Analyzing the Current and Future Potential Distributions of Visceral Leishmaniasis in Kenya: A Spatial Model Approach

Patrick, Maureen Nzilani

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Statistical Science of Strathmore University

Strathmore Institute of Mathematical Sciences

Strathmore University

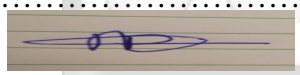
Nairobi, Kenya

June 2025

Declaration

I declare that this work has not been previously submitted and approved for award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Name: **Patrick Maureen Nzilani**

Signature: 

Date: May 28, 2025

Approval

The dissertation of Patrick Maureen Nzilani was reviewed and approved by the following:

Dr. Evans Otieno Omondi

Principal Supervisor,

Institute of Mathematical Sciences, Strathmore University.

Dr. Kennedy Senagi

Co-Supervisor,

Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu

Dean,

Institute of Mathematical Sciences, Strathmore University.

Prof. Bernard Shibwabo

Director,

Office of Graduate Studies, Strathmore University.

Abstract

Visceral Leishmaniasis (VL), is a vector-borne disease transmitted through sandfly bites and caused by protozoa parasites. The disease is mainly concentrated in poor and marginalized regions around the world. Currently, the global burden of the disease is borne by eastern African countries that share approximately 73% of the global VL cases. Kenya is one of the East African countries that shares in the huge burden of VL. While numerous models have been instrumental in comprehending VL transmission dynamics, persistent knowledge gaps underscore the necessity for more nuanced modeling. This study used a Bayesian spatial model with a specific focus on environmental, weather, population, and treatment data to analyze and predict VL transmission patterns in Turkana, Kenya. The findings from this study determined that age group 6-18 years, age group 0-5 years, age group 31-44 years, age group 19-30 years, sex (male), minimum temperature, land use land cover (LULC), and proximity to healthcare were significant risk factors for VL in Turkana County. This study concludes that there is an interplay of various factors that drive the transmission dynamics of the disease in Turkana County, thus emphasizing the need for targeted interventions to mitigate disease transmission.

KEY WORDS: Neglected tropical disease, Bayesian spatial model, Visceral Leishmaniasis, Future hotspots

Table of contents

List of figures vii

List of tables viii

List of abbreviations ix

List of publication x

List of conference xi

Acknowledgement xii

Dedication xiii

1 Introduction 1

1.1 Background to the study 1

1.2 Statement of the problem 3

1.3 Research objectives 3

1.3.1 General objective 3

1.3.2 Specific objectives 4

1.4 Justification of the study 4

1.5 Significance of the study 4

2 Literature Review 5

2.1 Introduction 5

2.2 Overview of VL models 5

2.3 Summary of VL models 9

2.4 Summary 12
2.5 Current research 13

3 Methodology **14**
3.1 Introduction 14
3.2 Study area 14
3.3 Data set 14
3.4 Data analysis 15
3.5 Model 16
3.6 Evaluation of model performance 19

4 Results and Interpretation **21**
4.1 Introduction 21
4.2 Exploratory data analysis 21
4.3 Correlation 23
4.4 Bayesian spatial model 24
4.5 Model performance 25
4.6 Prediction and mapping future hotspots for VL 26

5 Discussions, Conclusions and Recommendations **28**
5.1 Introduction 28
5.2 Discussion 28
5.3 Conclusion 31
5.4 Recommendations 31
5.4.1 Recommenation for future studies 31
5.4.2 Policy recommendations 32
5.5 Limitations 32

References **33**

Appendix A Ethical approval **41**

Appendix B Similarity index **42**

Appendix C R code 43

Appendix: Full R Code 43

List of figures

Figure 3.1: A map of Kenya showing the location of Turkana County. 15

Figure 4.1: Correlation Matrix 24

Figure 4.2: December 2024 prediction and MAP 27

List of tables

Table 2.1: Summary of Studies on Spatio-Temporal Analysis of VL 10

Table 3.1: Description of the variables used for exploratory data analysis and modeling. 16

Table 4.1: The number and percentage distribution of positive visceral leishmaniasis cases diagnosed and treated in Turkana County hospitals, including the patients' original residence. 22

Table 4.2: Distribution of positive and negative visceral leishmaniasis cases by age group and gender 23

Table 4.3: Summary statistics 25

Table 4.4: Bayesian spatial model performance 26

List of abbreviations

EDA	Exploratory Data Analysis
LULC	Land Use Land Cover
NDVI	Normalized Difference Vegetation Index
VL	Visceral Leishmaniasis

List of publication

1. Spatial Statistical Modelling of Visceral Leishmaniasis to Unearth its Distribution Dynamics and Predict High-Risk Landscapes in Turkana County in Kenya. Manuscript submitted to the Infectious Disease Modelling journal; currently under peer review by the editorial team.

List of conference

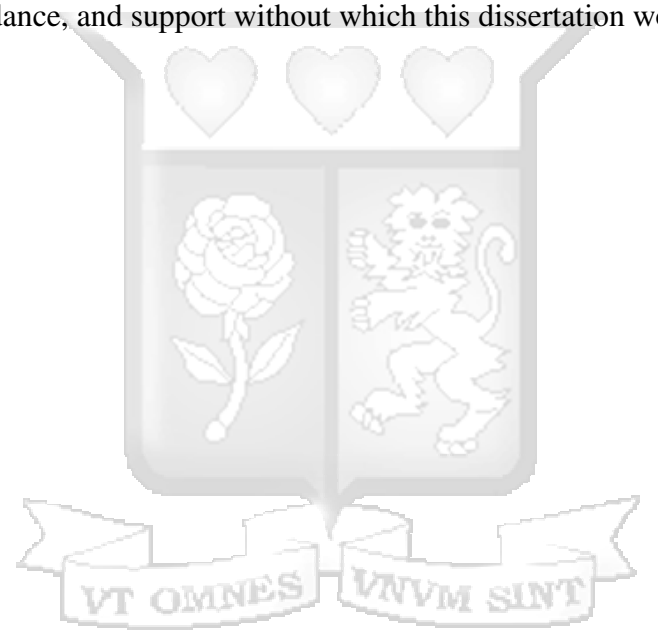
1. 15th KASH Conference: Kenya Medical Research Institute (KEMRI) Annual Scientific and Health (KASH) Conference, held at Safari Park Hotel, Nairobi, Kenya, 11–14 February 2025.

Acknowledgement

First and above all, I am grateful to the Almighty God for His provision, grace, and for granting me good health throughout the study period.

I am grateful Kimani Freeman and John Mwaura for their unwavering support and cheering me on throughout the study period.

I am indebted to my supervisors, Dr. Evans Otieno Omondi and Dr. Kennedy Senagi for their availability, kind guidance, and support without which this dissertation would not have been a success.



Dedication

This dissertation is dedicated to Kymora, Wangari, Mumbua, Pendo,Makena, Wanjiru,Nantale and Mama.

Chapter 1

Introduction

1.1 Background to the study

Visceral Leishmaniasis (VL) also known as kala-azar is a neglected tropical disease. The disease is transmitted through the bites of infected female phlebotomine sand flies, which feed on blood to produce eggs (World Health Organization, 2023). Over 70 animal species including humans can be hosts or reservoirs hosts of the leishmania parasite (World Health Organization, 2023).The transmission dynamics of VL extend beyond its causative agent and mode of transmission and also is influenced by a myriad of risk factors that play pivotal roles in its persistence and spread within populations. Environmental conditions, such as temperature, relative humidity, and proximity to natural vegetation, notably Acacia-Balanite trees, are strongly linked to the occurrence of the disease(Jones and Welburn, 2021a; Li and Zheng, 2019; Valero and Uriarte, 2020). In addition, socio-economic activities such as pastoralism and poor housing have been shown to be risk factors for the disease (Jones and Welburn, 2021a; Li and Zheng, 2019; Valero and Uriarte, 2020).

Global World Health Organization (WHO) estimates show that 50,000 to 90,000 new cases of VL occur annually. Ninety percent of those infections are concentrated in poor and impoverished communities of South East Asia, East Africa, and Brazil (Kanyina, 2020a). Despite VL being curable, it has a mortality rate of over 95% due to late diagnosis, poor treatments, poor case management and failure of treatment altogether (Alvar et al., 2012).

Recent reports shows that eastern Africa now bears the largest burden accounting for 73% of global VL cases, with children 15 years and below bearing half of that burden (World Health organization, 2024). Kenya, Ethiopia, South Sudan, Somalia, and Sudan rank among the top 10 countries globally for VL cases and deaths (Makau-Barasa et al., 2022). In Kenya, VL

remains a significant public health concern, particularly in marginalized and underserved arid and semi-arid areas. Eleven out of the 47 counties in Kenya are endemic, with Turkana, West Pokot, Baringo, Isiolo, Marsabit, Kitui, Wajir, and Garissa being the most affected (Grifferty et al., 2023). The disease was first documented in the country in 1935 in Mandera and Wajir counties Grifferty et al. (2023), and its transmission is influenced by a complex interplay of socio-economic, environmental, and climatic factors. Climate change, human migration, conflict, food insecurity, and weak health systems contribute to new VL foci, persistence, and sporadic outbreaks (Abdullahi et al., 2022; Jones and Welburn, 2021b; Makau-Barasa et al., 2022). Kenya's diverse landscape, characterized by varying ecological settings, provides suitable habitats for sandfly vectors and reservoir hosts, further facilitating disease spread.

Despite the high burden of VL, the actual disease prevalence and spatial distribution remain poorly documented due to factors such as low case detection, poor health records, and weak surveillance systems (Grifferty et al., 2023; World Health organization, 2024). Limited healthcare infrastructure and the high cost of treatment further hinder disease management. Although several epidemiological and clinical studies have explored diagnostic capacities, treatment modalities, and disease outbreaks Grifferty et al. (2023); Kanyina (2020b); Ouma and Mulambalah (2021); van Dijk et al. (2024), they often fail to capture the intricate interactions between human population density, socio-economic disparities, and climatic conditions that shape VL transmission. As a result, VL control efforts remain largely reactive rather than proactive, and asymptomatic individuals continue to play a role in disease transmission by acting as reservoirs for sandfly vectors (Government of Kenya, 2021). Given these challenges, the WHO Strategic Framework for VL elimination in eastern Africa emphasizes the need for robust disease surveillance (World Health Organization, 2024). To address this gap, this study employs spatial distribution models, including logistic regression and Bayesian approaches, to predict VL occurrence in Turkana County, considering patient diagnosis, treatment, demographic, weather, and environmental factors.

