

**ECOLOGICAL NICHE MODELING OF MALARIA VECTOR
DISTRIBUTION FOR CLIMATE CHANGE ADAPTATION IN
KENYA**

BY

KIMUYU, JACINTA SYOKAU

REG. NO: I80/91921/2013



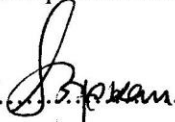
UNIVERSITY OF NAIROBI

**A Thesis Submitted in Fulfillment of the Requirement for Award of the
Degree of Doctor of Philosophy in Meteorology of the University of
Nairobi**

APRIL, 2015

DECLARATION

I declare that this thesis is my original work and has not been submitted elsewhere for examination, award of a degree or publication. Where other people's work or my own work has been used, this has properly been acknowledged and referenced in accordance with the University of Nairobi's requirements.

Signature  Date 21/04/2015

Jacinta Syokau Kimuyu


I80/91921/2013


Department of Meteorology

School of Physical Sciences

University of Nairobi

This thesis is submitted for examination with our approval as research supervisors:

	Signature	Date
Prof. Nzioka J. Muthama Department of Meteorology University of Nairobi P.O Box 30197-00100 Nairobi Kenya jmuthama@uonbi.ac.ke  <u>21/04/2015</u>

Dr. -Ing. Sammy M. Musyoka Department of Geospatial and Space Technology University of Nairobi P.O Box 30197-00100 Nairobi Kenya smmusyoka@uonbi.ac.ke  <u>21/4/2015</u>
---	---	------------------------------

ABSTRACT

This study employed Ecological Niche Modeling (ENM), a technique that encompasses a suite of tools that relate known occurrences of species or phenomena to raster geographic information system layers that summarize variation in several environmental dimensions. The spatial-temporal distributions of the main malaria vectors in Kenya were quantified using BIOCLIM and DOMAIN models to determine the relationship between vector distribution and climate change. The biological data used was from published sources (Okara et al., 2010 and MARA/ARMA, 1998), comprising of point samples for geo-referenced malaria vector occurrences. The climate data used was maximum temperature, minimum temperature and precipitation for current climate (1950-2000) and climate projection for HADCM3, CCCMA and SCIRO models of IPCC projected future climate under the A2a scenario by the years 2020, 2050 and 2080. The climate data was acquired in grid format from WorldClim global climate data which was further processed to generate 19 bioclimatic variables for Kenya.

The predictions showed that by the year 2020, the suitability areas for malaria vectors in Kenya will start to change from the current ecological suitability. Most areas where the malaria vectors are thriving currently will still remain suitable ecologies. New suitability zones will emerge in most counties ranging from low to very high suitability as shown by the predictions. By the year 2050, areas of suitability will expand at an alarming extend. The year 2080 has been predicted to show that the suitable ecologies will start to revert to the original areas of suitability as in the current climate. Therefore, climate change in Kenya will adversely affect the environment at an alarming rate by 2050, but beyond that there will be a level of stabilization, where further change will trigger reversal to the past climate.

For instance, BIOCLIM True or False prediction from HADCM3 by the year 2050 showed wide spread of malaria in counties like Narok, Kajiado, Kitui, Makueni, Machakos, Meru, Marsabit, Isiolo, Samburu, Baringo, West Pokot. Turkana county and Mandera among a few others will have some emerging isolated malaria hot spots. ENM prediction with HADCM3 future climate showed that Laikipia County will become unsuitable malaria ecology by the year 2050 and the case remains the same by the year 2080.

Validation results for prediction model performance showed that all the models used had errors in prediction as none of them had kappa =1 or AUC=1. The highest kappa ($k = 0.909$) and Area under ROC Curve (AUC=0.954) values were achieved from DOMAIN model with CCCMA projection by the year 2020. The lowest model performance values of $k = 0.427$ and AUC = 0.714 were obtained from BIOCLIM True or False model with HADCM3 projections by the year 2020.

The following conclusions were drawn from the Ecological Niche Modeling done using BIOCLIM, BIOCLIM True or False and DOMAIM prediction models: There is correlation between climate change as an explanatory variable and the distribution of main malaria vectors in Kenya. The spatial-temporal distribution of the main malaria vectors in Kenya varies under different IPCC future climate projections which are HADCM3, CCCMA and CSIRO. The future ecological niches for malaria vector occurrence in Kenya will extend from the current niches in most endemic areas, new hotspots will emerge and some suitable ecology will become unsuitable, resulting in varying areas from current climate predictions to projections by the year 2020, 2050 and 2080 under IPCC A2a scenario. Intervention strategies such as indoor or outdoor residual spraying, distribution of insecticide-treated mosquito nets (ITNs) and long-lasting insecticide-treated nets (LLINs) should be diversified in new emerging areas for disaster risk reduction and increase adaptive capacity and resilience among local communities.

DEDICATION

To *Samuel, Moses, Debra Ruth, Prince Leon*; and

To all mankind:

.....”*BE WORRIED, BE VERY WORRIED - climate change is not some vague future problem - it is already damaging the planet at an alarming rate*”..... (Special Report Global Warming *TIME*, April 3, 2006- www.time.com).

ACKNOWLEDGEMENT

First and foremost, thanks to God Almighty for this academic milestone; you are Ebenezer.

The author is indebted to the following people and organizations:

The research has been accomplished through financial assistance from Kenya National Council for Science and Technology, Ministry of Higher Education Science, Technology and Innovation Grant Fund (ST&I); Grant File No. NCST/ACC/003/005/154/4th CALL PhD.

The invaluable technical support offered by Digma Geosystems Ltd, Kenya is highly appreciated.

Sincere gratitude is extended to Prof. J.N. Muthama, my first supervisor for his commitment to this research from building the conceptual framework to conclusion. Many are the times I felt like resigning, but his constructive critique always brought some inspiration of thought.

To my second supervisor, Dr. –Ing. S.M. Musyoka, I am extremely grateful for your unrivaled fervor to work with Mathematical models, the underpinning of this research. It has been my great pleasure to tap from your Scientific and Mathematical intellectual property.

To Dr. G.O. Wayumba, Director, School of Surveying and Geospatial Science, Technical University of Kenya; I thank you dearly for mentorship and encouragement throughout the challenging research period. To Dr. S.O. Ayugi of Technical University of Kenya, I treasure all the priceless time you availed yourself to read the thesis and critique it at proximity.

To Mr. Victor Alegana of KEMRI, I am grateful and highly appreciate the research data links you offered.

To all my friends, I cannot hesitate to reckon with your moral support; many thanks for your understanding all the times I missed social gatherings and functions. I will be available soon.

Lastly but not the least, I am genuinely indebted to my entire family, especially to my little ones Debbie Ruth and Prince Leon; many are the times I could not be there for you, but patience pays; I will make up for the lost opportunities.

TABLE OF CONTENTS

DECLARATION.....	ERROR! BOOKMARK NOT DEFINED.
ABSTRACT.....	III
DEDICATION.....	V
ACKNOWLEDGEMENT.....	VI
TABLE OF CONTENTS	VII
LIST OF TABLES	X
LIST OF FIGURES	XI
LIST OF ACRONYMS	XV
CHAPTER 1: INTRODUCTION	1
1.1 Background.....	1
1.2 Statement of the Problem.....	6
1.3 Objective of the study	7
1.4 Justification and significance of the study	8
1.5 Scope and limitations	9
CHAPTER 2: LITERATURE REVIEW	12
2.1 Observed Climate Change and Variability	12
2.2 The Impact of Climate Change and Variability on Human health	23
2.3 Climate change and disease transmission in Africa.....	27
2.4 Models for malaria Early Warning System	28
2.5 Malaria Early Warning and Spatial mapping.....	33
2.6 Factors considered Ecological Niche Modeling (ENM).....	34
2.7 IPCC Reported Climate Change and Variability in African Region	39
2.8 Geostatistical Modeling and Prediction	41
2.8.1 Thiessen Polygons	43
2.8.2 Triangulation.....	45
2.8.3 Natural Neighbor Interpolation.....	46

2.8.4	Inverse Functions of Distance.....	47
2.8.5	Trend Surfaces	48
2.8.6	Splines.....	50
2.9	The Variogram.....	50
CHAPTER 3: METHODOLOGY		52
3.1	Data.....	52
3.1.1	Malaria Vector Data.....	53
3.1.2	Climate Data	55
3.2	Methodology.....	60
3.2.1	Selecting Prediction Models for Ecological Niche Modeling	61
3.2.2	Model Validation	64
3.2.2.1	The Presence/Absence Confusion Matrix	67
3.2.2.2	Selecting Thresholds of Occurrence.....	72
3.3	Choice of Future Climate Emission Scenario for Prediction.....	74
3.4	Testing Autocorrelation in point sample data.....	78
3.4.1	Point Autocorrelation in vectors1	83
3.4.2	Point Autocorrelation in vectors2.....	85
3.5	Climate Change and Malaria Vector Distributions in Kenya	88
3.5.1	Prediction and Mapping of Malaria Vector Ecological Niches	92
CHAPTER 4: RESULTS AND DISCUSSION.....		94
4.1	Ecological Niche Modeling (ENM) Results and Analysis	94
4.1.1	BIOCLIM Model Prediction with HADCM3, CCCMA and CSIRO Data	95
4.1.1.1	Generated BIOCLIM Prediction Results from HADCM3 Model data.....	97
4.1.1.2	Generated BIOCLIM Prediction Results from CCCMA Model data	98
4.1.1.3	Generated BIOCLIM Prediction Results from CSIRO Model	99
4.1.2	BIOCLIM True or False Model Prediction with HADCM3, CCCMA and CSIRO	100
4.1.2.1	Generated BIOCLIM True or False Prediction Results from HADCM3 Model....	101
4.1.2.2	Generated BIOCLIM True or False Prediction Results from CCCMA Model.....	102
4.1.2.3	Generated BIOCLIM True or False Prediction Results from CSIRO Model	103
4.1.3	DOMAIN Model Prediction with HADCM3, CCCMA and CSIRO Data.....	104
4.1.3.1	Generated DOMAIN Prediction Results from HADCM3 Model data	105
4.1.3.2	Generated DOMAIN Prediction Results from CCCMA Model data.....	106

4.1.3.3	Generated DOMAIN Prediction Results from CSIRO Model data	107
4.2	Model Output Results	108
4.2.1	Analysis and Discussion of BIOCLIM Model Prediction Results	108
4.2.1.1	Ecological Niche Area Analysis for BIOCLIM Model with HADCM3	115
4.2.1.2	Ecological Niche Area Analysis for BIOCLIM Model with CCCMA	118
4.2.1.3	Ecological Niche Area Analysis for BIOCLIM Model with CSIRO	121
4.2.1.4	Comparison of Ecological Niche Area from BIOCLIM Model with HADCM3, CCCMA and CSIRO climate projections	125
4.2.2	Analysis and Discussion of BIOCLIM True or False Model Prediction Results	127
4.2.3	Analysis and Discussion of DOMAIN Model Prediction Results.....	129
4.3	Evaluation of Prediction Model Performance.....	131
4.4	Discussion and Comparison of model validation outcomes	132
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS		135
5.1	Conclusions.....	135
5.2	Recommendations.....	139
5.3	Areas for Further Research	140
REFERENCES.....		142
APPENDIX A: MALARIA SPATIAL DATA		152
APPENDIX B: MODEL VALIDATION OUTCOMES		177
B.1	Validation Results for HADCM3 Predictions	178
B.1.1	HADCM3 BIOCLIM and BIOCLIM True or False Model Validation Results.....	178
B.1.2	HADCM3 DOMAIN Model Validation Results	180
B.2	Validation Results for CCCMA Predictions.....	182
B.2.1	CCCMA BIOCLIM and BIOCLIM True or False Model Validation Results	182
B.2.2	CCCMA DOMAIN Model Validation Results.....	184
B.3	Validation Results for CSIRO Predictions	185
B.3.1	CSIRO BIOCLIM and BIOCLIM True or False Model Validation Results.....	185
B.3.2	CSIRO DOMAIN Model Validation Results	187

LIST OF TABLES

Table 2. 1: Reported Temperature and Precipitation change.....	40
Table 3. 1 Description of acquired malaria vector data	53
Table 3. 2: WorldClim climate data (current and future).	58
Table 3. 3: Coding of bioclimatic variables.....	59
Table 3. 4: Presence / Absence Confusion Matrix.....	68
Table 3. 5: Measures of model predictive accuracy	69
Table 3. 6: Evaluation of Kappa strength of agreement	71
Table 3. 7: Interpretation of AUC values.....	72
Table 3. 8: Autocorrelation interpretation values	81
Table A. 1 Malaria Spatial Data from Okara et. al., 2010	153
Table A. 2 Malaria Spatial Data from MARA ARMA, 1998	168

LIST OF FIGURES

Figure 1. 1: A map of Kenya showing the county boundaries.....	11
Figure 2. 1: Comparison between global mean temperature anomalies (⁰ C) from observations (black) and AOGCM simulations forced with both anthropogenic and natural forcings(a) and (b) natural forcings only.....	17
Figure 2. 2: Scenarios A2, A1B and B1 Multi-model means of surface warming ; relative to 1980–1999 (source: IPCC, 2007: Figure 10.4).....	21
Figure 2. 3: Time series of globally averaged surface warming precipitation change	22
Figure 2. 4: Association between climate change and infectious diseases	26
Figure 2. 5: Thiessen Polygons from point data	44
Figure 3. 1: Map of Kenya showing malaria vector spatial distribution.....	54
Figure 3. 2: Flow diagram for the steps required in building a species distribution model.....	60
Figure 3. 3: ROC curves illustrating superior and random predictive ability.....	73
Figure 3. 4: Kenya population growth rate for selecting future climate scenario.....	76
Figure 3. 5: Probabilistic Population Projection to guide scenario selection: Total Population (thousands), based on the 2010 Revision of the World Population Prospects.....	77
Figure 3. 6: Malaria Vector distribution in Kenya; Vectors1 in red and Vectors2 in blue.....	83
Figure 3. 7: Statistics used in the determination of lag distance (0.05) for Vectors1.....	84
Figure 3. 8: Vector spatial autocorrelation using Geary and Moran Indices for Vectors1.....	85
Figure 3. 9: Statistics used in the determination of lag distance (0.05) for Vectors2.....	86
Figure 3. 10: Vector spatial autocorrelation using Geary and Moran Indices for Vectors2.....	87
Figure 3. 11: Difference in Seasonal Temperature Response in East Africa	91
Figure 3. 12: Difference in Seasonal Precipitation Response in East Africa.....	92
Figure 3. 13: Flow diagram for correlative malaria vector distribution modeling.	93

Figure 4. 1: BIOCLIM Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s HADCM3 2020(b), 2050(c) and 2080(d).....	97
Figure 4. 2: BIOCLIM Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CCCMA 2020(b), 2050(c) and 2080(d).....	98
Figure 4. 3: BIOCLIM Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CSIRO 2020(b), 2050(c) and 2080(d).....	99
Figure 4. 4: Bioclim True/False Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s HADCM3 2020(b), 2050(c) and 2080(d).....	101
Figure 4. 5: BIOCLIM True/False Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CCCMA 2020(b), 2050(c) and 2080(d).....	102
Figure 4. 6: BIOCLIM True/False Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CSIRO 2020(b), 2050(c) and 2080(d).....	103
Figure 4. 7: DOMAIN Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s HADCM3 2020(b), 2050(c) and 2080(d).....	105
Figure 4. 8: DOMAIN Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CCCMA 2020(b), 2050(c) and 2080(d).....	106

Figure 4. 9: DOMAIN Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CSIRO 2020(b), 2050(c) and 2080(d).....	107
Figure 4. 10: BIOCLIM predicted areas for non-suitable Ecological Niche from current climate, HADCM3, CCCMA and CSIRO.....	109
Figure 4. 11: BIOCLIM predicted areas of low suitability from current climate, HADCM3, CCCMA and CSIRO.....	110
Figure 4. 12: BIOCLIM predicted areas of medium suitability from current climate, HADCM3, CCCMA and CSIRO.....	111
Figure 4. 13: BIOCLIM predicted areas of high suitability from current climate, HADCM3, CCCMA and CSIRO.....	112
Figure 4. 14: BIOCLIM predicted areas of very high suitability from current climate, HADCM3, CCCMA and CSIRO.....	113
Figure 4. 15: BIOCLIM predicted areas of excellent suitability from current climate, HADCM3, CCCMA and CSIRO.....	114
Figure 4. 16: Areas for BIOCLIM predicted Ecologies with current climate	115
Figure 4. 17: Areas for BIOCLIM predicted Ecologies with HADCM3 by 2020	116
Figure 4. 18: Areas for BIOCLIM predicted Ecologies with HADCM3 by 2050	117
Figure 4. 19: Areas for BIOCLIM predicted Ecologies with HADCM3 by 2080	118
Figure 4. 20: Areas for BIOCLIM predicted Ecologies with CCCMA by 2020.....	119
Figure 4. 21: Areas for BIOCLIM predicted Ecologies with CCCMA by 2050.....	120
Figure 4. 22: Areas for BIOCLIM predicted Ecologies with CCCMA by 2080.....	121
Figure 4. 23: Areas for BIOCLIM predicted Ecologies with CSIRO by 2020.....	122
Figure 4. 24: Areas for BIOCLIM predicted Ecologies with CSIRO by 2050.....	123
Figure 4. 25: Areas for BIOCLIM predicted Ecologies with CSIRO by 2080.....	124

Figure 4. 26: Area Comparison of BIOCLIM predicted Ecological Niches from Current Climate (1950-2000), HADCM3, CCCMA and CSIRO.....	126
Figure 4. 27: Area Comparison of BIOCLIM True or False predicted Ecological Niches from current Climate (1950-2000), HADCM3, CCCMA and CSIRO.....	128
Figure 4.28: Area Comparison of DOMAIN predicted Ecological Niches from current Climate (1950-2000), HADCM3, CCCMA and CSIRO.....	130
Figure 4. 29: Comparison of prediction model performance using (a) AUC and (b) Kappa statistics.....	133
Figure 4. 30: Model Prediction Validation Comparison for BIOCLIM, BIOCLIM True or False and DOMAIN with HADCM3, CCCMA and CSIRO future climate projections	134
Fig. B. 1: BIOCLIM Model Validation Results for HADCM3 predicted malaria vector distribution cases, Bioclim (a), BIOCLIM True/False (b).....	180
Fig. B. 2: DOMAIN Model Validation Results for HADCM3 predicted malaria vector distribution cases, ROC (1a, 2a, 3a) and Kappa (1b, 2b, 3b).....	181
Fig. B. 3: BIOCLIM Model Validation Results for CCCMA predicted malaria vector distribution cases, Bioclim (a), BIOCLIM True/False (b).....	183
Fig. B. 4: DOMAIN Model Validation Results for CCCMA predicted malaria vector distribution cases, ROC (1a, 2a, 3a) and Kappa(1b, 2b, 3b).....	184
Fig. B. 5: BIOCLIM Model Validation Results for CSIRO predicted malaria vector distribution cases, BIOCLIM (a), BIOCLIM True/False (b)	186
Fig. B. 6: DOMAIN Model Validation Results for CSIRO predicted malaria vector distribution cases, ROC (1a, 2a, 3a) and Kappa (1b, 2b, 3b).....	187

LIST OF ACRONYMS

AMSD	climate change Adaptation, Mitigation Sustainable Development
AOGCM	Atmosphere Ocean Global Circulation Model
AR4	IPCC Fourth Assessment Report
ARIMA	Autoregressive integrated moving average
AUC	Area Under the Receiver Operating Characteristic Curve
BRT	Boosted Regression Trees;
CCCMA	Canadian Centre for Climate Modeling and Analysis
CCR	Correct Classification Rate
CEMs	Climate Envelope Models
CSIRO	Commonwealth Science and Industrial Research Organisation
DFA	Discriminant Function Analysis
ENM	Ecological Niche Modeling
ENSO	El Niño-Southern Oscillation
GAM	Generalized Additive Model;
GCMs	General Circulation Models
GDP	Gross Domestic Product
GIS	Geographic Information Systems
GLM	General Linear Model
GPS	Global Positioning System
HADCM3	Hadley Centre Coupled Model 3
HadCRUT 3	Hadley Centre/Climate Research Unit gridded surface temperature data set
HadRM3	Hadley Centre's regional climate model
ICPAC	IGAD Climate Prediction and Applications Centre
IPCC	Intergovernmental Panel on Climate Change
ITNs	Insecticide-Treated mosquito Nets (ITNs)
KEMRI	Kenya Medical Research Institute
LLINs	Long-Lasting Insecticide-treated Nets (LLINs)
LPDAAC	Land Processing Distributed Active Archive Center
LST	Land Surface Temperature
MARA/ARMA	Mapping Malaria Risk in Africa / <i>Atlas du Risque de la Malaria en Afrique</i>
MARS	Multivariate Adaptive Regression Splines

MEWS	Malaria Early Warning Systems
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NN	Neural Network
PRECIS	Providing Regional Climates for Impacts Studies
RCD	Regional climate downscaling
RCMs	Regional Climate Models
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RS	Remote Sensing
SAR	IPCC Second Assessment Report
SAT	Surface Air Temperature
SRES	Special Report on Emissions Scenarios storyline
SSTs	Sea Surface Temperatures
TAR	IPCC Third Assessment Report
TRMM	Tropical Rainfall Measuring Mission
USGS	United States Geological Survey
VBDs	Vector-borne diseases
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

1.1 Background

The Working Group II (WGII) of the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) has confirmed that global climate highlighted a wide range of predominantly negative impacts on human health (IPCC, 2007). Global warming increases the risk of some infectious diseases, particularly those that appear in warm areas. Evaluation on the evidence regarding the observed changes in human health and regional climate change has linked temperature trends and extreme temperature changes to a range of infectious and non-infectious disease outcomes. Deadly diseases often associated with hot weather, for instance the West Nile virus, Cholera and Lyme diseases are spreading rapidly throughout North America and Europe because increased temperatures in these areas allow disease carriers like mosquitoes, ticks, and mice to thrive.

Vector-borne diseases (VBDs) which are known to be sensitive to temperature and rainfall have been studied over time. Mostly, vector-borne diseases refer to infections that are transmitted by the bite of arthropods which suck either human or animal blood for their feed like mosquitoes or ticks. Malaria being one of those diseases and a fatal killer is mainly caused by five distinct species of plasmodium parasite, namely; (*Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*, *Plasmodium knowlesi*). The disease is transmitted from one affected individual to another by *Anopheline* mosquitoes. An estimated 216 million episodes of malaria were reported in 2010 worldwide, mostly amongst children under 5 years in the African Region (WHO, 2010). The number of global malaria deaths was estimated to be 1,238,000 in 2010 (Murray et al., 2012).

Since The Fourth Assessment Report (AR4) of IPCC, more research work has been done to elucidate the role of local warming on malaria transmission in the East African highlands. However, the effort has been thwarted by inefficient time series data on levels of drug resistance and intensity of intervention strategies on vector control. Earlier research had failed to find a clear increase in temperatures accompanying increases in malaria transmission, but new studies with aggregated meteorological data over longer periods have confirmed increasing temperatures since 1979 (Omumbo et al., 2011; Stern et al., 2011).

The influence of temperature on malaria development appears to be non-linear, and is vector-specific (Alonso et al., 2011). The strongly non-linear response to temperature means that even modest warming may drive large increases in transmission of malaria, if conditions are otherwise suitable (Alonso et al., 2011; Pascual et al., 2006). In Kenya, analysis of environmental factors associated with the malaria vectors *Anopheles gambiae* and *A. funestus* found that abundance, distribution, and disease transmission are affected in different ways by precipitation and temperature (Kelly-Hope et al., 2009). Although the incidence of malaria has reduced over much of East Africa (Stern et al., 2011), increased variability in disease rates has been observed in some high altitude areas (Chaves et al., 2012).

The greatest effect of climate change on transmission is likely to be observed at the extremes of the range of temperatures at which transmission occurs (Githeko et al., 2000). At around 30–32⁰C, the malaria vector capacity to thrive can increase substantially owing to a reduction in the extrinsic incubation period, despite a reduction in the vector's survival rate. Malaria vectors and their parasites also have the potential to be affected by environmental changes resulting from climate change. Environmental changes, either natural phenomenon or through

human intervention, alter the ecological balance and context within which vectors and their parasites breed, develop, and transmit disease (Patz et al., 2000).

The devastating change in climate has contributed to raise average temperatures in the Central Highlands district of Kenya, allowing the disease to creep into higher altitude areas where the population has little or no immunity. In 1989, the average temperature in the Central Highlands was 17°C, with malaria completely absent from the region as the parasite which causes it can only mature above 18°C. Currently, with the temperatures averaging 19°C, mosquitoes are carrying the disease into high altitude areas and epidemics have begun to break out in humans.

In Kenya, the Division of Malaria Control has cited that 70% of the total population is at risk of malaria (WHO, 2003). The World Health Organization officials have reported that increases in temperature and rainfall could accelerate the spread of malaria to high altitude areas as the vector's survival, generally known to be above 18°C would be enhanced, thereby increasing the possibility of malaria transmission. Rainfall affects the availability and suitability of breeding habitats while temperature affects the rate of mosquito and malaria parasite development, suitability of habitats and mosquito blood feeding rates. The geographic distribution of malaria parasites transmitted by mosquitoes of the genus *Anopheles* is the result of a complex interaction of biogeography, including biotic and abiotic factors that vary in both time and space.

In the early 1990s, there was little awareness of health risks posed by global climate change. This reflected a general lack of understanding of how the disruption of biophysical and ecological systems might affect the long-term wellbeing and health of populations. There was

little awareness among natural scientists that changes in their particular objects of study, for instance climate conditions, biodiversity stocks, ecosystem productivity, among others, were of potential importance to human health. Indeed, this was well reflected in the meager reference to health risks in the first major report of United Nation's Intergovernmental Panel on Climate Change published in 1991.

Subsequently, the situation changed as it was portrayed in the IPCC Second Assessment Report published in 1996. In the report, a full chapter was devoted to the potential risks to human health. The Third Assessment Report (TAR) published in 2001 went further to include discussion of some early evidence of actual health impacts, along with assessing potential health impacts. The report also highlighted the anticipated health impacts by major geographic regions. In this Third Assessment Report, the IPCC concluded that; *“Overall, climate change is projected to increase threats to human health, particularly in lower income populations, predominantly within tropical/subtropical countries”* (IPCC, 2001).

Broadly, The Third Assessment Report summarized that a change in climate conditions can have three kinds of health impacts as outlined below:-

- Those that are relatively direct, usually caused by weather extremes.
- The health consequences of various processes of environmental change and ecological disruption that occur in response to climate change.
- The diverse health consequences; (traumatic, infectious, nutritional psychological and others) that occur in demoralized and displaced populations in the wake of climate induced economic dislocation, environmental decline, and conflict situations.

Consideration of the relationships between climate change and vector-borne diseases suggests that warmer temperature is likely to have two major kinds of closely related, potentially detectable, outcomes: changes in vectors *per se*, and changes in vector-borne disease outcomes (Kovats et al., 2001). The ecology of some disease vectors in Africa are likely to be altered by climate change, hence consequently affecting the spatial and temporal transmission of such diseases. There have been a number of studies that reported associations between inter-annual variability in temperature and malaria transmission in the African highlands. In the highland areas of Kenya, malaria prevalence has been associated with rainfall and unusually high maximum temperatures (Githeko and Ndegwa, 2001).

Forecasting the range shifts of species in response to climate change is typically done by assuming that a species' current distribution represents its total climate tolerance, and creating a future 'climate envelope' based on predicted shifts in those abiotic conditions (Hughes et al., 1996; Iverson and Prasad, 2001; Pearson and Dawson, 2003; McClean et al. 2005; Hamann and Wang, 2006; Gómez-Mendoza and Arriaga, 2007). When predicting species' distribution, the models that are used make use of associations between environmental variables and known occurrence of the species. These aides in identifying environmental conditions within which species' populations can be maintained.

Therefore, the spatial distribution of the species' suitable environments can be estimated across any region of study. The two approaches that can be applied to characterize environmental conditions that are suitable for species are either mechanistic or correlative. When mechanistic models are used, they incorporate physiologically limiting mechanisms in a species' tolerance to environmental conditions. On the other hand, use of correlative models aims to estimate the suitable environmental conditions for a species by associating known

species' occurrence records with suites of environmental variables that can reasonably be expected to affect the species' physiology and probability of persistence.

1.2 Statement of the Problem

The National Malaria Strategy (2009–2017), launched by Kenya's Ministry of Public Health and Sanitation, identifies epidemic preparedness and response as a key approach to the containment of epidemics in Kenya. The National Malaria Strategy was developed in line with the Government's first Medium-Term Plan of Kenya Vision 2030 together with the Millennium Development Goals. Also, the strategy was drawn in partnership with Roll Back Malaria partnership goals and targets for malaria control. One of the drawn key strategic objectives and targets of the National Malaria Strategy was: to ensure that all malaria epidemic prone districts have the capacity to detect and the preparedness to respond to malaria epidemics annually by 2010 through capacity strengthening for epidemic preparedness and response. The WHO's Roll Back Malaria initiative notes that forecasting and early warning can reinforce local preparedness and allow authorities and communities to use cost-effective and timely control options to prevent excessive deaths.

However, despite the known causal links between climate and malaria transmission dynamics, there is still much uncertainty about the potential impact of climate change on malaria at local and global scales (IPCC, 2007). This is because of several factors, namely; the paucity of concurrent detailed historical observations of climate and malaria, the complexity of malaria disease dynamics, and the importance of non-climatic factors in determining infection and infection outcomes. Given the large populations living in highland areas of East Africa, the limitations of the analyses conducted, and the significant health risks of epidemic malaria, further research is warranted.

There has been an attempt to model the prevalence and distribution of malaria epidemic in Kenya, (Okara et al., 2010; Githeko et al., 2014). These studies did not investigate on Ecological Niche Modeling to project the future impact of climate change on malaria vector distribution. There has been no attempt to correlate the vector presence data with the underlying environmental factors, and further investigate the impact of climate change on the vector ecologies. Alongside weather monitoring and seasonal climate forecasts, other epidemiological and environmental factors are not incorporated into algorithms for malaria modeling to predict and project the future dynamics of malaria prevalence with climate change. Moreover, with the unpredictable change in climate, research on prediction models that integrates climate data and spatial-temporal changes is at its infancy.

Therefore, this study seeks to address the existing gap in spatial-temporal modeling of malaria vectors by correlating the spatial distribution of main malaria vectors in Kenya with the underlying ecological conditions with respect to climate change. There is urgent need to spatially integrate timely environmental data, biological data and the IPCC projected climate change scenarios to quantify the prevalence and distribution of malaria.

1.3 Objective of the study

The main objective of the study is to investigate spatial-temporal effect of climate change on the distribution of main malaria vectors in Kenya. The specific objectives were:-

- To model the spatial-temporal distribution of the main malaria vectors in Kenya.
- To determine through investigation the relationship between vector distribution and climate change.
- To project the future ecological niches for malaria vector occurrence in Kenya.
- To compare predictions from different IPCC future climate scenarios.

The research questions that guided the research were structured as follows:

- Is climate change a contributory factor to changing dynamics of malaria vector distribution in Kenya?
- What parameters determine the most appropriate spatial model for malaria distribution and prediction analysis?
- How does climate projection model and scenario change affect the predicted distribution of malaria vectors?

1.4 Justification and significance of the study

Climate change has the potential to significantly impact the distribution of malaria vectors. A number of species have been affected by recent climate change, with ranges expanding towards higher latitudes and longitudes (Parmesan and Yohe, 2003; Root et al., 2003). Further, recent studies have shown that it is challenging to attribute health outcomes to climate change or variability *per se* (WHO, 2003). Consequently, it is difficult to identify climate-related thresholds for human health.

Limited inclusions of different developmental scenarios in human health projections, and also the limited understanding of extent, rate of occurrence and adaptation measures of human populations to a changing climate justify the study. Important to note is the difficulty in generalizing health outcomes from one setting to another especially at the event of many diseases occurring with local transmission dynamics that cannot be represented in simple relationships. However, there are limited region-specific projections of climate change with significant impact on human health, and consideration of multiple, interacting and multi-causal health outcomes.

1.5 Scope and limitations

The research was conducted in Kenya, in order to model malaria prevalence in the entire country. Model generation was done based on IPCC projected climate for Hadley Centre Coupled Model 3 (HADCM3), Canadian Centre for Climate Modeling and Analysis (CCCMA) and Commonwealth Science and Industrial Research Organisation (CSIRO) models under A2a scenario. Biological data on malaria vector spatial distribution in Kenya was used to provide the presence data required for Ecological Niche Modeling. This was due to the fact that malaria prevalence and occurrence has taken new dimensions with changing climate, thus the need to quantify the role of climate change in altering malaria vector ecologies.

The PRECIS (Providing Regional Climates for Impacts Studies) model which has been customized for East Africa provides predicted rainfall and temperature patterns. PRECIS is essentially a regional climate modeling system. It is based on the third generation of the Hadley Centre's regional climate model (HadRM3), together with user-friendly data processing and visualization interface (Jones et al., 2004). The model's flexible design allows for applications in any region of the world. Just like any other regional climate models (RCMs), PRECIS is driven by boundary conditions simulated by General Circulation Models (GCMs). The RCMs do not replace GCMs, but they are powerful tools used together with the GCMs in order to add fine-scale detail to their broad-scale projections.

Global minimum and maximum mean temperature and precipitation data from WorldCLIM was acquired and used to extract the data for Kenya. Further processing was done to generate bioclimatic factors which are used in Ecological Niche Modeling. Climate envelopes are used to predict current ecological suitability of malaria vector survival and project future spatial-

temporal models. Climate change projections are simulated by Global Climate Models (GCMs). Different GCMs exist and for the same GCM, variants can be found that try to simulate climate change conditions related to different anthropogenic behaviors for the next century. The future trend analysis has been based on the climate projections (IPCC, 2001) Special Report on Emissions Scenarios storyline (SRES). The SRES describe the relationships between the forces driving greenhouse gas and aerosol emissions and their evolution during the 21st century. Figure 1.1 below shows the area of study.



Figure 1. 1: A map of Kenya showing the county boundaries.

CHAPTER 2: LITERATURE REVIEW

The organization of this chapter was done in eight sub-sections with intention to explain the observed changes in climate and the relevant studies that have been done to relate climate change and malaria prevalence.

2.1 Observed Climate Change and Variability

The history of the Earth has shown that the climate is always changing; from warm periods when the dinosaurs flourished, to the many ice ages when glaciers covered much of the land. The climate system is a complex, interactive system that consists of the atmosphere, land surface, snow and ice, oceans and other water bodies, and living things (IPCC, 2007). The climate system evolves in time under the influence of its own internal dynamics and due to changes in external factors that affect climate, called ‘forcings’. External forcings include natural phenomena such as volcanic eruptions and solar variations, as well as human-induced changes in atmospheric composition. The three fundamental ways that change the radiation balance of the Earth include:

- by changing the incoming solar radiation (e.g., by changes in Earth’s orbit or in the Sun itself);
- by changing the fraction of solar radiation that is reflected, called ‘albedo’ (e.g., by changes in cloud cover, atmospheric particles or vegetation); and
- by altering the long wave radiation from Earth back towards space (e.g., by changing greenhouse gas concentrations).

Therefore, ‘climate change’ refers to a change in the state of the climate that can be identified (e.g., using statistical tests) by changes in the mean and/or the variability of its properties, and that persists for an extended period, typically decades or longer (IPCC, 2007). Climate change may be due to internal processes and/or external forcings. Some external influences,

such as changes in solar radiation and volcanism, occur naturally and contribute to the total natural variability of the climate system.

Other external changes, such as the change in composition of the atmosphere that began with the industrial revolution, are the result of human activity. Internal variability is present on all time scales. Atmospheric processes that generate internal variability are known to operate on time scales ranging from virtually instantaneous (e.g., condensation of water vapour in clouds) up to years (e.g., troposphere-stratosphere or inter-hemispheric exchange). Other components of the climate system, such as the ocean and the large ice sheets, tend to operate on longer time scales. These components produce internal variability of their own accord and also integrate variability from the rapidly varying atmosphere (Hasselmann, 1976). In addition, internal variability is produced by coupled interactions between components, such as is the case with the El-Niño Southern Oscillation.

Distinguishing between the effects of external influences and internal climate variability requires careful comparison between observed changes and those that are expected to result from external forcing. These expectations are based on physical understanding of the climate system. Physical understanding is based on physical principles. This understanding can take the form of conceptual models or it might be quantified with climate models that are driven with physically based forcing histories. An array of climate models is used to quantify expectations in this way, ranging from simple energy balance models to models of intermediate complexity to comprehensive coupled climate models such as those that contributed to the multi-model data set (MMD) archive at the Program for Climate Model Diagnosis and Intercomparison (PCMDI).

Climate models are based on well-established physical principles that have been demonstrated to reproduce observed features of the recent and past climate changes. They attempt to simulate the behavior of the climate, in order to understand the key physical, chemical and biological processes which govern climate. There is considerable confidence that Atmosphere-Ocean General Circulation Models (AOGCMs) provide credible quantitative estimates of future climate change, particularly at continental and larger scales (IPCC, 2007). Atmosphere-Ocean Global Circulation Models (AOGCM), also known as 'coupled atmosphere-ocean models', are global climate models that model both atmospheric and ocean processes and interactions between them. The term Global Climate Model (GCM) is used to refer to climate models that reflect both atmosphere and ocean processes and feedbacks.

