SPATIAL ANALYSIS AND MODELING OF STRIGA WEEDS IN WESTERN KENYA REGION

by

JAPHETH NYANDORO GICHANA

Submitted on 5/9/2019 in partial fulfillment of the requirements for the award of the degree of Bachelor of Science in Geomatic Engineering and Geospatial Information Systems.

Department of Geomatic Engineering and Geospatial Information Systems (GEGIS)

DECLARATION

I declare that this project is my own work and has not been submitted by anybody else in any other university for the award of any degree to the best of my knowledge.

Sign………………………………… Date………………………………………..

Student name JAPHETH NYANDORO

Registration number EN281-1472/2014

Department of Geomatic Engineering and Geospatial Information Systems (GEGIS)

Jomo Kenyatta University of Agriculture and Technology

CERTIFICATION

This project has been submitted for examination with my approval as the candidate's supervisor.

Sign……………………………. Date ……………………………………………

Mr. Joshua Mwaura

GEGIS

ACKNOWLEDGEMENTS

First, I'd like to thank the source of all good things, God.

I thank the lecturers at Jomo Kenyatta University of Agriculture and Technology within the GEGIS department for their critique and constant guidance. More so, I am heavily indebted to my committed supervisor Mr. Joshua Mwaura whose assistance all through the study period lead to the realization of this project.

I want to acknowledge the Agricultural Improvement Support Services (AGRISS) NGO for providing me with the data required for this research on striga in western Kenya region and for their input in helping me understand the drivers and causes of the weed in the region.

Lastly, I express deep gratitude to my entire family for the love, concern, encouragement and financial support.

ABSTRACT

In developing countries, millions of people especially in rural areas depend on maize, sorghum, and other cereal crops for food security and as a source of income. In addition to this, these people contribute a huge percentage to the national food basket and hence the national food security. Despite of this, production of these crops is seriously affected by constraints such as weeds, pests, diseases, declining soil fertility among others. Chief among these constraints is the parasitic weed, Striga spp. that has been affecting cereal crop production in Sub-Saharan Africa for more than seven decades. Part of the reason this weed has existed this long, is a lack of understanding of the invasion drivers of the weed as well as a lack of scientific tools for predicting its future occurrence both spatially and temporally. However, with the advent of GIS-based species distribution modeling tools, this problem can be solved by combining various environmental, climatic and spatial components in order to understand and forecast species invasions.

This study focusses on the Western Kenya region which is the major cereal crop production region in the country. By the use of various geostatistical modeling tools, different environmental and climatic variables are spatially analyzed to determine their relationship with the occurrence of Striga weeds in the region. Once this was realized, spatial and temporal models were developed to show different areas that are at risk of invasion by striga based on the above factors, and to predict future risk areas based on forecasted climatic data.

TABLE OF CONTENTS

List of Figures

List of Tables

Acronyms and abbreviations

AATF – African Agricultural Technology Foundation

ASK - Agricultural Society of Kenya

CIMMYT – International Maize and Wheat Improvement Center

KARI – Kenya Agricultural Research Institute

CHIRPS – Climate Hazards Group InfraRed Precipitation with Station

RF – Random Forest

GWR – Geographically Weighted Regression

OLS - Ordinary Least Squares

1 Introduction

1.1 Background

A weed, also known as an invasive species, is an organism that causes ecological or economic harm in a new environment where it is not native. Striga or witchweed is a parasitic weed that plagues an estimated 2.4 million hectares of smallholder farmland in Sub-Saharan Africa (AATF, 2011). There are more than 20 different striga species in Africa alone but the five common species include Striga hermonthica, Striga asiatica, Striga aspera, Striga gesnerioides, and Striga forbesii. Striga hermonthica and Striga asiatica are the two common species of the witch weed in East Africa with the former common around Lake Victoria Basin while the latter is mainly found in the coastal areas.

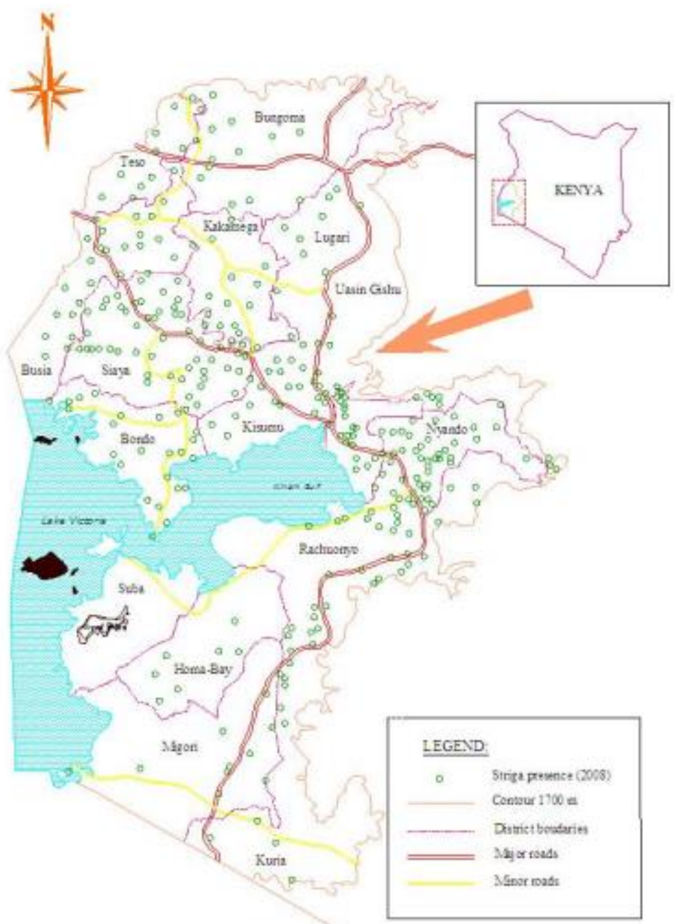


Figure 1. Part of western Kenya map showing *Striga hermonthica* infestations(De Groote et al., 2008)

Striga mainly affects cereal crops which include maize, sorghum, millet and rice, but it has also been reported to affect plants in the Grass family such as sugarcane, Napier and other native grasses. A mature striga plant produces large amounts of striga seeds which are triggered to germination once they are close to the above host crops. Otherwise, the seeds can remain dormant in the soil for over 20 years awaiting favorable conditions for their germination (De Groote et al., 2008).

According to the International Crops Research Institute for Semi-Arid Tropics (1996), striga attaches itself to the maize or sorghum roots from which it extracts its moisture and nutrient requirements for several weeks while living underground. During this period the weed produces toxic chemicals that begin to stunt and discolor the host plant. After this, the striga shoot emerges from the soil, producing fleshy green stems and narrow leaves, growing to a height of 0.5 to 1m (Woomer, Striga, & Project, n.d.).

Next, it produces several small purple flowers that later form capsules which contain 50,000 to 200,000 small seeds for a mature striga plant (*strides*, 2008). After the host crop dies, so too does the striga plant, causing the capsules to burst and the seeds therein to spread around the nearby soil. The seeds can be spread to other surrounding farms by wind, animals, and even humans as they move from one farm to another. The above cycle then repeats itself in other crops or planting seasons. In the long run, the striga plants inhibit plant growth, reduce crop yields and in extreme cases cause total crop loss.

1.2 Motivation and Problem Statement

The invasion of plant ecosystems and croplands by alien plant species is a cause of concern worldwide due to their negative environmental and economic impacts. More specifically, according to (Atera, Ishii, Onyango, Itoh, & Azuma, 2013) Striga spp is considered to be the greatest biological constraint to food production in Sub-Saharan Africa causing yield losses in cereals ranging from 15% under favorable conditions to 100% where several stress factors are involved.

The striga weed has been in existence in the fields of farmers within the Lake Victoria basin from as early as 1936 (Khan et al., 2006). However, the problem has intensified in the last two to three decades mainly due to cereal monocropping that has been brought about by the increase in population in the Western Kenya region (Oswald, 2005). Cereal monocropping exerts a strain on the nutrients present in the soil which consequently reduces soil fertility. Low soil fertility is one of the major drivers for striga infestation hence exacerbating the situation even further.

Figure 2: A striga-infested maize farm in the region

The weed is reported to be infesting about 217,000 ha of land in Kenya, leading to an annual crop yield loss of 53 million USD (Sanginga, N. and Woomer, 2009). These losses are contingent on the intensity of striga infestation, the host species and genotype, soil nutritional status, and climate among other factors (Rodenburg et al., 2015).

The western Kenya region is a major food-producing region of the country contributing more than 50% to the national food basket. In addition to this, cereals, especially maize are a major staple crop, a source of income and employment for millions of farmers in the region. However, cereal production in the region is plagued by a series of production constraints that hamper the ability of the farmers to meet their farming potential and for the government to meet its objectives for agricultural transformation.

Chief among these production constraints is striga infestation, which according to (Woomer and Savala 2009), has infested farmer's fields in Western Kenya with an average of 161 million seeds per ha of land translating to 3 parasitic stems per maize plant. Other similar studies in the region also show that there is an average of 14 striga weeds per square meter of farmland in the region (MacOpiyo, Vitale, & Sanders, 2010). Cereal consumption was approximately 4.6 million tonnes in 2017 (Ministry of Agriculture) while the production was 3.71 million tonnes. This shows that cereal production in the country lags far behind the required amount hence the need for action to be taken to beef up the current production.

1.3 Research Identification

The main aim of this research is to investigate the probable influence of environmental and climatic variables on striga occurrence and model areas that are at risk of infestation by the weed.

This is achievable through the following specific objectives:

1.3.1 Research Objectives

1. To carry out spatial analysis to identify and quantify the factors that allow the growth of striga weed.
2. To develop spatial-temporal models showing the risk areas of infestation of striga weeds in western Kenya region.
3. To develop a weed modeling system for automating the process of weed modeling based on machine-learning models.

1.3.2 Research Questions

The following questions are formulated with respect to the aforementioned objectives:

1. What factors contribute to the growth of striga weed species and to what extent do they contribute?
2. What are the high-risk and low-risk areas for striga infestation in the region?
3. How can I automatically determine the risk areas for other invasive plant species?

1.4 Study Outline

This thesis is organized into six chapters. Chapter 2 discusses the literature review that was done in relation to all aspects of this study. Chapter 3 discusses the materials and methods that were applied in achieving the objectives of this study. Chapter 4 describes the results obtained in the study with respect to the objectives. Chapters 5 and 6 discuss results with respect to objectives of the study placing them in the broader context of science and draws conclusions to my study in reference to results obtained and objectives of my study and providing recommendations to subsequent studies respectively.

2 Literature Review

2.1 Origin and Occurrence of Striga

There are 9 main Striga species found in Kenya (Table 1). Striga hermonthica and S. asiatica which are the most common in Kenya are believed to have originated in the North Eastern part of Africa most specifically in the Nubian hills of Sudan and Semien mountains of Ethiopia (Atera et al., 2013). These areas are also believed to be the origin of cereals such as sorghum and finger millet which are part of striga's host crops. S. gesnerioides is believed to have originated from West Africa. Due to human activities and movement over the years, the species has spread to other parts of Sub-Saharan Africa. S. hermonthica is considered to be the most predominant weed in western Kenya (MacOpiyo et al., 2010).