1.2 Statement of the problem

VL currently constitutes a critical public health challenge in Kenya. Eleven out of the 47 counties in Kenya are endemic for VL. However, the most affected counties include Turkana, West Pokot, Baringo, Isiolo, Marsabit, Kitui, Wajir, and Garissa (Grifferty et al., 2023). However, despite efforts to control the disease, factors such as increase in population and movement, climate, environment, and urbanization have been found to influence the endemicity of the disease. Thus, the disease appears to be emerging and reemerging within and outside the known foci points. As a result, the exact status of the VL endemicity and the factors that influence its transmission dynamics are not adequately understood. Traditional epidemiological methods face limitations in capturing the complex interactions between environmental, socio-economic, and population density factors driving VL transmission. Consequently, there is a critical need for more sophisticated modeling approaches to predict the current and future potential distributions of VL in Kenya.

This study seeks to use a Bayesian spatial model to help understand VL's transmission dynamics. This model is particularly useful as it offers a perspective as to how population, environmental, and weather factors interact within a spatially explicit framework, and thus capture the complex dynamics of VL transmission in Kenya. By integrating these diverse factors into a unified modeling framework, this study endeavors to provide insights into the current distribution of VL in Kenya and to project the future spread of the disease.

1.3 Research objectives

1.3.1 General objective

The main objective of the study is to design a Bayesian spatial model that incorporates population, environmental, and weather factors to analyze and evaluate the spread of Visceral Leishmaniasis in Kenya.

1.3.2 Specific objectives

1. To develop a Bayesian spatial model to analyze and evaluate the current spread of VL in Turkana, Kenya

2. To predict and map future potential infection risk zones of VL in Turkana, Kenya

1.4 Justification of the study

The study's importance lies in its potential to address the gaps in our understanding of visceral leishmaniasis in Kenya and inform evidence-based strategies for disease control and prevention. By developing a Bayesian spatial model, the study aims to aims to access the interactions between population, environmental, and weather factors within a spatially-explicit framework. This approach offers a nuanced understanding of VL transmission dynamics, allowing for the integration of diverse data sources and the exploration of different scenarios. Furthermore, with the looming threat of climate change altering environmental conditions, the study's insights into the impact of climate variability on VL transmission are of particular relevance. Ultimately, the study seeks to contribute to the broader efforts aimed at mitigating the burden of VL in Kenya and improving the health outcomes of affected populations.

1.5 Significance of the study

This study is expected to analyze the distribution of VL in Turkana County and further predict and map future infection risk zones in the county. By utilizing a spatial distribution model with demographic, environmental, population, and hospital data, the study provides valuable insights for targeted interventions. Additionally, the study supports the development of future risk maps that are accessible to stakeholders and inform them about emerging hotspot areas. This proactive approach strengthens disease monitoring, facilitates timely response measures, and ultimately contributes to reducing VL transmission and burden in Turkana, County.

Chapter 2

Literature Review

2.1 Introduction

This literature review explores diverse models utilized in understanding the transmission dynamics of VL. This review synthesizes findings to explain the complex relationship between hosts, vectors, population, weather, and environmental factors influencing the disease spread. Through a comprehensive analysis of these modeling approaches, this review seeks to give a nuanced understanding of the strengths and limitations of various models as well as the implications for public health interventions, shedding light on the intricate dynamics of VL transmission and the challenges researchers face in accurately capturing and predicting its complexities.

2.2 Overview of VL models

Zheng, Canjun and Wang, Liping and Li, Yi and Zhou, Xiao-Nong (2020) conducted a spatio-temporal analysis to explore Visceral Leishmaniasis between 2004 and 2018 in China. Over the 15 years, the study found that nearly 90% of the VL cases were concentrated in Gansu, Sichuan, and Xinjiang. The study also uncovered that meteorological factors such as ambient temperature were associated with VL risk. Despite the study being able to successfully identify risk areas for VL in China, the study was unable to identify the distributions, density, and seasonal characteristics of how humans were affected by sandflies. Knowledge of the habits of sandflies was also limited which limited the analysis of the transmission of the disease.

Rajabi et al. (2017) conducted a study to identify Visceral Leishmaniasis-Susceptible Areas using Spatial Modelling in the Southern Caucasus region of Iran, Azerbaijan, and Armenia. The study utilized knowledge and data-driven methods namely weights of evidence, fuzzy logic, and logistic regression. The study showed that fuzzy logic was the best method for identifying the susceptible areas for VL incidence. Findings from the study also revealed that riverside, irrigated farming, and orchard areas contributed greatly to susceptibility to the disease. In the rural nomadic areas, livestock was shown to have the potential to contribute and maintain vector density. The study areas were also noted to be underdeveloped and undeveloped and thus it was expensive and time-consuming to collect data on sandfly density and reservoir. As a result, the inadequate data limited the validation of the study model and restricted the study from fully exploring all dynamics that are associated with the spread of the disease.

Moirano et al. (2022) investigated the spatio-temporal patterns and meteorological-climatic factors influencing the occurrence of visceral leishmaniasis (VL) in Italy over eight years. Bayesian and Poisson models were utilized to analyze the spatial patterns. The study determined that climatic change influenced the spread of VL in Italy. The spread of the disease was found to be heterogeneous in Italy with the Sicilian Islands experiencing the highest incidence rates. In particular, an increase in temperature was associated with an increase in the spread of VL cases. This is because an increase in temperature tends to shorten vector development time and increase vector biting rate, vectorial capacity, and parasite replication within the vector. An increase in precipitation was associated with a decrease in the incidences of VL. However, Moirano is keen to point out that the association between VL and precipitation depends on the geographical area under study. The study's short period of disease observation and the small number of incident cases of the disease limited the study's statistical power.

Ben-Ahmed (2021) examined the spatial distribution of childhood visceral leishmaniasis (VL) in Tunisia. The study conducted a spatial analysis in which he looked at precipitation, continentality index, and pluviometric coefficient of Emberger as predictors to model the geographical distribution of VL. Precipitation was found to have a positive influence on VL incidence

rates. The study established a negative association between VL incidence rates and both the continentality index and the pluviometric coefficient of Emberger. The cross-sectional study concentrates on the period 2006-2016. Therefore, although it gives valuable information on the spatial distribution during this period, the study fails to show how this distribution could change over time. The study focuses on children under 5 years of age and thus limits the study's ability to explore how VL affects Tunisian people in general.

Galgamuwa et al. (2018) used multiple regression and autoregression analysis to explore the spatial distribution and seasonal variations of VL in Sri Lanka for the period 2009-2016. The study deduced that in some districts where VL was endemic, the disease was correlated with average wind speed, humidity, sun hours, and maximum temperature. Further, it was noted that the disease had a seasonal peak from July to September in the north-central region and in the southern region the disease peak was recorded from October to December. The study experienced a few limitations such as the true burden of the disease was not adequately captured in districts where the disease was localized in specific clusters. Since the disease cases were reported only at the district level, the study could not capture more detailed correlations that might have emerged if data were available for subdivisions within the districts. Further, the weather variables were sourced from a single station and failed to account for micro-climatic and habitat variations within districts. Consequently, the results of the study make critical assumptions on critical environmental and spatial factors, which could lead to distortion of the disease epidemiology.

Kesari et al. (2010) conducted a cross-sectional study in areas both endemic and non-endemic for VL in India. Their study utilized multilevel logistic regression analysis to identify household characteristics that influenced the suitability of VL in the areas of study. Findings from their study pointed out that mixed dwelling spaces (i.e areas where cattle sheds were close to people's houses), mud-plastered walls, and areas of study influenced the presence of vectors that caused VL. This study highlighted the importance of sandflies surveillance which is crucial in formulating a strategy for controlling the disease. However, it would have been paramount for

the study to also incorporate environmental and geographical factors to assess the relationship between the vectors and risk of infections in humans.

In Sudan Nackers et al. (2015) examined households and individual risk factors of VL in 24 endemic villages. While adopting an exploratory and multi-logistic regression analysis their study identified a myriad of risk factors associated with VL such as household size, sleeping location (whether indoor or outside), outdoor activity during the rainy season, type of animal repellent used, presence of Acacia nilotica, presence of dogs among other factors. The findings of this study give valuable insights and practical suggestions for guiding future interventions of VL in Sudan. However, due to the exploratory approach used by the study, some findings from the study cannot be significantly substantiated as true risk factors associated with VL in the region.

Jiang et al. (2021a) explored spatio-temporal risk factors associated with VL in 11 provinces of China to understand the disease spatial heterogeneity and predict future risk zones of the disease. Their study made use of boosted regression trees and spatial mapping techniques to explore how various geographical, ecological, socio-economic, and climate spatial covariates influenced the spread of VL. Their research determined that elevation, minimum temperature, relative humidity, and annual accumulated precipitation were the greatest contributors to the spatial heterogeneity of VL.

Subramanian et al. (2024) used a Bayesian model to analyze spatio-temporal patterns of VL in two endemic states in India from 2013 to 2022. Their study identified weather variables associated with the disease, including positive associations with mean temperature, minimum temperature, precipitation, land surface temperature, isothermality, and negative associations with maximum temperature. Additionally, soil moisture and the enhanced vegetation index emerged as crucial environmental factors influencing VL, while population density was the only demographic factor linked to the disease. This study underscore the complexity and

multifaceted nature of VL transmission and the need for further research in specific contexts.

To examine the recurrent nature of VL in Central China Zhao et al. (2021) scrutinized spatiotemporal and driving factors associated with VL in the region. Their work used a general additive model to investigate environmental, meteorological, and socio-economic covariates influence on the spread of VL. The drivers associated with VL in the study included; changes in grassland/forests, humidity, temperature, and population density. The study points out that vectors and dogs have been established as important driving factors of VL in the region by other researchers. However, due to data constraints, their study was not able to integrate these variables in the study.

Bhunia et al. (2010) research explored how altitude, temperature, humidity, rainfall, and the normalized difference vegetation index (NDVI) affected the incidence of VL in four Indian states particularly; Bihar, West Bengal, Uttar Pradesh, and Jharkhand during the period 2005-2007. This study made use of topography modeling (digital elevation model) to explore the relationship between the various independent variables and the incidence of the disease. Their investigation revealed that the disease was prevalent in regions characterized by low NDVI values (0.03–0.015), rainfall ranging from 1154 to 1834 mm, altitudes below 150 meters, and maximum and minimum temperatures within the range 25-29°C and 16-20°C respectively.