The use of AOGCMs has a limitation in projecting climate change at the regional and sub-regional level. This is due to the significant differences in climate occurrence at a scale below the resolution of the AOGCMs. Fully coupled GCM are typically on a resolution of 2-3° (~200-300 km) because they are computationally intensive. There has been an unprecedented, joint modeling effort by many groups worldwide. Currently, multi-model means are the basis of climate change projections. Quantitative assessment of differences between models, and in some instances, estimates of the probability of change of important climate system parameters has been reported to complement expert judgment.

Climate models provide better understanding of the climate system, clearly highlighting the past climate by comparing records of instrumental and paleo-climatic observations. The analysis enables the prediction of future climate change. As reported by Working Group I (WGI) in the Fifth Assessment report (AR5) (IPCC, 2013), over the 1951–2010 period, the

observed Global Mean Surface Temperature (GMST) increased by approximately 0.6°C. GHG increases likely contributed 0.5°C to 1.3°C, other anthropogenic forcings likely contributed -0.6°C to 0.1°C and natural forcings likely contributed -0.1°C to 0.1°C to observed GMST trends over this period. Internal variability likely contributed -0.1°C to 0.1°C to observed trends over this period (Knutson et al., 2013). GMST warmed strongly over the period 1900–1940, followed by a period with little trend, and strong warming since the mid-1970s (Section 2.4.3, Figure 10.1) (IPCC, 2013). Currently, these temperatures exceed the upper limit of natural (historical) variability. Climate is often defined as ‘average weather’, and it is usually described in terms of the mean and variability of temperature, precipitation and wind over a period of time, ranging from months to millions of years, the classical period being 30 years (IPCC, 2007).

Global climate models (GCMs) have been used extensively to project warming in the 21st Century due to mankind’s greenhouse gas pollution of the atmosphere. As reported by Meehl et al., (IPCC, 2007), an expert assessment based on the combination of available constraints from observations and the strength of known feedbacks simulated in the models used to produce the climate change projections have indicated that the equilibrium global mean Surface Air Temperature (SAT) warming for a doubling of atmospheric carbon dioxide (CO₂), or ‘equilibrium climate sensitivity’, is *likely* to lie in the range of 2°C to 4.5°C, with a most likely value of about 3°C. Equilibrium climate sensitivity is *very likely* larger than 1.5°C.

For fundamental physical reasons, as well as data limitations, values substantially higher than 4.5°C still cannot be excluded, but agreement with observations and proxy data is generally worse for those high values than for values in the 2°C to 4.5°C range. Climatologists assess

that most of that recent increase is due to human influence. Human societies have had long experience of naturally-occurring climatic vicissitudes due to mankind's greenhouse gas pollution of the atmosphere. The apparent change in world climate would influence the functioning of many ecosystems and their member species. Likewise, there would be impacts on human health.

The following diagram in Figure 2.1 shows an illustration of global mean near-surface temperature over the twentieth century from observations (black), and as obtained from 58 simulations produced by 14 different climate models driven by both natural and human-caused factors that influence climate (yellow). The multi-model ensemble mean of all these runs is also shown (red). As reported by Brohan et al., (2006), the vertical grey lines indicate the timing of major volcanic eruptions, (Hadley Centre/Climate Research Unit gridded surface temperature data set, (HadCRUT 3)). All data are shown as global mean temperature anomalies relative to the period 1901 to 1950, as observed (black), (HadCRUT3), and, in (a) as obtained from 58 simulations produced by 14 models with both anthropogenic and natural forcings.

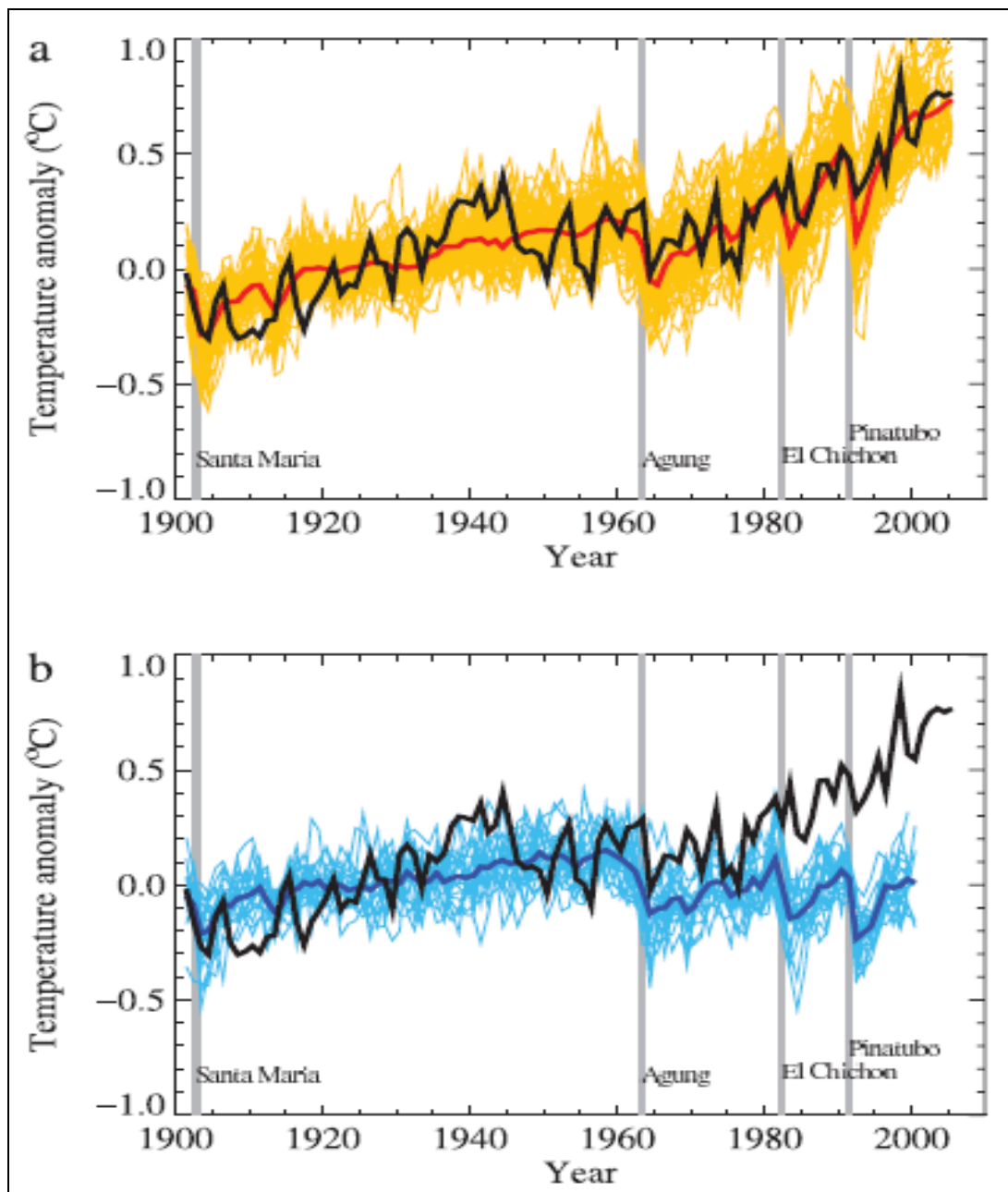


Figure 2.1: Comparison between global mean temperature anomalies ($^{\circ}\text{C}$) from observations (black) and AOGCM simulations forced with both anthropogenic and natural forcings(a) and (b) natural forcings only (Source: IPCC, 2007: Figure 9.5).

All data in Figure 2.1 are shown as global mean temperature anomalies relative to the period 1901 to 1950, as observed (black, Hadley Centre/Climatic Research Unit gridded surface temperature data set (HadCRUT3); Brohan et al., 2006) and, in (a) as obtained from 58 simulations produced by 14 models with both anthropogenic and natural forcings.

The multi-model ensemble mean is shown as a thick red curve and individual simulations are shown as thin yellow curves. Vertical grey lines indicate the timing of major volcanic events. Those simulations that ended before 2005 were extended to 2005 by using the first few years of the IPCC Special Report on Emission Scenarios (SRES) A1B scenario simulations that continued from the respective 20th-century simulations, where they were available. The simulated global mean temperature anomalies in (b) are from 19 simulations produced by five models with natural forcings only. The multi-model ensemble mean is shown as a thick blue curve and individual simulations are shown as thin blue curves. Simulations are selected that do not exhibit excessive drift in their control simulations (no more than 0.2°C per century). Each simulation was sampled so that coverage corresponds to that of the observations.

Modeling climate change scenarios has enabled identification of gaps in the development of regional and sub-regional scenarios. Global circulation models and regional climate models can be applied usefully to identify a range of uncertainties allowing strategic policy-making for adaptation. Regional climate downscaling (RCD) techniques, including both dynamical and statistical approaches, are being used to provide higher-resolution climate information than is available directly from contemporary global climate models. However, it is important that these techniques, together with the results they produce, be applied appropriately. The strengths and weaknesses should also be understood. This requires a better evaluation and quantification of the performance of the different techniques for application to specific problems.

For a future warmer climate, the current generation of models indicates that precipitation generally increases in the areas of regional tropical precipitation maxima (such as the

monsoon regimes) and over the tropical Pacific in particular, with general decreases in the subtropics, and increases at high latitudes as a consequence of a general intensification of the global hydrological cycle. Globally averaged mean water vapour, evaporation and precipitation are projected to increase. Intensity of precipitation events is projected to increase, particularly in tropical and high latitude areas that experience increases in mean precipitation. Even in areas where mean precipitation decreases (most subtropical and mid-latitude regions), precipitation intensity is projected to increase but there would be longer periods between rainfall events. There is a tendency for drying of the mid-continental areas during summer, indicating a greater risk of droughts in those regions. Precipitation extremes increase more than does the mean in most tropical and mid- and high-latitude areas (IPCC, 2007).

All models assessed in IPCC 2007, for all the non-mitigation scenarios considered, project increases in global mean surface air temperature (SAT) continuing over the 21st century, driven mainly by increases in anthropogenic greenhouse gas concentrations, with the warming proportional to the associated radiative forcing. There is close agreement of globally averaged SAT multi-model mean warming for the early 21st century for concentrations derived from the three non-mitigated IPCC Special Report on Emission Scenarios (SRES: B1, A1B and A2) including only anthropogenic forcing run by the AOGCMs (warming averaged for 2011 to 2030 compared to that of 1980 to 1999 is between $+0.64^{\circ}\text{C}$ and $+0.69^{\circ}\text{C}$, with a range of only 0.05°C).

Thus, this warming rate is affected little by different scenario assumptions or different model sensitivities, and is consistent with that observed for the past few decades. Possible future variations in natural forcings, for example, a large volcanic eruption, could change those

values somewhat, but about half of the early 21st-century warming is committed in the sense that it would occur even if atmospheric concentrations were held fixed at year 2000 values. By mid-century (2046–2065), the choice of scenario becomes more important for the magnitude of multi-model globally averaged SAT warming, with values of +1.3°C, +1.8°C and +1.7°C from the AOGCMs for B1, A1B and A2, respectively. About a third of that warming is projected to be due to climate change that is already committed. By late century (2090–2099), differences between scenarios are large, and only about 20% of that warming arises from climate change that is already committed (IPCC, 2007).

Scenarios are not simple projections, but are stories that present alternative images of how the future might unfold. Scenarios can help explore potential consequences of the interplay of multiple variables if they are carefully handled, and thereby help us to make considered and comprehensive decisions. The IPCC scenarios, contained in the Special Report on Emissions Scenarios (SRES), make projections into the next century and beyond and assume that climate change will be linear and involve gradual warming. But events of the last five years have overtaken the initial SRES scenarios. Climate has changed faster and more unpredictably than the scenarios outlined. Many of the phenomena assumed to lie decades in the future are already well underway. This faster pace of change on many fronts indicates that more sector-specific, short-term scenarios are needed.

Multi-model means of surface warming (relative to 1980–1999) for the scenarios A2, A1B and B1, are shown in the Figure 2.2 below as continuations of the 20th-century simulation. Values beyond 2100 are for the stabilization scenarios. Linear trends from the corresponding control runs have been removed from these time series. Lines show the multi-model means and shading denotes the ± 1 standard deviation range of individual model annual means. The

concept that the climate system exhibits commitment when radioactive forcing has changed is mainly due to the thermal inertia of the oceans.

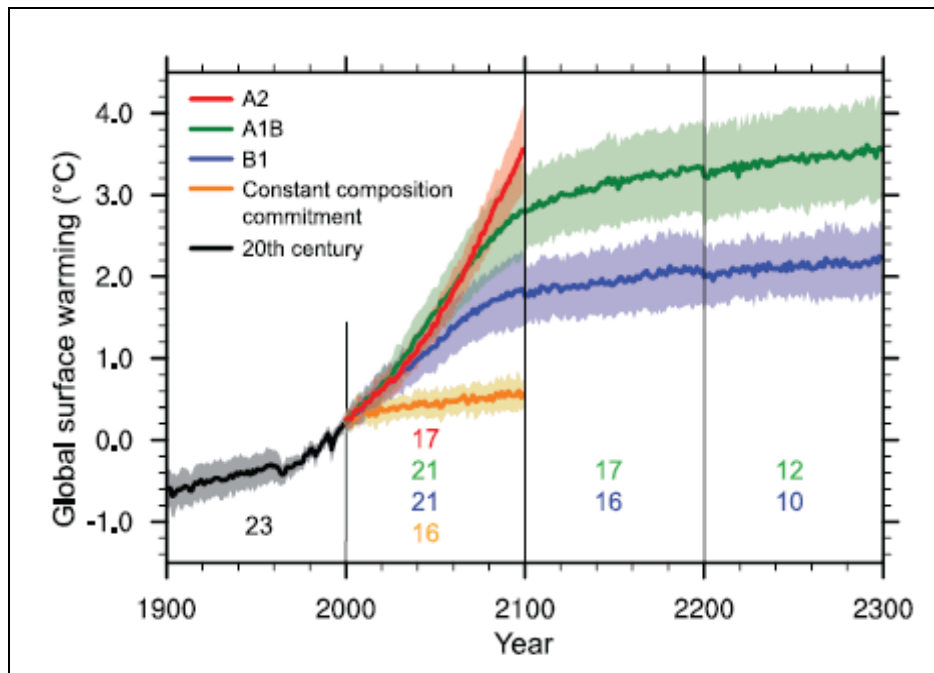


Figure 2. 2: Scenarios A2, A1B and B1 Multi-model means of surface warming ; relative to 1980–1999 (source: IPCC, 2007: Figure 10.4).

Global mean precipitation increases in all scenarios as shown in Figure 2.3, the right column, indicating an intensification of the hydrological cycle. Douville et al. (2002) show that this is associated with increased water-holding capacity of the atmosphere in addition to other processes. The multi-model mean varies approximately in proportion to the mean warming, though uncertainties in future hydrological cycle behaviour arise due in part to the different responses of tropical precipitation across models (Douville et al., 2005).

When expressed as a percentage of the mean simulated change for 1980 to 1999 (2.83 mm/day), the rate varies from about 1.4% °C⁻¹ in A2 to 2.3% °C⁻¹ in the constant composition commitment experiment. These increases are less than increases in extreme precipitation events, consistent with energetic constraints. Figure 2.3 below shows the time-

series of globally averaged surface warming. The numbers in parentheses following the scenario name represent the number of simulations shown. Values are annual means, relative to the 1980 to 1999 average from the corresponding 20th-century simulations, with any linear trends in the corresponding control run simulations removed. A three-point smoothing was applied. Multi-model (ensemble) mean series are marked with black dots (IPCC, 2007).

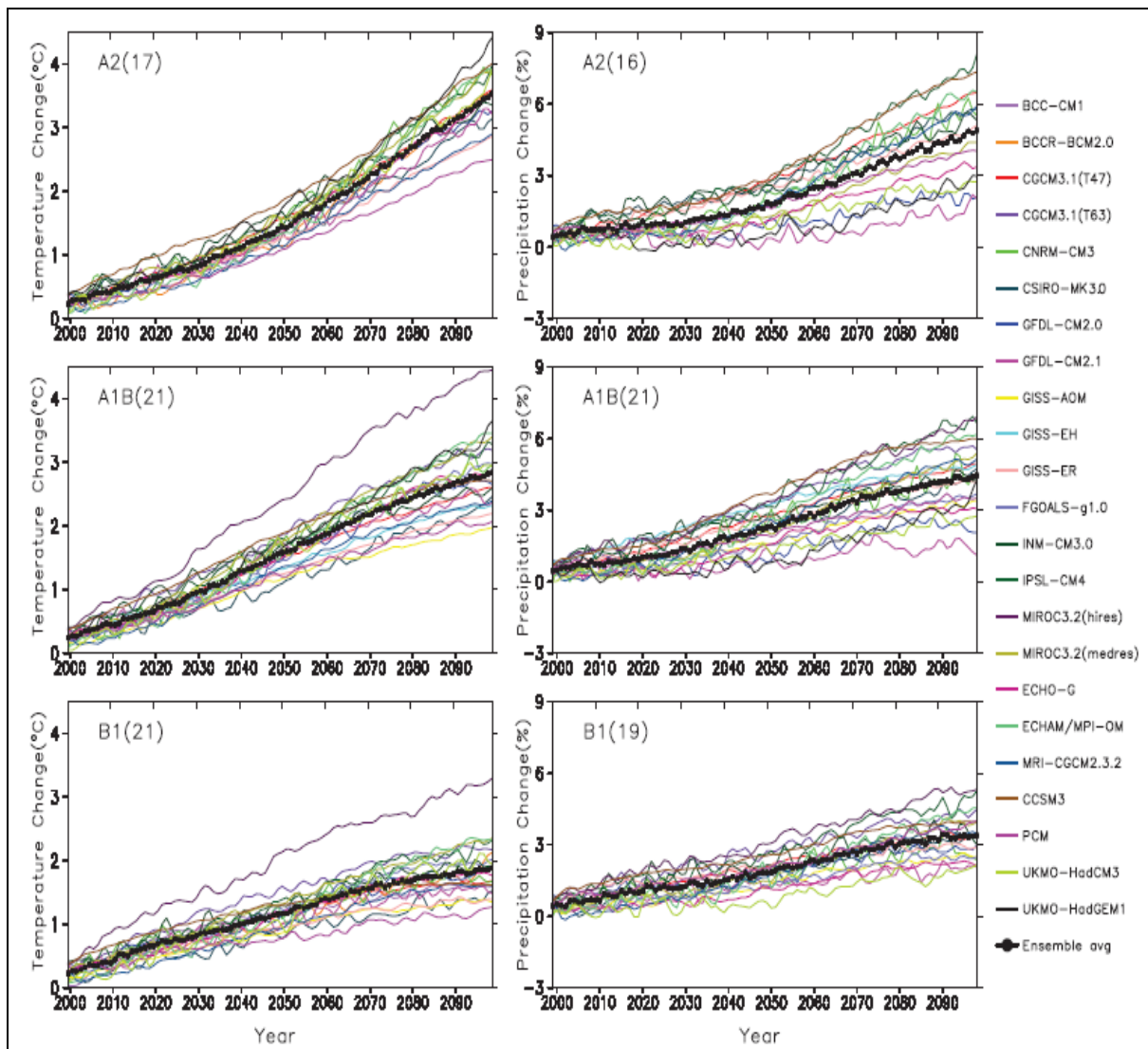


Figure 2.3: Time series of globally averaged (left) surface warming (surface air temperature change, °C) and (right) precipitation change (%) from the various global coupled models for the scenarios A2 (top), A1B (middle) and B1 (bottom) ; (source: IPCC, AR4, Fig. 10.5).

As reported by Wigley (2005), there are three specific definitions of climate change commitment:

- (i) the ‘constant composition commitment’, which denotes the further change of temperature (‘constant composition temperature commitment’ or ‘committed warming’), sea level (‘constant composition sea level commitment’) or any other quantity in the climate system, since the time the composition of the atmosphere, and hence the radiative forcing, has been held at a constant value;
- (ii) the ‘constant emission commitment’, which denotes the further change of, for example, temperature (‘constant emission temperature commitment’) since the time the greenhouse gas emissions have been held at a constant value; and
- (iii) the ‘zero emission commitment’, which denotes the further change of, for example, temperature (‘zero emission temperature commitment’) since the time the greenhouse gas emissions have been set to zero.

2.2 The Impact of Climate Change and Variability on Human health

Climate variability and change over the recent decades has already affected some health outcomes. Indeed, the World Health Organization reported that, in 2000, climate change was estimated to have been responsible for approximately 2.4% of worldwide diarrhea, and 6% of malaria in some middle-income countries (WHO, 2003). The first detectable changes in human health have been reported to be alterations in the geographic range and seasonality of certain infectious diseases. Vector-borne infections like malaria and dengue fever, and food-borne infections like salmonellosis peak in the warmer months. Warmer average temperatures combined with increased climate variability alter the pattern of exposure to thermal extremes and resultant health impacts in both summer and winter.

Unprecedented, the world population is presently encountering unfamiliar human-induced changes in the lower and middle atmospheres, and world-wide depletion of various other natural systems (e.g. soil fertility, aquifers, ocean fisheries, and biodiversity in general). Beyond the early recognition that such changes would affect economic activities, infrastructure and managed ecosystems, there is great recognition that global climate change poses risks to human population health. This topic is currently emerging as a major theme in population health research, social policy development and advocacy. Indeed, considerations of global climate-environmental hazards to human health have become a central role in the sustainability transition debate.

The IPCC TAR, 2001, highlighted the following findings:-

- An increase in the frequency or intensity of heat waves will increase the mortality and morbidity, principally in older age groups and among the urban poor.
- Any regional increases in climate extremes associated with climate change would cause deaths, and injuries, population displacement, and adverse effects on food production, freshwater availability and quality, and would increase the risk of infectious diseases, particularly in low-income countries.
- In some settings, the impacts of climate change may cause social disruptions, economic decline, and displacement of populations. The health impacts associated with socio-economic dislocation and population displacement are substantial.
- Changes in climate, including changes in climate variability, would affect many vector-borne infections. Populations at the margins of the current distribution of diseases might be particularly affected.
- Climate change represents an additional pressure on the world's food supply system and is expected to increase yields at higher latitudes and decrease yields at

lower latitudes. This would increase the number of undernourished people in the low-income world, unless major food redistribution is planned around the world.

- Assuming that the current emission levels continue, air quality in many large urban areas will deteriorate. Increase in exposure to ozone and other air pollutants could increase morbidity and mortality.

There are substantial differences between developed and developing countries in the incidence of various diseases. Over 40% of the population of the developing world, but only 2% of the developed world, is infected with at least one infectious or parasitic disease (Nathan et al, 1999). This wide disparity is mainly related to differences in socioeconomic conditions, including nutrition, sanitation, housing, working conditions, and availability of health care. Both developing-world and developed-world populations may experience increased rates of various infectious diseases if the climate changes. The association between climate change and infectious diseases is illustrated in Figure 2.4 below.

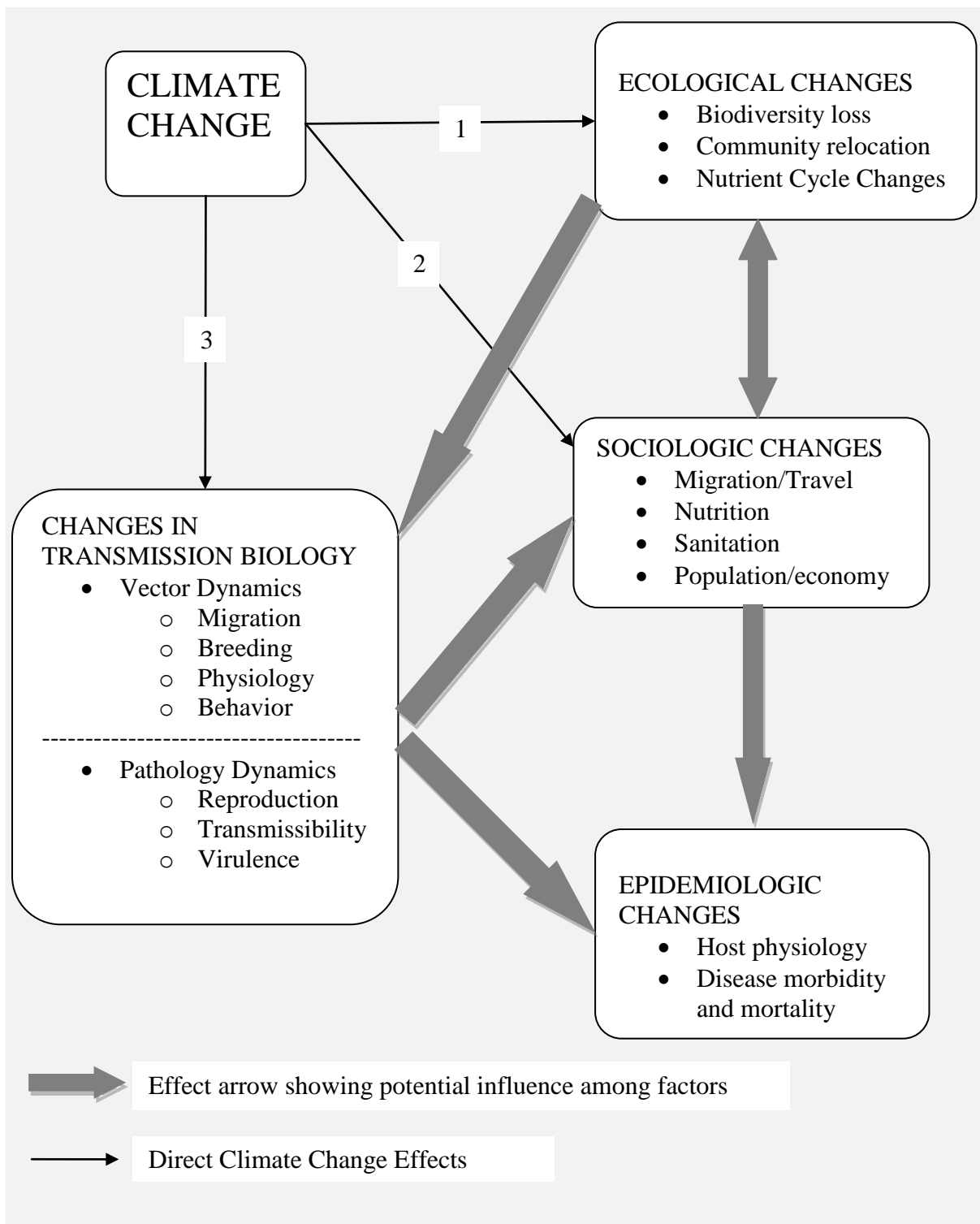


Figure 2. 4: Association between climate change and infectious diseases (source: Modified from IPCC, AR4)

2.3 Climate change and disease transmission in Africa

As reported in the fourth assessment report of IPCC, human health which is already compromised by a range of factors could be further negatively impacted by climate change and climate variability. The ecology of some disease vectors in Africa will possibly be altered, and consequently the spatial and temporal transmission of such diseases. Assessments of health in Africa show that many communities are already impacted by health stresses that are coupled to many causes, including poor nutrition, (IPCC, 2007). Hay et al., 2002 and Pascual et al., 2006 noted that the resurgence of malaria and links to climate and/ or other causal ‘drivers’ of change in the highlands of East Africa has recently attracted much attention and debate. For instance, in areas that have two rainy seasons, from March to May (MAM) and September to November (SON), there is an alternation with more rain falling in SON than previously the case in the northern sector of East Africa, (Schreck and Semazzi, 2004). The SON period is relatively warm, and higher rainfall is likely to increase malaria transmission. This is owing to the reduction in the duration required for larval development.

The spread of malaria into other new areas like the observations of malaria vector *Anopheles arabiensis*, in the central highlands of Kenya, where no malaria vectors have been observed previously, has been recorded, (Chen et al., 2006). Further insights into the observed warming trends are provided in four high-altitude sites in East Africa, from the end of 1970s onwards, (Pascual et al., 2006). Such trends contribute significantly to the biological implications of malaria vector population.

Micro-climate change caused by land-use changes, for instance, swamp reclamation for agricultural use and deforestation in the highlands of western Kenya, create suitable environment for the survival of *Anopheles gambiae* larvae, thus increasing the risk of

malaria, (Munga et al., 2006). Mosquito population rates and larval-to-pupal development have been observed to be significantly faster in farmland habitats than in swamp and forest habitats, (Munga et al., 2006). Malaria epidemics can also be triggered by floods in arid and semi-arid areas, (Thomson et al., 2006).

2.4 Models for malaria Early Warning System

Malaria early-warning systems based on vulnerability assessment and rainfall variability were used in India in the early part of the 20th century, (Gill, 1923). However, there has been a concerted effort in the recent years to develop such systems for epidemic prone areas in Africa. Rainfall anomalies are widely considered to be a major driver of inter-annual variability of malaria incidence in the semi-arid areas of Africa, (Connor et al., 1999). Analyses of time-series malaria and climate data have been conducted over the last century in many parts of the world, and have indicated that rainfall excess (or occasionally drought) is correlated with changes in malaria incidence in certain eco-epidemiologic settings, apparently as a result of its impact on the population dynamics of the *Anopheles* spp. mosquito vector, (Koenraadt et al., 2004). In some areas of the world where Sea Surface Temperatures (SSTs) in the Pacific and the El Niño-Southern Oscillation (ENSO) are important predictors of climate events, significant correlations between malaria incidence anomalies and SSTs have been observed (Kovats et al., 2003).

The early detection, containment and prevention of malaria epidemics are one of the four technical elements of the Global Malaria Control Strategy (WHO, 1993). The Roll Back Malaria (RBM) strategy (WHO, 1998) outlined a range of existing tools in time series analysis which can be useful in the development of a MEWS. As reported in RBM/WHO (2001), these tools include spectral analyses and autoregressive integrated moving average

(ARIMA). For instance, time series analysis was used to assess the contribution of climate to a malaria epidemic in Rwanda (Loevinsohn, 1994). The best fitting model from this study was developed using multiple linear regression. The model was developed with the following parameters: Monthly incidence = $4.32 + 1.64 \cdot \text{LN}(\text{monthly mean minimum temperature lagged by one month}) + 0.83 \cdot \text{LN}(\text{monthly mean minimum temperature lagged by two months}) + 5.34 \cdot 10^{-4} \cdot (\text{rainfall lagged by three months}) + 7.7 \cdot 10^{-4} \cdot (\text{rainfall lagged by four months})$.

According to Githeko et al., 2014, two general models were developed and validated for climate and ecosystem-based early malaria epidemic prediction models in East Africa. These models were referred to as the additive and the multiplicative models. The additive model was generated for the poorly drained U-shaped valley ecosystem while the multiplicative model was for the well drained V-shaped valleys. The additive model can be expressed as follows in Equation 2.1 (ibid):

$$\text{Epidemic Risk } ER^{i+4} = \frac{(T^i + R^{i+2})}{(T^m + R^m)} \times 100 \dots\dots\dots(2.1)$$

The additive model with exponential temperature effect can be expressed as in Equation 2.2:

$$\text{Epidemic Risk } ER^{i+3} = \frac{((T^i)^2 + R^{i+2})}{((T^m)^2 + R^m)} \times 100 \dots\dots\dots(2.2)$$

The multiplicative model is expressed as in Equation 2.3:

$$\text{Epidemic Risk } ER^{i+4} = \frac{(T^i \times R^{i+2})}{(T^m \times R^m)} \times 100 \dots\dots\dots(2.3)$$

The additive +18⁰C model can be written as Equation 2.4:

$$\text{Epidemic Risk } ER^{i+3} = \frac{\left((T-18)^i + (R^{i+2}) \right)}{(T-18)^m + R^m} \times 100 \dots\dots\dots(2.4)$$

Where;

T^i = temperature code at month i

R^{i+2} = Rainfall code at month $i + 2$

T^m = highest temperature anomaly code in the climatology data

R^m = maximum rainfall anomaly code in the climatology data

ER^{i+4} = epidemic risk at month $i + 4$

Malaria is a significant public health concern in Afghanistan, with approximately 60% of the population, or nearly 14 million people, living in malaria-endemic area. Afghanistan's diverse landscape contributes to the heterogeneous malaria distribution. Research aimed to understand the environmental effects on malaria transmission has been essential to the effective control of malaria in Afghanistan. Malaria transmission has been shown to be dependent on a number of environmental and meteorological variables (Safi et al., 2010). For countries in the tropics and the subtropics, rainfall is normally the most important variable, except for regions with high altitude where temperature may also be important.

In order to predict malaria risk in Afghanistan, malaria surveillance data collected from Afghan Ministry of Public Health and satellite-derived environmental data on precipitation, Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI) were used. Precipitation data was measured from the Tropical Rainfall Measuring Mission (TRMM), which is a joint mission between NASA and the Japan Aerospace Exploration

Agency (NASA GSFC, 2010). TRMM precipitation data with monthly resolution and at 0.25° from the GIOVANNI interface (Acker and Leptoukh, 2007). Both LST and NDVI were measured by Moderate Resolution Imaging Spectroradiometer (MODIS) aboard Terra and Aqua Spacecraft (NASA, 2009). The MODIS datasets are distributed by the Land Processing Distributed Active Archive Center (LPDAAC) of the United States Geological Survey (USGS).

The retrieved 8-day LST data was at 1 km spatial resolution with a sinusoidal projection, where the Normalized Difference Vegetation Index (NDVI) is at a monthly resolution. These two datasets were re-projected to a geographical projection compatible with the other datasets. In addition, the 8-day LST data was subsequently aggregated to monthly so as to synchronize with the malaria data temporal resolution. For each of the environmental variables, lagged variables were created. The lags were from the previous 1 month up to 4 months. All data was divided into two datasets, that is fit and prediction. The fit dataset was used in estimating the parameters, and the prediction dataset was reserved for forecast.

The modeling approaches used were; Neural Network (NN) and General Linear Model (GLM), specifically, Poisson regression. Neural Network modeling is capable of capturing the non-linear relationship. NN typically consists of interconnected nodes arranged in three major layers: input, hidden and output. In this study the number of nodes in the input layer was limited to three in order to reduce the model complexity. The input nodes are the environmental variables. The number of nodes in the hidden layer was limited to two for the same reason. There was only one node in the output layer that represents the level of monthly malaria case. Feed-forward and back-propagation neural network were employed.

The NN performance was measured by the Root Mean Square Error (RMSE) of both the fit and prediction dataset. Linear regression, which is a widely used method to predict the risk of infectious diseases, was used. Stepwise regression method was used in order to eliminate insignificant environmental variable predictors. As is the case for any infectious disease, the number of malaria cases in any given month depends on the number of cases in the previous months. Thus, the autocorrelation terms for the preceding four months were also used as input in the stepwise regression. In order to account for any possible components not represented in the environmental input parameters, sinusoidal terms were included, which consist of one factor for the trigonometric functions *SINE* and another for the *COSINE* component. Therefore, the monthly malaria case for each province can be written according to Equation 2.5 (ibid):

$$C(t) = \sum_{\tau=0}^4 EV(t-\tau) + \sum_{\tau=0}^4 AR(t-\tau) + S(t) + f(t) \dots\dots\dots (2.5)$$

Where $C(t)$ =Total cases in month t

$$EV(t) = T(t) + NDVI(t) + P(t) \dots\dots\dots(2.5.1)$$

$$AR(t) = C(t) \dots\dots\dots(2.5.2)$$

$$S(t) = \sin\left(\frac{2\pi t}{T}\right) + \cos\left(\frac{2\pi t}{T}\right) \dots\dots\dots(2.5.3)$$

$T(t)$ =Average temperature in month t

$NDVI(t)$ =Average *NDVI* in month t

$P(t)$ =Total precipitation in month t

$f(t)$ is the general trend accounting for factors other than environmental in nature such as improvement in public health support or population movements.

The regression model is thus:

$$C(t) = \sum_j \beta_j X_j(t) \dots\dots\dots(2.6)$$

Where; β_j is the weight for the predictor X_j that was selected using stepwise regression.

For the NN, the inputs were the previous month case, previous month LST and previous month NDVI.

2.5 Malaria Early Warning and Spatial mapping

The project for Mapping Malaria Risk in Africa / *Atlas du Risque de la Malaria en Afrique* (MARA/ARMA) collaboration was initiated to provide an Atlas of malaria for Africa, containing relevant information for rational and targeted implementation of malaria control (MARA/ARMA, 1998). The MARA/ARMA initiative was non-institutional and was operated in the spirit of an open collaboration. A group of dedicated African scientists based at institutions across the continent were involved in the project. The set objectives for the project were:-

- To map malaria risk in Africa through:-
 - i.) collection of published and unpublished malaria data.
 - ii.) spatial modeling of malaria distribution, seasonality and endemicity.
- To disseminate relevant information to national and international decision makers and other end users, in a range of useful formats.
- To develop capacity in malaria / health GIS.

Several malaria prevalence spatial models were generated in order to fulfil the set objectives.