Table 1: Striga spp distribution and occurrence in Kenya

Striga Species	Host Plants	Occurrence Area
S. asiatica	Maize, rice, sorghum, pearl millet, finger millet, sugarcane, wild grasses	Kilifi, Isiolo, Alupe, Daka Chom, Kiunga
S. bilabiate	Wild grasses	Naivasha, Chyulu hills, Rumbia, Kahawa, Mathews range
S. elegans	Wild grasses	Nairobi, Loitokitok, Laikipia, Rimuruti
S. forbesii	Sorghum, rice, maize, sugarcane	Narok, Mara plains, Kipini, Chyulu hills, Uasin Gishu plateau, Trans Nzoia
S. gesnerioides	Cowpea	Kilifi, Buna, Homa hills, Rongo, Nairobi, Naivasha
S. hermonthica	Maize, rice, sorghum, pearl millet, finger millet, sugarcane, wild grasses	Alupe, Churaimbo, Miwani, Bungoma, Kendu, Migori, Kuria, Nyamira, Siaya,

		Homabay
S. latericea	Sugarcane, wild grasses	Samburu, Mariakani, Kwale, Voi, Machakos, Sultan Hamud, Kilifi, Mwea
S. lutea	Wild grasses	Kwale, Shimba hills, Embu, Chyulu hills
S. pubiflora	Sugarcane, wild grasses	Kwale, Shimba hills, Voi

Source: (Atera et al., 2013)

2.2 Factors Influencing the occurrence of Striga

1. **Soil Nitrogen content**

   Previous studies that have been carried out on the relationship between Striga infestation and soil nitrogen have revealed a negative correlation between striga occurrence and soil nitrogen content. The higher the amount of soil nitrogen in the soil, the lower the number of striga plants and vice versa. According to (Kamara, Ekeleme, & Omoigui, 2009) application of Nitrogen fertilizers on fields that are heavily infested with striga led to a significant reduction on the number of striga plants overtime and an increase in crop yields. Other studies have also shown that intercropping striga-host crops such as maize with legumes such as beans and desmodium which help in fixing nitrogen into the soil reduces striga infestation considerably (Khan, Midega, Amudavi, Hassanali, & Pickett, 2008).

2. **Soil organic matter content**

   Soil organic matter also has a negative correlation with the number of striga plants in a given field. This is because as organic matter in the soil decomposes, various compounds including ethylene are released in the process. Ethylene helps to trigger the germination of striga seeds present in the soil. However, absence of host roots on which the germinating seeds can attach themselves to and draw nutrients from eventually leads to their death. This phenomenon is called ‘suicidal germination.’ Hence, applying organic manure in the soil a few weeks before planting begins can help deplete existing striga seeds in the soil and ensure they do not affect crops once planting has been done (Ayongwa, Stomph, Belder, Leffelaar, & Kuyper, 2011).

3. **Temperature**

   Plant growth is very sensitive to temperature and each plant species has, at any given stage in its life cycle a base temperature below which it will not grow and an optimum temperature within which it thrives. According to (Aflakpui, Gregory, & Froud-Williams, 1998) striga hermonthica requires a base temperature of 20°C, an optimum temperature of between 30°C to 35°C and a ceiling temperature of 48°C beyond which the seeds cannot germinate. Generally, the rate of germination of striga seeds tends to increase with an increase in land surface temperature.

4. **Precipitation**

   According to (Oswald, 2005), striga grows in regions between sea level to an altitude of 1600m and in production systems with annual precipitation rates of between 500 to 2000mm. Most places in the western Kenya region have fairly equivalent precipitation rates and hence this (among other factors) explains why striga infestation is high in the region.

Other striga influencing factors are summarized in the table below:

Table 2: Factors contributing to striga occurrence

FACTOR	**EFFECT**
Soil Nitrogen	Striga generally thrives in soils with low nitrogen content.
Soil Organic Matter	Striga generally thrives in soils with low soil organic matter content.
Temperature	Striga prevalence increases with increase in temperature up to a given threshold.
Precipitation	Striga proliferates in regions with precipitation rate ranging from 500 to 2000mm
Soil Texture	Soils with a lighter texture generally tend to favor the growth of striga weeds.
Elevation	Striga thrives in a wide range of altitude ranging from sea level up to an altitude

	of 1600m.

2.3 Management and control of striga

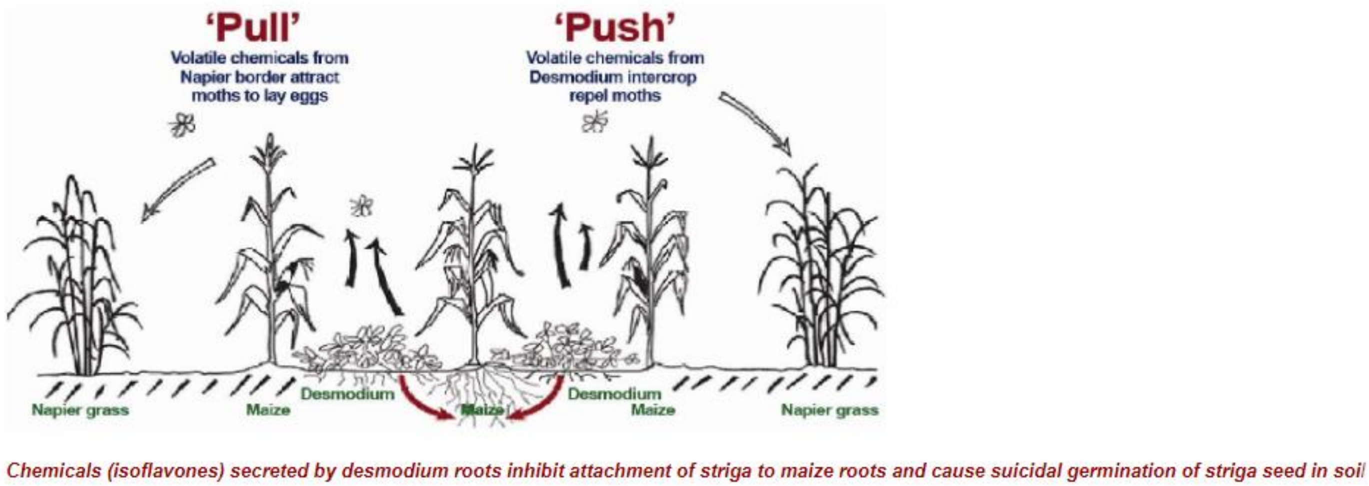
Striga control methods can be broadly classified into two: direct and indirect methods. Direct striga control methods are those that attack the weed directly and that have an immediate effect on the intensity and number of the weed in the field. Indirect methods are those that are aimed at the management of various farming practices in order to make the growth conditions of the weed less favorable (Oswald, 2005). Due to this, indirect methods often take a longer duration until an effect can be observed, but they control striga in a more sustainable way and increase crop yields overtime.

Direct striga control methods comprise of:

1. Planting striga-resistant host-crop varieties – CIMMYT in conjunction with KARI developed a hybrid maize variety called Imazapyr that has proven to be effective in controlling striga in the region. Imazapyr maize is coated with a special chemical that causes striga plants that attach to the maize roots to die off immediately (Oswald, 2005). Similarly, there are other resistant crop varieties that have been developed for sorghum (Study, n.d.), finger millet, rice and the other cereal crops to control striga. The main disadvantage of this method is that it is expensive and hence cannot be adopted by majority of the resource-poor farmers in the region.
2. Application of chemicals i.e. herbicides and ethylene – This is a chemical control method that involves spraying affected fields with special herbicides that kill the striga plants. Most of these herbicides are selective in nature and hence when applied, they will eliminate the weed and spare the cereal crop that’s beneficial to the farmer.
3. Hand weeding – This is a labor-intensive method that involves physically uprooting the individual striga plants from the field once they emerge. The method aims at ensuring the weed does not reach maturity and hence it does not produce seeds which if released may worsen the situation. The method is highly effective in controlling striga, however, its effect on crop yields is felt in the subsequent planting seasons.

Indirect striga control methods comprise of:

1. Crop rotation – This involves planting striga host crops in one season and then planting non-host crops in the following season. Rotating with non-host crops or Striga trap crops (crops which stimulate Striga germination but do not allow attachment of the parasite) interrupts the production of new Striga seeds and stimulates soil conditions which lead to the depletion of the Striga seed bank in the soil (Oswald & Ransom, 2001). Some of the crops that can be rotated with the above cereals include peanuts, soybean, beans, sunflower, pigeon pea and sesame.
2. Use of the 'push-pull' method – This is a novel strategy that was originally developed by ICIPE and that is used not only in striga management but also in crop pest management in resource poor farming systems. The method involves intercropping striga-host crops with repellent crops and an additional attractive trap plant that is planted around the intercrop. For instance, maize is intercropped with a legume, silverleaf desmodium (Desmodium uncinatum) after which napier grass (Pennisetum purpureum) is planted around the intercrop. Desmodium roots produce chemicals which stimulate striga seed germination (such as 4'',5''-dihydro-5,2',4'-trihydroxy-5''-isopropenylfurano-2'',3'';7,6-isoflavanone) and other chemicals which inhibit their attachment to maize roots (such as 4'',5''-dihydro-2'-methoxy-5,4'-dihydroxy-5''-isopropenylfurano-2'',3'';7,6-isoflavanone). This leads to suicidal germination of striga seeds which in turn reduces their seed bank over time (Khan et al., 2008). The desmodium also helps to replenish soil nitrogen into the soil which improves the nutrient-content and fertility of the soil and overtime increases crop yields (Khan et al., 2002).



Source: Push-pull.net

Figure 3: How push-pull method works

2.4 Economic Impacts of Striga

Despite the different strategies that have been adopted to control striga over the years, yield loss attributed to striga is still acute with reduction in cereal productivity ranging from 32% to 75% in severe cases (Hearne, 2009). In much broader scales, striga infestation is attributed to causing a loss of 30% to 50% of Africa's agricultural economy on 40% of its arable land (Amudavi, Khan, & Pickett, n.d.). In western Kenya region, a survey was conducted on 83 farms and it revealed that 73% of the farms are infested with Striga hermonthica.

The average loss of yield due to striga is 1.15, 1.10, 0.99 tons per hectare for maize, sorghum, and millet respectively (MacOpiyo et al., 2010). However, in areas with high striga densities, the damage can escalate to figures as high as 2.8 tons per hectare in maize and sorghum which are the most affected crops. Considering maize alone as per 2013, the above figures translate to a loss of 12.3% of the total 2.4 million metric tons of maize produced in Kenya during that year. This is equivalent to about 39.6 kg of maize loss per capita, which is about 20% of a normal person's annual food requirement (Atera et al., 2013).

Clearly, striga weed is a major threat not only to food security in the country but also on the economy. It requires innovative and collaborative actions from different disciplines in order to reclaim our Agricultural potential as a country.

2.5 Modeling of weeds using GIS

GIS has proved to be a significant tool in establishing relationships between various datasets by analyzing them spatially. This has been especially useful in the control of invasive plant and animal species that have become a cause of concern in many regions in the world due to their negative environmental and economic impacts. Invasive species lend themselves to GIS analyses because most of these species are spatial in nature.