2.3 Summary of VL models

Table 2.1 gives a summary of various models that have been explored in studying VL and gaps experienced by the studies

Table 2.1: Summary of Studies on Spatio-Temporal Analysis of VL

Study	**Key findings**	**Gaps**
Zheng, Canjun and Wang, Liping and Li, Yi and Zhou, Xiao-Nong (2020)	Spatio-temporal analysis of Visceral Leishmaniasis in China (2004–2018). 90% of cases in Gansu, Sichuan, and Xinjiang. Meteorological factors like temperature were associated with VL risk.	Lack of sufficient data prevented analysis of sandfly distributions and human-sandfly interactions. Limited knowledge of sandfly habits restricted transmission analysis.
Rajabi et al. (2017)	Identified VL-susceptible areas in the Southern Caucasus using weights of evidence, fuzzy logic, and logistic regression. Fuzzy logic performed best in identifying susceptible areas.	Underdeveloped study areas made sandfly density and reservoir data collection expensive and time-consuming. Limited data restricted model validation.
Moirano et al. (2022)	Investigated VL spatio-temporal patterns and meteorological factors in Italy using Bayesian and Poisson models. Found that climate change influenced VL spread.	Short observation period and small number of VL cases limited statistical power.
Ben-Ahmed (2021)	Studied spatial distribution of childhood VL in Tunisia. Precipitation increased VL incidence, while continentality index and pluviometric coefficient of Emberger had negative effects.	Lacked temporal distribution of VL and focused only on children under 5, limiting generalizability.

Study	Key findings	Gaps
Galgamuwa et al. (2018)	Used multiple regression and autoregression to analyze VL in Sri Lanka (2009–2016). Found correlations with wind speed, humidity, sun hours, and max temperature.	Did not capture full disease burden in cluster-concentrated areas. Weather data from a single station failed to account for microclimatic variations.
Kesari et al. (2010)	Used multilevel logistic regression to identify household characteristics influencing VL transmission. Mixed dwelling spaces, mud-plastered walls, and cattle shed proximity were key risk factors.	Did not incorporate environmental and geographical factors, limiting VL risk analysis.
Nackers et al. (2015)	Examined household and individual VL risk factors in 24 endemic villages. Identified risks: household size, sleeping location, outdoor activity, presence of Acacia nilotica and dogs.	Exploratory approach limited ability to confirm significant risk factors.
Jiang et al. (2021b)	Explored spatio-temporal VL risk factors in 11 Chinese provinces using boosted regression trees. Elevation, min temp, humidity, and precipitation contributed to spatial heterogeneity.	Did not capture long-term trends beyond the study period. Focused only on Western and Central China, leaving gaps in VL understanding elsewhere.
Subramanian et al. (2024)	Used a Bayesian model to analyze VL spatio-temporal patterns in two Indian states (2013–2022). Identified weather and environmental factors influencing VL transmission.	Considered environmental, bioclimatic, and demographic factors but excluded vector and reservoir host dynamics.

Study	Key findings	Gaps
Zhao et al. (2021)	Examined VL spatio-temporal drivers using a General Additive Model. Found effects of grassland/forest changes, humidity, temperature, and population density.	Lacked vector and dog data due to constraints, limiting understanding of VL transmission.
Bhunia et al. (2010)	Studied influence of altitude, temperature, humidity, rainfall, and NDVI on VL incidence in four Indian states. Found VL in low NDVI areas, specific rainfall/temperature ranges, and altitudes <150m.	Focused on topography but ignored socio-economic factors, human movement, and vector ecology.

2.4 Summary

Statistical models stand out as robust tools for modeling and mapping the spread of VL. These models enable a comprehensive understanding of VL distribution in areas with limited field data, inform targeted interventions, and assess the impact of various risk factors on disease patterns. Nonetheless, some of these models face persistent challenges in terms of data quality and availability, oversimplification of risk factors, and inherent uncertainty in modeling. These drawbacks underscore the importance of exploring other models such as the Bayesian spatial model and complementary methodologies to improve both the accuracy and reliability of VL distribution mapping and forecasts for successful public health initiatives.

2.5 Current research

This dissertation seeks to develop a Bayesian spatial model to address the pressing need for a comprehensive understanding of Leishmaniasis dynamics in Turkana, Kenya. Despite Bayesian models being successfully used in various regions to estimate the distribution of vector-borne disease in the Kenyan context, it remains largely unexplored. It is therefore paramount to employ a Bayesian spatial model approach to help explore the intricate relationship between population, weather and environmental factors, and the transmission of VL. By identifying high-risk areas and predicting potential shifts in VL distribution, this research will contribute crucial insights for public health planning, guiding targeted interventions, and aiding policymakers in developing strategies to mitigate the impact of Leishmaniasis in Turkana, Kenya.

Chapter 3

Methodology

3.1 Introduction

This section provides an overview of the methodology to be utilized by the study. Specifically, this chapter gives an overview of the data set, variables, model, and model performance.

3.2 Study area

Figure 3.1 shows the area of the study, Turkana County, located in the northwest of Kenya. Turkana is the second largest county in Kenya with an area of 77,000 km^2 covering approximately 13% of country's surface (Turkana County Government, 2024a).

3.3 Data set

Data on human VL cases was collected from 12 different public hospitals across Turkana County for the period 2019 and 2020. The secondary data on human VL cases was sourced International Centre of Insect Physiology and Ecology (icipe). A total of 1,674 records were collected. Patient treatment data collected included the unique (anonymized) identity number of the patient, the name of the hospital visited, the date the rk39 RDT test was done, the results of the test, and the date treatment was initiated. Demographic data captured at the hospital include the age of the patients, gender, and their area of residence. Considering the period when the data was collected, respective weather data (i.e., minimum temperature, average temperature, maximum temperature, humidity, and total precipitation) was sourced from Open-Meteo Open

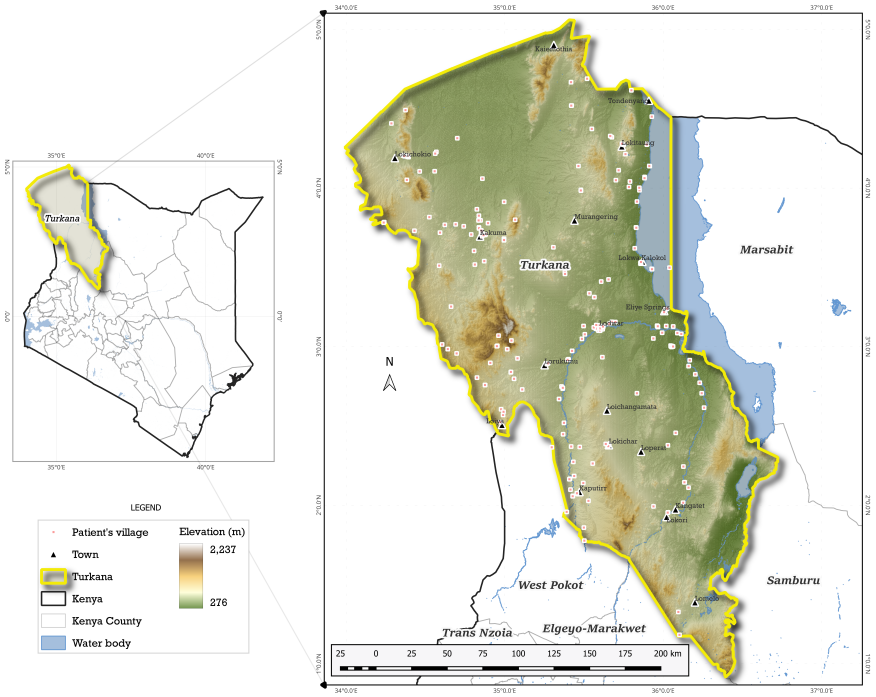


Figure 3.1: A map of Kenya showing the location of Turkana County.

Meteo (2024) and Envidat (Karger et al., 2019). Moreover, environmental data (i.e., canopy height, location of water body (e.g., rivers, lakes, wetlands, and reservoirs), LULC was obtained from National Aeronautics and Space Administration (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS) Google Earth Engine (2024)), and NDVI was collected from NASA MODIS Google Earth Engine (2024). Population density data was also extracted from Turkana raster maps from NASA Earth Data (CIESIN, 2024).

The set of variables used by the study is summarized in Table 3.1.

3.4 Data analysis

The initial data analysis stage comprised data pre-processing which involved cleaning the data set by removing outliers and imputing missing values for the age variable. The hospital dataset contained the month and year of diagnosis of VL patients. The study then lagged the dataset (based on the patient's date of diagnosis) up to 8 months to mimic the disease incubation period; which can be to 8 months (Piscopo and Mallia Azzopardi, 2007; Scarpini et al., 2022).The

Table 3.1: Description of the variables used for exploratory data analysis and modeling.

Item	Variable	Description
1	Hospital geographic coordinate	Tthe latitude and longitude of the hospital
2	Village geographic coordinate	The latitude and longitude of the village
3	Date of diagnosis	The date the patient was tested for VL
4	Distance to water bodies	The Euclidean distance between the coordinates of the patient's village and the nearest water body; measured in kilometers (km)
5	Gender	The sex of the patient; recorded as either male or female
6	Population density	The number of people per km^2
7	Minimum temperature	The lowest temperature recorded; measured in degrees Celsius (°C)
8	Maximum temperature	The highest temperature recorded; measured in °C
9	Mean temperature	The average temperature recorded; measured in °C
10	Elevation	The height above mean sea level; measured in meters (m)
11	Humidity	The amount of water vapor in the air; measured in percentage (%)
12	Precipitation	The total monthly rainfall; measured in millimeters (mm)
13	Age	The number of years the patient had lived at the time they visited the hospital
14	Proximity to healthcare	The Euclidean distance between the geographic coordinates of a patients' village and the hospital they visited; measured in km
15	Canopy height	The average height of trees; measured in meters (m)
16	NDVI	The difference between near-infrared and red reflectance divided by their sum
17	LULC	Classification of human activities and natural elements on a landscape such as tree cover, scrubland, grassland, cropland, built up, sparse vegetation, and wetland

environmental, population, and weather data was then merged with the hospital data based on the date of diagnosis for the patient and the respective lag dates. Exploratory data analysis (EDA) was then conducted to get a broader understanding of the variables utilized in the study. Correlation was also done to analyse the relationship between the variables. The study then fit a Bayesian spatial model to explore the distribution of VL during the study period and predict future distribution of the disease.