In one of the models, malaria survey data was analysed statistically against co-factors of

malaria such as climate and the presence of water bodies, thereby predicting malaria endemicity across the entire region. Another model that defines endemic malaria distribution in Africa was generated based on the biological constraints placed on the parasite and vector by temperature and rainfall. Malaria seasonality model was generated to define the duration of transmission season in months, including the start and end of transmission season. Such information is handy in intervention measures as different interventions are suitable under different situations of malaria seasonality. However, the created continental database of malaria survey results (MARA/ARMA, 1998) provides the opportunity for producing empirical models and maps of malaria distribution at regional and eventually at continental level, but not early warning system *per se*.

2.6 Factors considered Ecological Niche Modeling (ENM)

Climatic factors have been reported to greatly influence the pattern and level of malaria transmission globally. Temperature, rainfall and humidity are the most important climatic factors that directly affect malaria transmission. The ranges of minimum and maximum temperature greatly affect the development of the malaria parasite and its mosquito vector, which determines malaria transmission. Temperatures are higher around the equator and are known to vary minimally throughout the year. Temperatures decrease progressively as distance increases north or south of the equator. The warm and relatively constant temperature in tropical Africa is one of the reasons for high levels of malaria transmission near the equator. The right amount of rainfall is often an important factor in providing water collections that support vector breeding, which appears mainly after the rains. Therefore, the period following the rainy season is most conducive and malaria transmission rate is highest.

Temperature affects the malaria parasite life cycle. The time required for the parasite to complete its development in the mosquito can be shorter or longer than usual depending on the temperature. A decrease in temperature results in increase in the number of days necessary to complete the development for a given *Plasmodium* species. The time needed for the parasite to complete its development in the mosquito, decreases to less than 10 days as temperature increases from 21°C to 27°C, with 27°C being the optimum. The maximum temperature for parasite development is 40°C. Below 18°C, the life cycle of *Plasmodium falciparum* in the mosquito body is limited. The minimum temperatures are between 14–19°C, with *Plasmodium vivax* surviving at lower temperatures than *Plasmodium falciparum*. Malaria transmission in areas colder than 18°C can sometimes occur because the *Anopheles* often live in houses, which tend to be warmer than the outside temperature (OpenLearn Labspace website).

Temperature is also a determinant factor for the development of the mosquito larva as it matures more quickly at higher temperatures. Higher temperatures as well increase the number of blood meals taken and the number of eggs laid by the mosquitoes thus amplifying the number of mosquitoes in a given area. The minimum temperature for mosquito development is between 8–10°C; the optimum temperature is 25–27°C, and the maximum temperature for it is 40°C. Altitude influences the distribution and transmission of malaria indirectly, through its effect on temperature. Increase in altitude leads to temperature decreases, making the highlands colder as compared to the warmer lowlands. With increase in altitude to beyond 2,400 metres, the temperature does not go high enough to support malaria transmission and these areas are free of malaria (OpenLearn Labspace website).

Different *anopheline* mosquitoes prefer different types of water bodies in which they can breed. Water collections that support vector breeding appear mainly after the rains, and therefore malaria transmission is highest following the rainy season. Too much rainfall can flush away breeding habitats temporarily, but mosquitoes start breeding as soon as the rain stops. More often than not, flushing has a bigger impact on vector breeding habitats in the highlands and hilly areas than in the lowland plains. It is of essence to note that not all water collections are suitable for the mosquito life cycle. Rain water collections are the most important breeding ground, as the *anopheline* mosquitoes prefer to breed in fresh water collections created after the rainy season. Such water bodies may be clear or muddy (OpenLearn Labspace website).

The breeding habitat for different mosquitoes influences malaria transmission. Some mosquitoes breed in small pools that are partially or completely exposed to the sun, while others prefer to breed in shaded stagnant pools. Those adapted to breeding close to human settlements, and able to breed in wide ranges of environments, are better vectors of malaria than those that breed away from human habitation. *Anopheles gambiae* mosquitoes breed in a wide range of habitats, including small water collections such as hoof-prints, water-filled holes in rocks and trees, as well as dams, river beds and lake shores. Consequently they are responsible for much of the malaria transmission in Africa.

Less rainfall and drought can also prevent mosquito breeding and malaria transmission in other areas. Such places are usually covered by vegetation throughout the year and streams and rivers often flow rapidly. Lack of rains interrupt the flow of streams and pooling occurs along the stream. Thus a favorable environment for mosquito breeding occurs. Malaria vectors mainly breed in stagnant water collections, rarely in slightly moving waters and never

in rapidly flowing rivers and streams. In drier areas, rainfall can also affect malaria transmission indirectly through its effect on humidity. Vegetation cover increases after rainfall, which in turn increases the relative humidity of the environment.

Relative humidity affects the activity and survival of mosquitoes, hence malaria transmission. Mosquitoes should live at least 8–10 days to be able to transmit malaria. This is the duration of time required for the parasite to complete its development inside the gut of the mosquito host. Premature death of the mosquito leads to failure of parasite development and transmission. Mosquitoes' survival thrives under conditions of high humidity. They also become more active with increase in humidity. They are more active and prefer feeding during the night for the reason that the relative humidity of the environment is higher at night. If the average monthly relative humidity is below 60%, it is believed that the life of the mosquito is so short that very little or no malaria transmission is possible. Malaria transmission could be high due to the following factors:

- *Immediately following the rains*, the explanation for this is that plenty of water collections for vector breeding are available after the rainy season.
- *When the temperature is hot*, the reason is that temperature speeds up vector and parasite development.
- *When the rains fail and there is drought*, as rivers and small streams usually slow down into pools, thus stagnant water collections for vector breeding.
- *When the fields are covered with vegetation*, because when the vegetation cover is high the humidity increases; higher humidity helps the mosquito to live longer and transmit malaria (OpenLearn Labspace website).

Non-climatic factors are those that affect malaria transmission, but which are not related to the climate. These factors include, and not limited to; the type of vector, the type of parasite, drug resistance in parasites, environmental development and urbanization, population

movement and migration, the level of immunity to malaria in the human hosts, and insecticide resistance in mosquitoes. All these have a role in affecting the severity and incidence of malaria, over and above the climatic factors.

Mosquitoes need a blood meal to develop and reproduce, which can either come from humans or animals, with those that feed on both equally being weaker malaria vectors. The most efficient carriers of malaria are those that feed on humans than those that feed on animals. Not all mosquitoes transmit malaria but only *Anopheles* mosquitoes can carry the malaria parasite, with *Anopheles arabiensis*, being responsible for more than 95% of malaria transmissions. Different species of *Anopheles* mosquitoes differ in their capacity to transmit malaria, depending on the biology and behaviour of the mosquitoes. Mosquitoes in the *Anopheles gambiae* group, which includes *Anopheles arabiensis*, are the most efficient malaria vectors in the world. These mosquitoes are found only in Africa, thus vouching for the higher incidence of malaria in Africa as compared to other parts of the world.

Different human hosts are either immune, or non-immune to malaria, hence affecting the pattern of malaria transmission and the severity of the disease. Immune people often have a better chance of tolerating the effects of malaria and surviving the disease than non immune people. In highly endemic areas, children under five years of age and pregnant women have weak immunity to malaria infection and are the most at risk. Immunity to malaria develops slowly after several infections and children usually gain their immunity at least after five years. Certain population groups can either be infected by some types of malaria parasites, but not by others. For example most Africans south of the Sahara can get infected by *falciparum* malaria, but not by *vivax* malaria. *Falciparum* malaria which is the deadliest form of malaria is highly prevalent in the African continent. This explains why most of the deaths due to malaria occur in Africa (OpenLearn Labspace website).

2.7 IPCC Reported Climate Change and Variability in African Region

The Intergovernmental Panel on Climate Change (IPCC) evaluated the risk of climate change caused by human activity in TAR. In the future trend analysis described in TAR (IPCC, 2001), reference is made to Special Report on Emissions Scenarios storylines (SRES). The SRES describes the relationships between the forces driving greenhouse gas and aerosol emissions and their evolution during the 21st century.

Regional averages of temperature and precipitation projections for Africa from a set of 21 global models in the multi-model dataset (MMD) for the A1B scenario is shown in the Table 2.1 below which is extracted from IPCC 2007 report. The mean temperature and precipitation responses were first averaged for each model over all available realizations of the 1980 to 1999 period from the 20th Century Climate in Coupled Models (20C3M) simulations and the 2080 to 2099 period of A1B. Computing the difference between these two periods, the Table shows the minimum, maximum, median (50%), and 25% and 75% quartile values among the 21 models, for temperature (°C) and precipitation (%) change. Regions in which the middle half (25–75%) of this distribution is all of the same sign in the precipitation response are colored light brown for decreasing and light blue for increasing precipitation (IPCC, 2007).

Table 2. 1: Reported Temperature and Precipitation change: Regions are West Africa (WAF), East Africa (EAF), South Africa (SAF) and Sahara (SAH) (Source: Table 11.1, IPCC 2007, pg 854.

Region ^a	Season	Temperature Response (°C)						Precipitation Response (%)						Extreme Seasons (%)		
		Min	25	50	75	Max	T yrs	Min	25	50	75	Max	T yrs	Warm	Wet	Dry
AFRICA																
WAF	DJF	2.3	2.7	3.0	3.5	4.6	10	-16	-2	6	13	23		100	21	4
	MAM	1.7	2.8	3.5	3.6	4.8	10	-11	-7	-3	5	11		100		
12S,20W to	JJA	1.5	2.7	3.2	3.7	4.7	10	-18	-2	2	7	16		100	19	
	SON	1.9	2.5	3.3	3.7	4.7	10	-12	0	1	10	15		100	15	
22N,18E	Annual	1.8	2.7	3.3	3.6	4.7	10	-9	-2	2	7	13		100	22	
EAF	DJF	2.0	2.6	3.1	3.4	4.2	10	-3	6	13	16	33	55	100	25	1
	MAM	1.7	2.7	3.2	3.5	4.5	10	-9	2	6	9	20	>100	100	15	4
12S,22E to	JJA	1.6	2.7	3.4	3.6	4.7	10	-18	-2	4	7	16		100		
	SON	1.9	2.6	3.1	3.6	4.3	10	-10	3	7	13	38	95	100	21	3
18N,52E	Annual	1.8	2.5	3.2	3.4	4.3	10	-3	2	7	11	25	60	100	30	1
SAF	DJF	1.8	2.7	3.1	3.4	4.7	10	-6	-3	0	5	10		100	11	
	MAM	1.7	2.9	3.1	3.8	4.7	10	-25	-8	0	4	12		98		
35S,10E to	JJA	1.9	3.0	3.4	3.6	4.8	10	-43	-27	-23	-7	-3	70	100	1	23
	SON	2.1	3.0	3.7	4.0	5.0	10	-43	-20	-13	-8	3	90	100	1	20
12S,52E	Annual	1.9	2.9	3.4	3.7	4.8	10	-12	-9	-4	2	6		100	4	13
SAH	DJF	2.4	2.9	3.2	3.5	5.0	15	-47	-31	-18	-12	31	>100	97		12
	MAM	2.3	3.3	3.6	3.8	5.2	10	-42	-37	-18	-10	13	>100	100	2	21
18N,20E to	JJA	2.6	3.6	4.1	4.4	5.8	10	-53	-28	-4	16	74		100		
	SON	2.8	3.4	3.7	4.3	5.4	10	-52	-15	6	23	64		100		
30N,65E	Annual	2.6	3.2	3.6	4.0	5.4	10	-44	-24	-6	3	57		100		

Signal-to-noise ratios for these 20-year mean responses is indicated by first computing a consensus standard deviation of 20-year means, using those models that have at least three realizations of the 20C3M simulations and using all 20-year periods in the 20th century. The signal is assumed to increase linearly in time, and the time required for the median signal to reach 2.83 ($2 \times \sqrt{2}$) times the standard deviation is displayed as an estimate of when this signal is significant at the 95% level (IPCC, 2007). These estimates of the times for emergence of a clearly discernible signal are only shown for precipitation when the models are in general agreement on the sign of the response, as indicated by the coloring. The frequency (%) of extremely warm, wet and dry seasons, averaged over the models, is also presented.

Values are only shown when at least 14 out of the 21 models agree on an increase (bold) or a decrease in the extremes. A value of 5% indicates no change, as this is the nominal value for the control period by construction. The regions are defined by rectangular latitude/longitude boxes and the coordinates of the bottom left-hand and top right-hand corners of these are given in degrees in the first column under the region acronym (see table notes for full names of regions). Information is provided for land areas contained in the boxes except for the Small Islands regions where sea areas are used and for Antarctica where both land and sea areas are used (IPCC, 2007) on page 854.

2.8 Geostatistical Modeling and Prediction

The environment is continuous, but in general, properties at only a finite number of places can be measured, thus obtaining point samples. The samples deal with properties that vary in ways that are far from systematic and at all spatial scales. The environment extends more or less continuously in two dimensions. Its properties have arisen as the result of the actions and interactions of many different processes and factors. Each process might itself operate on several scales simultaneously, in a non-linear way, and with local positive feedback. The environment, which is the outcome of these processes, varies from place to place with great complexity and at many spatial scales, from micrometres to hundreds of kilometres. The major changes in the environment are obvious enough, but others are more subtle. When the variation at different spatial resolutions is described, insight can be gained into the processes and factors that cause or control it, and so predict in a spatial sense and manage resources.

For any qualitative data analysis, geostatistics enables estimation or prediction without bias and with minimum error. However, geostatistics can never provide complete information, of course, but, given the data, it can permit estimation of the probabilities that true values

exceed specified thresholds. Measurements are made on small volumes of material or areas a few centimetres to a few metres across, regarded as point samples or supports.

In some instances the supports are enlarged by taking several small volumes of material and mixing them to produce bulked samples. In others several measurements might be made over larger areas and averaged rather than recorded as single measurements. Even so, these supports are generally very much smaller than the regions themselves and are separated from one another by distances several orders of magnitude larger than their own diameters. Nevertheless, they must represent the regions, preferably without bias.

An additional feature of the environment is that at some scale the values of its properties are positively related, that is autocorrelated. Places close to one another tend to have similar values, whereas ones that are farther apart differ more on average. Geostatistics expresses this intuitive knowledge quantitatively and then uses it for prediction. However, error in the estimates is inevitable, but by quantifying the spatial autocorrelation at the scale of interest, the errors can be minimized and also estimated.

Measurements of the environment are made on small bodies of material or supports separated from one another by relatively large distances. These constitute a sample from a continuum that cannot be recorded everywhere. The values that can be lying in the intervening space should be predicted in a spatial sense from their more or less sparse data. For instance, meteorologists can predict rainfall from their rain gauges, hydrogeologists predict flow properties in rock from their measurements in boreholes, mining engineers estimate ore grades from diamond drill cores, and agronomists estimate concentrations of elements in the

soil from auger samples. Therefore, mapping the spatial distributions of the variables provides a basis for quantitative and qualitative analysis.

Mostly, attempts of spatial prediction have been mathematical, based on geometry and some appreciation of the physical nature of the phenomena. Most of them take into account only systematic or deterministic variation, but not any error. In these respects, the methods fall short of what is needed practically. In geostatistical prediction, Kriging is the logically more conclusive than the other spatial prediction attempts in that it builds on them and overcomes their weaknesses. Nearly all the methods of prediction, including the simpler forms of Kriging, can be seen as weighted averages of data (Webster and Oliver, 2007). The general prediction formula is as shown in Equation 2.7:

$$Z^*(X_0) = \sum_{i=1}^N \lambda_i Z(X_i) \dots\dots\dots(2.7)$$

- where; $Z^*(X_0)$ = the prediction;
- X_0 = a target point for which we want a value, the $Z(X_i)$; $i = 1, 2, \dots, N$;
- N = total number of sample points;
- X_i = measured data;
- λ_i = assigned weights.

Assigning of weights for some common methods of prediction is done as described in the following sub-sections.

2.8.1 Thiessen Polygons

Thiessen (Voronoi) polygons define individual areas of influence around each point of a set of sample points. The boundaries of the Voronoi polygons define the area that is closest to each point relative to all other points. They are mathematically defined by the perpendicular

bisectors of the lines between all sampling points. These polygons can be used for instance to delineate a set of soil sampling points to define non overlapping polygons for each soil type. The following diagram in Figure 2.5 illustrates Thiessen polygons generated from a feature layer of point sample data.

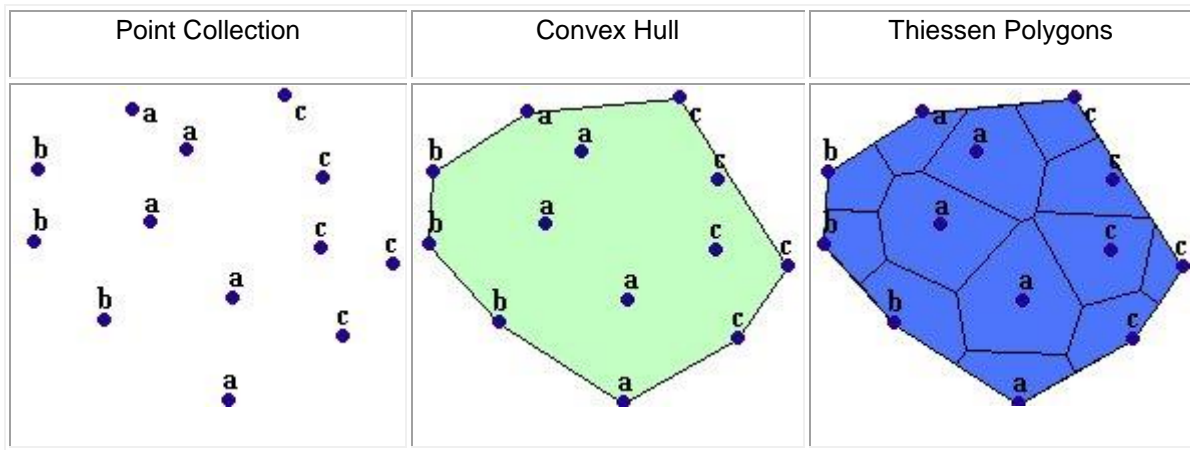


Figure 2. 5: Thiessen Polygons from point data

(Source: http://www.ian-ko.com/ET_GeoWizards/UserGuide/thiessenPolygons.htm)

In this method, the region sampled, R , is divided by perpendicular bisectors between the N sampling points into polygons or tiles, $V_i, i = 1, 2, \dots, N$, such that in each polygon all points are nearer to its enclosed sampling point X_i than to any other sampling point. The prediction at each point in V_i is the measured value at X_i , that is $Z^*(X_0) = Z(X_i)$. The weights are as in Equation 2.8 below:

$$\lambda_i = \begin{cases} \mathbf{1} & \text{if } X_i \in V_i, \\ \mathbf{0} & \text{otherwise} \end{cases} \dots\dots\dots(2.8)$$

This method has shortcomings because each prediction is based on just one measurement, there is no estimate of the error, and information from neighbouring points is ignored. When used for mapping, the result is crude in that the interpolated surface consists of a series of steps.

2.8.2 Triangulation

In this method, the sampling points are linked to their neighbors by straight lines to create triangles that do not contain any of the points. The measured values are envisaged as standing above the basal plane at a height proportional to those values so that the whole set of data forms a polyhedron consisting of more or less tilted triangular plates. The aim is to determine the height of the plate at X_0 from the apices of the triangle by linear interpolation. This can be represented as a weighted average and the weights are determined. In order to determine the weights, the coordinates of the three apices are denoted by $\{x_{11}, x_{12}\}$, $\{x_{21}, x_{22}\}$ and $\{x_{31}, x_{32}\}$; and those of the target point denoted by $\{x_{01}, x_{02}\}$. Then the weights are given by Equation 2.9 (Webster and Oliver, 2007):

$$\lambda_1 = \frac{(x_{01} - x_{31})(x_{22} - x_{32}) - (x_{02} - x_{32})(x_{21} - x_{31})}{(x_{11} - x_{31})(x_{22} - x_{32}) - (x_{12} - x_{32})(x_{21} - x_{31})} \dots\dots\dots(2.9)$$

The technique is simple and local, but the disadvantages are that, although it is somewhat better than the Thiessen method, each prediction makes no use of data further away, and there is again no measure of error. Unlike the Thiessen method, the resulting surface is continuous, but it has abrupt changes in gradient at the margins of the triangles. If the principal aim is to predict rather than to make a map with smooth isolines then the discontinuities in the derivative are immaterial. Another difficulty is that there is no obvious triangulation that is better than any other; even for a rectangular grid there are two options.

2.8.3 Natural Neighbor Interpolation

As reported by Webster and Oliver, (2007), Sibson (1981) combined the best features of the two methods above in what he called ‘natural neighbour interpolation’. The first step is to triangulate the data by Delauney’s method. The apices of the triangles are those sampling points in adjacent Dirichlet tiles. This triangulation is unique except where the data are on a regular rectangular grid. To determine the value at any other point, x_0 , that point is inserted into the tessellation, and its neighbours, the set T (the points within its bounding Dirichlet tiles), are used for the interpolation. Sibson (1981) referred to these points as ‘natural neighbours’. For each neighbour the area, A , of the portion of its original Dirichlet tile that became incorporated in the tile of the new point is calculated. These areas, when scaled to sum to 1, become the weights. This can be represented as in Equation 2.10:

$$\lambda_i = \frac{A_i}{\sum_{k=1}^N A_k} \text{ for all } i = 1, 2, \dots, N. \dots \dots \dots (2.10)$$

Where; λ_i = assigned weights,

A_i = overlap area corresponding to a known data point x_i

This means that if a point x_i is a natural neighbour, that is $x_i \in T$, then A_i has a value and the point carries a positive weight. If x_i is not a natural neighbor then it has no area in common with the target and its weight, λ_i , is zero. This interpolator is continuous and smooth except at the data points where its derivative is discontinuous. This was referred to as the *natural neighbour C^0 interpolant* (Sibson, 1981).

2.8.4 Inverse Functions of Distance

Methods based on inverse functions of distance, have the weights defined by the formula in equation 2.11 below:

$$\lambda_i = 1/|x_i - x_0|^\beta \text{ with } \beta > 0; \dots\dots\dots(2.11)$$

where: x_0 = a target point for which we want a prediction value;

x_i = measured data;

λ_i = assigned weights.

They are scaled so that they sum to 1, and the result is that data points near to the target point carry larger weight than those further away. The most popular choice of β is 2 so that the data are inversely weighted as the square of distance.

Similar to the triangulation, if x_0 coincides with any x_i then λ_i becomes infinite, the other weights are immaterial, and $z(x_0)$ takes the value $z(x_i)$. Therefore, interpolation is exact. A desirable feature of weighting by inverse squared distance is that the relative weights diminish rapidly as the distance increases, and so the interpolation is sensibly local. Further, because the weights never become zero there are no discontinuities. Its disadvantages are that the choice of the weighting function is arbitrary, and there is no measure of error. Moreover, it takes no account of the configuration of the sampling. So where data are clustered, two or more points may be at approximately the same distance and direction from x_0 , and each point will carry the same weight as an isolated point a similar distance away but in a different direction. This is clearly undesirable, and some implementations for mapping have elaborated the scheme to overcome this problem, such that the interpolated surface will have a gradient of zero at the data points, and maxima and minima can occur only there.

2.8.5 Trend Surfaces

Trend surfaces are a form of multiple regression in which the predictors are the spatial coordinates. For example in Equation 2.12:

$$z(x_1, x_2) = f(x_1, x_2) + \varepsilon, \dots\dots\dots(2.12)$$

where, $z(x_1, x_2)$ is the predicted value at $\{x_1, x_2\}$ and f denotes a function of the spatial coordinates there. The model contains an error term, ε , and in regression this is assumed to

be independently and identically distributed with mean 0 and variance σ_ε^2 . Plausible functions, usually simple polynomials such as planes, quadratics or cubics, are fitted by least squares to the spatial coordinates, and the resulting regression equation is used for the prediction. Thus for a plane the regression equation would be as shown in equation 2.13 below:

$$z = b_0 + b_1 x_1 + b_2 x_2, \dots\dots\dots(2.13)$$

On the other hand, for a quadratic surface the regression equation would be as shown in equation 2.14;

$$z = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2 + b_5 x_1 x_2, \dots\dots\dots(2.14)$$

where; x 's are the spatial coordinates and b 's are the coefficients.

The predictor can be expressed as a weighted average of the data used to obtain the trend. The spatial coordinates and their powers are represented by a matrix X with N rows for the N sampling points and as many columns as coefficients b to be estimated. For a first-order surface the spatial coordinates are written as a matrix as in Equation 2.15 below;

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{12} & x_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{N2} & x_{N2} \end{bmatrix} \dots\dots\dots(2.15)$$

In the matrix, the first column is a dummy variant of ones, and the recorded values of z at those places are vectors as in Equation 2.16;

$$z = \begin{bmatrix} z(x_1) \\ z(x_2) \\ \cdot \\ \cdot \\ z(x_N) \end{bmatrix} \dots\dots\dots(2.16)$$

The coefficients *b* are obtained from the matrix multiplication (Equation 3.12) and the predictions are then given by z_0^* (Equation 2.18), in which x_0 is the row vector $[1 \ x_{01} \ x_{02}]$. The weights are yielded by Equation (2.17) below. For a more complex surface the matrix *X* is simply extended by adding columns for the additional powers of x_1 and x_2 .

$$b = (x^T x)^{-1} x^T z, \dots\dots\dots(2.17)$$

$$z_0^* = x_0 b, \dots\dots\dots(2.18)$$

In most instances of trend surfaces, spatial variation is so complex that a polynomial of very high order is needed to describe it, and the resulting matrix equations are usually unstable. The residuals from the trend are auto-correlated, and so one of the assumptions of regression is violated. Thus, in a region containing both mountains and plain the prediction of topographic height on the plain will be determined by the much larger fluctuations in the mountains. Trend surfaces are not sufficiently local, and they do not return the values at data

points. Nevertheless, simple regression surfaces can represent long-range trend in some instances. The technique has its merits therefore in revealing long-range structures and filtering them to leave variation of shorter range that can be analyzed by other techniques.

2.8.6 Splines

A spline function also consists of polynomials, but each polynomial of degree p is local rather than global (Webster and Oliver, 2007). The polynomials describe pieces of a line or surface, and they are fitted together so that they join smoothly, and their $p-1$ derivatives are continuous. The places at which the pieces join are known as ‘knots’, and the choice of knots confers an arbitrariness on the technique. Splines can be constrained to pass through the data. Alternatively, by choosing knots away from the data points they can be fitted by least squares or some other method to produce smoothing splines. Typically the splines are of degree 3, referred to as cubic splines.

2.9 The Variogram

Most environmental variables are continuous; the stochastic processes that are believed to represent them are continuous, and so also are the autocovariance functions and variograms of a continuous lag. The variogram is a function of an underlying stochastic process and may be regarded as the theoretical variogram. It may be thought of as the average of the variograms from all possible realizations of the process. The regional variogram is the variogram of the particular realization in a finite region, R . It is the one that might be computed if complete information of the region could be available. The regional variogram does not necessarily represent the whole ensemble. A process that is second-order stationary might appear unbounded in a small region, especially if the distance across the region is

smaller than the correlation range. It is more or less accessible, depending on the effort we are prepared to devote to sampling the realization.

The experimental variogram is computed from data, $z(x_i)$, $i = 1, 2, \dots$, which constitute a sample from the region, R . It is also called the sample variogram. It necessarily applies to an actual realization, and it estimates the regional variogram for that realization. It is usually the only variogram that we may know. The variogram seeks to summarize the spatial relations in data. The true variogram representing the regional variation should be conti

CHAPTER 3: METHODOLOGY

The methodology employed in this research was Ecological Niche Modeling (ENM), a technique that encompasses a suite of tools that relate known occurrences of species or phenomena to raster geographic information system layers that summarize variation in several environmental dimensions. The spatial-temporal distributions of the main malaria vectors in Kenya were quantified using BIOCLIM and DOMAIN models to determine the relationship between vector distribution and climate change.

The biological data used was from published sources (Okara et al., 2010 and MARA/ARMA, 1998), comprising of point samples for geo-referenced malaria vector occurrences. The climate data used was maximum temperature, minimum temperature and precipitation for current climate (1950-2000) and climate projection for HADCM3, CCCMA and SCIRO models of IPCC projected future climate under the A2a scenario by the years 2020, 2050 and 2080. The climate data was acquired in grid format from WorldClim global climate data which was further processed to generate 19 bioclimatic variables for Kenya. Predictive modeling was done to correlate climate change as an explanatory variable to the changing dynamics of malaria vector distribution in Kenya.

3.1 Data

Correlative species distribution models require two types of data input:

- Biological data – describing the known species distribution.
- Environmental data – describing the landscape in which the species is found.

3.1.1 Malaria Vector Data

Malaria vector data was acquired from published sources. The details of the data are summarized in Table 3.1 below.

Table 3. 1: Description of acquired malaria vector data

Data Sources	<ul style="list-style-type: none"> • KEMRI (<i>Okara, et al.,2010</i>) • MARA ARMA project (1998).
Data types	<ul style="list-style-type: none"> • Database of Presence only data – records of localities where the malaria vectors have been observed with GPS reference. • Metadata – information about the databases.
Data Format	<ul style="list-style-type: none"> • Point vector data - representing points where malaria vectors have been observed. • Polygon vector data – polygons defining areas where malaria vectors are found. • Raster data – converted grid of cells representing ecology types.

The malaria vector data acquired, as amassed by Okara, et al., 2010 was primary empirical data from published and unpublished sources. This data was collated for the period from 1990 to 2009. Details recorded for each source in the database included the first author, year of publication, report type, survey location name, month and year of survey, the main Anopheles species reported as present and the sampling and identification methods used. Survey locations were geo-positioned using national digital place name archives and on-line geo-referencing resources. The geo-located species-presence data were displayed and described administratively, using first-level administrative units (province), and biologically, based on the predicted spatial margins of Plasmodium falciparum transmission intensity in Kenya for the year 2009 (Figure 3.1). Each geo-located survey site was assigned an urban or rural classification and attributed an altitude value.

A total of 498 spatially unique descriptions of *Anopheles* vector species across Kenya were identified. More than half (54%) of the sites surveyed were investigated since 2005. A total of 174 sites reported the presence of *Anopheles gambiae* complex without identification of sibling species. *Anopheles arabiensis* and *Anopheles funestus* were the most widely reported at 244 and 265 spatially unique sites respectively with the former showing the most ubiquitous distribution nationally. *Anopheles gambiae*, *Anopheles arabiensis*, *Anopheles funestus* and *Anopheles pharoensis* were reported at sites located in all the transmission intensity classes with more reports of *Anopheles gambiae* in the highest transmission intensity areas than the very low transmission areas (see Appendix A, Table A.1).

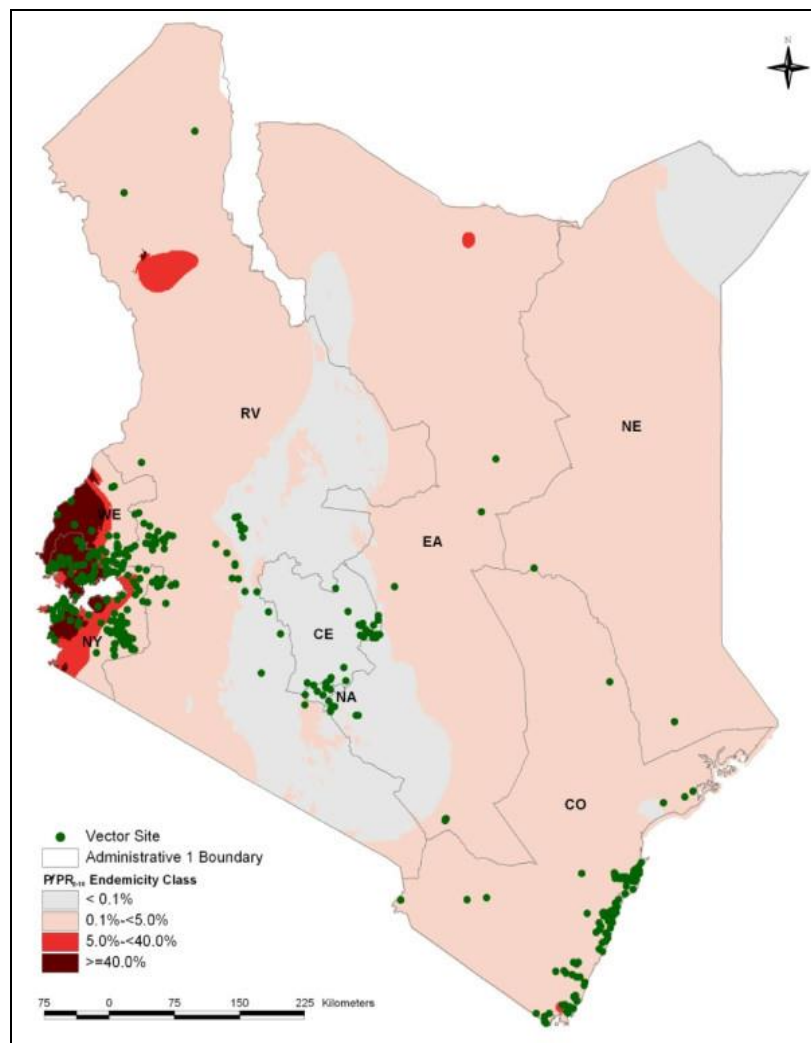


Figure 3. 1: Map of Kenya showing the spatial distribution of 498 unique survey sites displayed over first –level administrative boundaries (provinces) and the predicted endemicity classes.(Source: Okara, et al., 2010. Malaria journal, DOI: 10.1186/1475-2875-9-69).

The survey data for MARA/ARMA project (1998) had been geo-referenced so as to determine the latitude and longitude to be applied in geospatial analysis. The geospatial data for the survey sites was published in some reports, while in others, only the name of the area or settlement where the survey was carried out was provided. In such cases, as reported in the data collection process, the geographical location had to be obtained by either reading it off a topographical map, or by using digital maps and databases, such as the African Data Sampler (World Resources Institute, 1995) the Geoname Gazetteer (GDE Systems Inc., 1995) or Encarta (Microsoft, 1998). The database contained various data classes from different countries in Africa. For the purpose of this research, entomology data for Kenya was extracted from the published database (see Appendix 1, Table A.2).

3.1.2 Climate Data

The Climate data used was acquired from WorldClim climate surfaces for the globe. It covered the duration from 1950 – 2000 giving scenarios for future climate change for the globe and reconstructed palaeoclimates. The ESRI grid data format was extracted for geostatistical modeling. Further processing of the data was done to generate the 19 bioclimatic variables that were used for ecological niche modeling. These bioclimatic variables represent:

- annual trends (e.g., mean annual temperature, annual precipitation);
- seasonality (e.g., annual range in temperature and precipitation); and
- extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters).

The acquired data was adopted as developed by Hijmans, et al., 2005 and comprised the interpolated climate surfaces for global land areas, at a spatial resolution of 30 arc seconds (often referred to as 1-km spatial resolution). The climate elements that were considered in

the interpolation included monthly precipitation and mean, minimum, and maximum temperature. The input data used were gathered from a variety of sources and, where possible, were restricted to records from the 1950–2000 period. As reported by Hijmans, et al., 2005, the thin-plate smoothing spline algorithm implemented in the ANUSPLIN package for interpolation, using latitude, longitude, and elevation as independent variables was used to generate the high resolution data.

Any possible uncertainty arising from the input data and the interpolation by mapping weather station density, elevation bias in the weather stations, and elevation variation within grid cells and through data partitioning and cross validation was quantified. Uncertainty in climate surfaces usually stems from the quality of the input data and the interpolation method used. The geographic variation in uncertainty is illustrated by carrying out a number of complimentary analyses. Elevation bias tended to be negative (stations lower than expected) at high latitudes but positive in the tropics. Uncertainty was reported as highest in mountainous and in poorly sampled areas. Data partitioning showed high uncertainty of the surfaces on isolated islands, e.g. in the Pacific (Hijmans, et al., 2005).

The ANUSPLIN-SPLINA program, which uses every station as a data point, was used to generate interpolated climate surfaces. Second-order spline was fitted using latitude, longitude, and elevation as independent variables; due to the fact that this produced the lowest overall cross-validation errors as compared with some other settings like for instance, the third-order spline. SPLINA fits a continuous surface to the points, but, as in regression, the surface does not necessarily go through every observed point. The climate models created by ANUSPLIN-SPLINA are continuous surfaces that can be interrogated for any specific location and elevation within the specified domain of interpolation. The resolution of the data

thus only depends on the grid that is used to sample the continuous prediction (with the ANUSPLIN-LAPGRD program), and the higher the resolution of the grid, the better it represents the modeled climate data.

The fabricated climate surfaces data had the following advantages as compared to any previous global climatologies: the data are at a higher spatial resolution (400 times greater or more); more weather station records were used; improved elevation data were used; and more information about spatial patterns of uncertainty in the data is available. However, due to the overall low density of available climate stations, the generated surfaces do not capture of all variation that may occur at a resolution of 1 km, particularly of precipitation in mountainous areas.

The IPCC Third Assessment Report data for future climate projections calibrated and statistically downscaled using the WorldClim data for 'current' conditions was acquired for use in modeling the distribution of malaria vectors in Kenya. The current climate (1950-2000) data was obtained in ESRI grids format. These grids can be used in any ESRI software platform, hence exported to any other software for modeling. The datasets are summarized in Table 3.2 below.

Table 3. 2: WorldClim climate data (current and future).