GIS leverages the use of geostatistical modeling tools which help to understand invasion drivers, predict species range expansion or contractions in relation to natural and anthropogenic drivers, and to guide early detections (Capinha & Anastácio, 2011). The most common tools in disease modelling research includes ArcGIS/QGIS for visualization and geo-processing of the factors, geostatistical tools such as OLS, GWR, Kriging, and SPSS are used for determining significance of the variables used, and ERDAS for land use classification and extraction of various land indices (Palaniyandi, Anand, & Maniyosai, 2014).

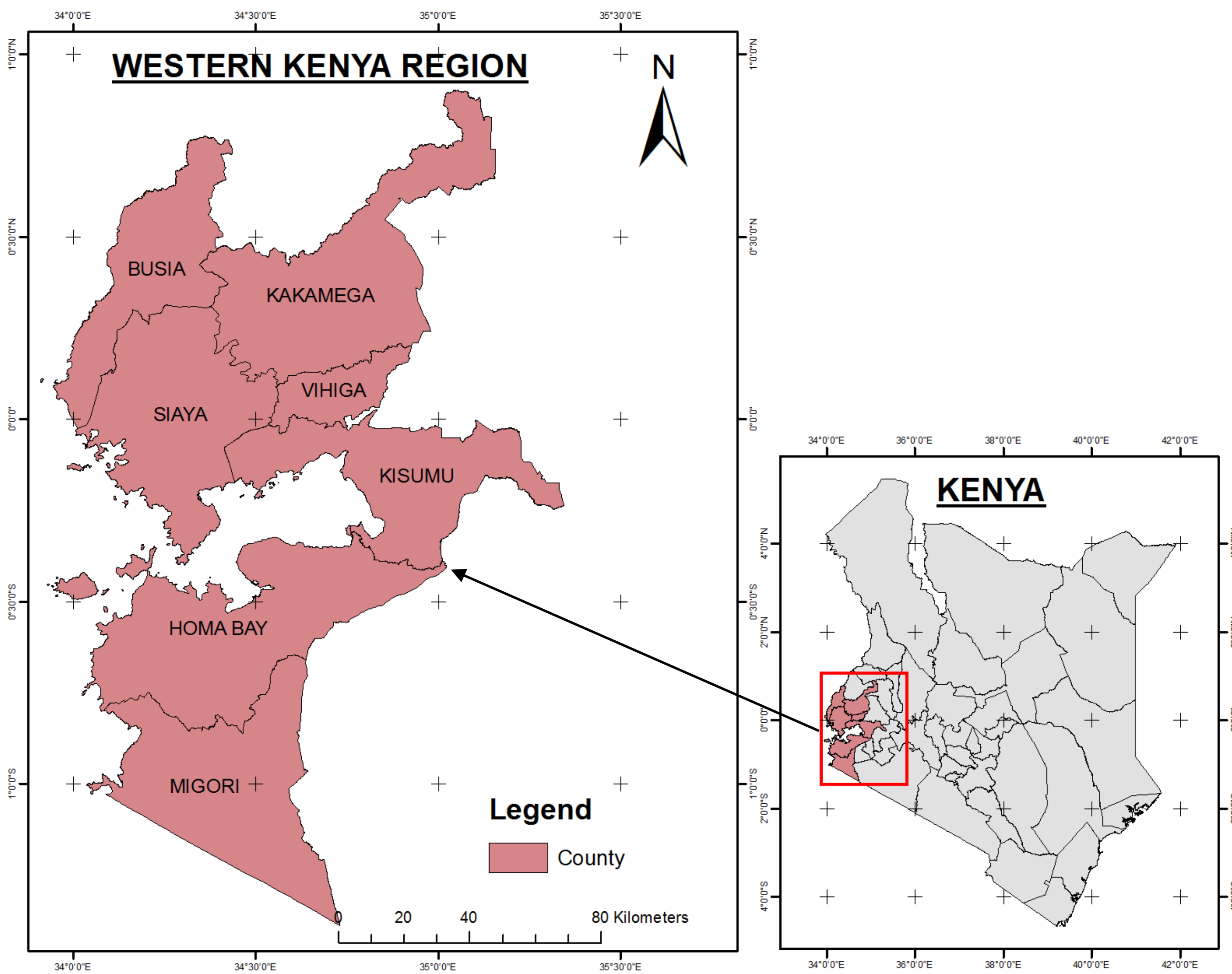
GIS was used as an early warning tool for forecasting weed distributions in the US at county level based on various climatic variables (Jarnevich et al., n.d.). The findings of the study were taken up by the local governments and used in the formulation of policies and measures to counter the likely effects of the weeds.

3 Materials and Methods

3.1 Study Area

The study area for this research is the western Kenya region comprising of seven counties: Migori (2586 km²), Homabay (3155 km²), Kisumu (2085 km²), Siaya (2496 km²), Busia (1628 km²), Vihiga (531 km²), and Kakamega (3034 km²). The above regions have a cumulative area of 15515.3 km² and a population of 6.652 million people (Kenya National Bureau of Statistics, census 2009).

Figure 4: Study Area Map



3.2 Data

Data and data sources utilized in this study are:

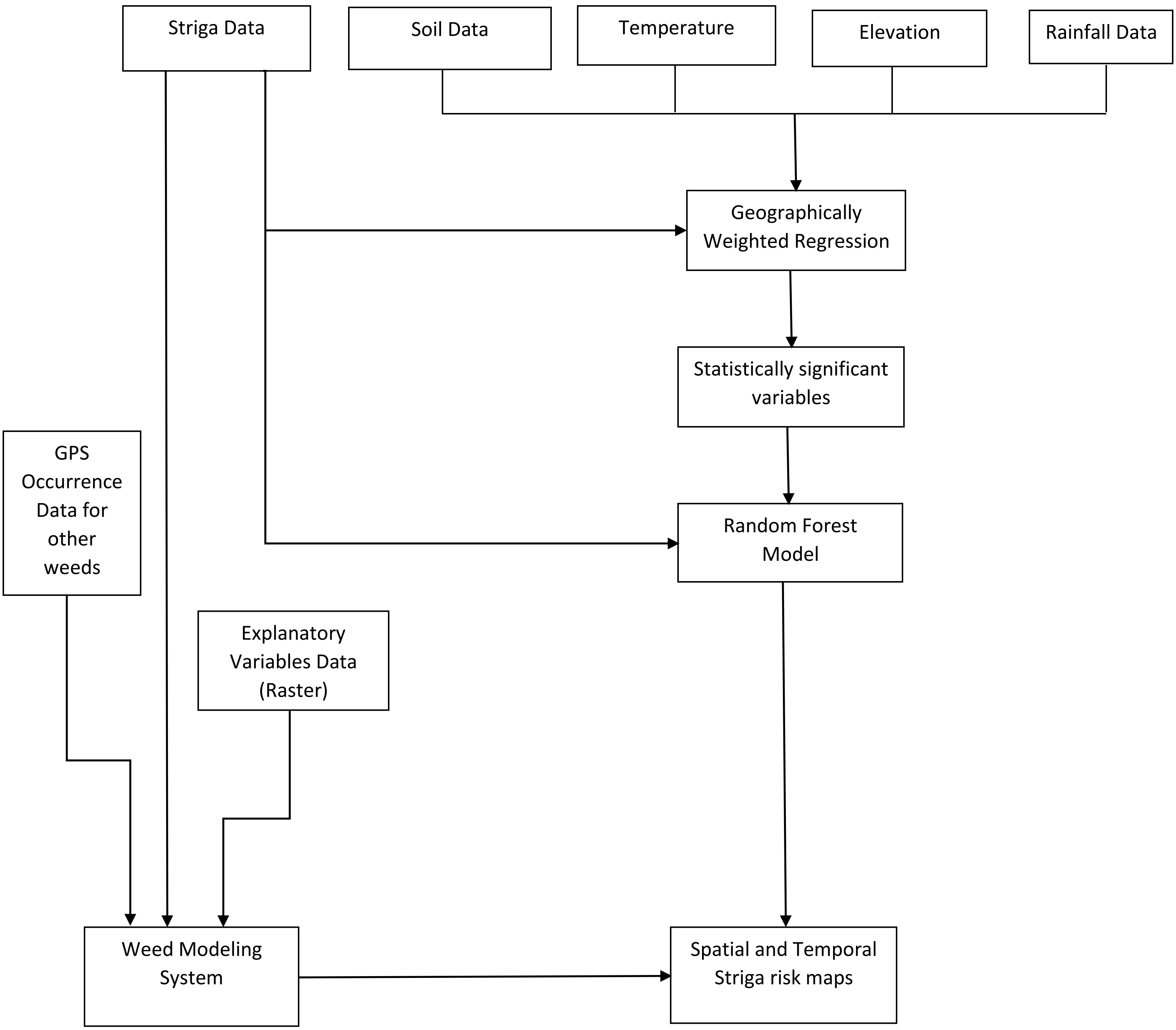
Table 3: Data Sources

DATA	DATA SOURCE
Striga occurrence data	GPS coordinates collected during fieldwork by AGRISS NGO based in the region.
Soil Nitrogen	AFRISOILS
Soil Organic Matter	AFRISOILS
Soil Texture	AFRISOILS
Temperature data	MODIS
Rainfall Data	CHIRPS Monthly averages
Elevation	SRTM 30m

3.3 Methodology

3.3.1 Conceptual Framework

Figure 5: Conceptual Framework



3.3.2 Data Collection

The first step involved the collection of data as listed before. The precipitation dataset was collected from CHIRPS in raster format for the months of July, August and September 2018. This is because the striga data that I was to use in modeling was collected during these three months. The precipitation data was in form of monthly averages and hence it was averaged for the three months using the raster calculator tool after which it was clipped to the study area.

Soil datasets (Nitrogen, Organic matter, and texture) were obtained from Afrisoils.net in raster format at a spatial resolution of 250m each. The data was then clipped to the study area for use. Administrative boundaries for the seven counties that constitute the western Kenya region were obtained from the Kenya open data portal.

Striga occurrence data was obtained from AGRISS which is a non-governmental organization based in Homabay county. This data was collected using a GPS during the period of July, August, and September 2018 when most cereal crops in the region were maturing in the field. The coordinates represent individual farms where striga was observed and the density of infestation of the weed in those farms. The data provided was in an Excel file and is shown below (additional columns show the socio-economic status of the farmers and striga control methods that were being used at the moment of data collection).

Figure 6: Striga Data Collected

County	Sub County	Farm Longitude	Farm Latitude	Date of collection
Homabay	Homabay Town	34.47625	-0.57482	18/07/2018
Homabay	Ndhiwa	34.23859	-0.791251667	19/07/2018
Kisumu	Kobongo	34.9901	-0.15066	24/07/2018
Busia	Nambale	34.22648	0.444015	30/07/2018
Migori	Kamsaki	34.47108	-0.964326667	3/9/2018
Vihiga	Muhanda	34.67828	0.055143	25/06/2018
Siaya	Ugenya	34.15675	0.24832	19/07/2018

3.3.3 Geographically Weighted Regression

Data Aggregation - The factors used were derived from several factors. Each factor was a unique dataset on its own. The factors, from literature review were found to be around 7. They include: Precipitation, Temperature, Soil organic matter, Soil Nitrogen, Soil Texture, Elevation, and NDVI. All these datasets were in raster format. Using the Extract grid values at points tool, raster values were extracted at each Striga occurrence point. The extracted values were all merged to one shapefile. This would make it easier to perform the analysis.