3.5 Model

The Bayesian spatial model integrates spatial correlation into data analysis across geographical regions, making it particularly effective for data exhibiting spatial dependence, where nearby locations share similar outcomes due to unobserved factors. Bayesian inference estimates

posterior distributions by updating prior distributions with observed data. While Markov Chain Monte Carlo (MCMC) methods have traditionally been used for this purpose, they can be computationally intensive, especially in high-dimensional parameter spaces like spatial data analysis, due to long burn-in periods and the need for subsampling to ensure convergence. To address these limitations, More efficient methods such as integrated nested Laplace approximations (INLA) and stochastic partial differential equation (SPDE) approaches have emerged. INLA, in particular, provides a computationally efficient alternative for latent Gaussian models, enhancing the feasibility and speed of model fitting in complex, high-dimensional analyses (Moraga, 2019). The model has numerous applications in various fields such as disease modeling Damien et al. (2022); Kang et al. (2016); Wang et al. (2024) and environmental studies (Gacutan et al., 2024; Roth et al., 2023).

In this study, the Bayesian spatial model was used to explore the distribution of VL and its predict future distribution in Turkana, county. VL has an incubation period of 8 months. Thus the Bayesian spatial model was fit for the 8 months (that is the month of diagnosis and 7 months before the diagnosis). To account for spatial autocorrelation in the study, a spatial covariate function that represented the location of the patients was factored into the model. The month of infection was also accounted for in the model as a random effect to account for the monthly variability of the disease. The study further assumes that each individual is characterized by the true probability $P(\mathbf{x}_i)$ of having the disease. Given this probability, the outcome variable Y_i, representing whether an individual i has the disease ($Y_i = 1$) or not ($Y_i = 0$), follows a binomial distribution:

$$Y_i \mid P(\mathbf{x}_i) \sim \text{Binomial}(1, P(\mathbf{x}_i)). \tag{3.1}$$

The probability $P(\mathbf{x}_i)$ of having the disease is then related to its linear predictor through the logit link function:

$$\text{logit}(P(\mathbf{x}_i)) = \beta_0 + \mathbf{D}\beta + S(\mathbf{x}_i) + u_i, \tag{3.2}$$

where β_0 denotes the intercept, $\mathbf{D}$ is a design matrix with rows corresponding to the covariate data for each individual, and

$$\begin{aligned}\beta = (&\beta_{\text{Temp Min}}, \beta_{\text{Sex}}, \beta_{\text{NDVI}}, \beta_{\text{LULC}}, \beta_{\text{Age Group}},\\ &\beta_{\text{Proximity to Healthcare}}, \beta_{\text{Mean Humidity}}, \beta_{\text{Canopy Height}},\\ &\beta_{\text{Total Precipitation}}, \beta_{\text{Population Density}}, \beta_{\text{Distance to Water}}),\end{aligned}$$

where each β represents the effect of the corresponding covariate. u_i represents the random effect corresponding to the month of infection. $S(\mathbf{x}_i)$ is a spatial random covariate associated with the location of $\mathbf{x}_i$ and follows a zero-mean Gaussian process with Matérn covariance function Moraga (2019).

$$\text{Cov}(S(\mathbf{x}_i), S(\mathbf{x}_j)) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\kappa\|\mathbf{x}_i - \mathbf{x}_j\|\right)^{\nu} K_{\nu}\left(\kappa\|\mathbf{x}_i - \mathbf{x}_j\|\right), \tag{3.3}$$

In this equation, K_ν refers to the modified Bessel function of the second kind, with order ν, where $\nu > 0$ represents the smoothness parameter. The parameter σ^2 denotes the variance, and $\kappa > 0$ is related to the practical range $\rho = \sqrt{\frac{8\nu}{\kappa}}$, which corresponds to the distance at which the spatial correlation diminishes to approximately 0.1 Moraga (2019). Introducing this term incorporates spatially correlated random effects, implying that prevalence estimates at a given location are more likely to be influenced by neighboring points, which is a common feature in geospatial data analysis. INLA computes approximations to the marginal posteriors of the parameters $\mathbf{w} := (\beta_0, \beta, S)$ and their corresponding hyper-parameters ϕ using the following integrals Equation 3.4 and Equation 3.5 Moraga et al. (2021):

$$\pi(w_a \mid \mathbf{x}) = \int \pi(w_a \mid \phi, \mathbf{x})\pi(\phi \mid \mathbf{x}) d\phi, \tag{3.4}$$

$$\pi(\phi_m \mid \mathbf{x}) = \int \pi(\phi \mid \mathbf{x}) d\phi_{-m}. \tag{3.5}$$

For $\pi(\phi \mid \mathbf{x})$, the Laplace approximation

$$\tilde{\pi}(\phi \mid \mathbf{x}) \propto \pi(\mathbf{w}, \phi, \mathbf{x}) \tilde{\pi}_G(\mathbf{w} \mid \phi, \mathbf{x}) \Big|_{\mathbf{w}=\mathbf{w}^*(\phi)} \tag{3.6}$$

is used, where $\tilde{\pi}_G(\mathbf{w} \mid \phi, \mathbf{x})$ is the Gaussian approximation to $\mathbf{w} \mid \phi, \mathbf{x}$, and $\mathbf{w}^*(\phi) = \arg\max_{\mathbf{w}} \pi_G(\mathbf{w} \mid \phi, \mathbf{x})$ is the posterior mode.

For $\pi(w_a \mid \phi, \mathbf{x})$, several choices exist, such as the Laplace approximation Moraga et al. (2021):

$$\tilde{\pi}(w_a \mid \phi, \mathbf{x}) \propto \pi(\mathbf{w}, \phi, \mathbf{x}) \tilde{\pi}_G(\mathbf{w}_{-a} \mid w_a, \phi, \mathbf{x}) \Big|_{w_a = w^*_{-a}(w_a, \phi)}. \tag{3.7}$$

The numerical integration is used to yield the following approximations to (3.4) and (3.5) Moraga et al. (2021):

$$\tilde{\pi}(w_a \mid \mathbf{x}) = \sum_m \tilde{\pi}(w_a \mid \phi_m, \mathbf{x}) \tilde{\pi}(\phi_m \mid \mathbf{x}) \Delta_m, \tag{3.8}$$

$$\tilde{\pi}(\phi_n \mid \mathbf{x}) = \sum_p \tilde{\pi}(\phi_p \mid \mathbf{x}) \Delta_p, \tag{3.9}$$

where Δ_m, Δ_p represent the area weights corresponding to ϕ_m, ϕ_p, respectively.

This study used the INLA and SPDE approaches to model the Bayesian model and incorporate spatial dependencies in the data efficiently.

The INLA and SPDE approaches were implemented in R-INLA package version 24.06.27 in R statistical software.

3.6 Evaluation of model performance

In this study 80% of the data were used for training the model and 20% reserved for testing. The performance of the models was evaluated using accuracy, precision, and recall which are some of the most commonly used evaluation metrics (De Diego et al., 2022; Sokolova and

Lapalme, 2009; Vujović et al., 2021).

Accuracy measures the overall correctness of the model, by assessing the proportion of true VL and non-VL cases that were correctly identified. Accuracy is mathematically defined as;

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.10}$$

Precision measures the proportion of true VL cases (true positives) among all cases that were predicted as positive VL cases. Precision is mathematically denied as;

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3.11}$$

Recall measures the ability of the model to correctly identify all actual VL cases. Recall is mathematically defined as;

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3.12}$$

Chapter 4

Results and Interpretation

4.1 Introduction

This chapter presents the findings of the study, including exploratory data analysis, correlation analysis, spatial autocorrelation, the results of the Bayesian model and model performance. The purpose of these analyses is to uncover patterns, relationships, and spatial dependencies in the data, which are crucial for understanding the underlying dynamics of VL transmission in Turkana county. EDA provides a summary of the dataset through descriptive statistics offering insights into the distribution of the disease across different locations and also the distribution of the disease across different age groups and sex. Correlation analysis examines the relationships between variables to identify potential associations, while spatial autocorrelation assesses the extent to which similar values cluster together geographically. The Bayesian model results offers insights as to which variables are significant in explaining VL infections. The chapter also includes an interpretation of the findings, emphasizing their relevance to the study objectives. This study assumes that the presence of VL infection indicates the presence of sandflies responsible for transmitting the infection, either in the same area or nearby.

4.2 Exploratory data analysis

Data on VL cases were summarized to show the area of residence of the patients. Table 4.1 shows that the majority of the VL cases came from Turkana Sub-Counties. Turkana central, Turkana West and Loima Sub-Counties had the highest number of VL cases. However, it was noted that a few cases of VL, although reported and treated at Turkana hospitals, the patients'

origin of residence was from neighboring counties and countries.

Table 4.1: The number and percentage distribution of positive visceral leishmaniasis cases diagnosed and treated in Turkana County hospitals, including the patients' original residence.

Patient's county/country/sub-county	**Number of VL cases**	**Percent**
a) Neighboring Countries		
Ethiopia	20	3.3%
Uganda	2	0.3%
b) Turkana Sub-Counties		
Loima	104	17.2%
Turkana Central	222	36.7%
Turkana East	8	1.3%
Turkana North	64	10.6%
Turkana South	46	7.6%
Turkana West	128	21.2%
c) Neighboring Counties		
West Pokot	7	1.2%
Marsabit	4	0.7%
Total	**605**	**100%**

Table 4.2 (Section A) shows the distribution of VL cases by gender and age. Of the total VL cases, 74.54% were male, while 25.46% were female. This implied that the number of males infected was roughly three times greater than that of females. Among the male cases, the highest proportions were found in the 6-18 years age group (60.53%), and the lowest proportion was the 45 years and above (4.21%). In contrast, the distribution of VL cases in females was concentrated in the 6-18 years age group (44.81%) and the group that was least infected was 45 years and above (5.19%) age group.

Table 4.2 (Section B) revealed that 66.00% of the patients who tested negative were male, while

34.00% were female, suggesting that males tested negative at nearly twice the rate of females.