Down loaded Climate Grids		
Current (1950-2000)		Precipitation: prec_30s_esri.zip Maximum temperature: tmax_30s_esri.zip Minimum temperature: tmin_30s_esri.zip
Climate Model	By (Year)	IPCC Projected Future Climate Grids
HADCM3	2020	Precipitation: wc_30s_HADCM3_A2a_2020_prec.zip Maximum temperature: wc_HADCM3_A2a_2020_tmax.zip Minimum temperature: wc_HADCM3_A2a_2020_tmin.zip
	2050	Precipitation: wc_30s_HADCM3_A2a_2050_prec.zip Maximum temperature: wc_HADCM3_A2a_2050_tmax.zip Minimum temperature: wc_HADCM3_A2a_2050_tmin.zip
	2080	Precipitation: wc_30s_HADCM3_A2a_2080_prec.zip Maximum temperature: wc_HADCM3_A2a_2080_tmax.zip Minimum temperature: wc_HADCM3_A2a_2080_tmin.zip
CCCMA	2020	Precipitation: wc_30s_CCCMA_A2a_2020_prec.zip Maximum temperature: wc_CCCMA_A2a_2020_tmax.zip Minimum temperature: wc_CCCMA_A2a_2020_tmin.zip
	2050	Precipitation: wc_30s_CCCMA_A2a_2050_prec.zip Maximum temperature: wc_CCCMA_A2a_2050_tmax.zip Minimum temperature: wc_CCCMA_A2a_2050_tmin.zip
	2080	Precipitation: wc_30s_CCCMA_A2a_2080_prec.zip Maximum temperature: wc_CCCMA_A2a_2080_tmax.zip Minimum temperature: wc_CCCMA_A2a_2080_tmin.zip
CSIRO	2020	Precipitation: wc_30s_CSIRO_A2a_2020_prec.zip Maximum temperature: wc_CSIRO_A2a_2020_tmax.zip Minimum temperature: wc_CSIRO_A2a_2020_tmin.zip
	2050	Precipitation: wc_30s_CSIRO_A2a_2050_prec.zip Maximum temperature: wc_CSIRO_A2a_2050_tmax.zip Minimum temperature: wc_CSIRO_A2a_2050_tmin.zip
	2080	Precipitation: wc_30s_CSIRO_A2a_2080_prec.zip Maximum temperature: wc_CSIRO_A2a_2080_tmax.zip Minimum temperature: wc_CSIRO_A2a_2080_tmin.zip

After acquisition of the global climate surfaces from WorldClim database, further processing was done to extract by mask the national grid for Kenya. Bioclimatic variables were generated in DIVA-GIS eventually to be used in ecological niche modeling. The default coding for the bioclimatic variables was as shown in Table 3.3.

Table 3. 3: Coding of bioclimatic variables

Abbreviation	Description of bioclimatic variable
BIO1	Annual Mean Temperature.
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp)).
BIO3	Isothermality (BIO2/BIO7) (* 100).
BIO4	Temperature Seasonality (standard deviation *100).
BIO5	Max Temperature of Warmest Month.
BIO6	Min Temperature of Coldest Month.
BIO7	Temperature Annual Range (BIO5-BIO6).
BIO8	Mean Temperature of Wettest Quarter.
BIO9	Mean Temperature of Driest Quarter.
BIO10	Mean Temperature of Warmest Quarter.
BIO11	Mean Temperature of Coldest Quarter.
BIO12	Annual Precipitation.
BIO13	Precipitation of Wettest Month.
BIO14	Precipitation of Driest Month.
BIO15	Precipitation Seasonality (Coefficient of Variation).
BIO16	Precipitation of Wettest Quarter.
BIO17	Precipitation of Driest Quarter.
BIO18	Precipitation of Warmest Quarter.
BIO19	Precipitation of Coldest Quarter.

3.2 Methodology

In order to model the spatial-temporal distribution of malaria vectors in Kenya, the model-building process entailed model selection, model fitting, and model validation. These three basic steps are applied iteratively until an appropriate model for the data to be analyzed has been developed. In the model selection step, plots of the data, process knowledge and assumptions about the process are used to determine the form of the model that can be fit to the data. Figure 3.2 below shows the model-building process.

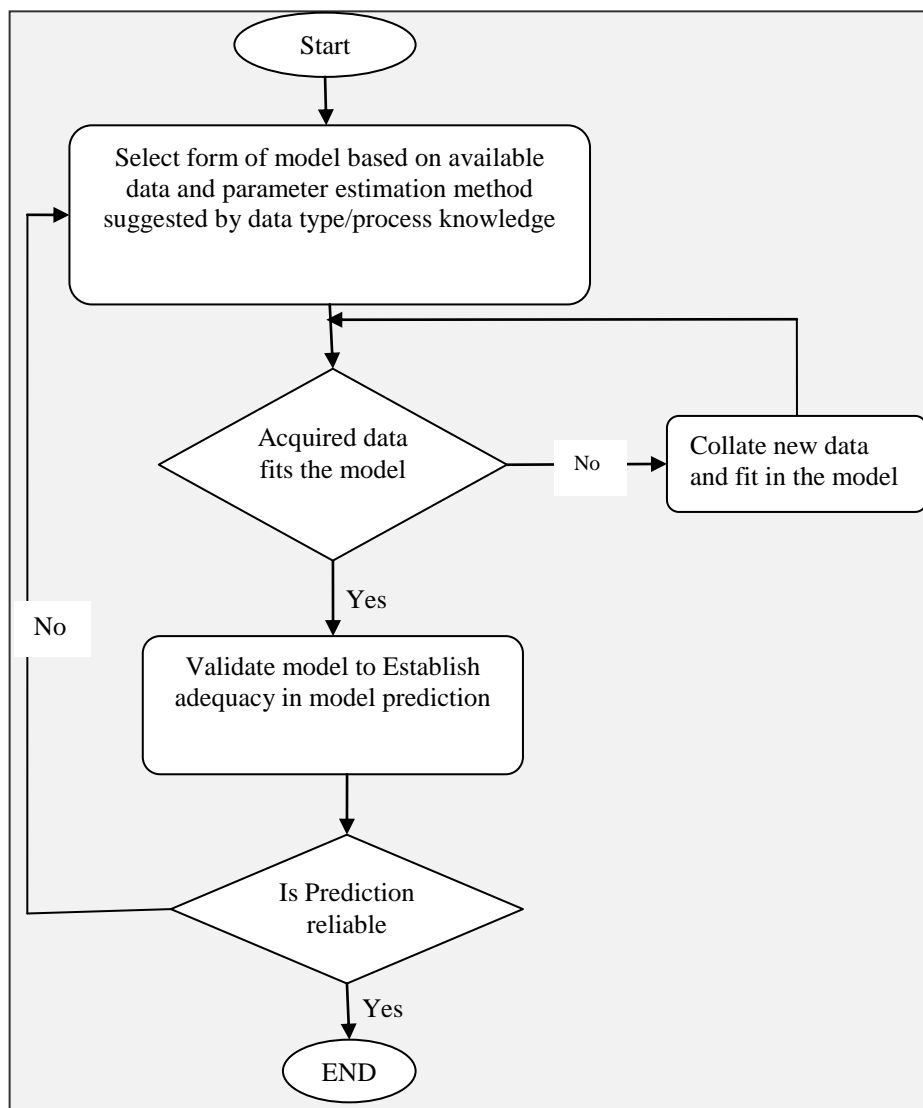


Figure 3. 2: Flow diagram for the steps required in building a species distribution model.

Further to model selection, using the selected model and possibly information about the data, an appropriate model-fitting method is used to estimate the unknown parameters in the model. When the parameter estimates have been made, the model is then carefully assessed. The underlying assumptions of the analysis are quantified for plausibility. Validity of the assumptions renders the model useful so as to answer the scientific or engineering questions that prompted the modeling effort. In case the model validation identifies problems with the current model, then the modeling process is repeated using information from the model validation step to select and/or fit an improved model.

3.2.1 Selecting Prediction Models for Ecological Niche Modeling

There exist varieties of environmental modeling approaches that are capable of being applied to generate species distribution in current and future ecologies under climate change scenarios. The selection of appropriate modeling methodology largely depends on the type of data available. The different published models used in environmental modeling include the following (Pearson, 2007):

- ♦ BIOCLIM / DOMAIN Models are useful in modeling presence only malaria vector distribution data and prediction is made without any reference to other samples in the study area. The Software used to implement it is Diva GIS which is freeware.
- ♦ MAXENT / ENFA models use “background” environmental data for the entire study area; focus on how the environment where the malaria vectors are known to occur relates to the environment across the study area (‘the background’); and importantly, the occurrence localities are included as part of the background.
- ♦ BIOMOD Multiple Methods sample ‘pseudo-absences’ from the study area. In principle, any presence-absence algorithm can be implemented using pseudo-absences; the aim here being to assess differences between the occurrence localities

and a set of localities chosen from the study area that are used in place of real absence data. The occurrence localities within the set of pseudo-absences are not included.

- ♦ Models for presence and absence or pseudo-absence are implemented in R- GIS Software which is freeware. These models include:
 - GLM – Generalized Linear Model;
 - GAM – Generalized Additive Model;
 - BRT – Boosted Regression Trees;
 - MARS – Multivariate Adaptive Regression Splines.

For this research, BIOCLIM and DOMAIN models were adopted based on the type of data available to predict the ecological niches of Malaria vector distribution in Kenya. This criterion for model selection was arrived at since presence only biological data was collated. Models that predict distributions of species by combining known occurrence records with digital layers of environmental variables have much potential for application in ecological niche modeling. The most applied strategy for estimating the actual or potential geographic distribution of a species is by characterizing the environmental conditions that are suitable for the species to thrive in certain ecology, and then identify where suitable environments are distributed in space. The environmental conditions that are suitable for a species can be characterized either by use of a *mechanistic* or a *correlative* approach.

Mechanistic models aim to incorporate physiologically limiting mechanisms in a species' tolerance to environmental conditions. For instance, a species response to environmental variables like daily precipitation, mean daily temperatures and night length can be used to model the distribution using mechanistic models of factors such as phenology, frost injury and reproductive success. Such mechanistic models would require thorough understanding of

the physiological response of species to environmental factors. Therefore, in terms of applicability, such models are difficult to develop for all but the most well understood species.

Conversely, correlative models usually aim to estimate the environmental conditions that are suitable for a species by associating known species' occurrence records with suites of environmental variables that can reasonably be expected to affect the species' physiology and probability of persistence. The central assertion in such approach is that the observed distribution of a species would provide useful information regarding the environmental requirements of that particular species. Since spatially unequivocal occurrence records for a large number of species are available, the vast majority of species' distribution models are correlative. Therefore, this fusion has embarked on the correlative approach to malaria vector distribution modeling.

The extracted climate data can be used to predict new suitable ecologies where a species is likely to occur, or would survive if it was brought there. BIOCLIM prediction results into different types of areas that are mapped according to percentile range of suitability. The areas are classified in terms of their suitability ranging from not suitable to excellent areas. BIOCLIM true/false makes use of binary numbers. All areas that are within the envelope described by the data points, cut off beyond a certain user defined percentile, are mapped as "true" (1) and all other areas are mapped as "false" (0).

The Domain procedure calculates the Gower distance statistic between each cell on the map and each point, using the values of the 19 climate variables. The distance between point *A* and grid cell *B* for a single climate variable is calculated as the absolute difference in the

values of that variable divided by the range of the variable across all points. The Gower distance is then the mean over all climate variables (Equation 3.1).

$$d_{AB} = \frac{1}{p} \sum_{k=1}^p \frac{|A_k - B_k|}{range(k)} \dots\dots\dots(3.1)$$

The Domain similarity statistic is calculated as in Equation 3.2 below:

$$D = 1 - d_{AB} \dots\dots\dots(3.2)$$

The maximum similarity between a chosen grid cell and all sample points is mapped. In DIVA GIS software that is used to implement the prediction, this value is then multiplied by 100. A good match is thus a high number (e.g., above 95). Domain true/false selects all areas that are within the specified threshold and are mapped as “true” (1) while all other areas are mapped as “false” (0).

3.2.2 Model Validation

The prediction models used in this study were validated to assess the accuracy of models’ predictions. The data was split into test data (evaluation data) and calibration data (training data). Training data was used to build the model. In order to test predictive performance, test data was used to run the model predictions. BIOCLIM and DOMAIN models were used in ecological studies as tools for extrapolating field measurements and integrating complex ecological information over space and time.

BIOCLIM and DOMAIN predictive models have been used in this research as a means for estimating spatial-temporal patterns of malaria vector distribution in Kenya with climate data from HADCM3, CCCMA and CSIRO under A2a scenario. The potential impact of climate change on malaria suitable ecologies has been predicted and an assessment of the predictive accuracy is fundamental. Section 3.1.2.1 provides further explanations on the evaluation of BIOCLIM and DOMAIN model performance.

Ideally, qualitative and quantitative assessment of model performance assists in determining the suitability of the model for specific applications and may help to identify those aspects of the model that may require improvement. In the event that different modeling techniques have been explored, an assessment of model performance provides a basis for comparing alternative methodologies in order to achieve the most probable accuracy. In species distribution modeling, predictor variables are typically organized as raster (grid) type files. Each predictor should be a 'raster' representing a variable of interest. Bioclimatic Variables derived from mean, maximum and minimum precipitation and temperature over decades were generated and stored in GIS ESRI grid format.

In univariate statistical modeling, dependent variables typically consist of field measurements performed at known locations and the independent variables are derived from the raster files containing continuous data. Univariate regression models are used to relate the dependent and independent variables. However, the means for compiling data, synthesizing these data and developing predictive models to relate ecological functions to quantifiable landscape characteristics entails complex interrelationships. Since the factors that determine the values of environmental variables are numerous, largely unknown in detail, and interact with a complexity that we cannot disentangle, we can regard their outcomes as random.

If a stochastic view is adopted, then at each point in space there is not just one value for a property but a whole set of values. The observed value there can be regarded as one drawn at random according to some law, from some probability distribution. This means that at each point in space there is variation, a concept that has no place in classical estimation. Thus, at a point x a property, $Z(x)$, is treated as a random variable with a mean, μ , a variance, σ^2 , and higher-order moments, and a cumulative distribution function (cdf). It has a full probability

distribution, and it is from this that the actual value can be drawn. By knowing approximately what that distribution might be, values at unrecorded places can be estimated from data in the neighborhood and errors can be put on the estimates.

Multivariate statistical techniques that can evaluate many variables simultaneously were used to predict the malaria vector distribution. Multiple regression analysis relates the dependent variable statistically to several independent variables, resulting in an equation of the form shown in equation 3.3 below:

$$z = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots \dots \dots (3.3)$$

where a_0, a_1, a_2, \dots are the coefficients of the regression and x_1, x_2, x_3, \dots are the independent variables. Discriminant Function Analysis (DFA) has been used with GIS-derived vector distribution data to develop multivariate ecological model predictions. Most environmental variables, such as the temperature and precipitation, are continuous. For these a value $z(x)$ can be thought of as one of an infinite number of possible values, with a cdf that is the probability that Z takes any value less than or equal to a particular value z_c (Equation 3.4);

$$F\{Z(x; z)\} = Prob[Z(x) \leq z_c] \text{ for all } z \dots \dots \dots (3.4)$$

The probability $F\{Z(x; z)\}$ takes values between 0 and 1. Its derivative is the probability density function, the pdf (Equation 3.5);

$$f\{Z(x)\} = \frac{dF\{Z(x; z)\}}{dz} \dots \dots \dots (3.5)$$

The distribution may be bounded, as in the case of a proportion or percentage, but the most useful assumption is that it is not, so that $-\infty \leq Z(x) \leq +\infty$. The given description represents an individual point x and it applies to the infinitely many points in the space; at each point x_i ;

$i=1, 2, \dots, Z(x_i)$ has its own distribution and cdf. The range of possible values constitutes an ensemble, and one member of the ensemble is the realization. The set of random variables, $Z(x_1), Z(x_2), \dots$; constitute a random function, a random process, or a stochastic process. The set of actual values of Z that comprise the realization of the random function referred to as a regionalized variable. A region is regarded as made up of a population of units, so that a random function $Z(x)$ can be taken as a super population, with an infinite number of units in space and an infinite number of values of Z at each point in the space. Thus it is said to be doubly infinite.

3.2.2.1 The Presence/Absence Confusion Matrix

The confusion matrix was used to summarize the predictive performance of the adopted ecological prediction models. Binary model predictions (i.e. predictions of suitable and unsuitable rather than probabilities) are required in order to complete the confusion matrix. The confusion matrix records the frequencies of each of the four types of predictions from analysis of test data as follows:

- (a) True positive - model predicts presence of an event and test data confirms;
- (b) False positive - model predicts presence of an event but test data show absence;
- (c) False negative - model predicts absence of an event and test data show presence;
- (d) True negative - model predicts and test data show absence of an event.

The models used generated presence–absence predictions. These models are referred to as presence–absence models, and are usually evaluated by comparing the predictions with a set of validation sites and constructing a confusion matrix that records the number of true positive (a), false positive (b), false negative (c) and true negative (d) cases predicted by the model. Presence data are records of occurrence of species in certain time in space. Absence data are complex as a species can fail to be identified in ecology where it thrives, thus

recording a false absence. There errors encountered were either errors of omission and commission in identifying these records. The confusion matrix was applied as shown in Table 3.4 below:

Table 3. 4: Presence / Absence Confusion Matrix

	Validated Data Set	
Model Presence	a-True +ve	b-False +ve
Model Absence	c- False -ve	d-True -ve

An error matrix was used to evaluate the predictive accuracy of BIOCLIM and DOMAIN models; *a* being the number of cells for which presence was correctly predicted by the model; *b*, number of cells for which the species was not found but the model predicted presence; *c*, number of cells for which the species was found but the model predicted absence; and *d*, number of cells for which absence was correctly predicted by the model. The columns are reference data, or known classifications as the sites were evaluated on the ground while the rows are model predictions. Models generating non-dichotomous scores on an ordinal scale are often evaluated by applying a certain threshold to transform the scores into a dichotomous set of presence–absence predictions, and constructing a corresponding confusion matrix. One simple measure of accuracy that can be derived from the confusion matrix is the proportion of correctly predicted sites, which is the overall accuracy. Model performance is a measure of how well a model explains an independent dataset.

Measures of predictive accuracy were calculated from the 2×2 error matrix. Overall accuracy gives the rate of correctly classified cells. Sensitivity is the probability that the model will correctly classify a presence. Specificity is the probability that the model will correctly classify an absence. The kappa statistic normalizes the overall accuracy by the

accuracy that might have occurred by chance alone. The kappa statistic is a chance corrected measure of agreement between two sets of categorized data. Kappa result ranges from 0 to 1 and the higher the value of Kappa, the stronger the agreement. If Kappa = 1, then there is perfect agreement and if Kappa = 0, then there is no agreement.

Therefore, kappa was used as a measure of the agreement between model predictions and existing reality, thus determining if the values contained in an error matrix represented truly significant result better than random. Omission error is the proportion of actual presences that are not predicted properly while commission error is a measure of all absences placed where the species is not. Correct Classification Rate (CCR) is a measure of how many of our test points are correctly predicted versus not. In all formulae $n = a + b + c + d$. The parameters used to measure the predictive accuracy of correlative models used are summarized in Table 3.5 below:

Table 3. 5: Measures of model predictive accuracy

Performance measure	Formula
Sensitivity	$\frac{a}{a + c}$
Specificity	$\frac{d}{b + d}$
Kappa statistic	$\frac{\left(\frac{a + d}{n}\right) - \frac{(a + b)(a + c) + (c + d)(d + b)}{n^2}}{1 - \frac{(a + b)(a + c) + (c + d)(d + b)}{n^2}}$
Omission error	$\frac{c}{a + c}$
Commission error	$\frac{b}{b + d}$
Overall accuracy / Correct Classification Rate (CCR)	$\frac{a + d}{n}$

The measure of overall accuracy has been criticized for ascribing high accuracies for rare species, thus rendering measures of sensitivity and specificity most applicable. Theoretically, kappa responds in a unimodal fashion in prevalence, and that the level of prevalence found to maximize Kappa depends on the ratio between sensitivity and specificity. The dependency of Kappa on prevalence eventually introduces statistical artifacts to the estimates of model predictive accuracy. Sensitivity is the proportion of observed presences that are predicted as such, and therefore quantifies omission errors. Specificity on the other hand is the proportion of observed absences that are predicted as such, and therefore quantifies commission errors.

Sensitivity and specificity are independent of each other when compared across models. In addition, they are independent of prevalence, which can be elaborated as the proportion of sites in which the species was recorded as present. Thus, Kappa can be represented in the form shown in Equation 3.6 below:

$$Kappa = \frac{P_0 - P_e}{1 - P_e}, P_0 = p \cdot s_n + (1 - p) \cdot s_p, P_e = -2(s_n + s_p - 1)p(1 - p) + p_0 \dots \dots \dots (3.6)$$

where p, s_n and s_p are prevalence, sensitivity and specificity respectively, p_0 is the observed accuracy, while p_e is the accuracy expected to occur by chance. The extremism of Kappa at p satisfies both $(s_n - s_p)p^2 - 2(1 - s_p)p + (1 - s_p) = 0$ and $0 \leq p \leq 1$. When $s_n + s_p - 1 > 0$, the extremism is a maximum and when $s_n + s_p - 1 < 0$, the extremism is a minimum. The former case characterizes models with performance better than random. Therefore, the prevalence that maximizes the Kappa score of a given model is a function of the sensitivity and specificity of the models. The Kappa statistic ranges from -1 to +1, with values of +1 indicating perfect agreement and values of ≤ 0 indicates a performance no better than random (Cohen, 1960). The k value can be interpreted as shown in Table 3.6 below (Altman, 1991):

Table 3. 6: Evaluation of Kappa strength of agreement

Value of K	Strength of agreement
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very good

Alternatively, ROC curve can be used as a method for assessing the accuracy of prediction models (Fielding and Bell, 1997). ROC curves are constructed by using all possible thresholds to classify the predictions into confusion matrices, obtaining sensitivity and specificity for each matrix, and then plotting true positive rate (sensitivity) against the corresponding proportion of false positives rates (1- specificity). Therefore, an ROC curve gives the following information;

- It shows the tradeoff between sensitivity and specificity any increase in sensitivity will be accompanied by a decrease in specificity (Pearce and Ferrier, 2000).
- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
- The area under the ROC Curve (AUC) is a measure of test accuracy, often used as a single threshold-independent measure for model performance.

The accuracy of the test depends on how well the test separates the group being tested into those with and without the predicted results. Accuracy was measured by the area under the ROC curve. Two methods are commonly used to compute the AUC:

- a non-parametric method based on constructing trapezoids under the curve as an approximation of area; and

- a parametric method using a maximum likelihood estimator to fit a smooth curve to the data points.

For a perfect test, an AUC value of 1 is obtained; an AUC value of 0.5 represents a worthless test. A rough guide for classifying the accuracy of a model prediction test is the traditional academic point system as shown in Table 3.7 below:

Table 3. 7: Interpretation of AUC values

AUC Value	Interpretation
0.90 – 1	excellent (A)
0.80 - 0.90	good (B)
0.70 - 0.80	good (B)
0.60 - 0.70	poor (D)
0.50 - 0.60	fail (F)

3.2.2.2 Selecting Thresholds of Occurrence

In order to test model performance using statistics derived from the confusion matrix, it was important to convert continuous model output into binary predictions of ‘present’ or ‘absent’, by setting a threshold probability value above which the species is predicted to be present. Several different methods have been used in selecting thresholds of occurrence (Pearson, 2007). One approach is to use the lowest predicted value of environmental suitability or probability of presence across the set of sites where malaria vectors have been detected. This method assumes that malaria vector presence is restricted to locations equally or more suitable than those at which the species has been observed. The approach seeks to identify the minimum areas in which malaria vectors occur whilst ensuring that no localities at which the vectors have been observed has been omitted (i.e. omission rate=0 and sensitivity = 1). The approach has been found to be appropriate for presence-only data type (Pearson, 2007).

When the model output is continuous, it is often important to derive a test statistic that provides a single measure of predictive performance across the full range of possible thresholds. This can be achieved using a statistic known as AUC: the Area under the Receiver Operating Characteristic (ROC) Curve. The ROC curve is defined by plotting sensitivity against '1 – specificity' across the range of possible thresholds, then the AUC test is derived from the ROC Curve. Sensitivity and specificity are used as these two measures take into account all four elements of the confusion matrix (true and false presences and absences). Conventionally, it is common to subtract specificity from 1 (i.e. 1 – specificity) so that both sensitivity and specificity vary in the same direction when the decision threshold is adjusted (Pearce and Ferrier, 2000).

The ROC curve thus describes the relationship between the proportion of observed presences correctly predicted (sensitivity) and the proportion of observed absences incorrectly predicted (1 – specificity). Therefore, a model that predicts perfectly will generate an ROC curve that follows the left axis and top of the plot, whilst a model with predictions that are no better than random, meaning that it is unable to classify accurately sites at which the species is present and absent, will generate a ROC curve that follows the 1:1 line (see illustration in Figure 3.3 below).

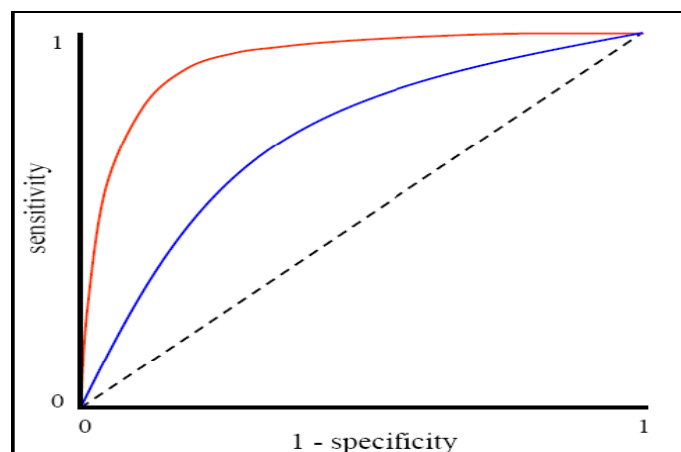


Figure 3. 3: ROC curves illustrating superior predictive ability (red) and random predictive ability (dashed black line, 1:1)

3.3 Choice of Future Climate Emission Scenario for Prediction

IPCC has investigated several storylines but the data used to create the ESRI grids used the SRES A2a and B2a story-line scenarios. The storylines represent different demographic, social, economic, technological, and environmental developments that diverge in increasingly irreversible ways. Both A2a and B2a storylines describe a ‘regionalization’ meaning a heterogeneous world development opposed to a ‘globalization’ of a homogeneous world development as described in the A1 and B1 storyline family.

The future climate scenario A2a describes a highly heterogeneous future world with regionally oriented economies. The main driving forces are a high rate of population growth, increased energy use, land-use changes and slow technological change. Conversely, the B2a storyline is also regionally oriented but with a general evolution towards environmental protection and social equity. In Comparison to A2a, B2a has a lower rate of population growth, a smaller increase in GDP but more diverse technological changes and slower land-use changes. A2a storylines was used to describe the projected climate change in Kenya.

The A2a storyline seemed more appropriate since demographic data has portrayed an upward trend for population growth rate in Kenya, hence adoption of the A2a storyline to model the future climate change impact on the malaria vector ecologies. Population growth rate can be defined as the average annual percent change in the population, resulting from a surplus (or deficit) of births over deaths and the balance of migrants entering and leaving a country. The growth rate may either be positive or negative and is a major factor in determining the magnitude of burden that would be imposed on a country by the changing needs of its people. Rapid population growth can be viewed as threatening by neighboring countries.

Kenya's population stands at 38.6 million according to the 2009 National Population and Housing Census results as reported by Kenya National Bureau of Statistics, (2010). This growth was a tremendous increase on the 1979 census, whose tally stood at 15.3million, and rising to 21.4 million ten years later. The 1999 Census results were not released to the public but it was estimated that the population stood at 28.7 million.

According to statistics, an average kenyan woman gives birth to 4 to 6 babies. Projections indicate that the country's population will stand at 51.3 million by 2025, if the current rate of population growth is maintained. These figures will put Kenya as one of the most populous countries in Africa. The figures indicate that the country's population is growing at the rate of 1 million people per year. The population growth rate in Kenya was reported to be at 2.44 percent in the year 2011, higher than the world's which is at 1.2 percent. The high population growth rate can be attributed mostly to poor family planning methods and ignorance.

According to the 2010 revision of the World Population Prospects the total population for Kenya was 40.5m in 2010, compared to only 6m in 1950. The proportion of children below the age of 15 in 2010 was 42.5%, 54.9% was between 15 and 65 years of age, while 2.7% was 65 years or older. The charts in Figure 3.4 and 3.5 have shown estimates and probabilistic projections of the total population for Kenya which were generated by United Nations, Department of Economic and Social Affairs, among other 195 countries and areas with a population of 100,000 or more in 2010. The projections are based on the probabilistic projections of total fertility and life expectancy at birth, based on estimates of the 2010 Revision of the World Population Prospects.

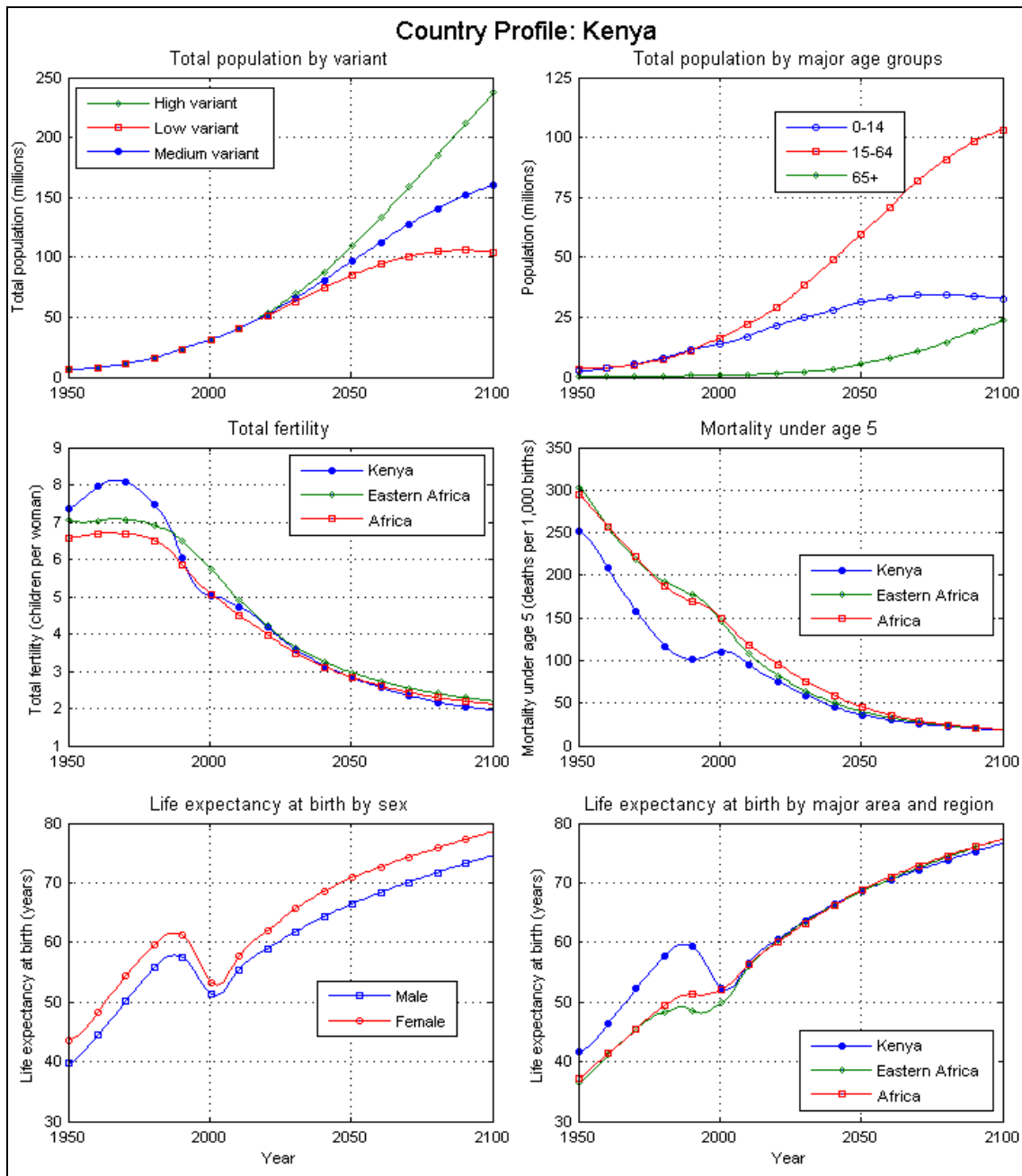


Figure 3. 4: Kenya population growth rate for selecting future climate scenario

(Source: United Nations, Department of Economic and Social Affairs, Population Division (2011): World Population Prospects: The 2010 Revision. New York (Updated 24 October 2012)).

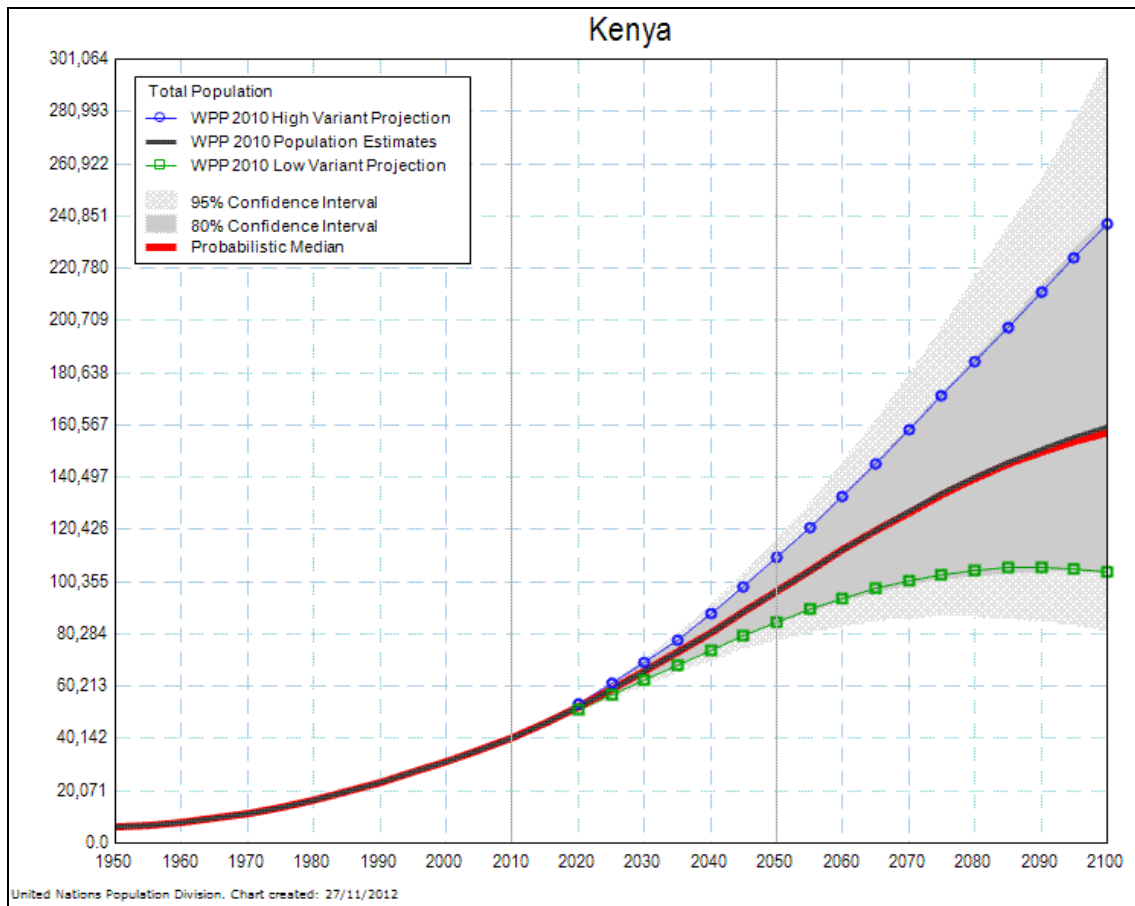


Figure 3. 5: Probabilistic Population Projection to guide scenario selection: Total Population (thousands), based on the 2010 Revision of the World Population Prospects.

(Source: United Nations, Department of Economic and Social Affairs, Population Division (2010): Population projections using probabilistic projections of total fertility and life expectancy at birth, based on a Bayesian Hierarchical Model (BHM). New York (internal data set)).

These probabilistic projections of total fertility and life expectancy at birth were carried out with a Bayesian Hierarchical Model. The figure displays the deterministic high and low variant of the 2010 Revision of the World Population Prospects, as well as the probabilistic median and the 95 as well as 80 percent confidence intervals of the probabilistic projections. It should be noted that most likely there is a (small) difference between the median of these probabilistic population projections and the medium variant of the official 2010 Revision of the World Population Prospects. This is due to the fact that the population projections

displayed here includes *probabilistic* projections of life expectancy at birth, while the 2010 Revision was based on a *deterministic* projection of life expectancy at birth.