Determining the dependent and independent variables- All the 7 factors listed above form the independent variables (explanatory variables). Their values will each be statistically analyzed against the dependent variable to determine its significance in the study. The dependent variable in this case is the number of observed striga occurrences in the Western Kenya region.

Defining the Unique ID (UID)-The GWR regression needs a unique ID field in the attribute table in order to work. The integer field contains a different value for every feature in the input feature class. This field is created by using the field calculator in the attribute table then it computes and replicates the values from the object ID field to the unique ID field. This makes the data ready to run the GWR regression.

- *Linear Regression Equation* – GWR is a local spatial linear regression method that allows the relationships being modeled to vary spatially across the study area. Hence, the method is more robust in analyzing spatial relationships compared to Ordinary Least Squares which is a global regression method. GWR is a classical refinement of the OLS equation which is given by:

$$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots \beta_7 X_7 + \varepsilon}$$

Where:

- Dependent variable (**Y**) – What you're trying to model or predict.
- Explanatory variable (**X**) – Variable that influence or help explain the dependent variable.
- Coefficients (**β**) – Values computed by the regression tool, reflecting the relationship and strength of each explanatory variable to the dependent variable.
- Residuals (**ε**) – The portion of the dependent variable that isn't explained by the model i.e. the model's under or over predictions.

Running the GWR - The GWR tool is found in the spatial statistic tool. The input feature is specified, the unique ID field, the explanatory variable and the dependent variable. The output is also specified. Generation of the report is optional, however its recommended. The GWR is the run to compute the regression.

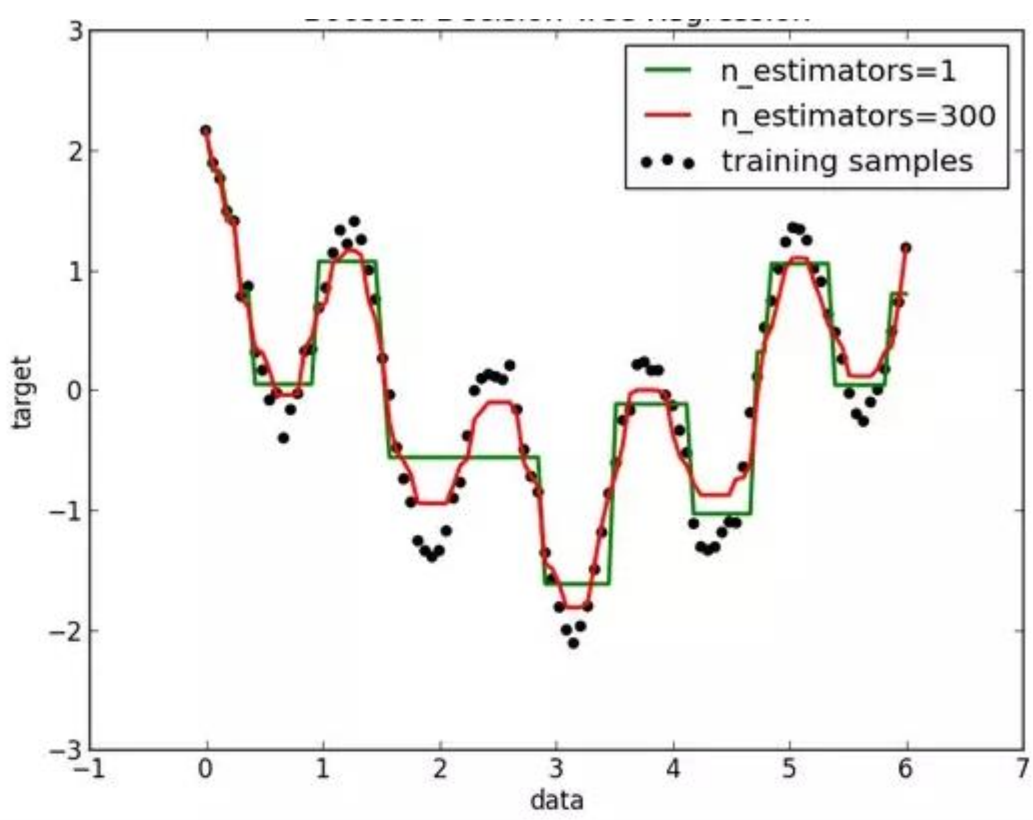
Results interpretation - The generated report contains most of the results needed to draw conclusions. Results interpretation in GWR is comprehensive and requires a lot of background knowledge on regression to understand the results. The reported is generated in PDF format.

3.3.4 Spatial Prediction Models from Random Forest

Random forest is a machine-learning based classification and regression algorithm that uses bagged decision tree models to predict the occurrence of a given variable in a given location. RF is a robust spatial regression model for spatial predictions; however, it does not have an equation. The nature of random forest algorithm inherently leads to the destruction of any simple mathematical representation.

RF works by building a large number of decision trees that act as an ensemble. Each individual tree tries to fit a regression model to the target variable using each of the independent variables. Then for each independent variable, the data is split randomly at several split points. Random forest then calculates the Sum of Squared Error (SSE) at each split point between the predicted value and the actual value. The variable resulting in the minimum SSE is selected for the node and used for the final prediction above the others. Then this process is recursively continued till the entire data is covered.

Figure 7: Sample random forest model curve



The above figure shows that random forest is a non-linear regression model and hence it identifies the best relationship between the dependent and independent variable and creates a line that best fits that relationship.

Instead of just averaging the predictions of each decision tree, RF uses two concepts that give it the name random:

Random sampling of training observations when building trees.

Random subsets of features for splitting nodes.

In conclusion, RF builds multiple decision trees and merges their predictions together to get a more accurate and stable prediction rather than relying on individual decision trees. Each tree in a random forest learns from a random sample of the training observations provided.

3.3.5 Future prediction of potential risk areas

Prediction involves forecasting the probable number of striga densities expected in the next few planting seasons. This requires one to determine the number of observed striga occurrences seasonally for the past several years and use the time series analysis to determine the approximate number of cases expected in the next season. However, since this data was not available, a different approach was applied as described below.

Prediction can also involve using seasonal weather forecast to determine the amount of precipitation and temperature expected in the next season. Using this forecasted data, it can be combined just like before with the other environmental variables to develop models of the predicted risk areas expected in the next season. Prediction in this case works with an assumption that the changes in elevation, soil Nitrogen and Soil Organic Matter are negligible hence kept constant. Prediction has a tremendous significance in this study because decision makers can be equipped with sufficient information on how to well to deal with the situation that's coming. This enables proper planning and effective management beforehand so as to appropriately control the spread and infection of striga in the coming seasons.

4 Results

4.1 Regression results

Both GWR and OLS regression methods were used. OLS was used to get a rough estimation of the relationship between the different explanatory variables with the dependent variable on a global scale. After this GWR was run successfully for each explanatory variable to fine-tune the results. The output for GWR gave two results relevant to this study; an output feature class and a message window containing the statistical report of the results. This is illustrated below.

Figure 8: OLS Results

Summary of OLS Results

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-118.652615	396.263570	-0.299429	0.765019	401.422287	-0.295581	0.767950	--------
SOILNITROG	-14.105255	5.569174	-2.532737	0.012298*	5.475783	-2.575934	0.010918*	5.067152
RAINFALL	0.326918	0.072897	4.484672	0.000016*	0.059375	5.505991	0.000000*	1.573120
SOILORGANI	-3.428914	0.655698	-5.229413	0.000001*	0.652449	-5.255455	0.000001*	4.983831
TEMPERATUR	13.202758	13.963156	0.945543	0.345834	13.929336	0.947838	0.344668	66.234827
ELEVATION	0.008645	0.077251	0.111911	0.911029	0.082898	0.104288	0.917065	60.627756
SOILTEXTUR	3.540376	0.929342	3.809550	0.000208*	1.067653	3.316037	0.001145*	1.835091

OLS Diagnostics

Input Features:	Striga	Dependent Variable:	STRIGA_OCC
Number of Observations:	163	Akaike's Information Criterion (AICc) [d]:	1403.318148
Multiple R-Squared [d]:	0.788755	Adjusted R-Squared [d]:	0.780630
Joint F-Statistic [e]:	97.079551	Prob(>F), (6,156) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	594.978763	Prob(>chi-squared), (6) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	33.901043	Prob(>chi-squared), (6) degrees of freedom:	0.000007*
Jarque-Bera Statistic [g]:	2.108620	Prob(>chi-squared), (2) degrees of freedom:	0.348433

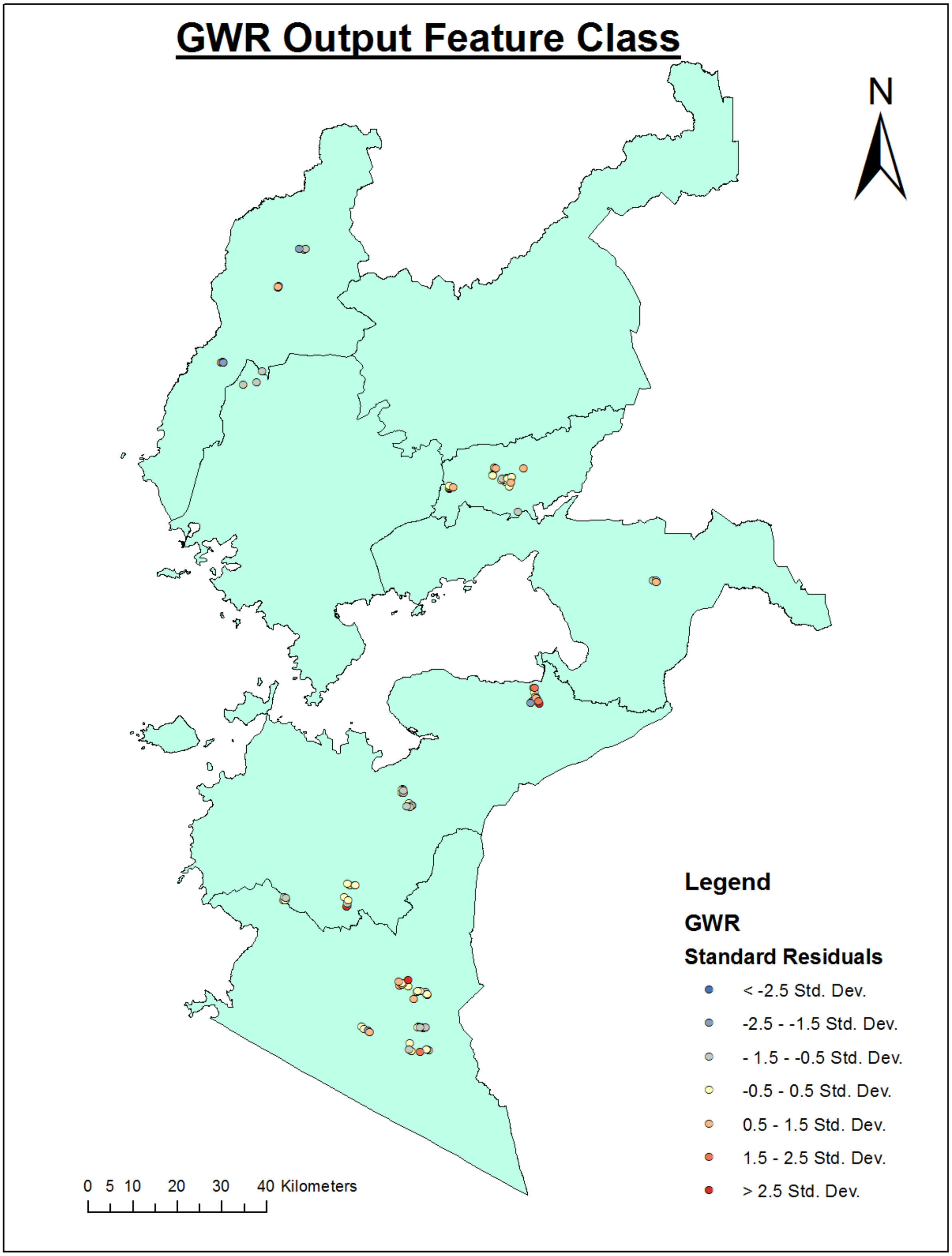


Figure 9: GWR Feature Class

```
Neighbors            : 47
ResidualSquares      : 25295.045158250927
EffectiveNumber      : 15.03862511774958
Sigma                : 13.20965818317063
AICc                 : 1295.9082874308306
R2                   : 0.8862998910501344
R2Adjusted           : 0.8752887288926906
```