Table 4.2: Distribution of positive and negative visceral leishmaniasis cases by age group and gender

a) Positive

Age Group	Male		Female		VL Cases	
	Count	%	Count	%	Count	%
0–5 years	70	15.52%	42	27.27%	112	18.51%
6–18 years	**273**	**60.53%**	**69**	**44.81%**	**342**	**56.53%**
19–30 years	66	14.63%	22	14.29%	88	14.55%
31–44 years	23	5.10%	13	8.44%	36	5.95%
45+ years	19	4.21%	8	5.19%	27	4.46%
Total	**451 (74.54%)**	**100%**	**154 (25.46%)**	**100%**	**605**	**100**

b) Negative

Age Group	Male		Female		Negative Patients	
	Count	%	Count	%	Count	%
0–5 years	93	20.22%	51	21.52%	144	20.66%
6–18 years	**219**	**47.61%**	**92**	**38.82%**	**311**	**44.62%**
19–30 years	67	14.56%	50	21.10%	117	16.79%
31–44 years	31	6.74%	19	8.01%	50	7.17%
45+ years	50	10.87%	25	10.55%	75	10.76%
Total	**460 (66.00%)**	**100%**	**237 (34.00%)**	**100%**	**697**	**100**

4.3 Correlation

Figure 4.1 shows the results of the correlation of the variables used by the study. The correlation findings showed that there was a high correlation ($r \pm 0.7$) among maximum temperature, mean temperature, minimum temperature, and elevation. Due to the high correlation elevation, mean and maximum temperature were dropped from further analysis.

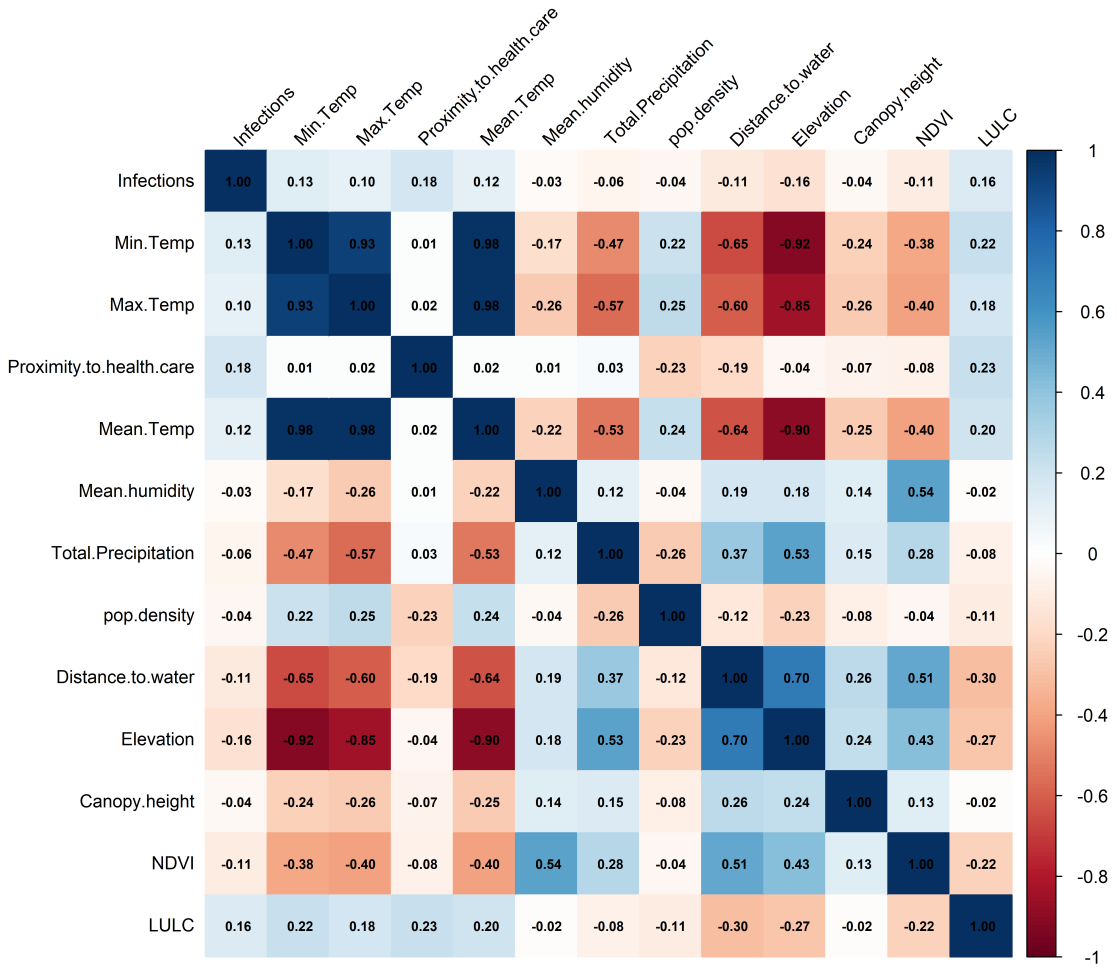


Figure 4.1: Correlation Matrix

4.4 Bayesian spatial model

For all possible 8 months of incubation with VL, Lag 2 (that is, the infection occurring 2 months before the diagnosis of the patients) was found to be the most optimal. The Bayesian spatial model was then used to identify significant variables that act as risk factors for the occurrence of VL in Turkana. The Bayesian spatial model relies on the 95% confidence interval to evaluate the significance of the predictor variables. A 95% confidence interval indicates the range within which we expect the true parameter to lie 95% of the time. If the value zero (0) falls in this range, it suggests no significant effect of the predictor on the outcome.Table 4.3 highlights that significant factors associated with VL infections include sex (male), proximity to healthcare facilities,

age groups (0-5 years, 6-18 years, 19-30 years, 31-44 years), minimum temperature, and LULC.