3.4 Testing Autocorrelation in point sample data

Several techniques were applied to test the autocorrelation in malaria vector point samples. For instance, autocorrelation was used which is the correlation between elements of a series and others from the same series separated from them by a given interval. Spatial autocorrelation is found to exist when the values of nearby geographical objects are more similar than those of objects that are further away. Spatial autocorrelation is concerned with the degree to which objects or activities at some place on the earth's surface are similar to other objects or activities located nearby. Its existence is reflected in the proposition which Tobler (1970) referred to as the "first law of geography: everything is related to everything else, but near things are more related than distant things." Therefore, it becomes impossible to imagine a world in which spatial autocorrelation could be absent. There could be no regions of any kind, since the variation of all phenomena would have to occur independently of location, and places in the same neighborhood would be as different as places a continent apart.

Covariance and correlation are measures of the similarity between two different variables (X, Y) . The magnitude of the covariance, $Cov(X, Y)$, increases with increasing similarity in the patterns of variation of the two variables about their respective means. The correlation coefficient, $\rho(X, Y)$, ranges from 1 for perfect positive correlation to -1 for perfect negative correlation and is in the vicinity of 0 for uncorrelated variables. Equations 3.7 and 3.8 are used to compute covariance and correlation coefficient respectively.

$$Cov(X,Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \dots\dots\dots(3.7)$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \dots\dots\dots(3.8)$$

Spatial autocorrelation can be interpreted as a descriptive index, measuring aspects of the way things are distributed in space. However, it can also be seen as a causal process, measuring the degree of influence exerted by something over its neighbors. The spatial pattern of a distribution is defined by the arrangement of individual entities in space and the geographic relationships that exists among them. A prerequisite to understanding the complicated spatial processes underlying the distribution of a phenomenon is the capability of evaluating the spatial patterns. Spatial autocorrelation indicates the extent to which the occurrence of one feature is influenced by similar features in the adjacent area.

Therefore, statistics of spatial autocorrelation provide a useful indicator of spatial patterns. Different indicators of spatial autocorrelation are available, but those applied to test the vector point distribution are the global indicators of spatial association, join count statistics; Moran's I (Moran, 1948) and Geary's c (Geary, 1954). The resulting *correlogram* illustrates autocorrelation at each lag distance. Membership in a given distance class is defined by assigning a weight to each pair of points in the analysis; typically this weight is a simple indicator function, taking on a value of 1 if within the distance class, else 0.

Geary's index (Geary, 1954; 1968) is a measure of spatial autocorrelation for area objects and interval data. Attribute similarity c_{ij} is calculated from the squared difference in value, as shown in Equation 3.9. A Geary's c value of zero indicates positive spatial autocorrelation.

Negative spatial autocorrelation is indicated by a value greater than one. A Geary's c value of one suggests that no spatial autocorrelation is present.

$$C = [(n - 1) / 2 \sum \sum W_{ij}] \cdot [\sum \sum W_{ij} (x_i - x_j)^2 / \sum (x_i - \bar{x})^2] \dots\dots\dots(3.9)$$

Where: n = the number of observations;

W_{ij} = the distance (spatial lag) between pixels i and j ;

x_i = the DN value at location i ;

x_j = the DN value at location j ;

Moran's index (Moran, 1948) provides an alternative to Geary's for the same data context, and in most applications both are equally satisfactory. Perhaps the only obvious advantage of one over the other is that the Moran index is arranged so that its extremes match the earlier intuitive notions of positive and negative correlation, whereas the Geary index uses a more confusing scale. The Moran index is positive when nearby areas tend to be similar in attributes, negative when they tend to be more dissimilar than one might expect, and approximately zero when attribute values are arranged randomly and independently in space. The attribute similarity measure used by the Moran index makes it analogous to a covariance between the values of a pair of objects (see Equation 3.10).

$$I = \frac{\sum_i \sum_j W_{ij} C_{ij}}{s^2 \sum_i \sum_j W_{ij}} \dots\dots\dots(3.10)$$

Where: s^2 denotes the sample variance, $\sum_i (z_i - \bar{z})^2 / n$,

W_{ij} = the distance (spatial lag) between pixels i and j ; and

$$C_{ij} = (z_i - \bar{z})(z_j - \bar{z}) \dots\dots\dots(3.10a)$$

The *Autocorrelation* option in DIVA-GIS calculates this relationship by using two common statistics, the Geary and Moran indices. The results obtained are analyzed using the interpretation values depicted in Table 3.8. The spatial autocorrelation functionality implemented in DIVA-GIS is based on the “Rookcase” software by Sawada (1999).

Table 3. 8: Autocorrelation interpretation values (rather than o^* , the precise expectation is $-1/(n-1)$)

Geary index	Moran index	Interpretation
$0 < c < 1$	$i > 0^*$	Autocorrelation exists, data are clustered
$c=1$	$i = 0^*$	Independent, at random
$C > 1$	$i < 0^*$	Negative autocorrelation

Spatial autocorrelation is important as an index for it provides a type of information about a spatially distributed phenomenon that could not be available in any other form of statistical analysis, and which can be vital to correct interpretation. Spatial autocorrelation may be classified as either positive or negative depending on how the values appear. Positive spatial autocorrelation has all similar values appearing together, while negative spatial autocorrelation has dissimilar values appearing in close association. A positive spatial autocorrelation refers to a map pattern where geographic features of similar value tend to cluster on a map, whereas a negative spatial autocorrelation indicates a map pattern in which geographic units of similar values scatter throughout the map. When no statistically significant spatial autocorrelation exists, the pattern of spatial distribution is considered random.

The value of a variable at a given location tends to be similar to the values of that variable in nearby locations in the positive case. In deduction, if the value of some variable is low in a given location, the presence of spatial autocorrelation indicates that nearby values are also low. Conversely, negative spatial autocorrelation usually is characterized by dissimilar variant values in nearby locations.

The biological data used is for malaria vector distribution in Kenya. It has the spatial component in latitude and longitude obtained using Global Positioning System (GPS), as well as the attribute data for each point. The point data represents locations where any species of malaria vector was observed to be present. The point data sets were from MARA ARMA project, 1998 (vectors1 in red) and that from KEMRI published by Okara et al., 2010 (vectors2 in blue). Figure 3.6 below shows the spatial distribution of malaria vectors from the two data sets.

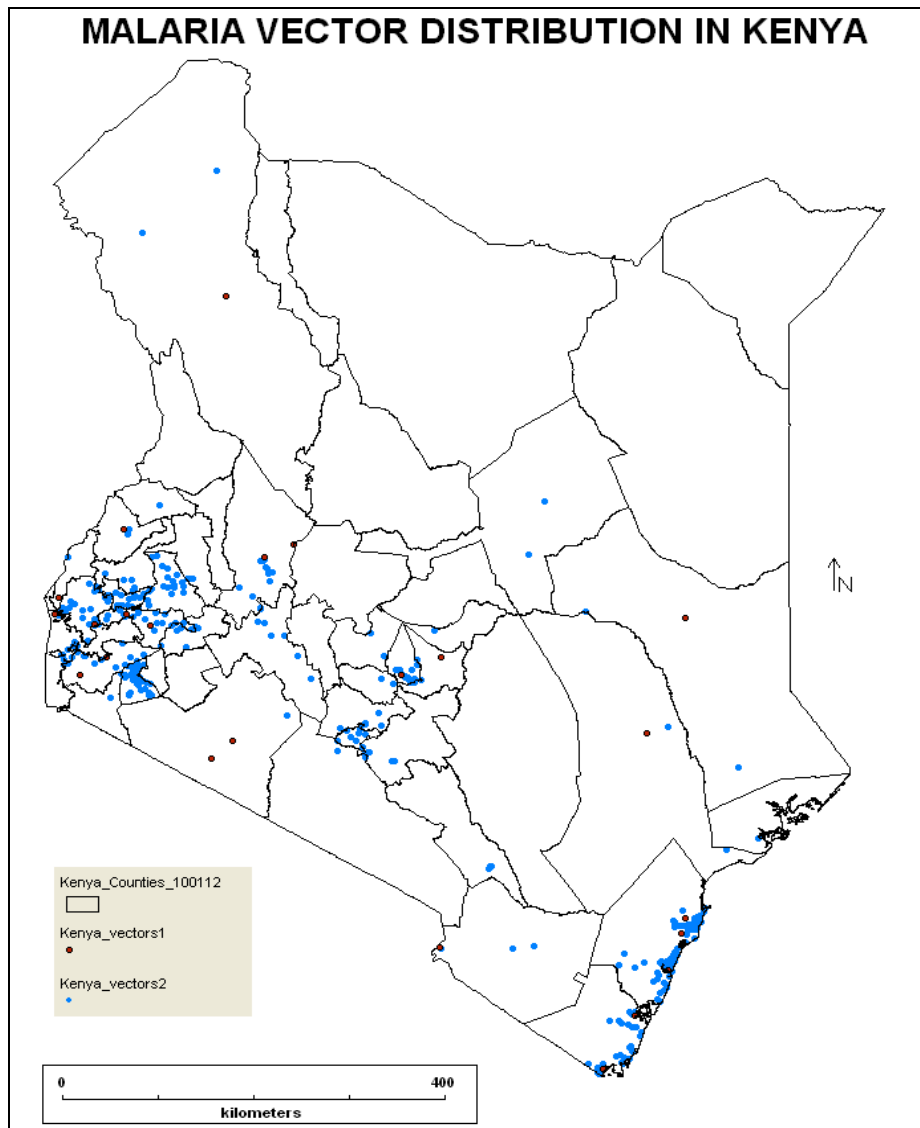


Figure 3. 6: Malaria Vector distribution in Kenya; Vectors1 in red (MARA/ARMA, 1998) and Vectors2 in blue (Okara et al., 2010)

3.4.1 Point Autocorrelation in vectors1

In order to calculate the autocorrelation for malaria vector points in vectors1, it was imperative to define which neighboring points to consider. This was done by specifying the separation distance between vector points referred to as lag (or neighborhood) distance, which was set as 0.05km. The lag distance was based on the assumption that malaria vectors can travel an average distance of 0.05km radius hence neighborhood polygons can emerge. Only pairs of points within the lag distance were considered for calculation. Figure 3.7

below shows the parameters used to determine autocorrelation of malaria vector point samples in vectors1.

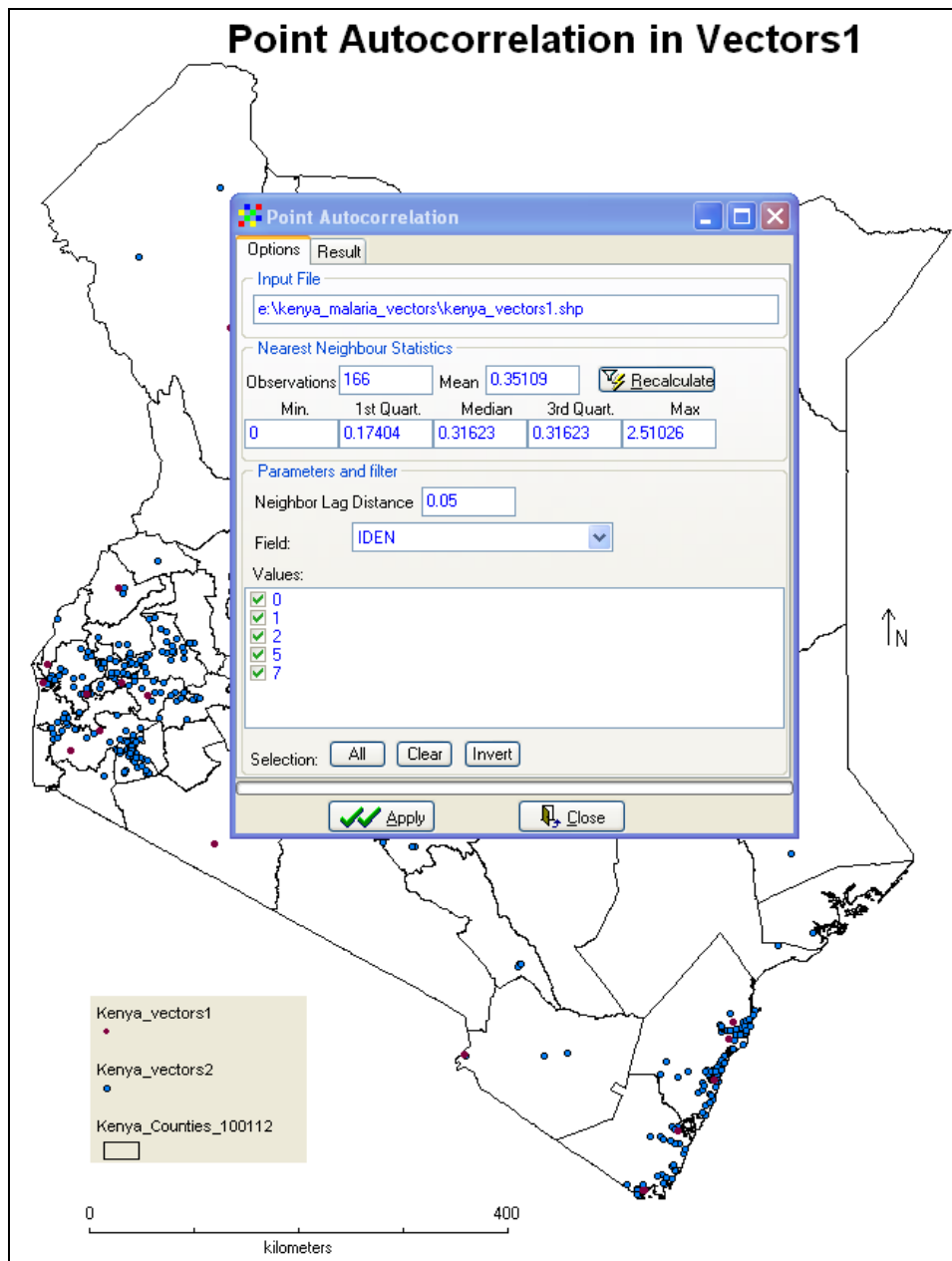


Figure 3. 7: Statistics used in the determination of lag distance (0.05) for Vectors1.

Point autocorrelation in vectors1 was analyzed with a lag distance of 0.05km between neighboring points in the data sample. With the lag distance of 0.05km only 24 pairs could be generated from a total of 166 vector points. The Geary's index value obtained was $c = 5.9948$ falling under $c > 1$ in Geary's index and Moran's index value was $i = 0.26726$ falling

under $i < 0^*$ in Moran's index (Figure 3.8). This shows that autocorrelation exists and the data are clustered. A negative autocorrelation exists in vector1 data, explaining the sparse distribution within the spatial space.

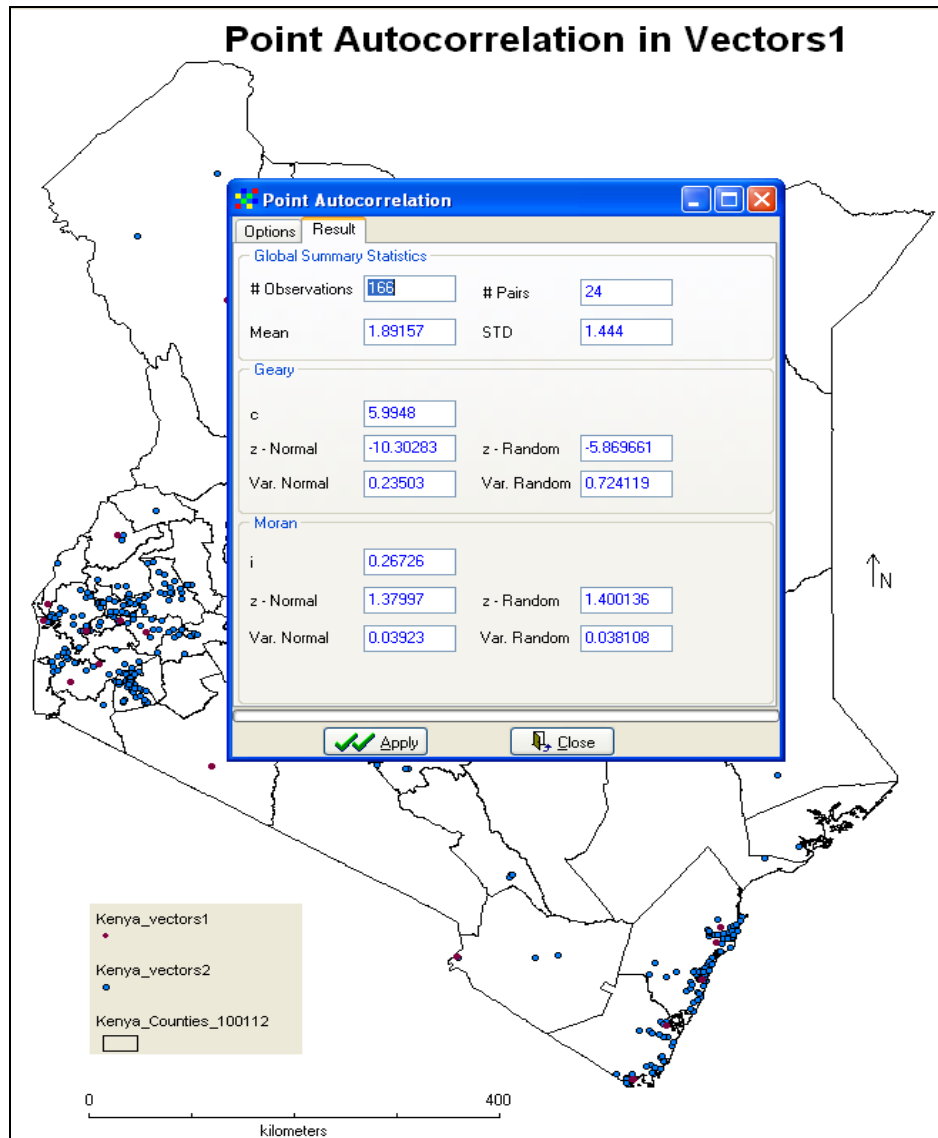


Figure 3.8: Vector spatial autocorrelation using Geary ($c=5.9948$) and Moran ($i=0.26726$) Indices for Vectors1.

3.4.2 Point Autocorrelation in vectors2

Point autocorrelation in vectors2 was determined with similar lag distance of 0.05km between neighboring points as in vectors1 above. Only pairs of points within the lag distance were considered for calculation. From a total number of 408 vector point observations, 607

pairs were obtained. To determine malaria vector point autocorrelation, a number of statistics were applied as shown in the Figure 3.9. These statistics include the number of observations (points) and statistics describing the distances among all pairs of points (min, max, mean, median and first and third quartile).

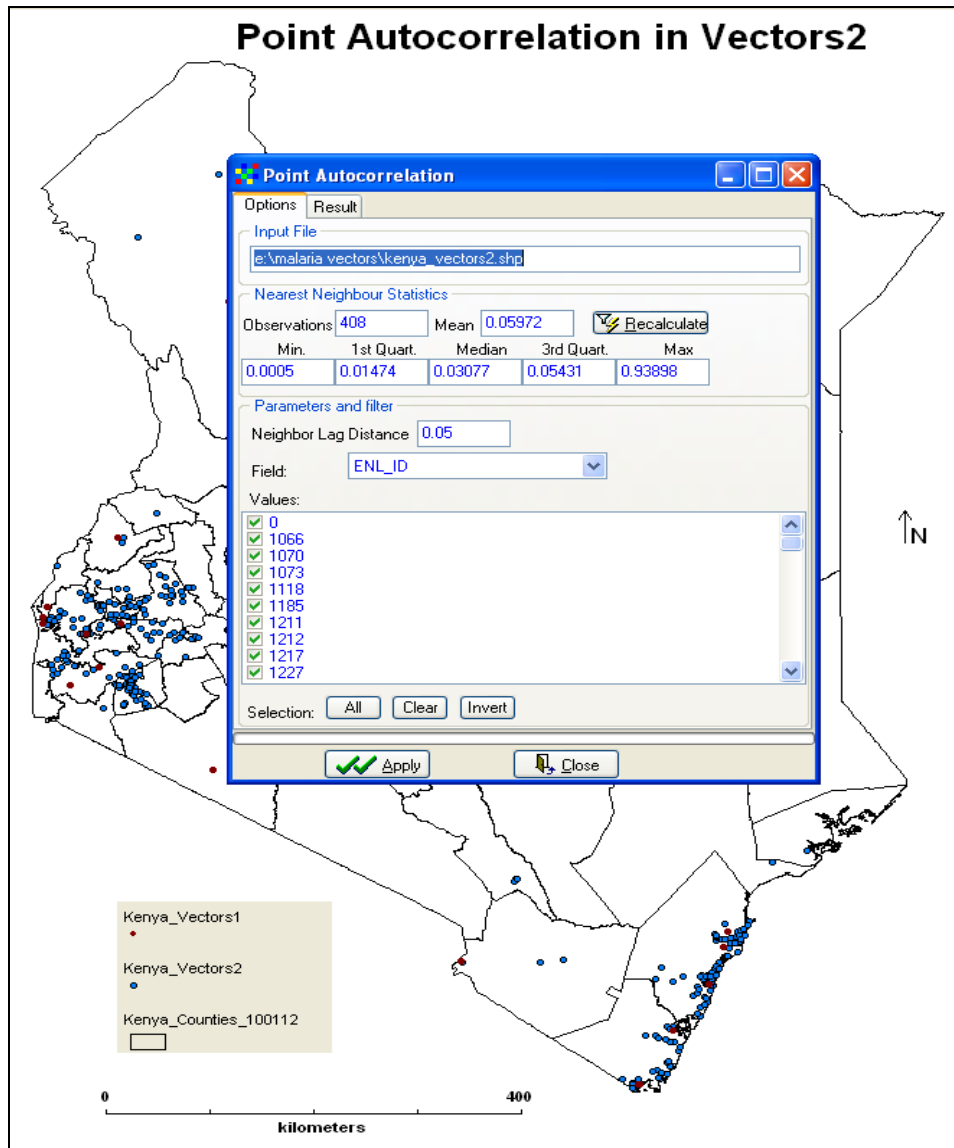


Figure 3. 9: Statistics used in the determination of lag distance (0.05) for Vectors2.

The autocorrelation in vector2 points was calculated and the results are as shown in Figure 3.10 below. The value obtained for c is 0.52751, falling under $0 < c < 1$ in Geary's index and

$i = 0.41394$, falling under $i > 0$ in Moran's index. This shows that autocorrelation exists and the data are clustered.

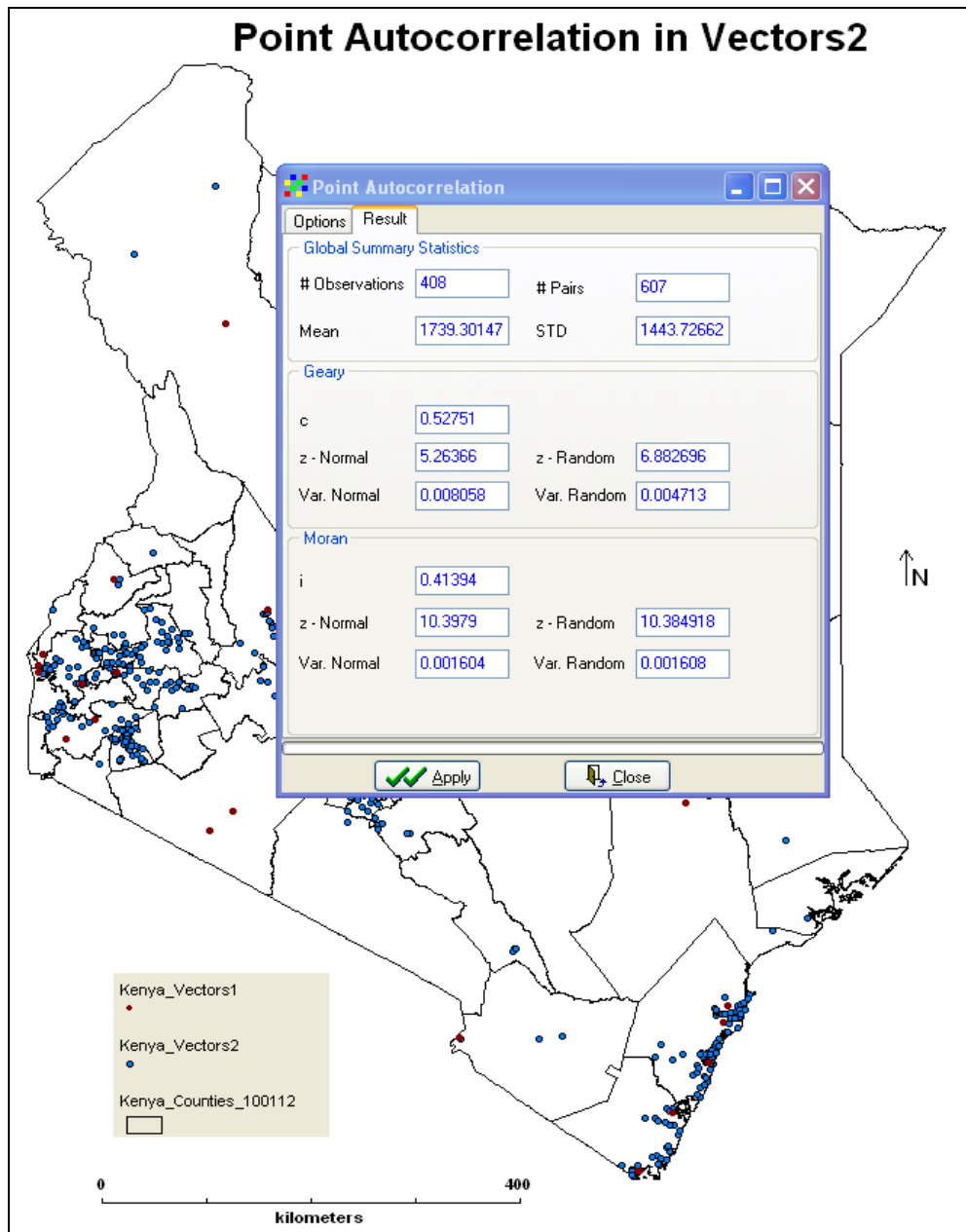


Figure 3. 10: Vector spatial autocorrelation using Geary ($c=0.52751$) and Moran ($i=0.41394$) Indices for Vectors2.

3.5 Climate Change and Malaria Vector Distributions in Kenya

In order to predict the change in malaria vector distribution in Kenya, there should be clear evidence of change in climate in the region likely to affect the ecology. During the twentieth century, world average surface temperature was reported to have increased by approximately 0.6°C, and approximately two-thirds of that warming has occurred since 1975 (IPCC 2007). Climatologists forecast further warming, along with changes in precipitation and climatic variability, during the coming century and beyond. Their forecasts are based on increasingly sophisticated global climate models, applied to plausible future scenarios of global greenhouse gas emissions that take into account alternative trajectories for demographic, economic and technological changes and evolving patterns of governance. There is *likely* to be an increase in annual mean rainfall in East Africa (IPCC 2007).

As reported by the working group III of AR4, future greenhouse gas (GHG) emissions are the product of very complex dynamic systems, determined by driving forces such as demographic development, socio-economic development, and technological change. Their future evolution is highly uncertain. Scenarios are alternative images of how the future might unfold and are an appropriate tool with which to analyze how driving forces may influence future emission outcomes and to assess the associated uncertainties.

Different scenarios were used to project the future global climate. Each storyline assumes a distinctly different direction for future developments, such that the four storylines differ in increasingly irreversible ways. Together they describe divergent futures that encompass a significant portion of the underlying uncertainties in the main driving forces. They cover a wide range of key “future” characteristics such as demographic change, economic development, and technological change. For this reason, their plausibility or feasibility should

not be considered solely on the basis of an extrapolation of current economic, technological, and social trends. The four storylines as described in IPCC, 2007 are as follows:

- The A1 storyline and scenario family describes a future world of very rapid economic growth, global population that peaks in mid-century and declines thereafter, and the rapid introduction of new and more efficient technologies. Major underlying themes are convergence among regions, capacity building, and increased cultural and social interactions, with a substantial reduction in regional differences in per capita income. The A1 scenario family develops into three groups that describe alternative directions of technological change in the energy system. The three A1 groups are distinguished by their technological emphasis: fossil intensive (A1FI), non-fossil energy sources (A1T), or a balance across all sources (A1B).
- The A2 storyline and scenario family describes a very heterogeneous world. The underlying theme is self-reliance and preservation of local identities. Fertility patterns across regions converge very slowly, which results in continuously increasing global population. Economic development is primarily regionally oriented and per capita economic growth and technological change is more fragmented and slower than in other storylines.
- The B1 storyline and scenario family describes a convergent world with the same global population that peaks in midcentury and declines thereafter, as in the A1 storyline, but with rapid changes in economic structures toward a service and information economy, with reductions in material intensity, and the introduction of clean and resource-efficient technologies. The emphasis is on global solutions to economic, social, and environmental sustainability, including improved equity, but without additional climate initiatives.

- The B2 storyline and scenario family describes a world in which the emphasis is on local solutions to economic, social, and environmental sustainability. It is a world with continuously increasing global population at a rate lower than A2, intermediate levels of economic development, and less rapid and more diverse technological change than in the B1 and A1 storylines. While the scenario is also oriented toward environmental protection and social equity, it focuses on local and regional levels.

Kenya is in East Africa (EA) region thus the projected changes apply. As shown in the Table 2.1 above, the reported minimum and maximum annual temperature change is 1.8°C and 4.3°C respectively. The quartile temperature response in all the quarters has portrayed an increasing trend. Figure 3.11 below has been generated to visualize the relationship in seasonal temperature response in East Africa over the given periodic difference. From the generated seasonal temperature response charts, it is evident that the season that will portray the greatest difference in temperature for the 1980-1999 and 2080-2099 climatic periods is June, July, August (JJA). This season is a quarter that comes immediately after the rainy quarter March, April, May (MAM), hence favorable for malaria vector transmission. The season with least temperature response difference is December, January, February (DJF) which is a rainy season, hence not conducive for malaria vector breeding and transmission. These two seasons provide the most limiting factors for malaria vector ENM.

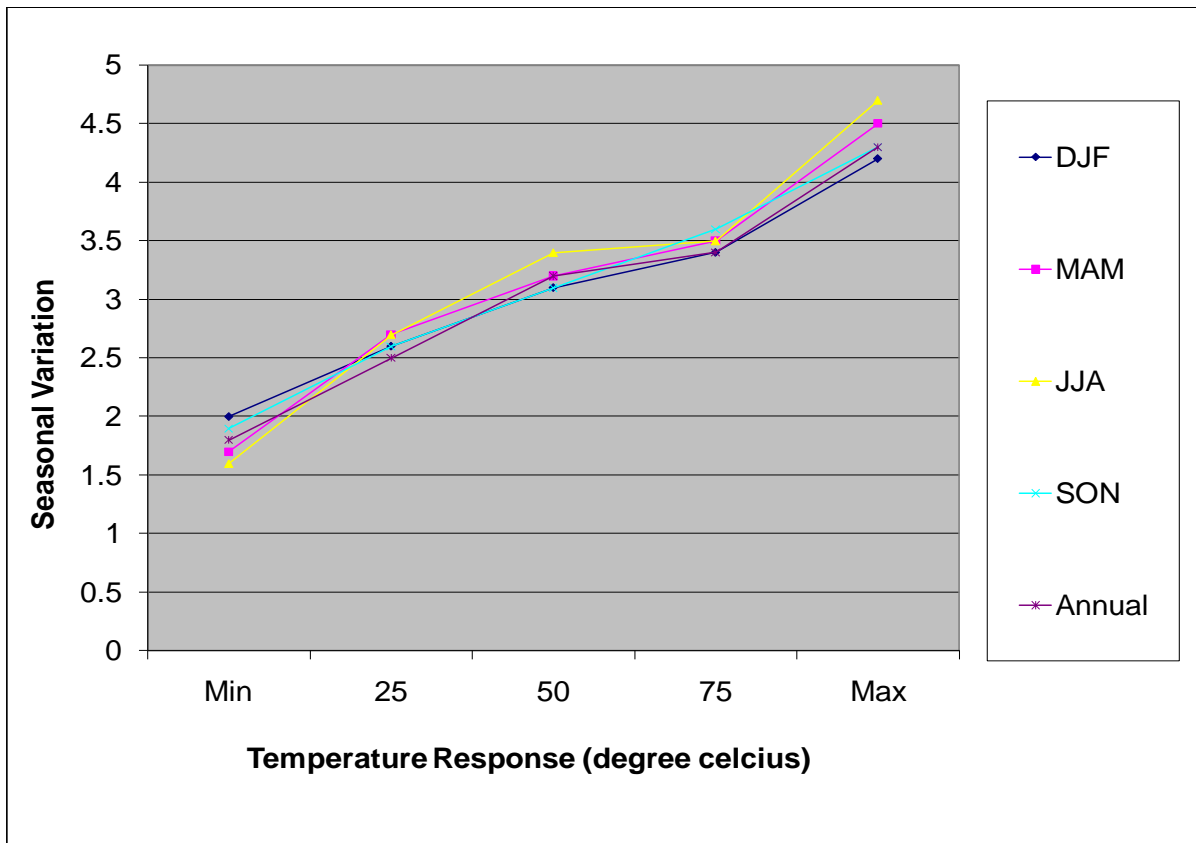


Figure 3. 11: Difference in Seasonal Temperature Response in East Africa for the periods (1980 – 1999) and (2080 – 2099)

The precipitation response (%) in East Africa was portrayed as increasing in all the seasons after differencing between 1980-1999 and 2080-2099 climatic periods. Figure 3.12 has been generated to visualize the differences and compare the seasons. The season with the least difference is JJA while DJF has the highest up to 75% quartile value, then September, October, November (SON) season portrayed the highest precipitation difference up to maximum.

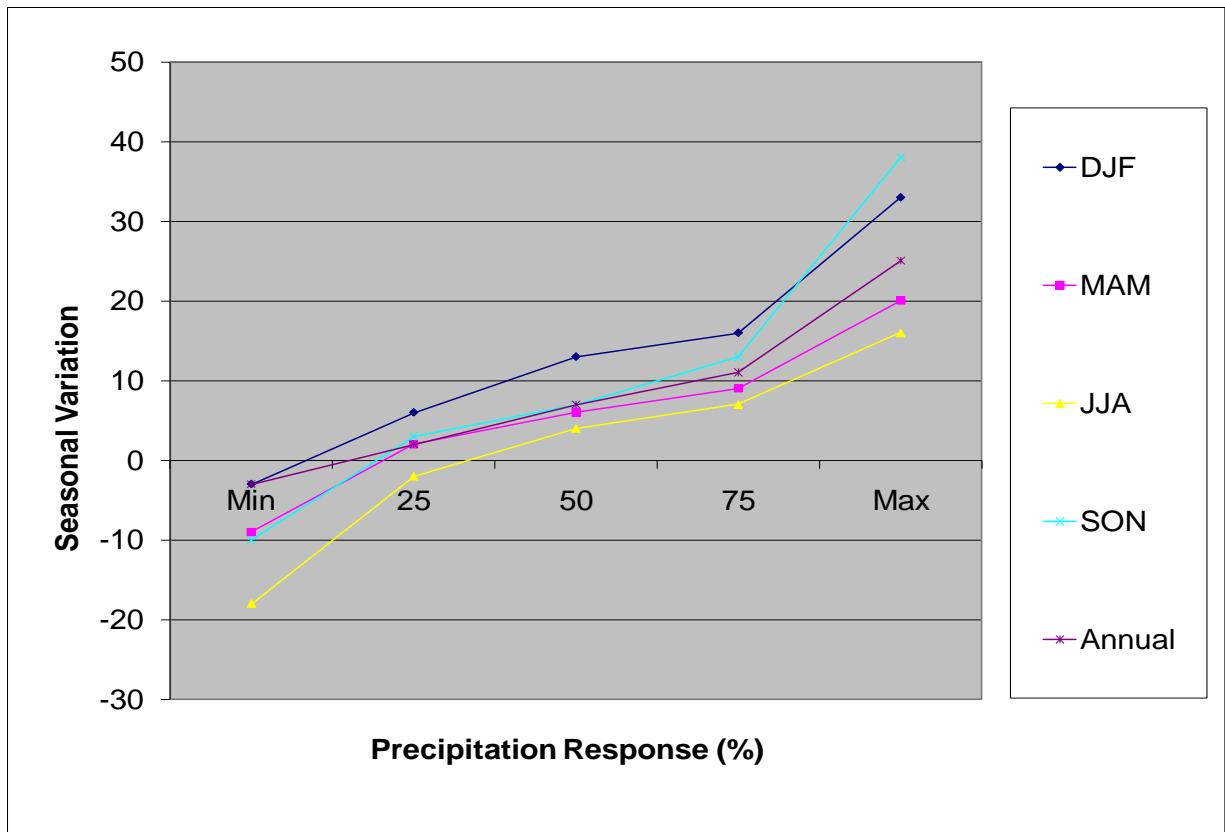


Figure 3. 12: Difference in Seasonal Precipitation Response in East Africa for the periods (1980 – 1999) and (2080 – 2099)

3.5.1 Prediction and Mapping of Malaria Vector Ecological Niches

The principal steps that have been executed in order to build and validate a correlative species' distribution model for malaria vectors in Kenya are outlined in Figure 3.13 below. The two types of model input data were known species' occurrence records, and a suite of environmental variables. Raw environmental variables, such as precipitation and temperature records collected from weather stations, had been processed to generate model inputs, further used to produce bioclimatic variables that were thought to have a direct physiological role in limiting the ability of the species to survive.