Figure 10: GWR results

4.2 Identification of Significant variables

From the regression analysis done in the previous section and from literature review, the following factors were considered in the study:

Figure 11: Soil Nitrogen map

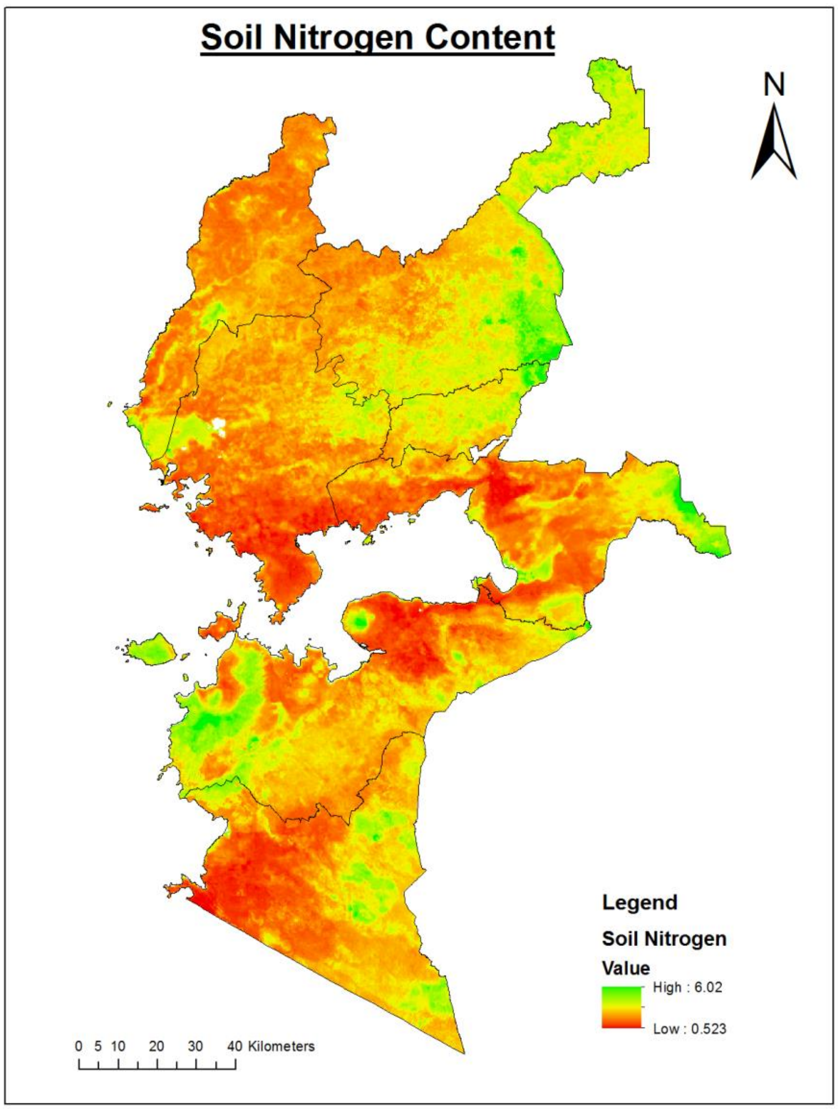


Figure 12: Soil Organic matter map

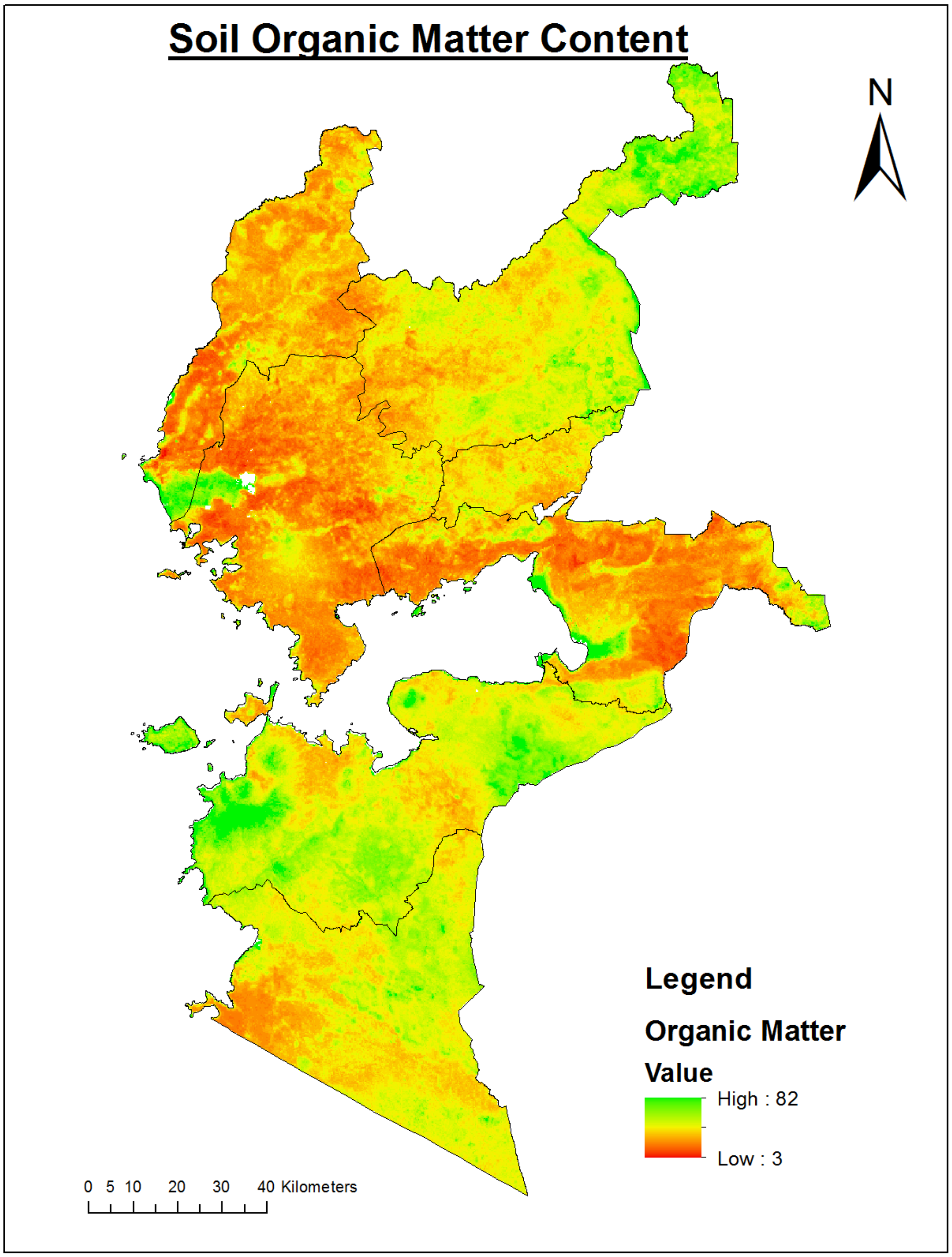


Figure 13: Soil Texture map

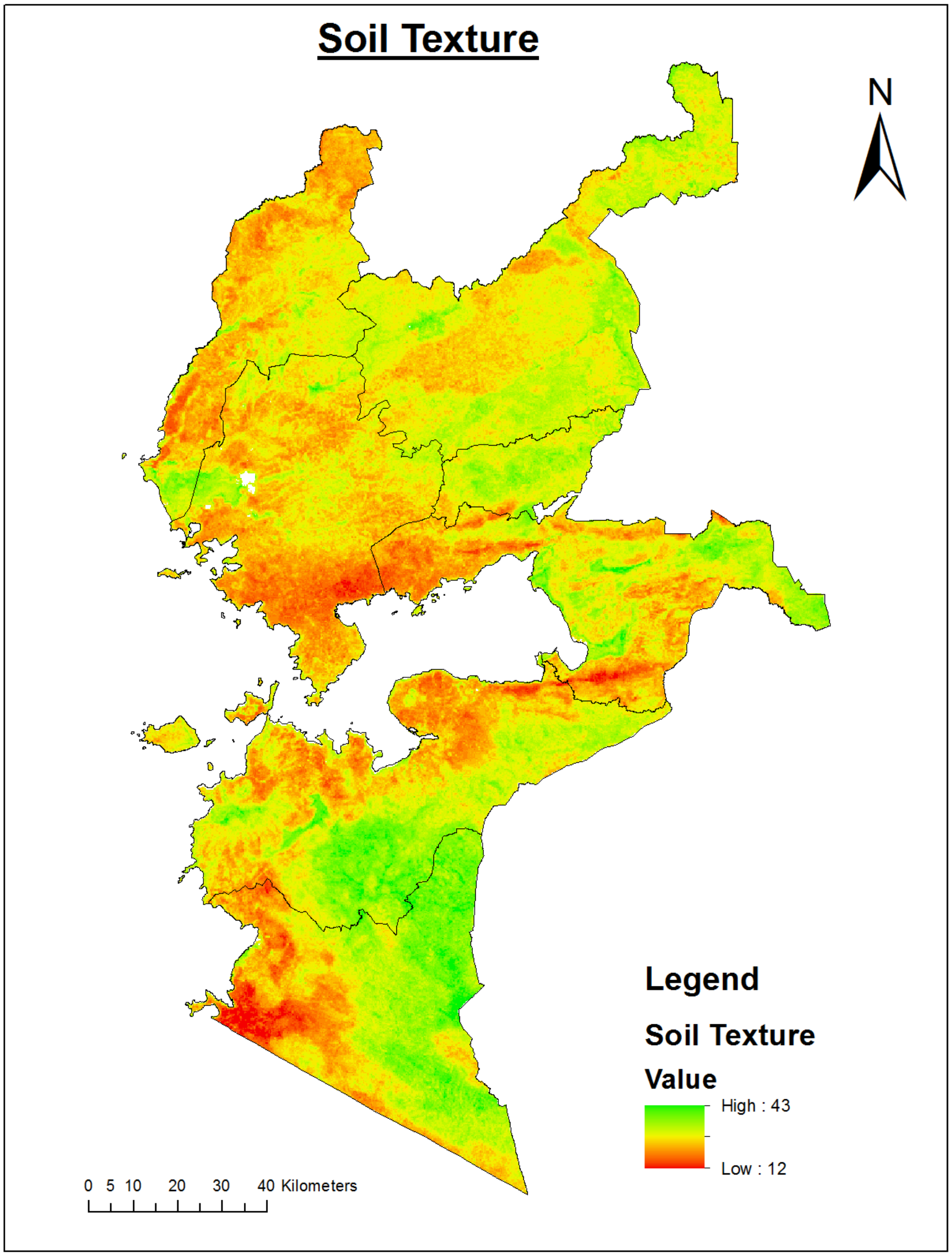