Table 4.3: Summary statistics

```
## Fixed effects:
##                   mean    sd 0.025quant 0.5quant 0.975quant   mode kld
## (Intercept)     -2.908 0.829    -4.528   -2.910     -1.274 -2.910    0
## SEXMALE          0.331 0.148     0.040    0.331      0.622  0.331    0
## distance         0.006 0.003     0.000    0.006      0.009  0.006    0
## age_group0-5     0.979 0.302     0.388    0.979      1.571  0.979    0
## age_group19-30   0.916 0.312     0.382    0.916      1.529  0.916    0
## age_group31-44   0.955 0.369     0.226    0.954      1.679  0.954    0
## age_group6-18    1.313 0.280     0.762    1.312      1.862  1.312    0
## tempmin          0.139 0.059     0.026    0.139      0.254  0.139    0
## NDVI            -1.442 0.804    -3.585   -1.430      0.103 -1.430    0
## mean_humidity    0.011 0.009    -0.007    0.011      0.030  0.011    0
## total_precip    -1.028 31.619  -63.029   -1.028     60.973 -1.028    0
## population_dens. 0.000 0.001    -0.002    0.000      0.001  0.000    0
## Distance_to_W..  0.005 0.004    -0.003    0.004      0.013  0.004    0
## forest_height    0.015 0.017    -0.019    0.015      0.050  0.015    0
## LULC             0.057 0.028     0.002    0.057      0.111  0.057    0

## Random effects:
##   Name   Model
##  spatial SPDE2 model
##  month   IID model

## Model hyperparameters:
##                       mean     sd 0.025quant 0.5quant 0.975quant   mode
## Theta1 for spatial -4.04 1.22e+00    -6.57   -4.00      -1.76  -3.82
## Theta2 for spatial  3.15 7.18e-01     3.13    3.13       4.64   3.02
## Precision for month 21819.83 2.38e+04  1445.63  14359.03  84948.78 3939.40

## Deviance Information Criterion (DIC) ............... 1365.19
## Deviance Information Criterion (DIC, saturated) ..... 1363.69
## Effective number of parameters .....................   30.01

## Watanabe-Akaike information criterion (WAIC) ........ 1365.80
## Effective number of parameters .....................   29.57

## Marginal log-Likelihood: -765.89
## is computed
## Posterior summaries for the linear predictor and the fitted values are computed
## (Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```

4.5 Model performance

The model performance was evaluated using accuracy, precision and recall metrics. The higher the numbers of accuracy, precision and recall the better the model is at performance. Table 4.4 shows the Bayesian spatial model at lag 2 has the highest performance scores with accuracy, precision, and recall of 0.7423, 0.7155, and 0.7094 respectively.

Table 4.4: Bayesian spatial model performance

	Bayesian spatial model		
Lag	**Accuracy**	**Precision**	**Recall**
1	0.7038	0.6724	0.6667
2	**0.7423**	**0.7155**	**0.7094**
3	0.7385	0.7207	0.6838
4	0.7038	0.6818	0.6410
5	0.7231	0.7064	0.6581
6	0.7154	0.6972	0.6496
7	0.7154	0.6972	0.6496
8	0.7269	0.7130	0.6581

4.6 Prediction and mapping future hotspots for VL

Figure 4.2 presents the predicted risk map for visceral leishmaniasis (VL) transmission in Turkana County for December 2024. This map was generated using key environmental and demographic variables, including minimum temperature, land use/land cover (LULC), population density, total precipitation, normalized difference vegetation index (NDVI), and distance to water bodies. These variables formed the basis of the predictive model, which was subsequently deployed through a web-based interface maintained by the International Centre of Insect Physiology and Ecology (ICIPE). The variables age group, sex, and proximity to healthcare facilities were excluded from the prediction phase due to the unavailability of their future values. Risk levels were categorized based on the predicted probabilities of VL transmission: areas with probabilities ranging from 0 to 0.5 were classified as low risk; those with probabilities between 0.5 and 0.7 were considered moderate risk; and areas exceeding a probability of 0.7 were identified as high-risk zones. These classifications are visually represented on the map through distinct color-coded regions, facilitating interpretation and risk communication.

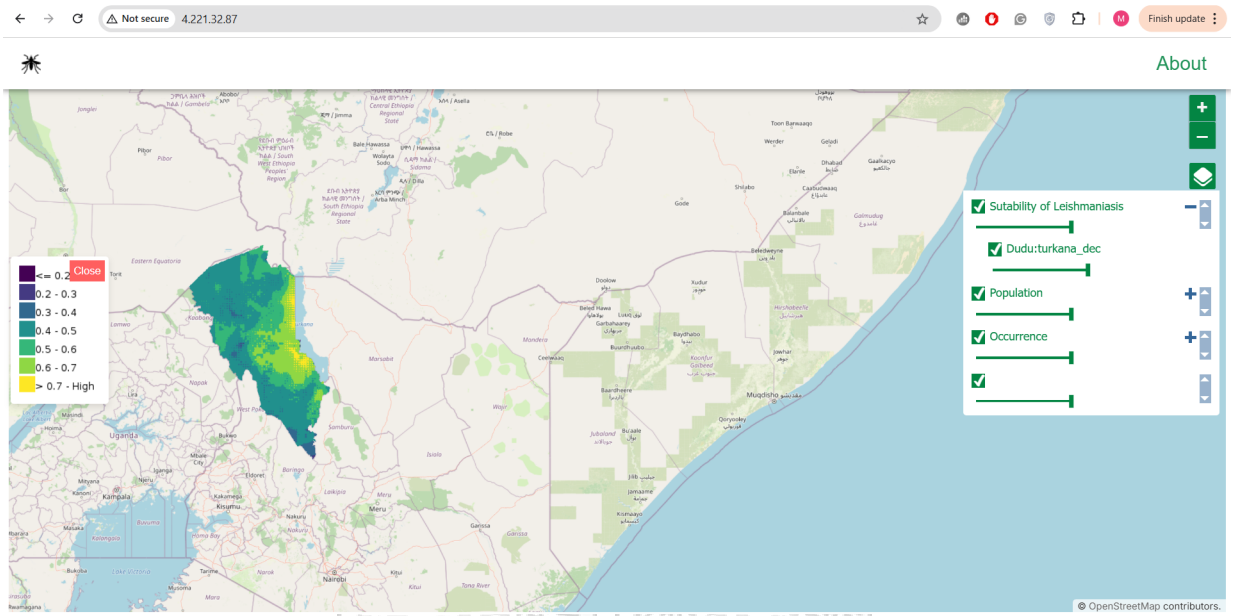


Figure 4.2: December 2024 prediction and MAP

Chapter 5

Discussions, Conclusions and Recommendations

5.1 Introduction

The objective of this research was to use a Bayesian spatial model that incorporated population, environmental, and weather factors to analyze and evaluate the spread of VL. In addition, the study sought to map the future potential distribution of VL in Turkana County. To carry out this study, we collated data from different sources, cleaned it, and pre-processed it to make it ready for EDA and modeling. The Bayesian spatial model was then used to analyse, evaluate, and predict the future distribution of VL in Turkana County.

5.2 Discussion

This study established that proximity to healthcare was a positive and significant variable in explaining VL infections implying that people close to the healthcare centers had a better chance at diagnosis and treatment centers. However, this study notes that access to healthcare in the county faces significant challenges as patients are required to travel an average distance of 48.87 km to access a healthcare facility. Further, the county is dominated by marginalized nomadic pastoralist communities, is sparsely served with only 14 VL treatment centers, and its socio-economic conditions further limit healthcare access (Baraka and Romanus, 2022; Turkana County Government, 2024b). Generally, in Kenya, long travel distances significantly hinder access to care, with patients often taking up to 90 days to reach treatment centers from

the onset of symptoms (World Health Organization, 2024). Similar delays are reported in Ethiopia and Sudan, and even longer in Eritrea and Somalia (World Health Organization, 2024). The situation is dire in Somalia, where insecurity limits access, and in South Sudan, where flooding and poor infrastructure make travel nearly impossible. As a result, reported VL cases in eastern Africa often underrepresent the true disease burden, reflecting only those who can reach treatment centers (World Health Organization, 2024).

The study also established that LULC was a significant predictor of VL. It was noted that VL cases were primarily concentrated in areas with tree cover, scrubland, water bodies (such as River Turkwel, River Kerio, and Lake Turkana), and grassland. Grasslands and scrublands offer suitable habitats for both vectors and rodent hosts. These environments, shaped by a combination of ecological and human factors, support sandfly populations. Low-vegetation areas, including those with herbaceous plants and shrubs, provide ideal conditions for sandfly development. The accumulation of organic matter and litter in these areas creates breeding sites for immature sandflies, while nectar sources sustain adult sandflies (Martín et al., 2020; Parras et al., 2012; Ramos et al., 2014). In urban areas, the presence of domestic or barnyard animals and accumulated waste can attract sandflies, offering breeding grounds and shelter for adults, which in turn increases human-vector interactions (Martín et al., 2020). Acacia trees, the dominant tree cover in the region, likely contribute to a higher abundance of sandflies, putting humans at risk of bites during daily activities such as herding or fetching water (Hassaballa et al., 2021). Additionally, seasonal climate variations influence pasture growth and grazing patterns, which may further increase the risk of exposure for herders (Gao and Cao, 2019).

The age groups 0-5 years, 6-18 years, 19-30 years, 31-44 years were significantly associated with VL. The descriptive statistics showed that the highest number of VL-positive cases were found in patients aged 6–18 years, accounting for up to 56%, while the lowest percentage (4.46%) occurred in patients aged 45 years and above. The 0-5 years age group made up 18.51%, and together, children under 18 years accounted for 75.04% of all positive cases. This implied that generally children under the age of 18 years old, were the most likely to seek

diagnosis and test positive for VL. The population in Turkana County is largely composed of children under 19 years old, representing approximately 58% of the population, who are more susceptible to VL exposure (Kenya National Bureau of Statistics, 2019). Other studies Andrade et al. (2020); Bern et al. (2010); Bi et al. (2020); da Silva Santana Cruz et al. (2021) note that children are at higher risk of contracting the disease due to immunological immaturity, poor nutrition, and poverty. Turkana County is the poorest in Kenya with a poverty index of 79.4% , and faces a high risk of VL due to a combination of poverty, malnutrition, and cultural practices (United Nations High Commissioner for Refugees, 2024). A large portion of the population depends on pastoralism, and children are especially vulnerable as they spend many hours herding livestock, increasing their exposure to sandfly bites Abdullahi et al. (2022). Malnutrition is prevalent among children and teenagers, which makes them particularly susceptible to severe forms of VL (Bi et al., 2020; Dye and Williams, 1993).

The study further established that males were significantly affected by VL Other studies, findings have also uncovered the same phenomena in (the neighboring) Marsabit County in Kenya Kanyina (2020b), Metemma Hospital in Ethiopia Terefe et al. (2015), India and Nepal Cloots et al. (2020), and urban endemic regions in Brazil (Andrade et al., 2020; da Silva Santana Cruz et al., 2021). Immune responses to parasitic infections, like VL, can differ between males and females due to their sex hormones. Hormones such as testosterone, estradiol, and progesterone affect how the immune system responds to these infections. Testosterone, in particular, can reduce inflammation and promote cell death, leading to weaker immune responses. This results in males being more vulnerable to infections and having poor control over parasitic loads compared to females (de Araújo Albuquerque et al., 2021; Sellau et al., 2024; Snider et al., 2009). In addition traditional gender roles which are still prevalent in Turkana County put men at a higher risk for VL compared to females. From a young age, men typically engage in outdoor activities like livestock trading, fishing, carpentry, pastoralism, and construction. In contrast, women mainly handle indoor tasks such as caregiving, cooking, and cleaning (J. and V., 2008). Because males spend more time outdoors, they have greater exposure to sandfly bites, putting them at a higher risk of VL infection. Therefore, males are more likely to participate in outdoor activities thus increasing their exposure to sandfly bites and consequently being at

higher risk of VL infection (Andrade et al., 2020; da Silva Santana Cruz et al., 2021).

5.3 Conclusion

The findings of this study demonstrate the adaptability of the Bayesian spatial model in analyzing VL transmission patterns and identifying significant risk factors. The application of the model enabled the identification of statistically significant explanatory variables influencing VL occurrence in Turkana. This model was able to reliably capture spatial dependencies and random effects influencing the disease while at the same time ensuring statistical robustness. Additionally, evaluating multiple incubation periods allowed for optimal lag selection, with Lag 2 (infection occurring two months before diagnosis) emerging as the best lag for explaining VL cases. The model selection process was guided by the performance metrics accuracy, precision, and recall which ensured the identification of the most effective predictive framework. Notably, the Bayesian spatial model at Lag 2 exhibited the highest performance score. The outcomes of this study provide valuable insights, particularly in identifying high-risk demographic groups, such as males and children under 18, as well as environmental and population factors like proximity to healthcare, minimum temperature, and LULC. These findings highlight the model's potential to inform targeted interventions and resource allocation strategies for VL control.

5.4 Recommendations

5.4.1 Recommenation for future studies

This study demonstrates that the Bayesian spatial model provides a practical framework for understanding VL transmission dynamics and identifying high-risk populations. However, this study operated under the assumption that the presence of VL cases indicated the presence of the vector within the location or nearby. It is paramount that future studies incorporate

vector data to refine transmission models and improve their predictive accuracy. Integrating entomological data, such as sandfly abundance and habitat suitability, is key to strengthening the understanding of VL epidemiology.

5.4.2 Policy recommendations

Public health authorities should implement regular vector surveillance programs to monitor sandfly abundance and their role in VL transmission. This will help identify high-risk areas and enable more targeted and effective vector control measures. Additionally, the Kenyan government should enhance data collection by integrating epidemiological and vector information, ensuring that the disease control strategies are based on comprehensive and accurate data. Public health policies should incorporate predictive modeling tools to optimize resource allocation, allowing VL prevention and control programs to be more data-driven and thus effective in reducing disease transmission.

5.5 Limitations

This study faced limitations in acquiring environmental and population data at an ideal spatial and temporal resolution, which may have affected the precision of the analysis. Additionally, the study was constrained by the lack of vector data, with the study relying on the assumption that VL cases indicate the presence of sandflies in the area or in the neighborhood.Incorporating vector data in future studies would provide a more comprehensive understanding of transmission dynamics and improve model performance.

References

Abdullahi, B., Mutiso, J., Maloba, F., Macharia, J., Riongoita, M., and Gicheru, M. (2022). Climate change and environmental influence on prevalence of visceral leishmaniasis in west pokot county, kenya. *Journal of Tropical Medicine*, 2022:1–6.

Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., Boer, M. d., and Team, W. L. C. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PloS one*, 7(5):e35671.