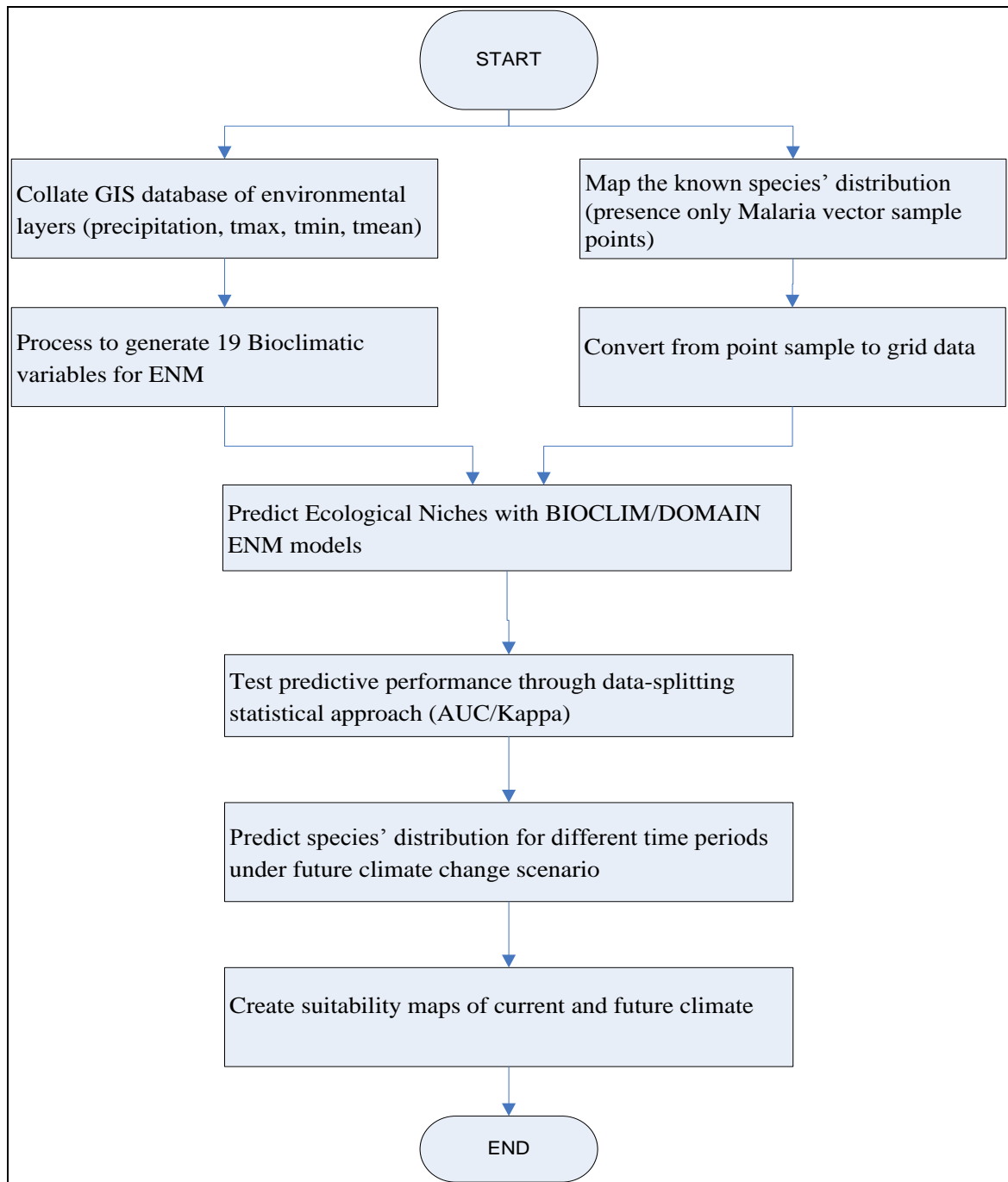


Figure 3. 13: Flow diagram for correlative malaria vector distribution modeling.

CHAPTER 4: RESULTS AND DISCUSSION

Malaria vector distribution modeling was done using BIOCLIM and DOMAIN species distribution models. The modeling effort quantified climate change as an explanatory variable in alteration of the suitable malaria vector ecologies in Kenya. Since two different data sets for published malaria vector geo-location had been acquired, there was need to test the autocorrelation of point data. Ecological Niche Modeling was done using the more clustered data set from Okara et al., 2010 which is denoted as Kenya_vectors2 in Figure 4.1. IPCC Future climate for HADCM3, CCCMA and CSIRO under A2a Scenario was used in prediction of malaria prevalence zones by 2020, 2050 and 2080.

4.1 Ecological Niche Modeling (ENM) Results and Analysis

Ecological Niche Modeling (ENM) was done to predict the geographic ranges of malaria vector distribution, and hence malaria parasite prevalence in Kenya. The predictive spatial risk mapping was achieved by use of Worldclim ESRI grid rasters for current climate from 1950 to 2000 and IPCC projected A2a scenario for 2020, 2050 and 2080, for three different climate models (HADCM3, CCCMA and CSIRO). Biological data for vector presence with spatial reference was obtained from existing publications. Geographic mapping of these vector points and climate surfaces was done with Diva - GIS software, which has ecological niche modeling functionalities. In order to achieve the set objectives, surface maps at 30 second resolution for minimum and maximum monthly mean temperature and precipitation were collected and re-processed to create the bioclimatic factors, eventually used to develop future suitability maps.

Bioclimatic variables were generated from three different future climate models, namely, HADCM3, CCCMA and CSIRO all of 30 arc seconds resolution, under A2a scenario. Ecological Niche Modeling was performed with each data set and predicted suitability maps produced using BIOCLIM, BIOCLIM True/False and DOMAIN predictions. The suitability maps for malaria vector prevalence were generated for visualization and arranged in succession from the maps for current climate (1950 – 2000) being the first, followed by the predicted prevalence by the year 2020, then 2050 and finally 2080 predictions.

4.1.1 BIOCLIM Model Prediction with HADCM3, CCCMA and CSIRO Data

Ecological Niche Modeling was done using BIOCLIM prediction model to create climate envelope for malaria vectors in Kenya. The environmental data used in ecological suitability analysis was from HADCM3, CCCMA and CSIRO future climate under A2a scenarios. Ecological suitability prediction was based on the assumptions that:

- malaria vector presence is restricted to locations equally or more suitable than those at which the species has been observed; and
- most environmental variables are continuous.

Discriminant Function Analysis (DFA) was applied to map different classes of suitable areas for malaria vectors to thrive. Areas completely outside the 0-100 percentile envelope for one or more climate variables get a code “0”. The cells within the 5-95 percentile get a code “3”, those outside this range but within the 2.5-97.5 percentile get a code “2”, and the ones outside this but within the 0 -100 percentile for all climate variables get a code “1”. Different color shades were used to denote the suitability classes. The approach therefore identifies the minimum areas in which malaria vectors occurs whilst ensuring that no localities at which the vectors have been observed are omitted (i.e. omission rate=0 and sensitivity = 1).

The areas found not to harbor the suitability to sustain malaria vectors were denoted by gray color. Areas of low suitability were denoted as dark green, while light green color was used to describe areas of medium potential for malaria vectors to thrive. Ecological zones with high potential were shown in yellow while orange color was used to denote those ecologies of very high potential. Red color was used to designate all the ecological zones with excellent geographical space for malaria vectors in Kenya. Maps for the predicted results using BIOCLIM model and the three IPCC future climates scenarios are shown in Figures 4.1 for HADCM3, 4.2 for CCCMA and 4.3 for CSIRO models all under A2a scenario.

4.1.1.1 Generated BIOCLIM Prediction Results from HADCM3 Model data

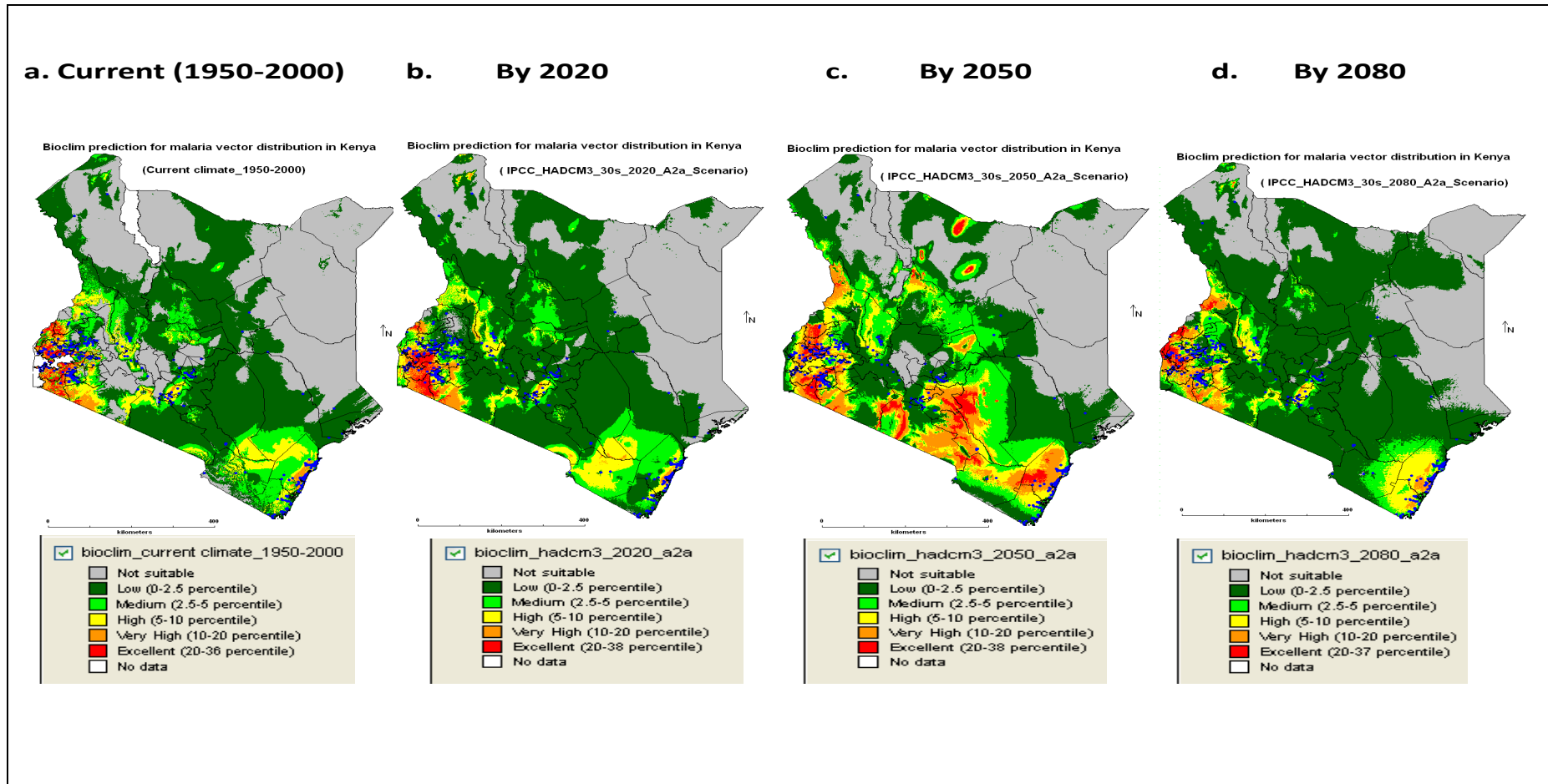


Figure 4. 1: BIOCLIM Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s HADCM3 2020(b), 2050(c) and 2080(d)

4.1.1.2 Generated BIOCLIM Prediction Results from CCCMA Model data

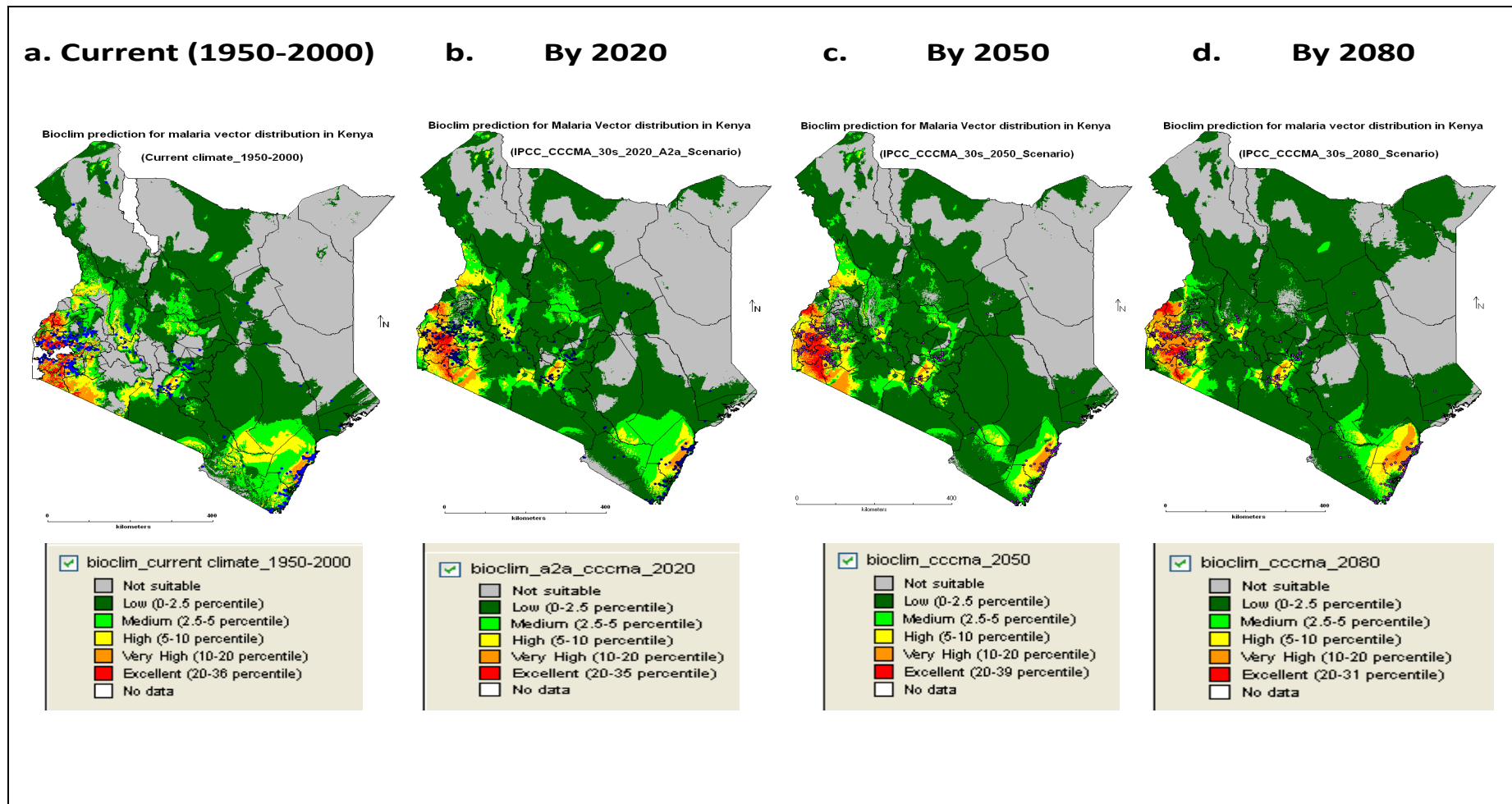


Figure 4. 2: BIOCLIM Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CCCMA 2020(b), 2050(c) and 2080(d)

4.1.1.3 Generated BIOCLIM Prediction Results from CSIRO Model data

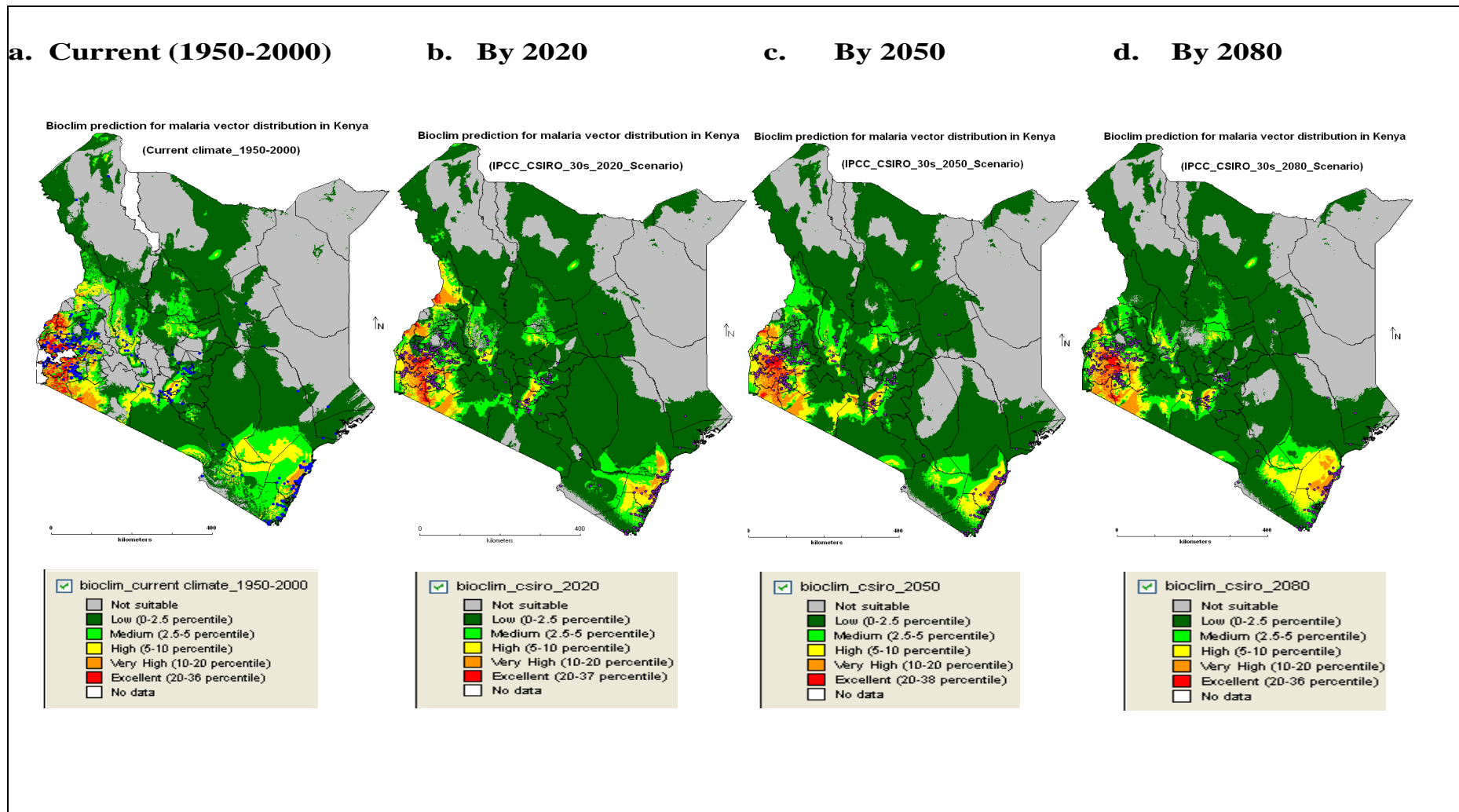


Figure 4. 3: BIOCLIM Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CSIRO 2020(b), 2050(c) and 2080(d)

4.1.2 BIOCLIM True or False Model Prediction with HADCM3, CCCMA and CSIRO Data

Malaria vector distribution modeling was done using BIOCLIM True/False species distribution model. All areas that fell within the envelope described by malaria presence sample points were cut off beyond a certain user defined percentile and mapped using binary categorization as “true” (1) and “false” (0). All the areas found to be suitable for malaria vectors to thrive were denoted with red color while those areas with no potential to sustain malaria vectors were shown in gray. Forecasting for malaria vector distribution was done with HADCM3, CCCMA and CSIRO IPCC future projected climate under A2a scenario. The generated maps had two ecological zones with gray color denoting all the unsuitable areas and red color denoting the areas of excellent ecologies for malaria vectors to thrive. The results are displayed in Figures 4.4 for HADCM3, 4.5 for CCCMA and 4.6 for CSIRO models.

4.1.2.1 Generated BIOCLIM True or False Prediction Results from HADCM3 Model data

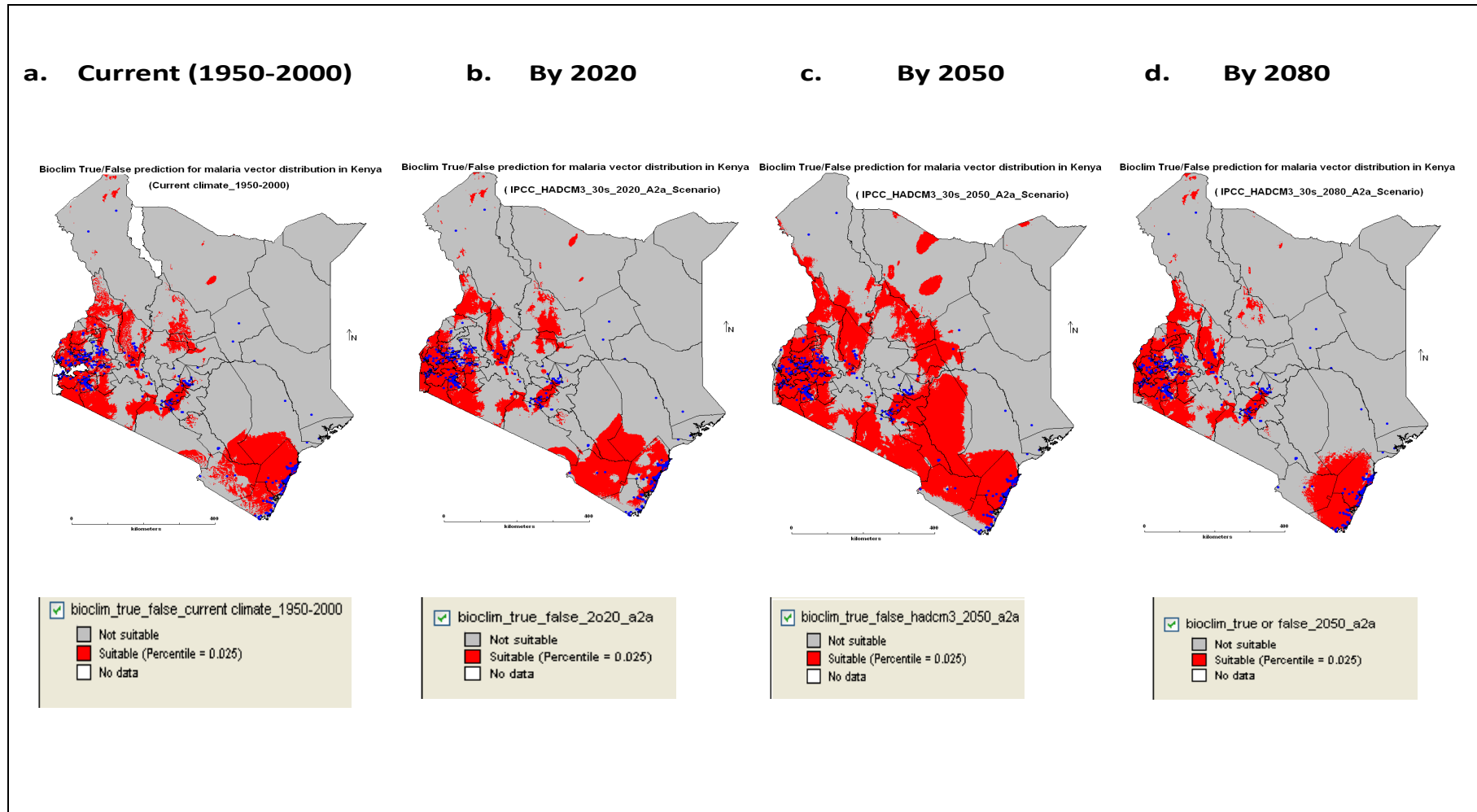


Figure 4. 4: BIOCLIM True/False Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s HADCM3 2020(b), 2050(c) and 2080(d)

4.1.2.2 Generated BIOCLIM True or False Prediction Results from CCCMA Model data

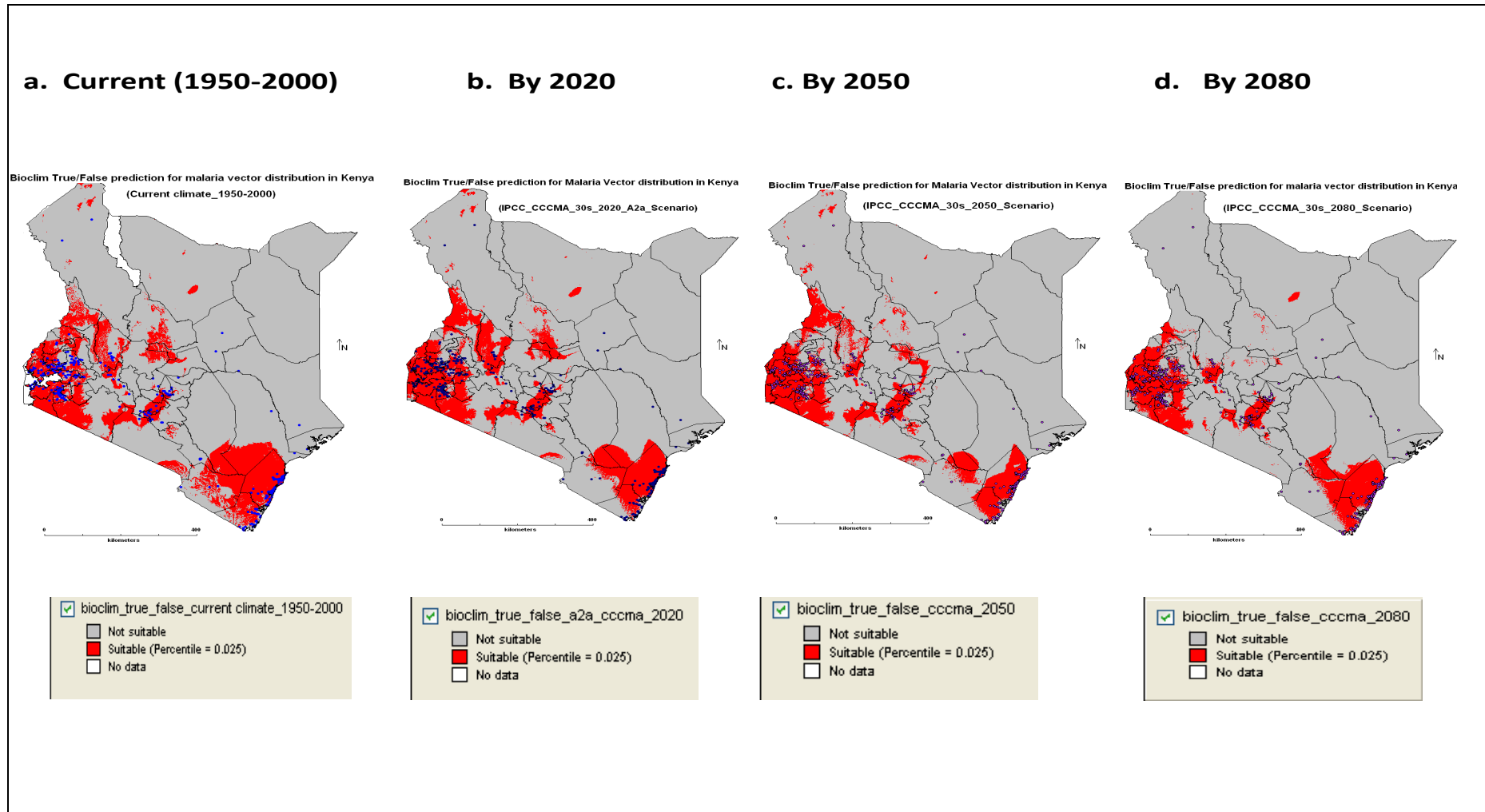


Figure 4. 5: BIOCLIM True/False Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CCCMA 2020(b), 2050(c) and 2080(d)

4.1.2.3 Generated BIOCLIM True or False Prediction Results from CSIRO Model data

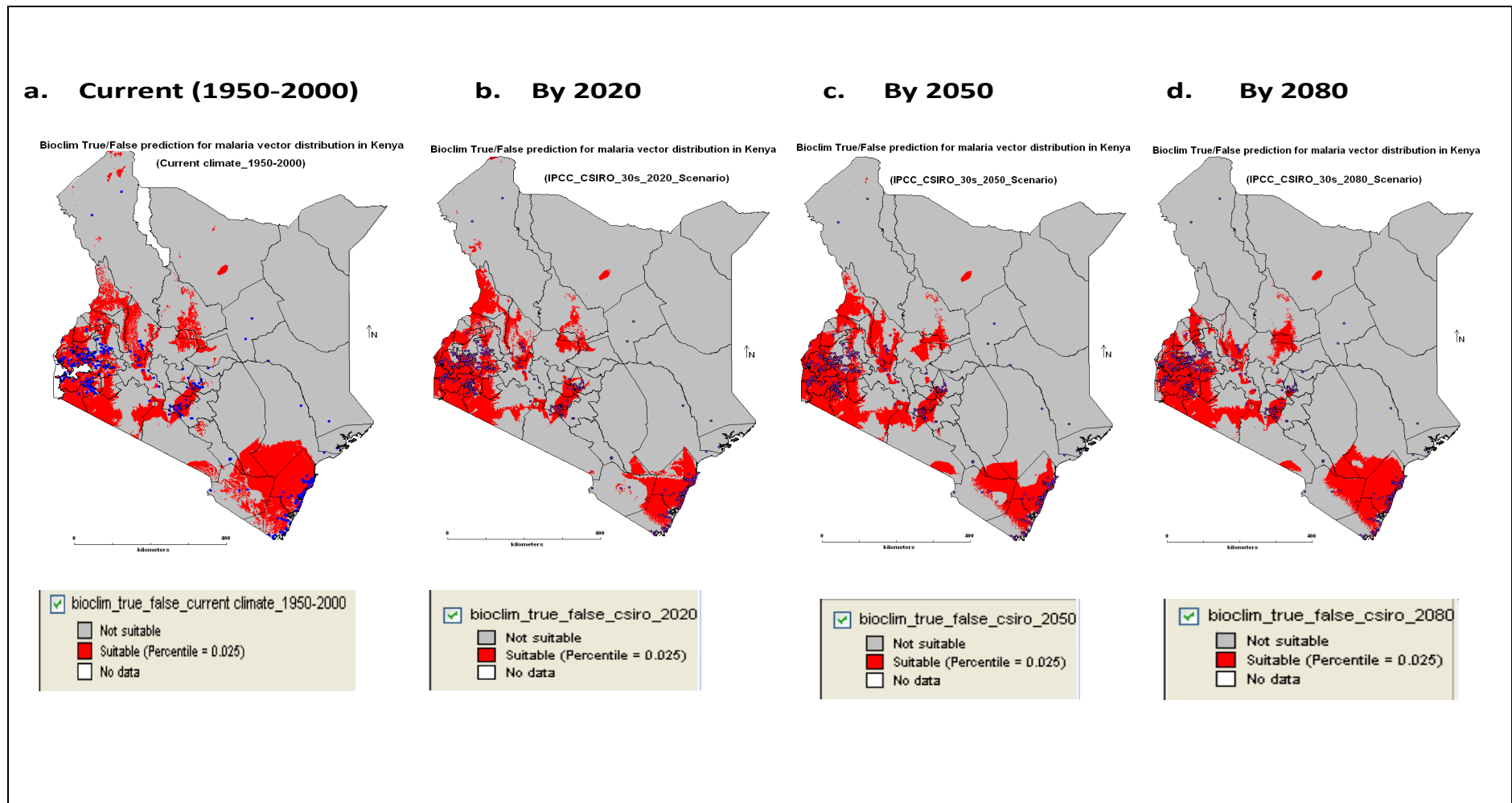


Figure 4. 6: BIOCLIM True/False Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CSIRO 2020(b), 2050(c) and 2080(d)

4.1.3 DOMAIN Model Prediction with HADCM3, CCCMA and CSIRO Data

Prediction of the ecological space for malaria vectors in Kenya was done using DOMAIN that calculates the Gower distance statistic characteristic between each cell on the map and each point, from the generated 19 climate variables for HADCM3, CCCMA and CSIRO IPCC future climate models under A2a scenario. The distance between any malaria vector sample point, say *A* and grid cell *B* for a single climate variable is calculated. This is obtained as the absolute difference in the values of that variable divided by the range of the variable across all points (Equation 3.1). The Gower distance is therefore the mean over all the 19 bioclimatic variables for each future climate model.

The generated prediction maps represented six different ecological zones. The areas that were found not to be suitable were denoted with gray color. Dark green color denoted the ecological zones of low suitability while light green denoted those areas found to be of medium suitability for malaria vectors to thrive. Areas denoted by yellow color were those found to harbour high potential while color orange denoted those areas found to harbour very high potential. The ecological zones that were found to be excellent for malaria vectors to thrive in were shown in color red. Suitability maps were generated for the three IPCC future projected climate change scenarios. The results are shown in Figures 4.7 for HADCM3, 4.8 for CCCMA and 4.9 for CSIRO models.