Figure 14: Elevation map

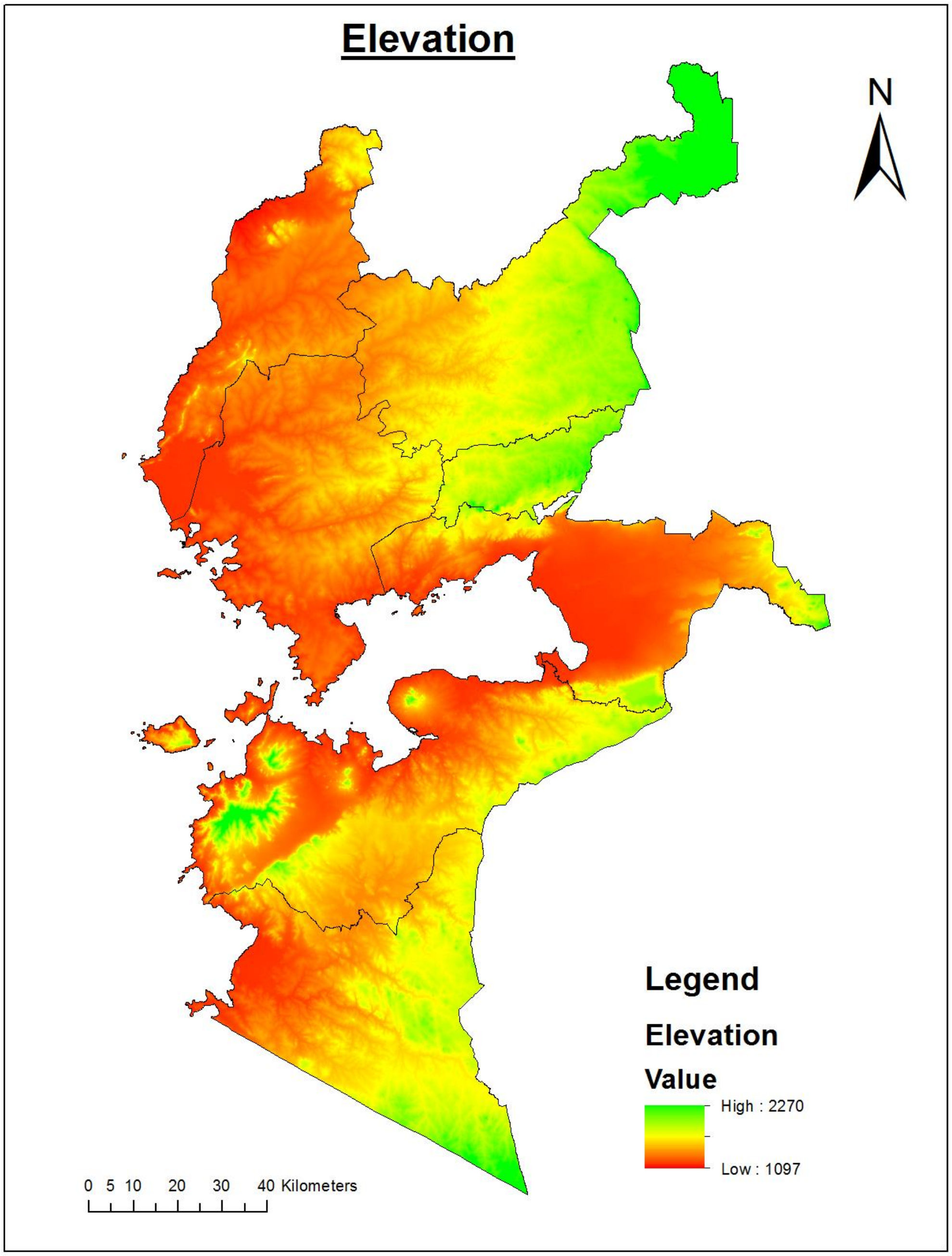
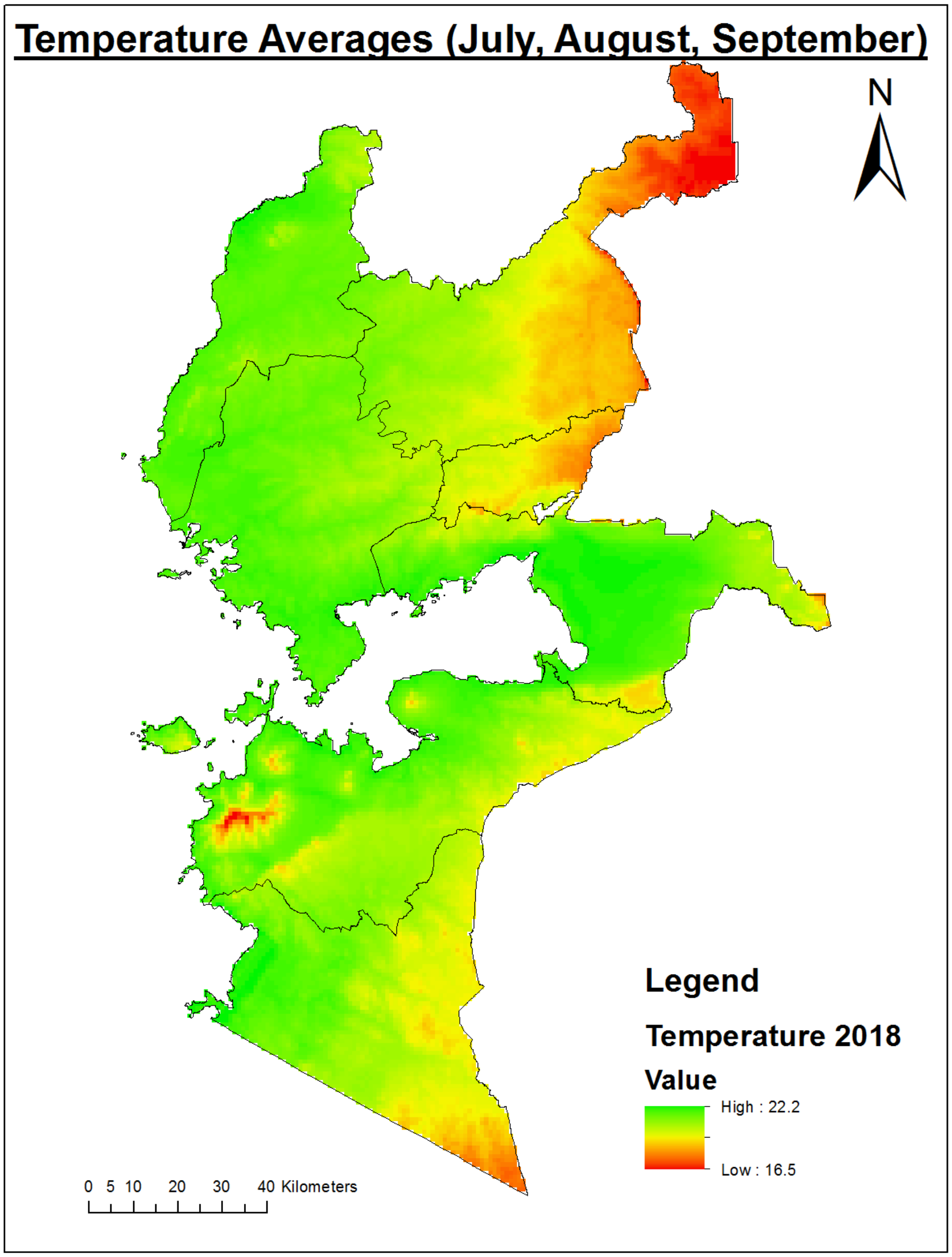


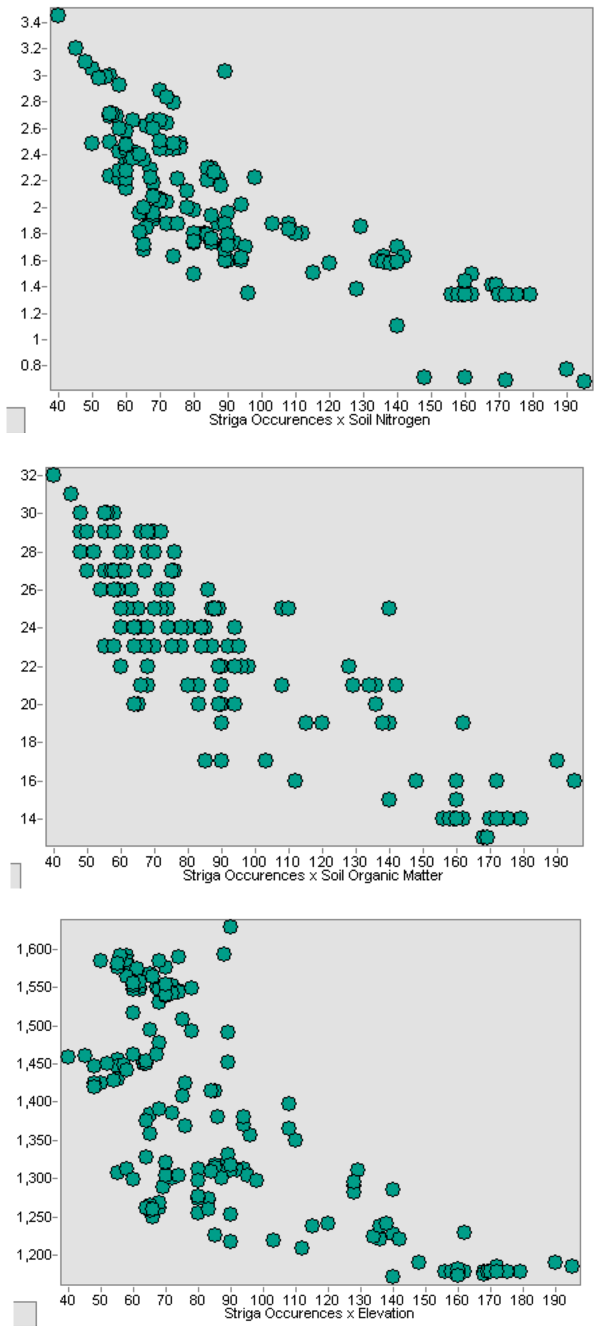
Figure 15: Temperature map



4.3 Scatterplot Matrices

To get a visual picture of the relationship between the dependent and independent variables, scatterplot matrices were generated with the dependent variable on the y-axis and the independent variable on the x-axis.