Andrade, A. W. F., Souza, C. D. F., and Carmo, R. F. (2020). Analysis of spatial clustering, time trend, social vulnerability and risk of human visceral leishmaniasis in an endemic area in brazil: an ecological study. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 114(8):575–584.

Baraka, M. and Romanus, O. (2022). Socio economic development in turkana west, kenya. Technical report, United Nations Human Settlements Programme.

Ben-Ahmed, K. (2021). Childhood visceral leishmaniasis in tunisia: A cross-sectional study in local spatial analysis. *Acta Tropica*, 214:105793.

Bern, C., Courtenay, O., and Alvar, J. (2010). Of cattle, sand flies and men: A systematic review of risk factor analyses for south asian visceral leishmaniasis and implications for elimination. *PLOS Neglected Tropical Diseases*, 4(2):1–9.

Bhunia, G. S., Kesari, S., Jeyaram, A., Kumar, V., and Das, P. (2010). Influence of topography on the endemicity of kala-azar: a study based on remote sensing and geographical information system. *Geospatial Health*, 4(2):155–165.

Bi, K., Chen, Y., Zhao, S., Ben-Arieh, D., and (John) Wu, C.-H. (2020). A new zoonotic visceral leishmaniasis dynamic transmission model with age-structure. *Chaos, Solitons & Fractals*, 133:109622.

CIESIN (2024). Gpwv4 population density - revision 11: Data download. "Accessed on 9-05-2024".

Cloots, K., Burza, S., Malaviya, P., Hasker, E., Kansal, S., Mollett, G., Chakravarty, J., Roy, N., Lal, B. K., and Rijal, S. (2020). Male predominance in reported visceral leishmaniasis cases: Nature or nurture? a comparison of population-based with health facility-reported data. *PLoS Neglected Tropical Diseases*, 14(1):e0007995.

da Silva Santana Cruz, C., Soeiro Barbosa, D., Oliveira, V. C., Cardoso, D. T., Guimarães, N. S., and Carneiro, M. (2021). Factors associated with human visceral leishmaniasis cases during urban epidemics in brazil: a systematic review. *Parasitology*, 148(6):639–647.

Damien, B. G., Sode, A. I., Bocossa, D., Elanga-Ndille, E., Aguemon, B., Corbel, V., Henry, M.-C., Glèlè Kakaï, R. L., and Remoué, F. (2022). Bayesian spatial modelling of malaria burden in two contrasted eco-epidemiological facies in benin (west africa): call for localized interventions. *BMC Public Health*, 22(1):1754.

de Araújo Albuquerque, L. P., da Silva, A. M., de Araújo Batista, F. M., de Souza Sene, I., Costa, D. L., and Costa, C. H. N. (2021). Influence of sex hormones on the immune response to leishmaniasis. *Parasite Immunology*, 43(10-11):e12874.

De Diego, I. M., Redondo, A. R., Fernández, R. R., Navarro, J., and Moguerza, J. M. (2022). General performance score for classification problems. *Applied Intelligence*, 52(10):12049–12063.

Dye, C. and Williams, B. G. (1993). Malnutrition, age and the risk of parasitic disease: visceral leishmaniasis revisited. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 254(1339):33–39.

Gacutan, J., Tait, H., and Clark, G. F. (2024). Assessing drivers of estuarine debris using a bayesian spatial modelling approach (inla-spde). *Estuarine, Coastal and Shelf Science*, 296:108592.

Galgamuwa, L. S., Dharmaratne, S. D., and Iddawela, D. (2018). Leishmaniasis in sri lanka: spatial distribution and seasonal variations from 2009 to 2016. *Parasites & vectors*, 11:1–10.

Gao, X. and Cao, Z. (2019). Meteorological conditions, elevation and land cover as predictors for the distribution analysis of visceral leishmaniasis in sinkiang province, mainland china. *Science of the Total Environment*, 646:1111–1116.

Google Earth Engine (2024). Modis land cover type yearly global 500m (mcd12q1).

Government of Kenya (2021). Kenya strategic plan for control of leishmaniasis 2021 - 2025. Technical report, Government of Kenya.

Grifferty, G., Shirley, H., O'Brien, K., Hirsch, J. L., Orriols, A. M., Amechi, K. L., Lo, J., Chanda, N., El Hamzaoui, S., Kahn, J., Yap, S. V., Watson, K. E., Curran, C., Atef AbdelAlim, A., Bose, N., Cilfone, A. L., and Wamai, R. (2023). The leishmaniases in kenya: A scoping review. *PLOS Neglected Tropical Diseases*, 17(6):e0011358.

Hassaballa, I. B., Sole, C. L., Cheseto, X., Torto, B., and Tchouassi, D. P. (2021). Afrotropical sand fly-host plant relationships in a leishmaniasis endemic area, Kenya. *PLoS Neglected Tropical Diseases*, 15(2):e0009041.

J., W. D. and V., B. J. ("2008"). Livelihood diversification opportunities for pastoralists in turkana, kenya. international livestock research institute. https://cgspace.cgiar.org/server/api/core/bitstreams/7a5df525-f62f-4007-a97a-376826db13e9/content. "Accessed on 21-07-2024".

Jiang, D., Ma, T., Hao, M., Qian, Y., Chen, S., Meng, Z., Wang, L., Zheng, C., Qi, X., Wang, Q., and Ding, F. (2021a). Spatiotemporal patterns and spatial risk factors for visceral leishmaniasis from 2007 to 2017 in western and central china: A modelling analysis. *Science of the Total Environment*, 764:144275.

Jiang, D., Ma, T., Hao, M., Qian, Y., Chen, S., Meng, Z., Wang, L., Zheng, C., Qi, X., Wang, Q., and Ding, F. (2021b). Spatiotemporal patterns and spatial risk factors for visceral leishmaniasis from 2007 to 2017 in Western and Central China: a modelling analysis. *Science of The Total Environment*, 764:144275.

Jones, C. M. and Welburn, S. C. (2021a). Leishmaniasis beyond east africa. *Frontiers in Veterinary Science*, 8:618766.

Jones, C. M. and Welburn, S. C. (2021b). Leishmaniasis beyond east Africa. *Frontiers in Veterinary Science*, 8:618766.

Kang, S. Y., Cramb, S. M., White, N. M., Ball, S. J., and Mengersen, K. L. (2016). Making the most of spatial information in health: a tutorial in bayesian disease mapping for areal data. *Geospatial Health*, 11(2):Article–number.

Kanyina, E. W. (2020a). Characterization of visceral leishmaniasis outbreak, marsabit county, kenya, 2014. *BMC Public Health*, 20:1–8.

Kanyina, E. W. (2020b). Characterization of visceral leishmaniasis outbreak, marsabit county, kenya, 2014. *BMC Public Health*, 20(1):1–8.

Karger, D. N., Schmatz, D., Dettling, G., and Zimmermann, N. E. (2019). High resolution monthly precipitation and temperature timeseries for the period 2006-2100. https://www.envidat.ch/dataset/chelsa_cmip5_ts. Accessed on 12-05-2024.

Kenya National Bureau of Statistics (2019). 2019 kenya population and housinng census. Technical report, Kenya National Bureau of Statistics.

Kesari, S., Bhunia, G. S., Kumar, V., Jeyaram, A., Ranjan, A., and Das, P. (2010). Study of house-level risk factors associated in the transmission of indian kala-azar. *Parasites & Vectors*, 3:1–5.

Li, Y. and Zheng, C. (2019). Associations between meteorological factors and visceral leishmaniasis outbreaks in jiashi county, xinjiang uygur autonomous region, china, 2005–2015. *International Journal of Environmental Research and Public Health*, 16(10):1775.

Makau-Barasa, L. K., Ochol, D., Yotebieng, K. A., Adera, C. B., and de Souza, D. K. (2022). Moving from control to elimination of visceral leishmaniasis in east africa. *Frontiers in Tropical Diseases*, 3:965609.

Martín, M. E., Stein, M., Willener, J. A., Kuruc, J. A., and Estallo, E. L. (2020). Landscape effects on the abundance of lutzomyia longipalpis and migonemyia migonei (diptera: Psychodidae: Phlebotominae) in corrientes city, northern argentina. *Acta Tropica*, 210:105576.

Moirano, G., Ellena, M., Mercogliano, P., Richiardi, L., and Maule, M. (2022). Spatio-temporal pattern and meteo-climatic determinants of visceral leishmaniasis in italy. *Tropical Medicine and Infectious Disease*, 7(11):337.

Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. CRC Press, Boca Raton, FL.

Moraga, P., Dean, C., Inoue, J., Morawiecki, P., Noureen, S. R., and Wang, F. (2021). Bayesian spatial modelling of geostatistical data using inla and spde methods: A case study predicting malaria risk in mozambique. *Spatial and Spatio-temporal Epidemiology*, 39:100440.

Nackers, F., Mueller, Y. K., Salih, N., Elhag, M. S., Elbadawi, M. E., Hammam, O., Mumina, A., Atia, A. A., Etard, J.-F., Ritmeijer, K., and Chappuis, F. (2015). Determinants of visceral leishmaniasis: A case-control study in gedaref state, sudan. *PLOS Neglected Tropical Diseases*, 9:1–16.

Open Meteo (2024). Open meteo. https://open-meteo.com/. "Accessed on 9-05-2024".

Ouma, F. F. and Mulambalah, C. S. (2021). Persistence and changing distribution of leishmaniases in kenya require a paradigm shift. *Journal of Parasitology Research*, 2021:1–6.

Parras, M. A., Rosa, J. R., Szelag, E. A., and Salomón, O. D. (2012). Identification of the natural breeding sites of sandflies (diptera: Psychodidae: Phlebotominae), potential vectors of leishmaniasis, in the province of chaco, argentina. *Memórias do Instituto Oswaldo Cruz*, 107(4):550–552.

Piscopo, T. V. and Mallia Azzopardi, C. (2007). Leishmaniasis. *Postgraduate Medical Journal*, 83(976):649–657.

Rajabi, M., Pilesjö, P., Bazmani, A., and Mansourian, A. (2017). Identification of visceral leishmaniasis-susceptible areas using spatial modelling in southern caucasus. *Zoonoses and public health*, 64(7):e5–e22.

Ramos, W. R., Medeiros, J. F., Julião, G. R., Ríos-Velásquez, C. M., Marialva, E. F., Desmouliére, S. J., Luz, S. L. B., and Pessoa, F. A. C. (2014). Anthropic effects on

sand fly (diptera: Psychodidae) abundance and diversity in an amazonian rural settlement, brazil. *Acta Tropica*, 139:44–52.

Roth, S., Lee, B. S., Nicholas, R. E., Keller, K., and Haran, M. (2023). Bayesian spatial models for projecting corn yields. *Remote Sensing*, 16(1):69.

Scarpini, S., Dondi, A., Totaro, C., Biagi, C., Melchionda, F., Zama, D., Pierantoni, L., Gennari, M., Campagna, C., Prete, A., and Lanari, M. (2022). Visceral leishmaniasis: Epidemiology, diagnosis, and treatment regimens in different geographical areas with a focus on pediatrics. *Microorganisms*, 10(10):1887.

Sellau, J., Hansen, C. S., Gálvez, R. I., Linnemann, L., Honecker, B., and Lotter, H. (2024). Immunological clues to sex differences in parasitic diseases. *Trends in Parasitology*, 40(11):1029–1041.

Snider, H., Lezama-Davila, C., Alexander, J., and Satoskar, A. R. (2009). Sex Hormones and Modulation of Immunity against Leishmaniasis. *Neuroimmunomodulation*, 16(2):106–113.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437.

Subramanian, S., Maheswari, R. U., Prabavathy, G., Khan, M. A., Brindha, B., Srividya, A., Kumar, A., Rahi, M., Nightingale, E. S., Medley, G. F., Cameron, M. M., Roy, N., and Jambulingam, P. (2024). Modelling spatiotemporal patterns of visceral leishmaniasis incidence in two endemic states in India using environment, bioclimatic and demographic data, 2013–2022. *PLOS Neglected Tropical Diseases*, 18:e0011946.

Terefe, Y., Afera, B., Bsrat, A., and Syoum, Z. (2015). Distribution of human leishmaniasis (vl) and its associated risk factors, in metemma, ethiopia. *Epidemiology Research International*, 2015(1):630812.

Turkana County Government (2024a). Home page - turkana county government. turkana county government - pamoja tujijenge. https://turkana.go.ke/. Accessed on 01-12-2024.

Turkana County Government ("2024"b). Turkana County Integrated Development Plan, 2013 - 2017. Turkana County Government, United Nations, Government of Kenya. https://www.devolution.go.ke/sites/default/files/2024-03/Turkana-CIDP-2013-2017.pdf. "Accessed on 01-06-2024".

United Nations High Commissioner for Refugees (2024). Kenya registered refugees and asylum-seekers. Technical report, United Nations High Commissioner for Refugees.

Valero, N. N. H. and Uriarte, M. (2020). Environmental and socioeconomic risk factors associated with visceral and cutaneous leishmaniasis: a systematic review. *Parasitology research*, 119(2):365–384.

van Dijk, N. J., Amer, S., Mwiti, D., Schallig, H. D. F. H., and Augustijn, E.-W. (2024). An epidemiological and spatiotemporal analysis of visceral leishmaniasis in west pokot, kenya, between 2018 and 2022. *BMC Infectious Diseases*, 24(1):1169.

Vujović, Ž. et al. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6):599–606.

Wang, Y., Chen, X., and Xue, F. (2024). A review of bayesian spatiotemporal models in spatial epidemiology. *ISPRS International Journal of Geo-Information*, 13(3):97.

World Health Organization (2023). Leishmaniasis. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis.

World Health organization (2024). New framework launched to eliminate visceral leishmaniasis in eastern africa. https://www.who.int/news/item/12-06-2024-new-framework-launched-to-eliminate-visceral-leishmaniasis-in-eastern-africa. "Accessed on 12-06-2024".

World Health Organization (2024). Strategic framework for the elimination of visceral leishmaniasis as a public health problem in eastern africa 2023–2030. Technical report, World Health Organization.

Zhao, Y., Jiang, D., Ding, F., Hao, M., Wang, Q., Chen, S., Xie, X., Zheng, C., and Ma, T. (2021). Recurrence and driving factors of visceral leishmaniasis in central China. *International Journal of Environmental Research and Public Health*, 18(18):9535.

Zheng, Canjun and Wang, Liping and Li, Yi and Zhou, Xiao-Nong (2020). Visceral leishmaniasis in northwest china from 2004 to 2018: a spatio-temporal analysis. *Infectious Diseases of Poverty*, 9(06):29–41.

Appendix A

Ethical approval

24th June 2024

Ms Patrick Maureen,
maureen.nzilani@strathmore.edu

Dear Ms Patrick,

RE: Analyzing the Current and Future Potential Distributions of Visceral Leishmaniasis in Kenya: A Spatial Model Approach

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2325/24.** The approval period is from **24th June 2024 to 23rd June 2025.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix B

Similarity index

Maureen_Final_Thesis_2025 (9)-pages.pdf

ORIGINALITY REPORT

15%	10%	11%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Paula Moraga, Christopher Dean, Joshua Inoue, Piotr Morawiecki, Shahzeb Raja Noureen, Fengpei Wang. "Bayesian spatial modelling of geostatistical data using INLA and SPDE methods: A case study predicting malaria risk in Mozambique", Spatial and Spatio-temporal Epidemiology, 2021 Publication	1%
2	www.frontiersin.org Internet Source	1%
3	www2.mdpi.com Internet Source	1%
4	etd.aau.edu.et Internet Source	1%
5	"Challenges and Solutions Against Visceral Leishmaniasis", Springer Science and Business Media LLC, 2023 Publication	<1%
6	journals.plos.org Internet Source	<1%
7	www.medrxiv.org Internet Source	<1%

Appendix C

R code

Appendix: Full R Code

Full code available at: https://github.com/mnzilani/Leishmaniasis/blob/main/Bayesian%20Full%20model%2060.Rmd

