per capita*. The World Bank, 2019. Web. 20 February 2020.

Thouless, Christopher, et al. "African elephant status report 2016." *Occasional Paper Series of the IUCN Species Survival Commission* 60 (2016).

UNDP. “Human Development Report 2019: Beyond Income, Beyond Averages, Beyond Today - Inequalities in Human Development in the 21st Century.” (2019).

United Nations Office on Drugs and Crime. *World wildlife crime report: Trafficking in protected species*. United Nations Office on Drugs and Crime (UNODC), 2016.

Watson, Fred GR, et al. "Human encroachment into protected area networks in Zambia: implications for large carnivore conservation." *Regional Environmental Change* 15.2 (2015): 415-429.

Wittemyer, George, et al. "Illegal killing for ivory drives global decline in African elephants." *Proceedings of the National Academy of Sciences* 111.36 (2014): 13117-13121.