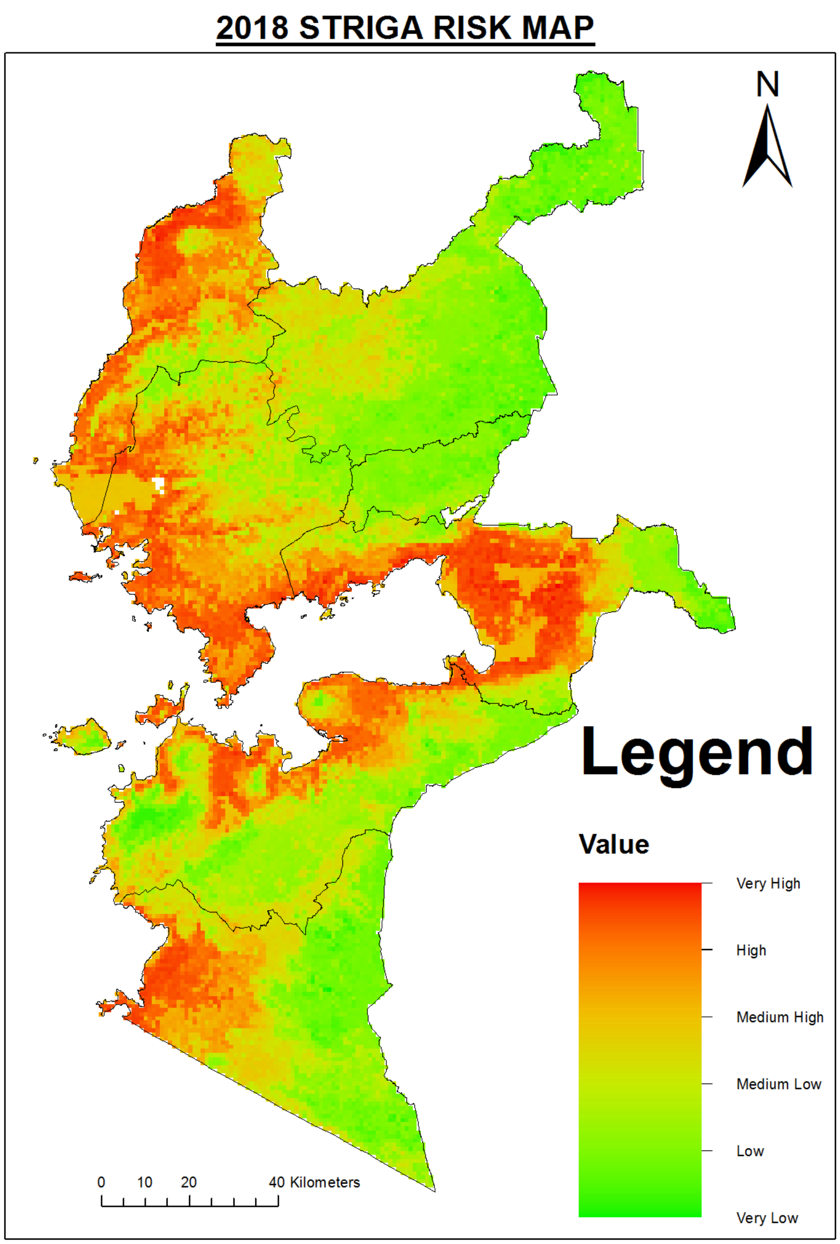
Figure 16: Scatterplots



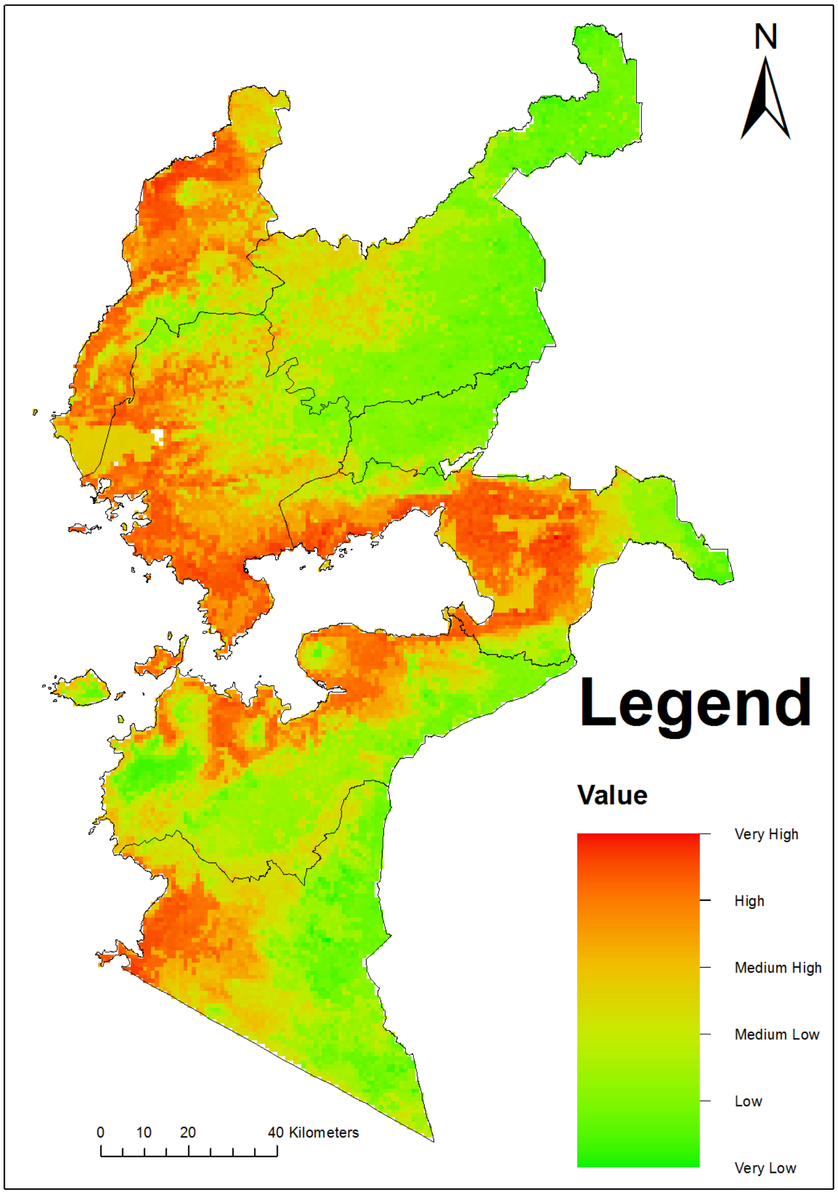
4.4 Spatial Prediction Models by Random Forest

Once the above variables were determined, each of them was resampled to a standard spatial resolution of 1km. This is because the explanatory variable with the coarsest resolution was 1km. Once the explanatory variables were resampled, they were raster stacked and then used for the RF modeling. Below are the spatial prediction models derived from random forest:

Figure 17: Striga risk maps



2019 STRIGA RISK MAP



4.5 Automated Weed Modeling System

A weed modeling system was developed to aid various agricultural organizations in the region to combat any other invasive weeds that may arise in the region. The inputs to the system include a geojson file with spatial data on the dependent variable and multiple raster files with data for the explanatory variables.

Figure 18: Weed modeling system inputs

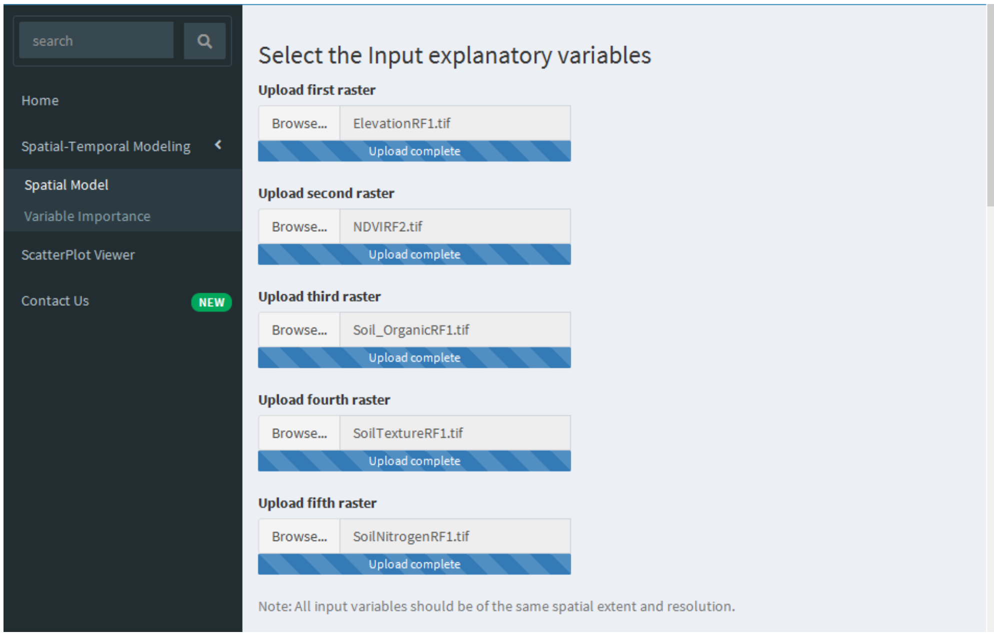
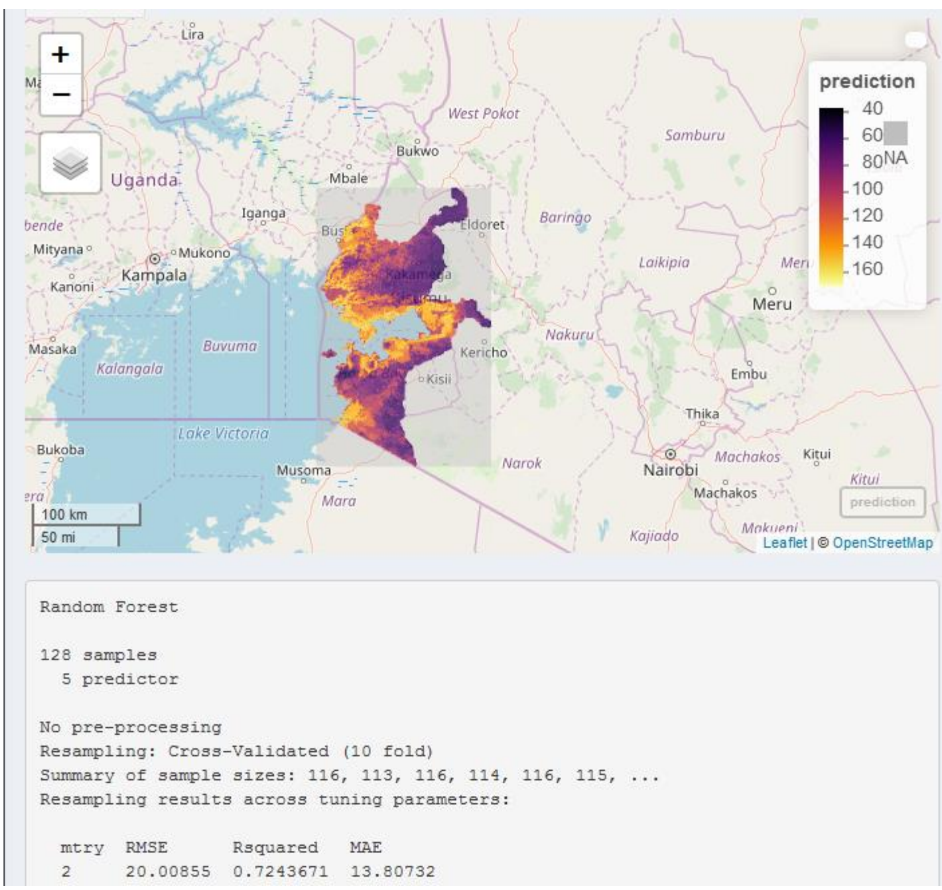


Figure 19: Weed modeling system outputs



5 Discussion

5.1 Regression Analysis

Regression analysis allows one to model, examine, and explore spatial relationships to better understand the factors behind observed spatial patterns, and to predict outcomes based on that understanding. Numerous regression methods exist but for this study, only OLS and GWR regression methods were used. OLS is a global regression method while GWR is a local, spatial regression method that allows the relationships being modeled to vary across the study area. There are two typical outputs generated from either of the two methods: an output feature class and a table showing various statistics/diagnostics of the analysis. From the tabular results shown in section 4.1 of this report the following can be discussed:

I) Coefficients – The coefficients column is used to show the relationship between the dependent and explanatory variables. A negative coefficient, as is the case with soil nitrogen and organic matter, means the relationship is negative while a positive coefficient as is the case with temperature and rainfall means the relationship is positive.