```
# Load Libraries
library(dplyr)
library(tidyr)
library(INLA)
library(ggplot2)
library(Metrics)
library(sp)
library(caret)
library(pROC)

# Load Data
df <- read.csv("C:/.../L1.csv")

# Create Age Group
df <- df %>% mutate(age_group = case_when(
  AGE >= 0 & AGE <= 5 ~ "0-5 years",
  AGE >= 5.01 & AGE <= 18 ~ "6-18 years",
  AGE >= 18.01 & AGE <= 30 ~ "19-30 years",
  AGE >= 30.01 & AGE <= 44 ~ "31-44 years",
  AGE > 44.01 ~ "45 years and above"
```

```
))

# Jitter Coordinates
df <- df %>% mutate(
  lat_jittered = lat + runif(n(), -1e-5, 1e-5),
  lon_jittered = lon + runif(n(), -1e-5, 1e-5)
)

df$age_group <- as.factor(df$age_group)
df$SEX <- as.factor(df$SEX)

# Define Coordinates
coords <- cbind(df$lat_jittered, df$lon_jittered)

# Split Train/Test
set.seed(123)
trainIndex <- createDataPartition(df$Infections, p = .8,
   list = FALSE)
trainData <- df[trainIndex, ]
testData <- df[-trainIndex, ]
trainCoords <- coords[trainIndex, ]
testCoords <- coords[-trainIndex, ]

# Create Mesh
mesh_train <- inla.mesh.2d(loc = trainCoords, max.edge = c
   (0.5, 2), cutoff = 0.1)
spde_train <- inla.spde2.matern(mesh = mesh_train, alpha =
   2)
A_train <- inla.spde.make.A(mesh_train, loc = trainCoords)
spde_index_train <- inla.spde.make.index(name = "spatial", n
   .spde = spde_train$n.spde)

```

```r
# Stack Data
stack_train <- inla.stack(
  data = list(y = trainData$Infections),
  A = list(A_train, 1),
  effects = list(spde_index_train, list(trainData))
)

# Define Formula
formula <- y ~ SEX + distance + age_group + tempmin + NDVI +
  mean_humidity + total_precip + population_density +
  Distance_to_Water_.km. + forest_height + LULC +
  f(spatial, model = spde_train) + f(month, model = "iid")

# Train Model
result_train <- inla(formula, family = "binomial",
  data = inla.stack.data(stack_train),
  control.predictor = list(A = inla.stack.A(stack_train)),
  control.compute = list(dic = TRUE, waic = TRUE)
)

# Test Preparation
testData <- testData %>% mutate(
  lat_jittered = lat + runif(n(), -1e-5, 1e-5),
  lon_jittered = lon + runif(n(), -1e-5, 1e-5)
)
testCoords <- cbind(testData$lat_jittered, testData$lon_
   jittered)
A_test <- inla.spde.make.A(mesh_train, loc = testCoords)

# Stack Test
stack_test <- inla.stack(
  data = list(y = testData$Infections),
```

```
  A = list(A_test, 1),
  effects = list(spde_index_train, list(testData))
)

# Predict
result_test <- inla(formula, family = "binomial",
  data = inla.stack.data(stack_test),
  control.predictor = list(A = inla.stack.A(stack_test),
     compute = TRUE),
  control.compute = list(dic = TRUE, waic = TRUE)
)

# Evaluation
predictions_test <- result_test$summary.fitted.values[1:nrow
   (testData), "mean"]
predicted_classes_test <- ifelse(predictions_test > 0.5, 1,
   0)
conf_matrix_test <- table(Predicted = predicted_classes_test
   , Actual = testData$Infections)
accuracy_test <- sum(diag(conf_matrix_test)) / sum(conf_
   matrix_test)
precision_test <- conf_matrix_test[2, 2] / sum(conf_matrix_
   test[2, ])
recall_test <- conf_matrix_test[2, 2] / sum(conf_matrix_test
   [, 2])

# Output Metrics
cat("Accuracy:", accuracy_test, "\n")
cat("Precision:", precision_test, "\n")
cat("Recall:", recall_test, "\n")

# ROC and AUC
```

```
roc_obj_test <- roc(testData$Infections, predictions_test)
auc_val_test <- auc(roc_obj_test)
cat("Test AUC:", auc_val_test, "\n")
plot(roc_obj_test, main = paste("ROC Curve (AUC =", round(
    auc_val_test, 2), ")"))
```

Listing 1: Bayesian INLA Model in R

Code for Prediction: https://github.com/mnzilani/Leishmaniasis/blob/main/Bayesian%20Predictions.R

Code for Interpolation: https://github.com/mnzilani/Leishmaniasis/blob/main/IDW_interpolation_code%20(1).R