4.1.3.1 Generated DOMAIN Prediction Results from HADCM3 Model data

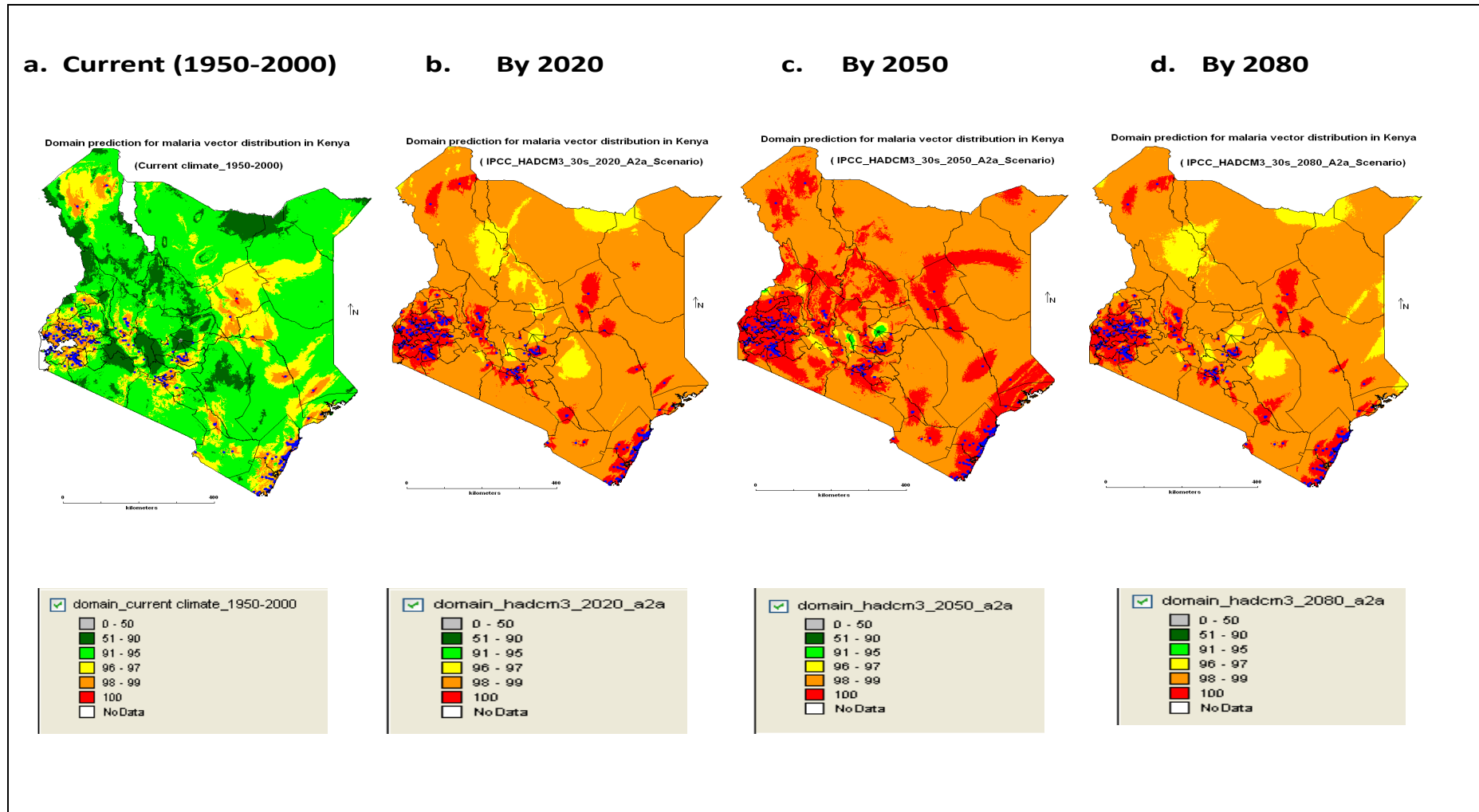


Figure 4. 7: DOMAIN Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s HADCM3 2020(b), 2050(c) and 2080(d)

4.1.3.2 Generated DOMAIN Prediction Results from CCCMA Model data

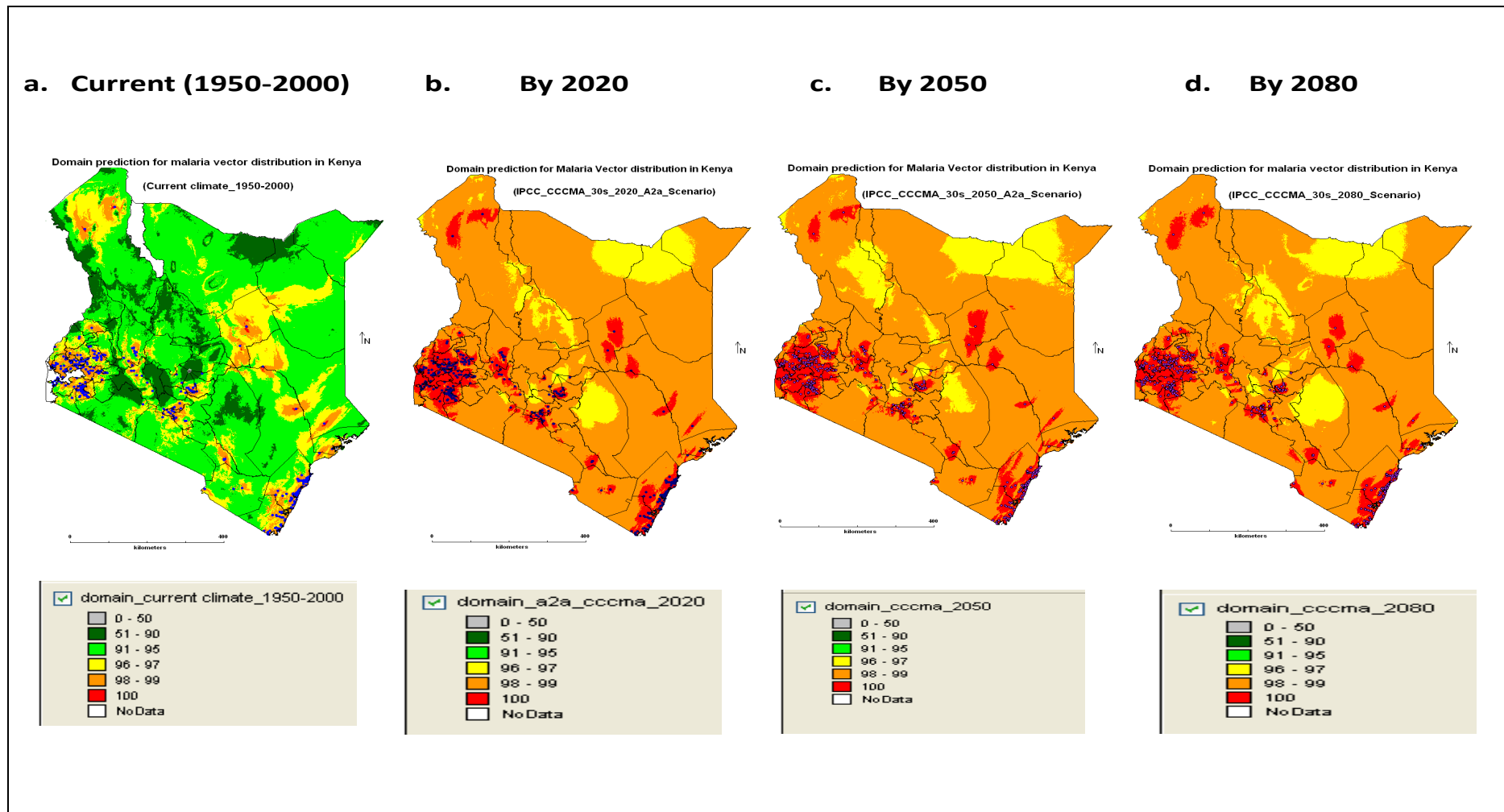


Figure 4. 8: DOMAIN Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CCCMA 2020(b), 2050(c) and 2080(d)

4.1.3.3 Generated DOMAIN Prediction Results from CSIRO Model data

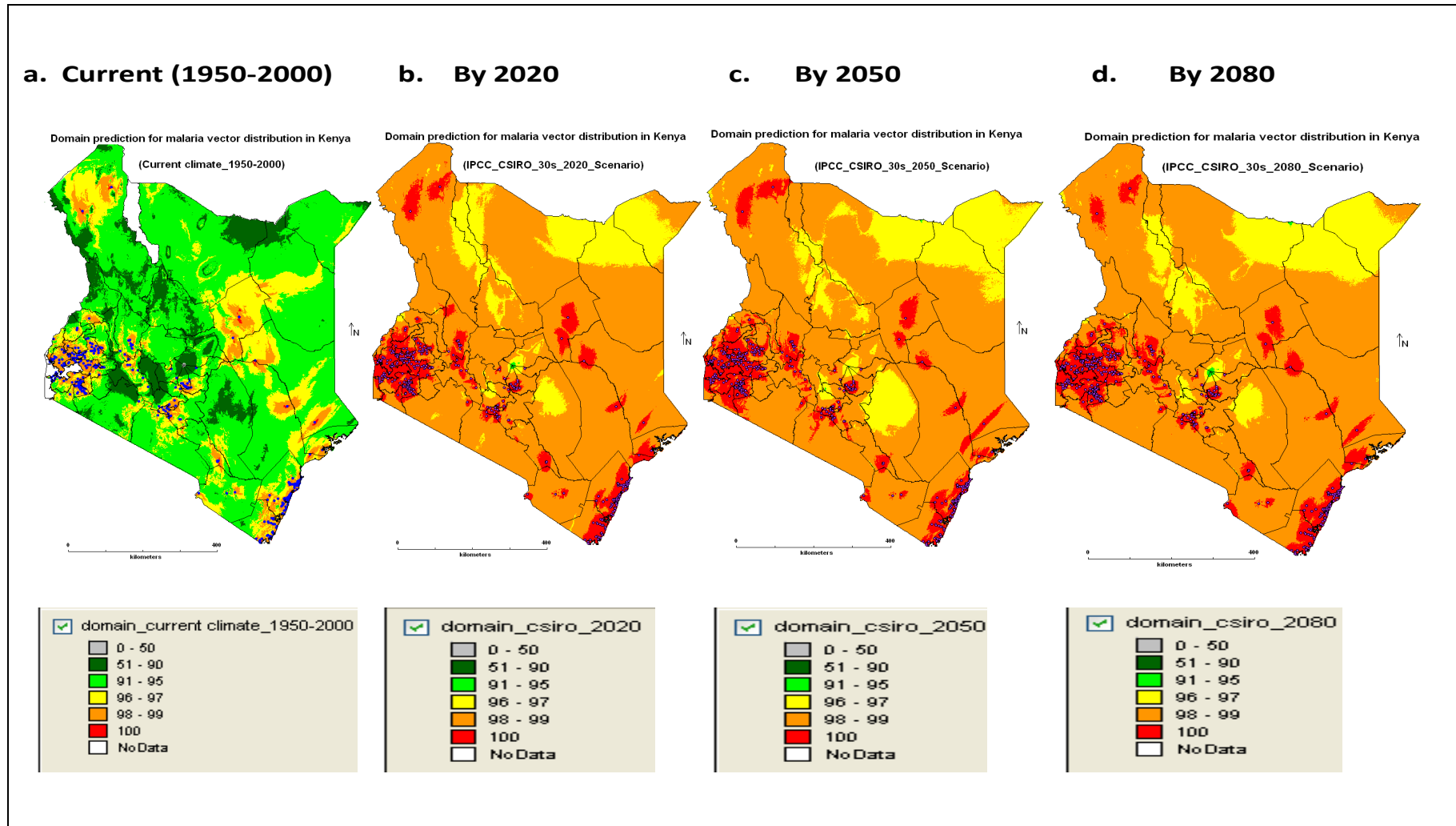


Figure 4. 9: DOMAIN Ecological Niche models for malaria vector distribution in Kenya for current climate 1950-2000(a) and IPCC projected 30s CSIRO 2020(b), 2050(c) and 2080(d)

4.2 Model Output Results

Ecological Niche Modeling was done with current climate data (1950 – 2000) and IPCC projected data for three climate models (HADCM3, CCCMA and CSIRO). The projected data was at 30 years' interval from the current climate. Thus, predictive modeling was done and mapped as current climate, projected by the years' 2020, 2050 and 2080. The predicted suitability maps generated from the three climate models have shown that climate change in Kenya will result to the spread of malaria vectors to new areas. Most current ecologies will still remain as viable malaria zones, with continuous extent to neighboring zones. Further change to climatic conditions will trigger extremely new hotspots that were not close to the current existing ecosystems. The results obtained in each prediction model are analyzed in sections 4.2.1 for BIOCLIM, 4.2.2 for BIOCLIM True or False and 4.2.3 for DOMAIN.

4.2.1 Analysis and Discussion of BIOCLIM Model Prediction Results

Prediction of suitable ecologies using BIOCLIM model resulted into six categorized classes where malaria vectors could be found. Representative bar graphs were generated for each BIOCLIM prediction with the three IPCC future projected climate scenarios. The areas for non-suitable ecological niches for malaria vectors to thrive in Kenya have been quantified based on current climate and IPCC future projections. Figure 4.10 below shows that the current climate (1950-2000) has the largest unsuitable area of 285,830 km² while CCCMA by 2080 prediction has the least unsuitable area of 180,984 km². Predicted non-suitable niches for HADCM3 and CSIRO projections have shown a similar trend of decrease of area from current to by the year 2020, with the area increasing by the year 2050, then further decrease by the year 2080. On the contrary, CCCMA projection has deviated from this trend but

portrays continuous decrease in ecological niche suitability from current climate through 2020 and 2050 up to by the year 2080.

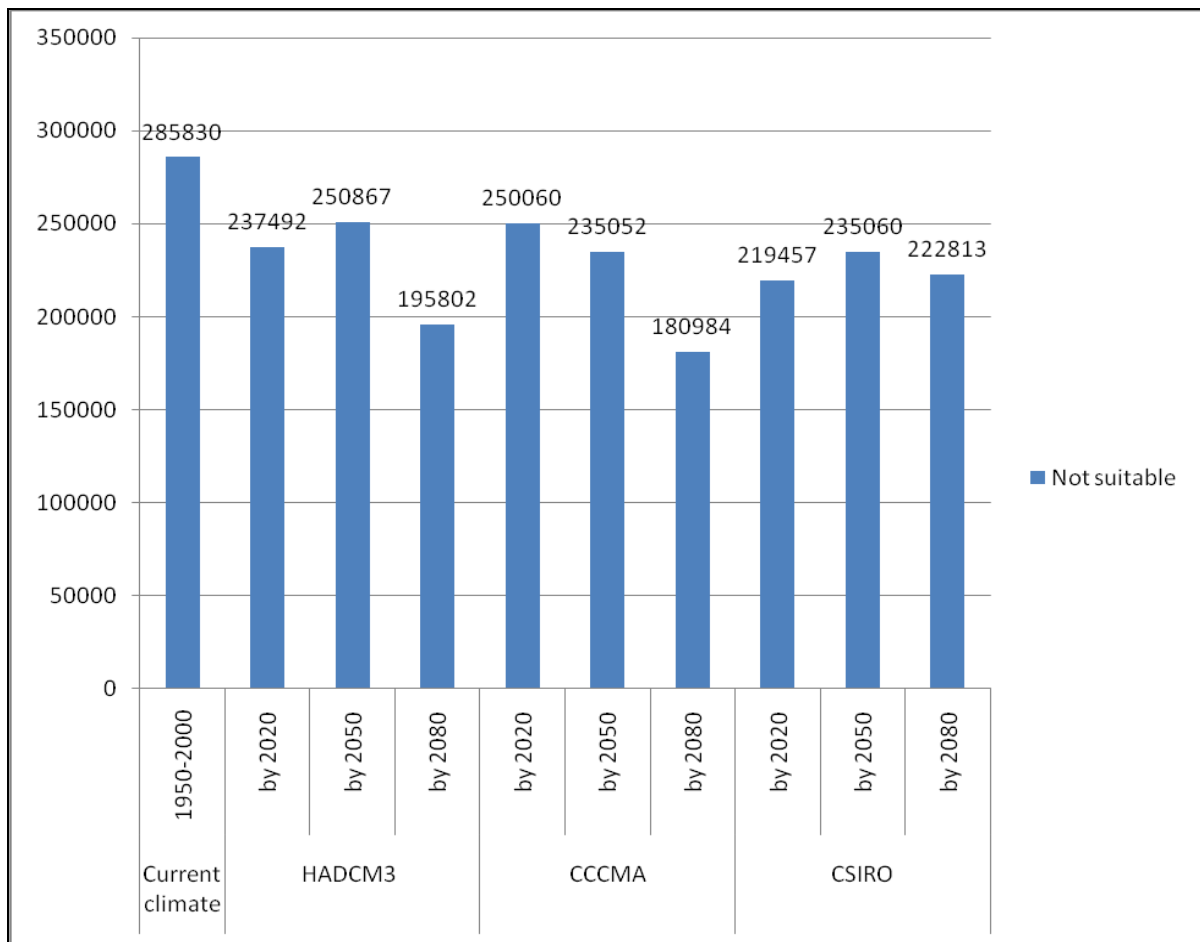


Figure 4. 10: BIOCLIM predicted areas for non-suitable Ecological Niche from current climate, HADCM3, CCCMA and CSIRO

The ecological niche with low potential to support the malaria vectors to thrive in Kenya was also quantified based on area for each BIOCLIM prediction with different future climate projections. Figure 4.11 below shows how the area changed from current climate (1950-2000) through 2020 and 2050, up to by the year 2080. The highest area of low suitability for malaria vector habitat in Kenya was 428,596 km² predicted with CCCMA projection by the year 2080. BIOCLIM prediction from HADCM3 projection by the year 2050 resulted into the least area of low suitability for malaria vectors to thrive in Kenya. Trend analysis showed that bioclim prediction for low suitability for HADCM3 and CSIRO projections will increase upto

by the year 2020, followed by decrease by the year 2050, then further increase upto 2080. On the contrary, BIOCLIM predictions with CCCMA projection portrayed continuous increase from current climate upto by the year 2080 of low suitability ecological niche.

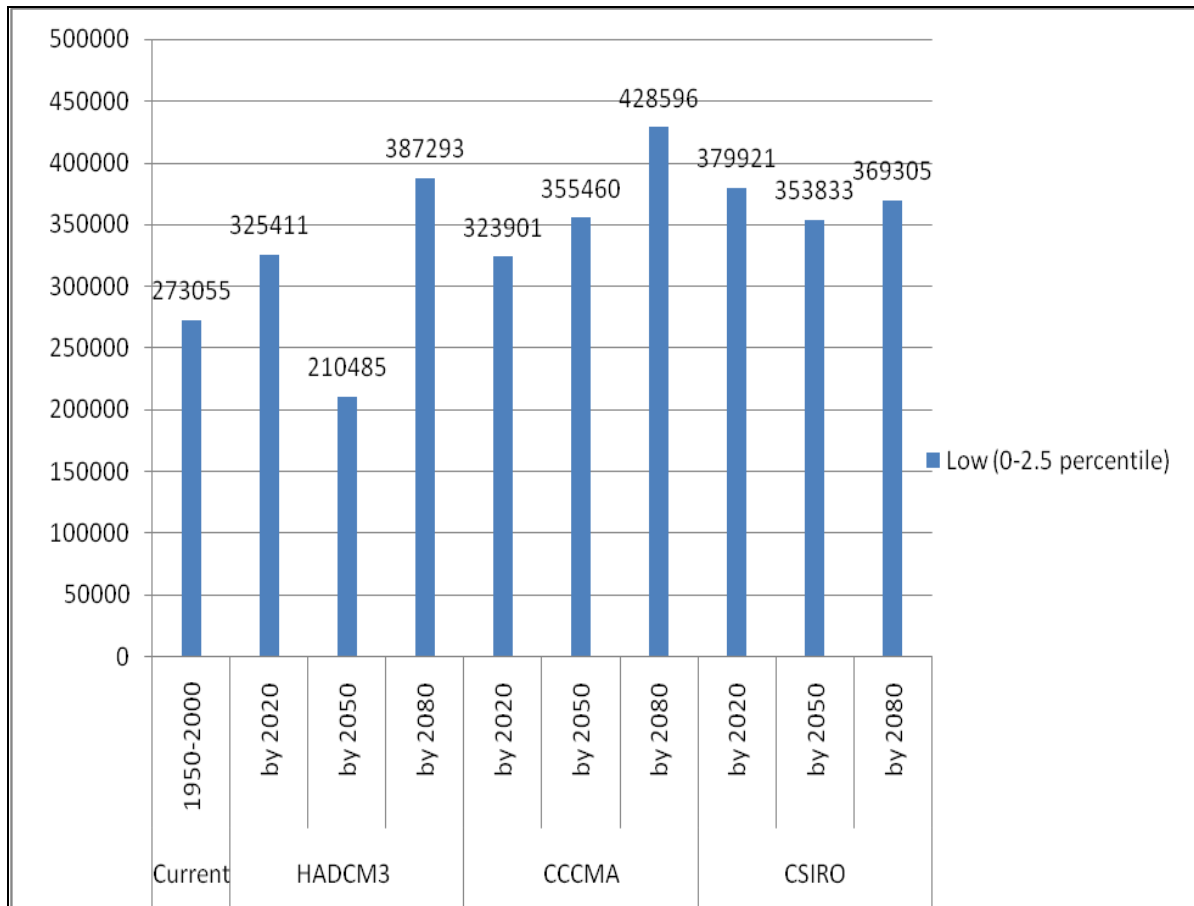


Figure 4. 11: BIOCLIM predicted areas of low suitability from current climate, HADCM3, CCCMA and CSIRO

Further investigation was done to study the trend in areas found to harbor medium potential for malaria vectors to thrive. Figure 4.12 below showed that the highest area with medium potential was 97,990 km² predicted by BIOCLIM with HADCM3 projection by the year 2050. The lowest area of medium potential was 46,046 km² predicted by BIOCLIM with CCCMA future climate scenario. When prediction was done with HADCM3 projection, the area of medium potential was found to increase from current climate through 2020 up to by

the year 2050, followed by drastic decline up to by the year 2080. Prediction with CCCMA projection resulted into an increase in area of medium suitability from current climate up to by the year 2020, beyond which there was decrease through the year 2050 up to by the year 2080. BIOCLIM prediction with CSIRO projection showed that areas of medium suitability will decline from the area predicted with current climate up to by the year 2020, increase moderately up to 2050, followed by decrease in the area up to by the year 2080.

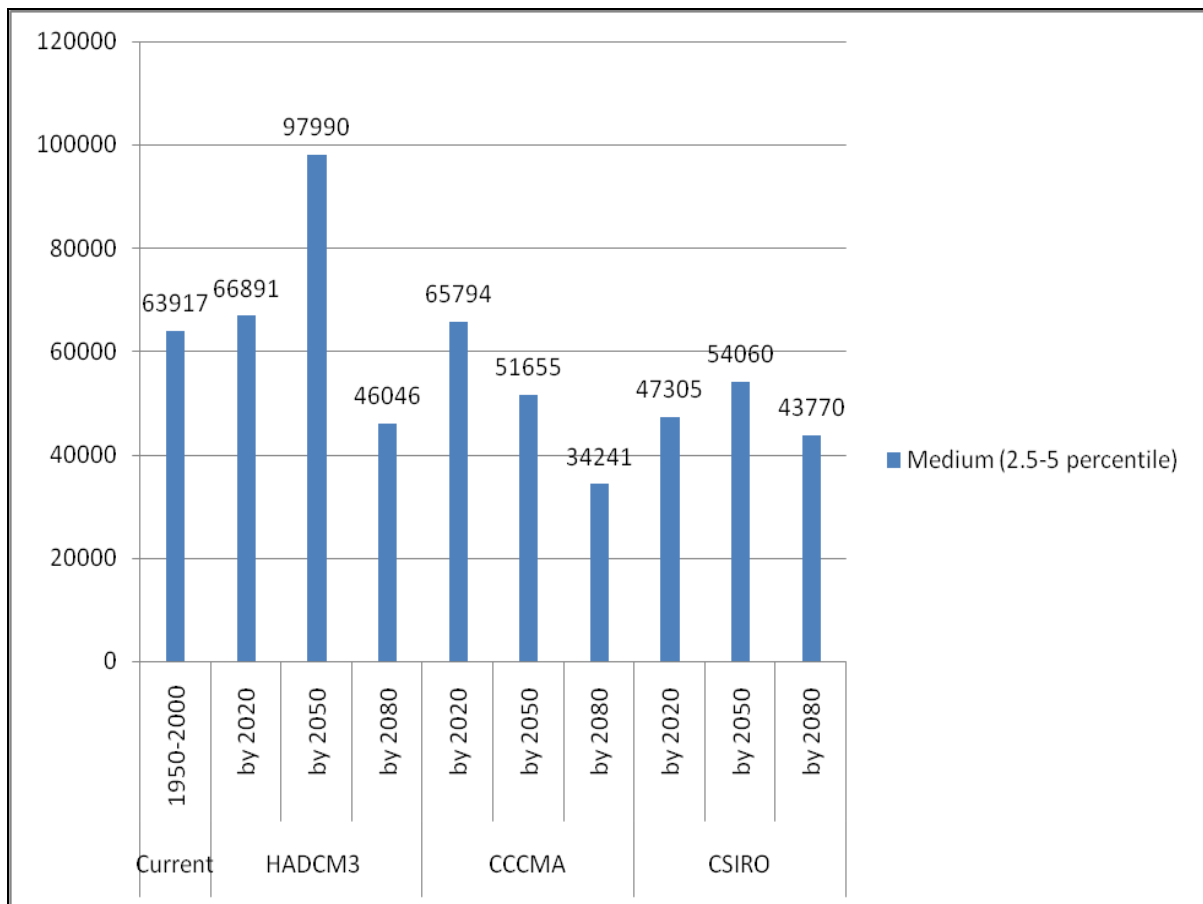


Figure 4. 12: BIOCLIM predicted areas of medium suitability from current climate, HADCM3, CCCMA and CSIRO

BIOCLIM prediction with HADCM3 projection by the year 2050 resulted into the highest area of 58,728 km² found to be of high suitability habitat for malaria vectors in Kenya. The lowest habitat area was 26,480 km² predicted with CCCMA projection by the year 2080. Trend analysis for predictions with HADCM3 projection showed that the ecological niche of

high potential for malaria vectors to thrive will increase moderately from the current climate prediction through the year 2020, followed by drastic increase up to the year 2050, then decrease drastically up to the year 2080. Predictions with the CSIRO projection showed that there will be very gradual decrease from the current climate prediction through by the year 2020, by 2050 and up to the year 2080. Conversely, predictions with CSIRO projection showed that there will be decreased areas of high suitability from the current climate prediction up to by the year 2020, followed by an increase of the high suitability areas through the year 2050, up to by the year 2080 (Figure 4.13).

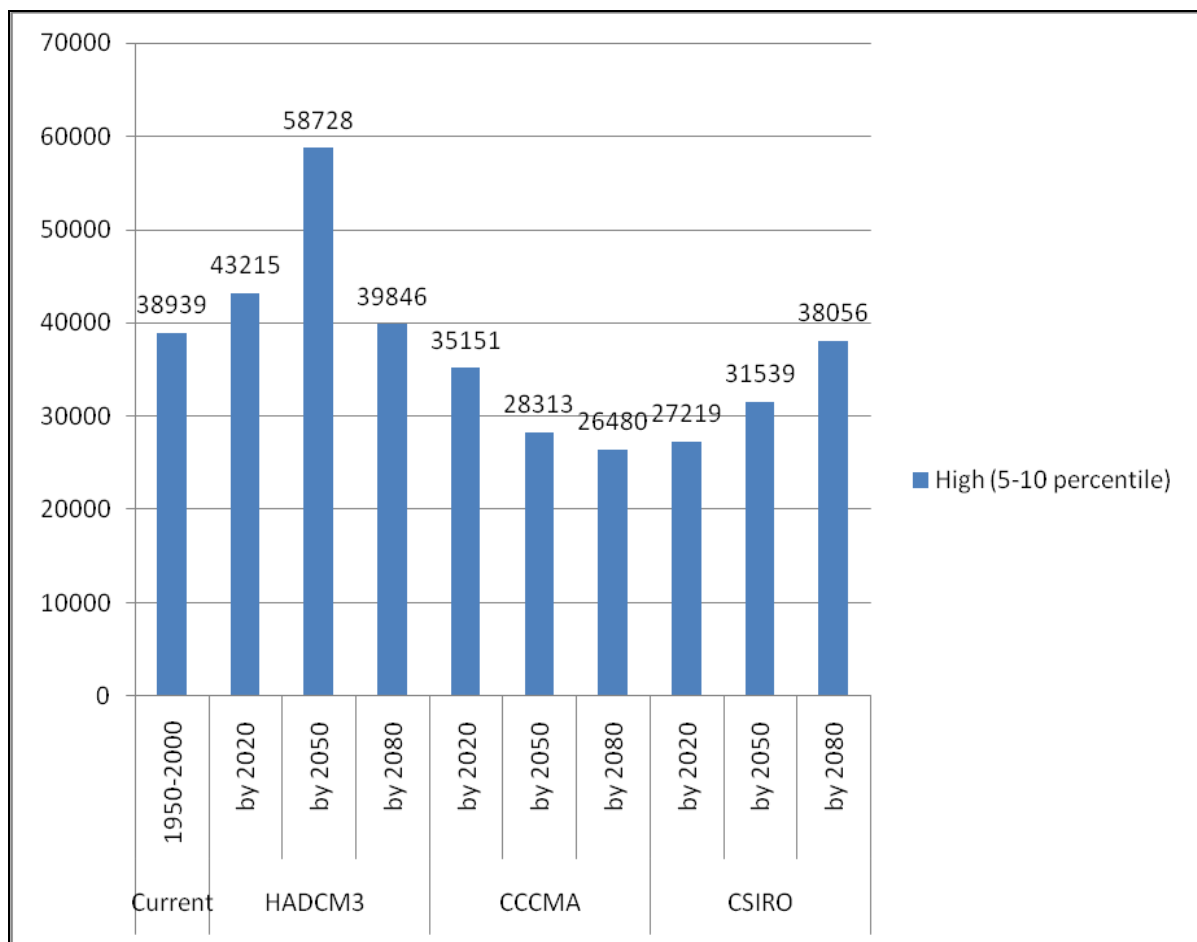


Figure 4. 13: BIOCLIM predicted areas of high suitability from current climate, HADCM3, CCCMA and CSIRO

The highest ecological niche of very high suitability for malaria vectors to thrive was predicted with HADCM3 projection by the year 2050 as shown in Figure 4.14 below. Apart

from the prediction with HADCM3 projection by the year 2050 which was very high with an area of 56,033 km², the rest of the predictions varied moderately, ranging from the lowest area of 11,861 resulting from current climate projection, to an area of 18,582 km² resulting from prediction with HADCM3 projection by the year 2080.

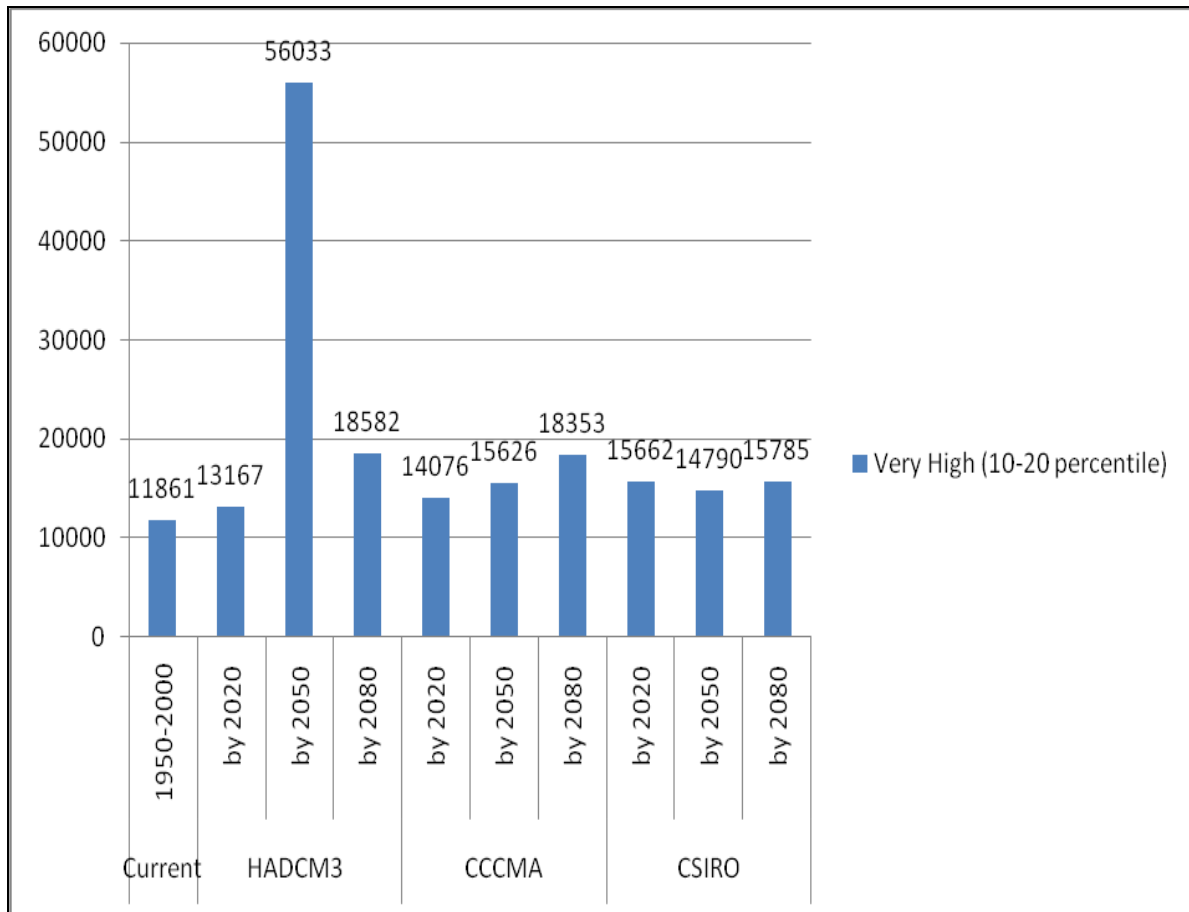


Figure 4. 14: BIOCLIM predicted areas of very high suitability from current climate, HADCM3, CCCMA and CSIRO

The ecological niche with the highest potential to sustain malaria vectors in Kenya was classified as area of excellent suitability. BIOCLIM prediction with HADCM3 projection produced the highest area of 18,618 km². Prediction with CSIRO projection by the year 2020 was 3,157km² which was the lowest area of very high suitability. Trend analysis in Figure 4.15 showed that there will be moderate increase of very high suitability from current prediction up to by the year 2020 as predicted with HADCM3 projection, followed by very

drastic increase up to by the year 2050, and further drastic decrease up to the year 2080. Prediction with CCCMA projection registered slow decrease from current climate prediction up to the year 2020, followed by moderate increase up to by the year 2050 and further decrease by the year 2080. The predicted areas with CSIRO projection changed moderately, with a decrease from current climate prediction to by the year 2020, then slight increase by the year 2050 and further slight decrease by the year 2080.

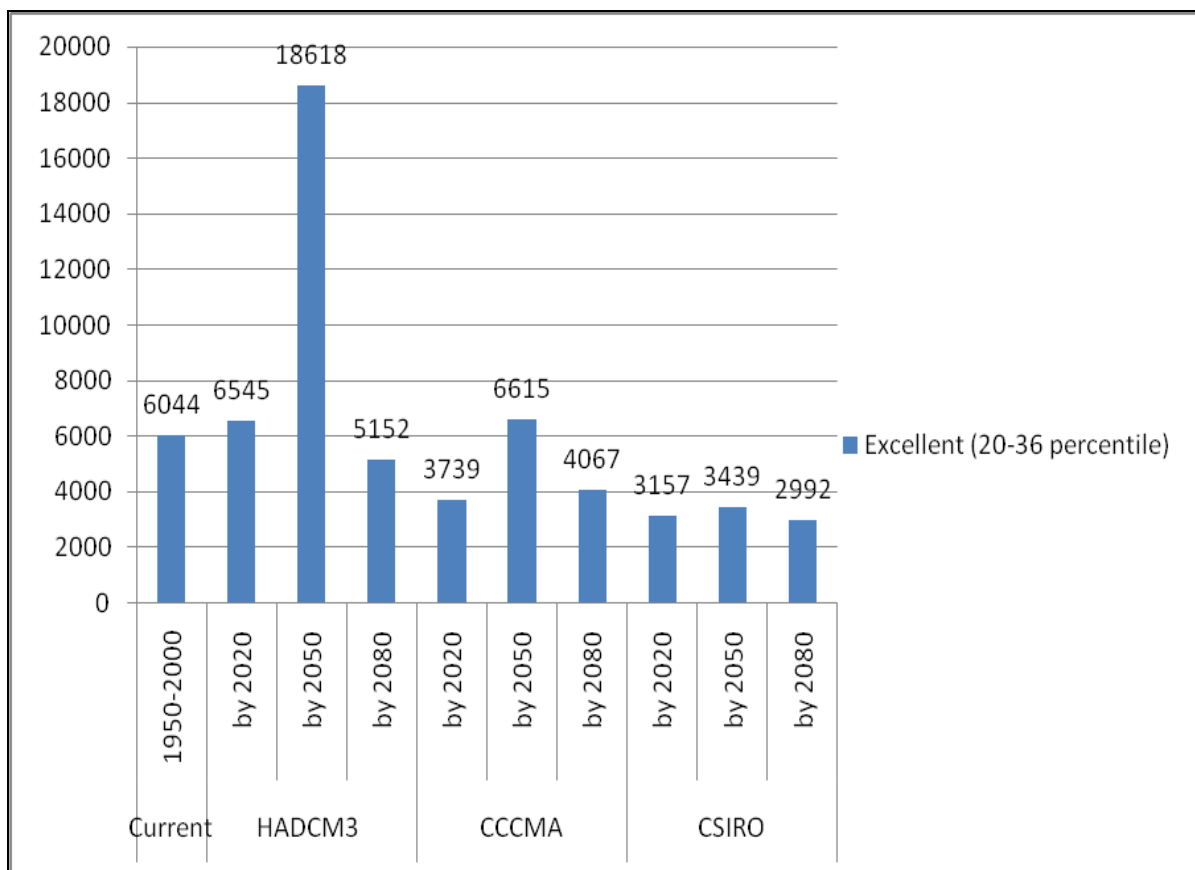


Figure 4. 15: BIOCLIM predicted areas of excellent suitability from current climate, HADCM3, CCCMA and CSIRO

Malaria vector suitability areas for all BIOCLIM predicted ecological niches done with the three IPCC future climate projections were analyzed in bar graphs representing each ecological niche. The colors for each bar corresponded to the colors of the predicted ecological niches in the earlier spatial models and they are shown in the following sections 4.1.1.1, 4.1.1.2 and 4.1.1.3 for HADCM3, CCCMA and CSIRO respectively.

4.2.1.1 Ecological Niche Area Analysis for BIOCLIM Model with HADCM3

BIOCLIM prediction with current climate (1950-2000) resulted in to areas of different suitability as habitats for malaria vectors in Kenya. These ecological niches were analyzed and the total area of non-suitability was found to be 285, 830 km². The total area found to be low in suitability was 273, 055 km². The ecological niche found to harbor medium potential occupied a total area of 63, 917 km² across Kenya. The area of high potential was 38, 939 km² while area of very high suitability was found to be 11, 861 km². The total area that was found to be of excellent suitability was 6, 044 km². Visualization for the ecological niche areas has been done in Figure 4.16 below.