II) VIF (Variance Inflation Factor) value – This column on the table is used to check for redundancy among the explanatory variables. If the VIF value for any of the variables is larger than about 7.5, it means one or more variables are telling the same story (redundant).

III) Probability and Robust probability – Two columns, probability and robust probability measure a variable's statistical significance. An asterisk next to the probability tells you the specific variable is significant. If a variable is not significant, then it is not helping the model, and unless theory tells us a particular variable is critical, we should remove it.

IV) R squared and Adjusted R squared values – These two values are used to check for model performance. The values range from 0 – 1.0 and they tell how much of the variation of the dependent variable was explained by the model. And so, the higher the R squared value the higher the model performance and vice versa. In addition to these two model performance indicators, the AIC (Akaike's Information Criterion) value can also be used to compare two or more models to determine the best of them. The model with the lowest AIC value represents the most superior model. However, all the factors have to be kept constant (dependent and independent/explanatory variables) in all the models being evaluated when using AIC value as a measure of model performance.

5.2 Susceptibility to striga infestation

In order to determine which areas are more susceptible to invasion by striga weed and those that are not, certain criteria were used as discussed in previous sections. The application of said criteria led to the classification of different parts of the Western Kenya region into various levels of susceptibility to striga invasion. These levels represent the likelihood based on various factors of there being invasion of striga weed in the respective areas. They are:

Very Low: 0 – 10% chance of invasion

Low: 11 – 30% chance of invasion

Medium Low: 31 – 50% chance of invasion

Medium High: 51 – 70% chance of invasion

High: 71 – 90% chance of invasion

Very High: 91 - 100% chance of invasion

The table below summarizes the requirements for the two extreme levels of invasion i.e. very low and very high.

Table 4: Invasion level criteria

Very Low	Very High
Has very high soil nitrogen content	Has very low soil nitrogen content
Has very high soil organic carbon	Has very low soil organic carbon
Has medium to high soil texture	Has very low soil texture

For the other classes in between the two extremes, the same criteria also apply but it falls in between the criteria that was used to determine the more extreme scenarios above.

5.3 Validation and testing of model results

The purpose of carrying out validation is to evaluate how accurate the model is in predicting striga occurrences in the region. To achieve this, 20% of the striga data acquired was set aside for validation purposes. Once this was done, the RF model was generated using 80% of the remaining data. Model predicted values were then extracted at the exact coordinates of the 20% of the test data. The extracted values were then exported in a table and merged with the test data.

Figure 20: Validation sample table

	A	B	C	D
1		Predicted	Actual Values	
2	1	59.39343	55	
3	2	63.45275	58	
4	3	64.8733	59	
5	4	69.24386	62	
6	5	72.52283	64	
7	6	73.1862	65	
8	7	87.1351	68	
9	8	90.97803	80	
10	9	103.5934	85	
11	10	109.8235	96	
12	11	127.0934	128	
13	12	144.5655	170	

The resultant excel file was imported into GeoDa software and the validation done. The validation accuracy was found to be 89%.

6 Conclusion

In conclusion, the study was able to analyze and identify the factors that may or may not contribute to the invasion of farmlands by striga weed and ultimately determine the risk of invasion throughout the Western Kenya region. It was determined that soil characteristics, specifically soil nitrogen, organic matter and texture are the major factors that influence the invasion of striga weed. Climatic variables such as temperature are equally important, though, to a lower extent. Regions around the lake region, especially the western parts of Kisumu, Homabay, and Siaya counties experience a higher risk of invasion while majority of Kakamega county and the eastern parts of Migori and Homabay county were found to have the least risk of invasion by striga. The results of this analysis were confirmed against observed striga data that was taken in the field and the validation accuracy was high.

An automated web-based weed modeling system was finally developed and tested to aid various governmental and non-governmental organizations in the region to easily and efficiently model any existing or new weed that may arise in the region presently or in the near future.

Finally, prediction of striga invasion both spatially and temporally will contribute a lot in establishing proper planning and management measures of controlling the weed. Continuous monitoring of soil characteristics in the region and weather forecasting for the next few seasons are essential for determining future risk prevalence of striga weeds in the Western Kenya region.

6.1 Recommendations

The following recommendations were made from the study:

1. Authorities and organizations within the region should take more action in supporting and encouraging farmers to focus on improving soil fertility in their farms through practices such as crop rotation, fertilizer application, intercropping among others.
2. Authorities within the region should carry out more frequent and extensive collection of spatial data on the occurrence of striga weed to aid in deeper scientific analysis of the problem with more accurate results.
3. Further research on the subject should be done to further quantify and analyze the risk of invasion of striga weeds and other weeds that are uprising in the region recently such as dodder weed.

7 References

Aflakpui, G. K. S., Gregory, P. J., & Froud-Williams, R. J. (1998). Effect of temperature on seed germination rate of Striga hermonthica (Del.) Benth. *Crop Protection*, *17*(2), 129–133. https://doi.org/10.1016/S0261-2194(97)00096-3

Amudavi, D., Khan, Z., & Pickett, J. (n.d.). *Enhancing the Push-Pull strategy*. 10–12.

Atera, E. A., Ishii, T., Onyango, J. C., Itoh, K., & Azuma, T. (2013). Striga Infestation in Kenya: Status, Distribution and Management Options. *Sustainable Agriculture Research*, *2*(2), 99. https://doi.org/10.5539/sar.v2n2p99

Ayongwa, G. C., Stomph, T. J., Belder, P., Leffelaar, P. A., & Kuyper, T. W. (2011). Organic matter and seed survival of Striga hermonthica - Mechanisms for seed depletion in the soil. *Crop Protection*, *30*(12), 1594–1600. https://doi.org/10.1016/j.cropro.2011.08.012

Capinha, C., & Anastácio, P. (2011). Assessing the environmental requirements of invaders using ensembles of distribution models. *Diversity and Distributions*, *17*(1), 13–24. https://doi.org/10.1111/j.1472-4642.2010.00727.x

De Groote, H., Wangare, L., Kanampiu, F., Odendo, M., Diallo, A., Karaya, H., & Friesen, D. (2008). The potential of a herbicide resistant maize technology for Striga control in Africa. *Agricultural Systems*, *97*(1–2), 83–94. https://doi.org/10.1016/j.agsy.2007.12.003

Hearne, S. J. (2009). Control - the striga conundrum. *Pest Management Science*, *65*(5), 603–614. https://doi.org/10.1002/ps.1735

Jarnevich, C. S., Holcombe, T. R., Barnett, D. T., Stohlgren, T. J., Kartesz, T., Jarnevich, C. S., … Kartesz, J. T. (n.d.). *Forecasting Weed Distributions using Climate Data : A GIS Early Warning Tool*. *3*(4), 365–375. https://doi.org/10.1614/IPSM-08-073.1

Kamara, A. Y., Ekeleme, F., & Omoigui, L. (2009). *Archives of Agronomy and Soil Science Influence of nitrogen fertilization on the performance of early and late maturing maize varieties under natural infestation with Striga hermonthica (Del .) Benth*. (December 2014), 37–41. https://doi.org/10.1080/03650340802342284

Khan, Z. R., Hassanali, A., Overholt, W., Khamis, T. M., Hooper, A. M., Pickett, J. A., … Woodcock, C. M. (2002). *CONTROL OF WITCHWEED Striga hermonthica BY INTERCROPPING WITH Desmodium spp ., AND THE MECHANISM DEFINED AS ALLELOPATHIC*. *28*(9), 1871–1885.

Khan, Z. R., Midega, C. A. O., Amudavi, D. M., Hassanali, A., & Pickett, J. A. (2008). On-farm evaluation of the "push-pull" technology for the control of stemborers and striga weed on maize in western Kenya. *Field Crops Research*, *106*(3), 224–233. https://doi.org/10.1016/j.fcr.2007.12.002

Khan, Z. R., Midega, C. A. O., Hassanali, A., Pickett, J. A., Wadhams, L. J., & Wanjoya, A. (2006). Management of witchweed, Striga hermonthica, and stemborers in sorghum, Sorghum bicolor, through intercropping with greenleaf desmodium, Desmodium intortum. *International Journal of Pest Management*, *52*(4), 297–302.

https://doi.org/10.1080/09670870600816991

MacOpiyo, . Laban, Vitale, J., & Sanders, J. (2010). *An ex-ante impact assessment of a Striga control programme in East Africa*. 123.

Oswald, A. (2005). Striga control - Technologies and their dissemination. *Crop Protection*, *24*(4), 333–342. https://doi.org/10.1016/j.cropro.2004.09.003

Oswald, A., & Ransom, J. K. (2001). *Striga control and improved farm productivity using crop rotation*. *20*, 113–120.

Palaniyandi, M., Anand, P. H., & Maniyosai, R. (2014). *Spatial cognition : A geospatial analysis of vector borne disease transmission and the environment , using rem .... Spatial cognition : a geospatial analysis of vector borne disease transmission and the environment ,*. (September).

Rodenburg, J., Cissoko, M., Kayeke, J., Dieng, I., Khan, Z. R., Midega, C. A. O., … Scholes, J. D. (2015). Do NERICA rice cultivars express resistance to Striga hermonthica (Del.) Benth. and Striga asiatica (L.) Kuntze under field conditions? *Field Crops Research*, *170*, 83–94. https://doi.org/10.1016/j.fcr.2014.10.010

Sanginga, N. and Woomer, P. L. (2009). Integrated Soil Fertility Management in Africa: Principles Practices and Development Process. In *Tropical Soil Biology and Fertility Institute of the International Centre for Tropical Agriculture. Nairobi.*

strides. (2008). (December 2007), 2008.

Study, F. (n.d.). *Feasibility Study on in Sorghum Striga Control*.

Woomer, P., Striga, A., & Project, M. (n.d.). *New Approaches to Controlling Striga Infestation*.