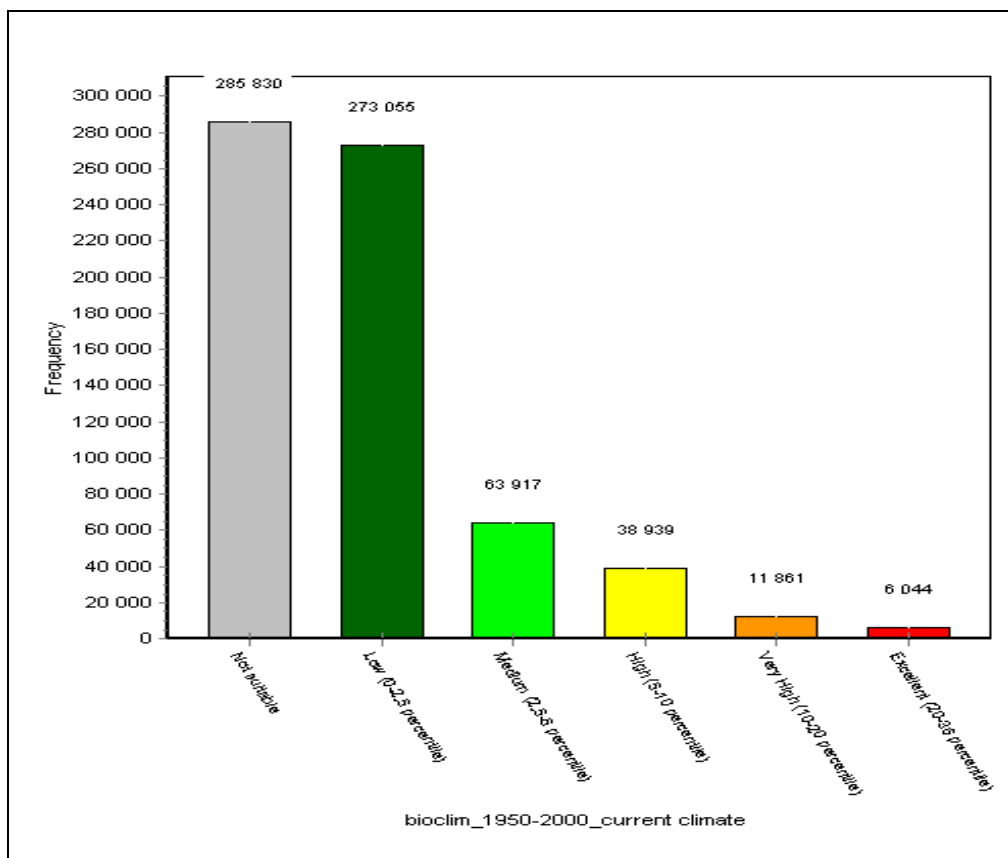


Figure 4. 16: Areas for BIOCLIM predicted Ecologies with current climate

When BIOCLIM prediction was done with climate data from HADCM3 projection by the year 2020, analysis of the ecological niches showed that the total area found to non-suitable as malaria vector habitat will be 237, 492 km². The total area found to be low in suitability was 325, 411km². The ecological niche found to harbor medium potential occupied a total area of 66, 891km² across Kenya. The area of high suitability will be 43, 215 km² while area of very high suitability was predicted to be 13, 167 km². Excellent ecological niche will be 6, 545 km²; slightly more than that in the current climate condition. Comparison of these ecological niche areas have been shown in Figure 4.17 below.

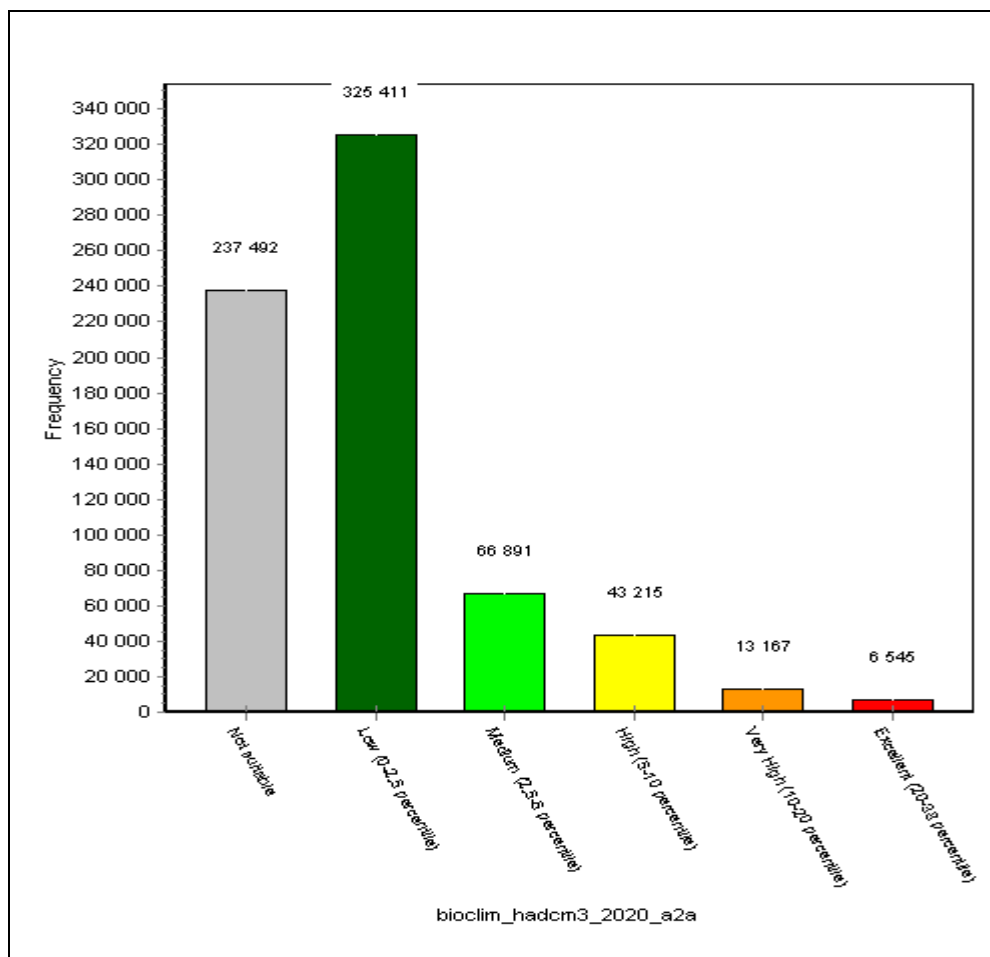


Figure 4. 17: Areas for BIOCLIM predicted Ecologies with HADCM3 by 2020

By the year 2050, the predicted ecological niches will emerge with different areas as follows: not suitable area (250, 867 km²), area of low suitability (210, 485 km²), medium suitability area (97, 990 km²), area of high (58, 728 km²) and area of very high suitability (56, 033 km²). The ecological niche that will be of excellent suitability for malaria vectors was predicted to be 18, 618 km². This area has increased from the previous prediction for by the year 2020. These areas can be visualized in Figure 4.18 below.

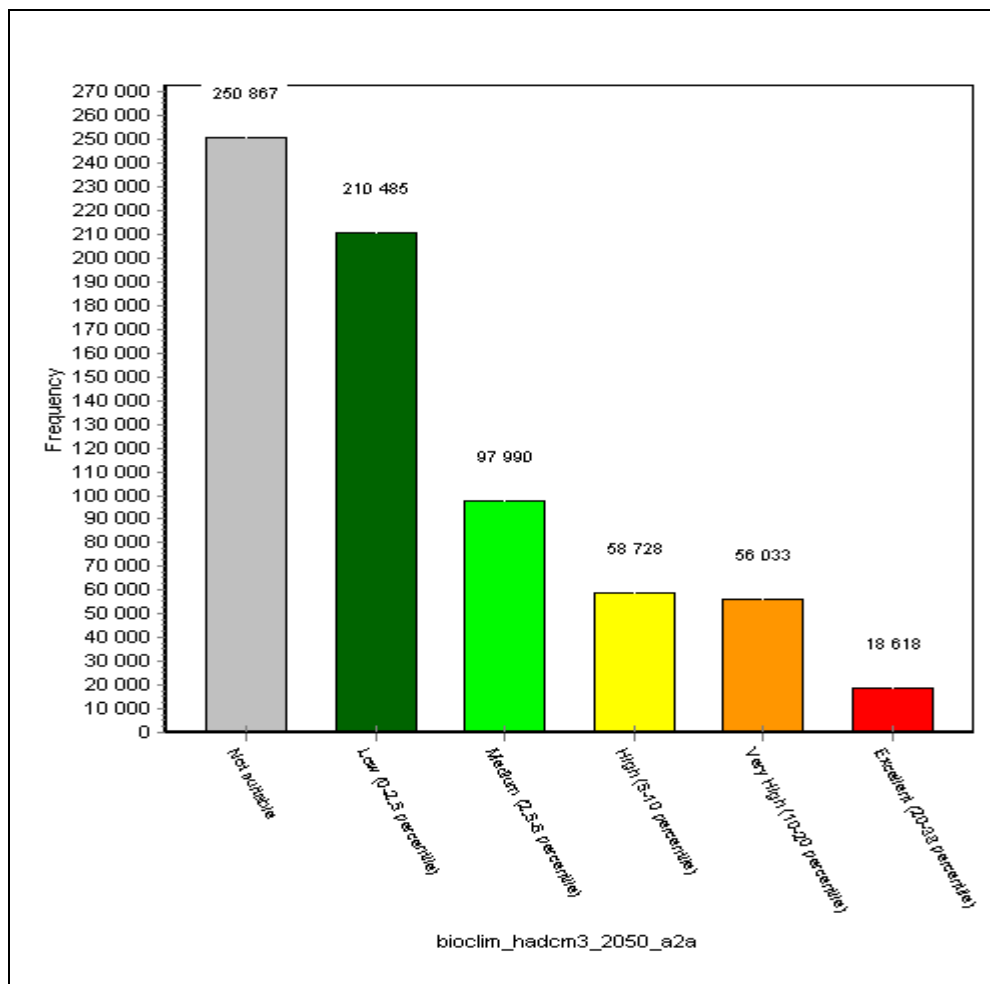


Figure 4. 18: Areas for BIOCLIM predicted Ecologies with HADCM3 by 2050

HADCM3 Future climate by the year 2080 was also used to spatially quantify the effect of climate on malaria vector distribution in Kenya. Different ecologies will emerge and the areas were analyzed as shown in Figure 4.19 below. Area of 195, 802 km² will be unsuitable, area of low suitability will be 387, 293 km², area of medium suitability will be 46, 046 km², area

of high suitability will be 39, 846 km² and area of very high suitability will be 18, 582 km². The area of excellent suitability will drop significantly to 5, 152 km², being the lowest excellent ecological niche predicted using BIOCLIM model with HADCM3 projection.

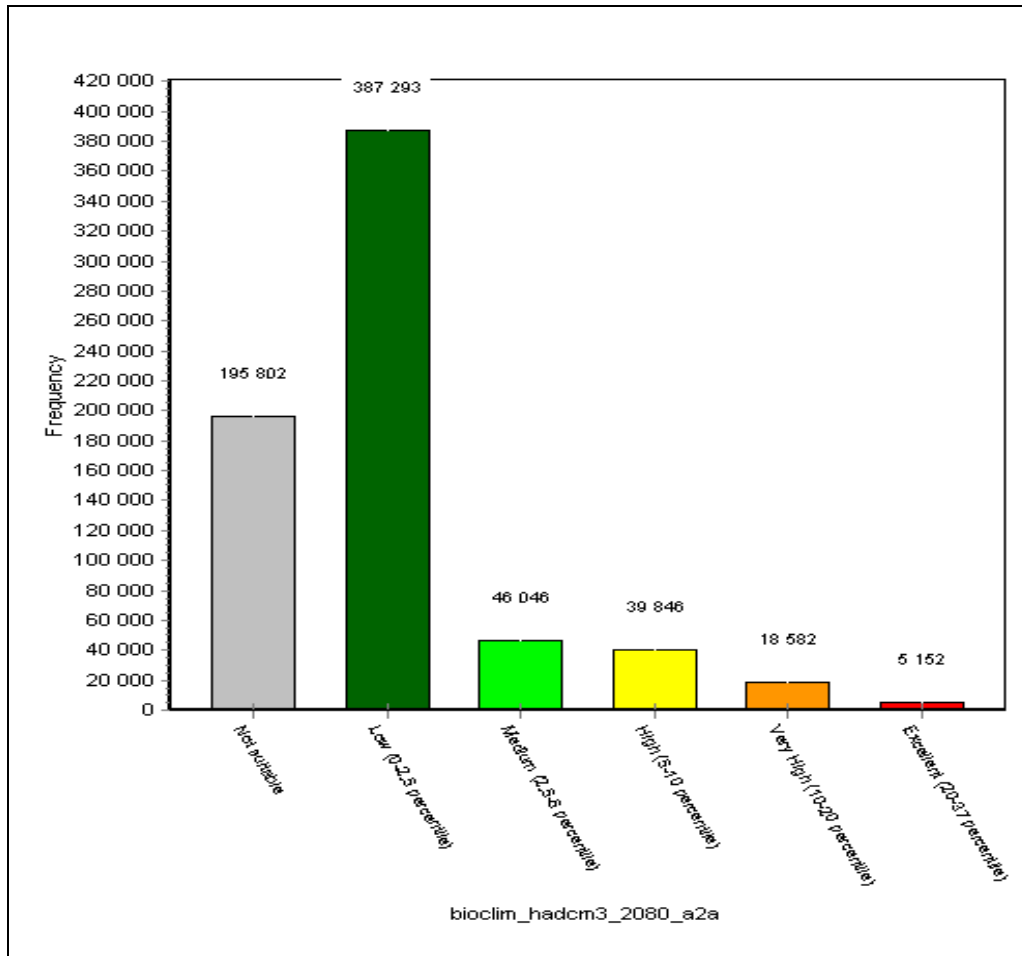


Figure 4. 19: Areas for BIOCLIM predicted Ecologies with HADCM3 by 2080

4.2.1.2 Ecological Niche Area Analysis for BIOCLIM Model with CCCMA

BIOCLIM prediction with CCCMA future projection resulted in to areas of different suitability as habitats for malaria vectors in Kenya. From the analyzed results, the total area of non-suitability will be 250, 060 km². The total area of low suitability will be 323,901 km². The medium potential area will be a total area of 65, 794 km² across Kenya. The area of high potential will be 35, 151 km² while area of very high suitability was found to be 14,076 km².

The total area that was found to be of excellent suitability was 3, 739 km². Visualization for the ecological niche areas has been done in Figure 4.20 below.

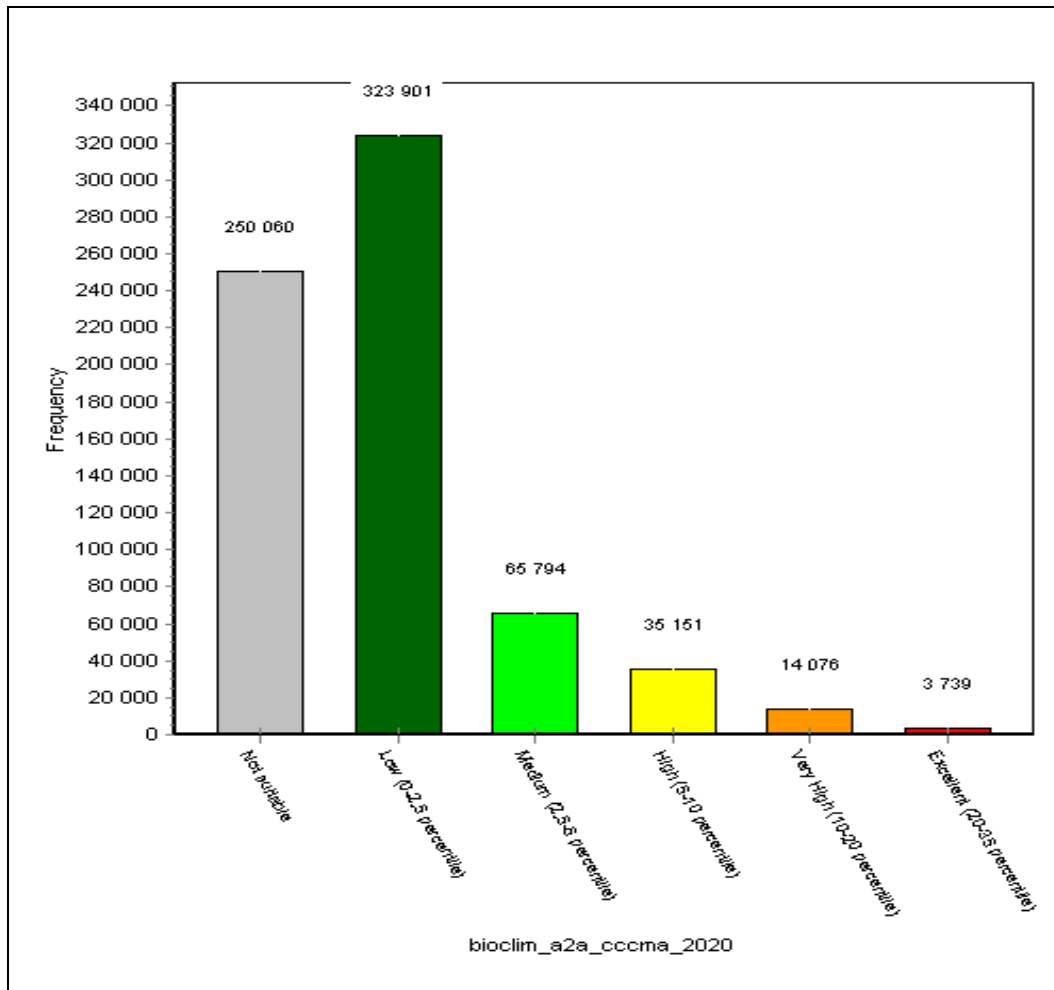


Figure 4. 20: Areas for BIOCLIM predicted Ecologies with CCCMA by 2020

Prediction of Ecological Niches with BIOCLIM model with climate data from CCCMA projection by the year 2050 showed that non-suitable malaria vector habitat areas will be 235, 052 km². The total area found to be low in suitability was predicted to be 355, 460 km². The ecological niche found to harbor medium potential occupied a total area of 51, 655km². The area of high suitability will be 28, 313 km² while area of very high suitability was predicted

to be 15, 626km². Excellent ecological niche will be 6, 615km². Comparison of these ecological niche areas have been shown in Figure 4.21 below.

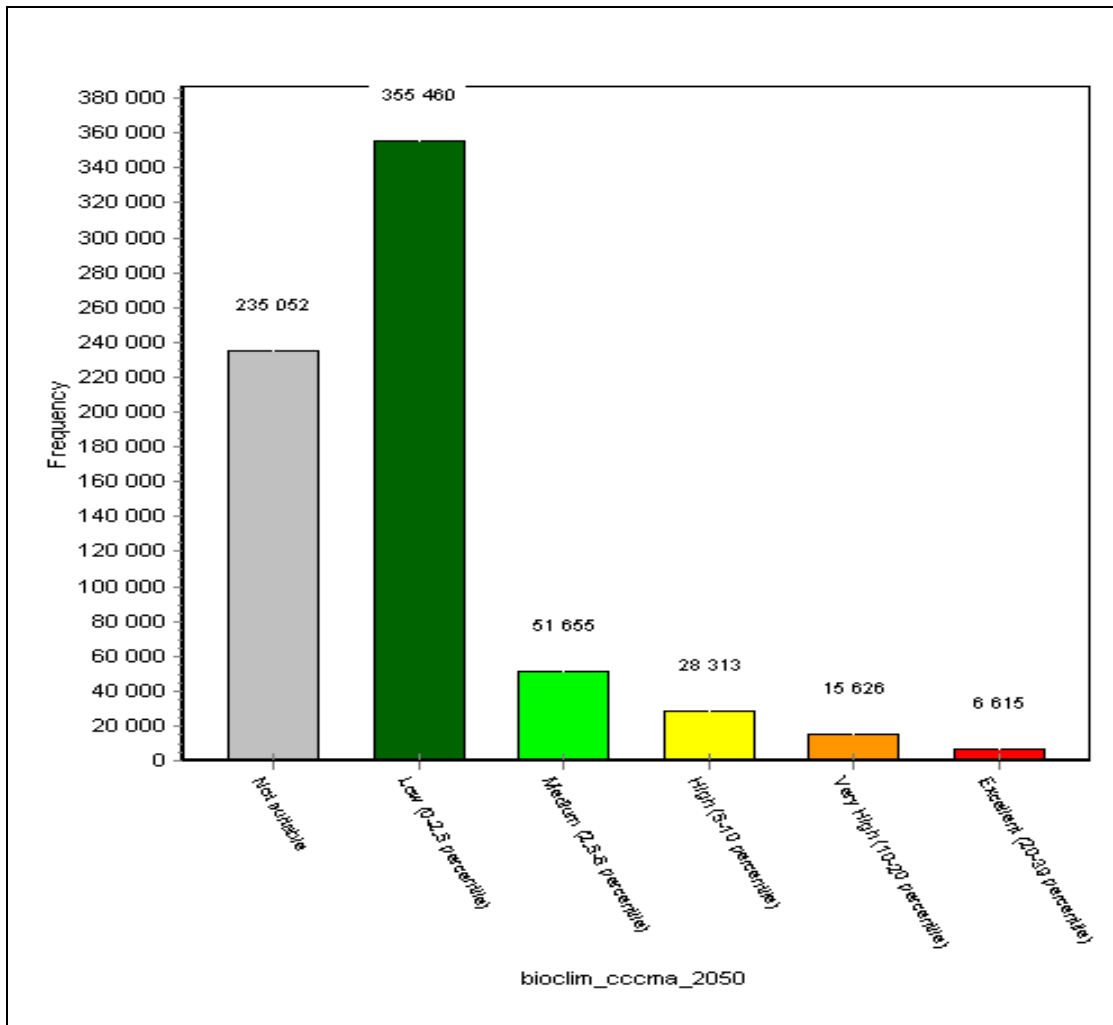


Figure 4. 21: Areas for BIOCLIM predicted Ecologies with CCCMA by 2050

By the year 2080, the BIOCLIM predicted with CCCMA climate projection will emerge with different areas of ecological niches as follows: not suitable area (180, 984 km²), area of low suitability (428, 596 km²), medium suitability area (34, 241 km²), area of high (26, 480 km²) and area of very high suitability (18, 353 km²). The ecological niche that will be of excellent

suitability for malaria vectors was predicted to be 4, 067 km². These areas can be visualized in Figure 4.22 below.

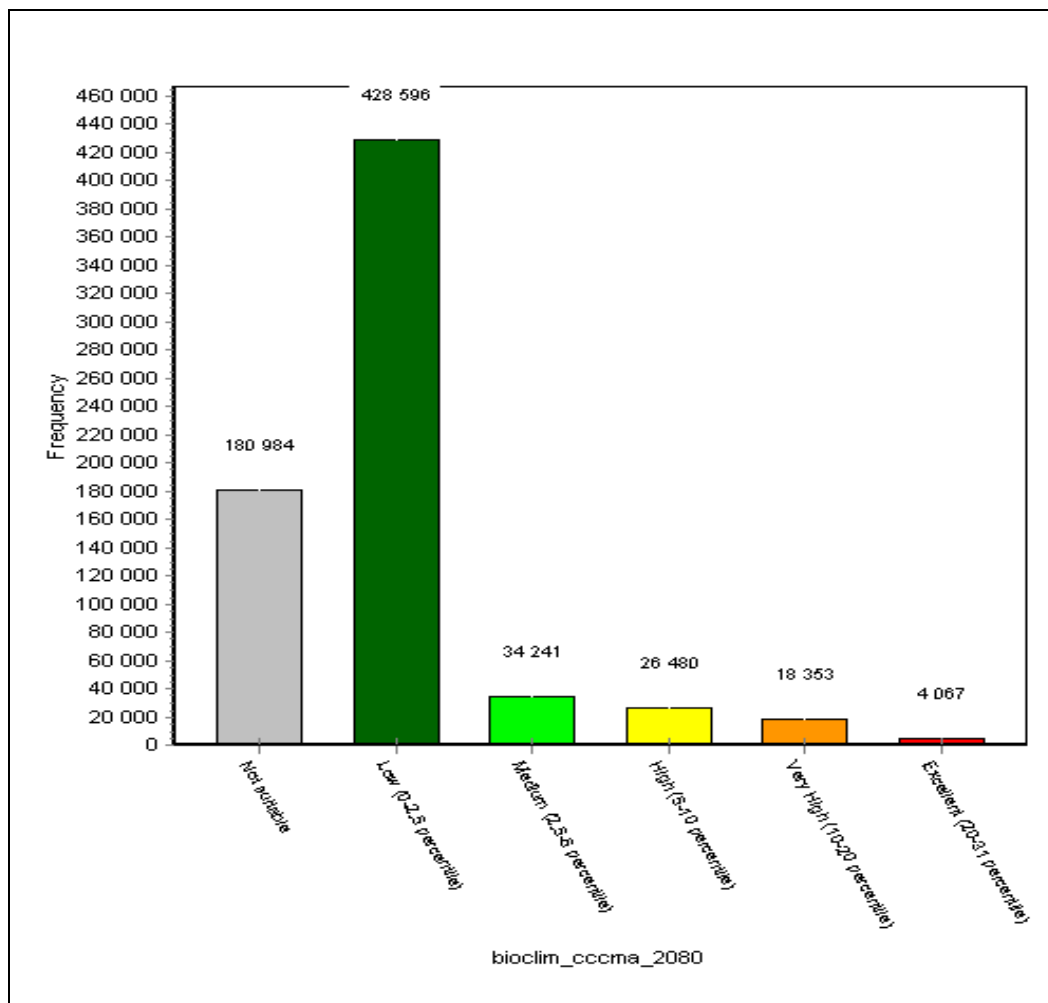


Figure 4. 22: Areas for BIOCLIM predicted Ecologies with CCCMA by 2080

4.2.1.3 Ecological Niche Area Analysis for BIOCLIM Model with CSIRO

When BIOCLIM prediction was done with climate data from CSIRO future projection, the following areas for six different ecological niches were obtained; Not suitable (219, 457 km²), Low suitability (379, 921 km²), Medium suitability area (47, 305 km²), High (27, 219 km²), Very High area (15, 662 km²) and Excellent suitability (3, 157 km²). These ecological zone areas have been shown Figure 4.23 below.

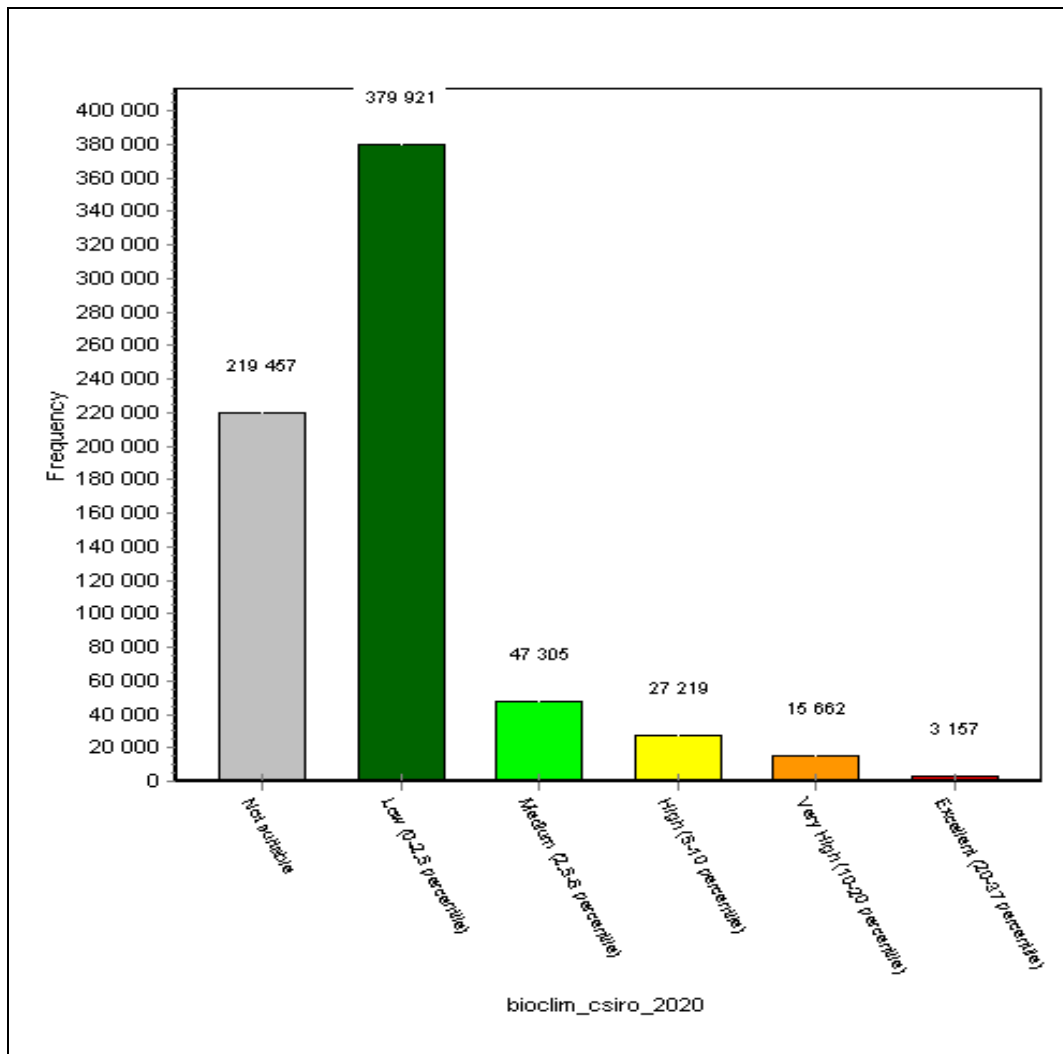


Figure 4. 23: Areas for BIOCLIM predicted Ecologies with CSIRO by 2020

By the year 2050 as projected by IPCC with CSIRO climate model, the BIOCLIM predicted ecological niches will emerge with different areas as follows: not suitable area (235, 060 km²), area of low suitability (353, 833 km²), medium suitability area (54,060 km²), area of high Suitability (31, 539km²) and area of very high suitability (14, 790 km²). The ecological niche that will be of excellent suitability for malaria vectors was predicted to be 3, 439 km². This area has increased from the previous prediction for by the year 2020 which was 3, 157 km². These areas can be visualized in Figure 4.24 below.

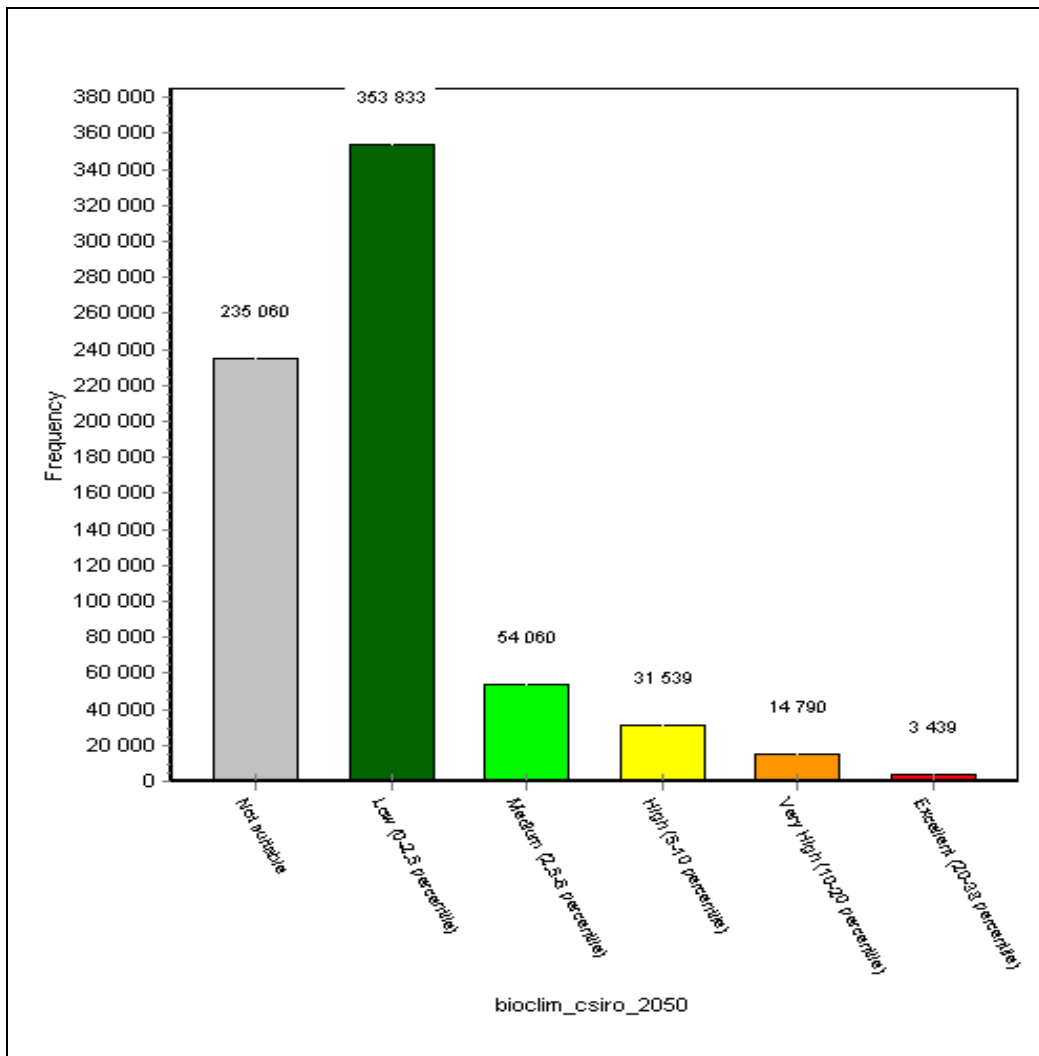


Figure 4. 24: Areas for BIOCLIM predicted Ecologies with CSIRO by 2050

BIOCLIM predicted with CCCMA climate projection by the year 2080 that ecological niches will emerge with different areas of as follows: not suitable area (222, 813 km²), area of low suitability (369,305 km²), medium suitability area (43, 770 km²), area of high (38, 056 km²) and area of very high suitability (15, 785 km²). The ecological niche that will be of excellent suitability for malaria vectors was predicted to be 2, 992 km². These areas can be visualized in Figure 4.25 below. Important to note is fact that this model has predicted the smallest excellent ecological niche with BIOCLIM model when all the three climate projections are considered.

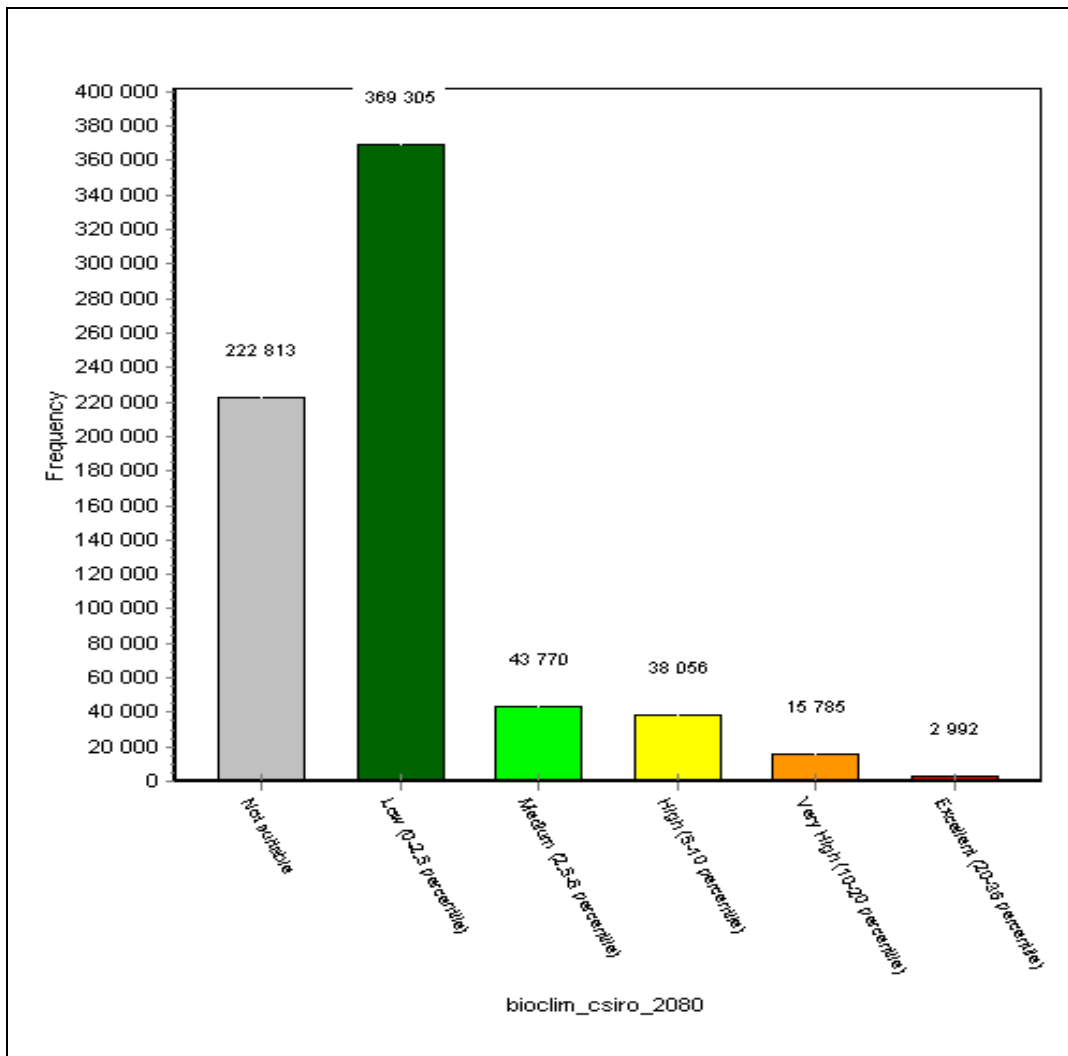


Figure 4. 25: Areas for BIOCLIM predicted Ecologies with CSIRO by 2080

4.2.1.4 Comparison of Ecological Niche Area from BIOCLIM Model with HADCM3, CCCMA and CSIRO climate projections

The areas comparison of the six predicted ecological niches resulting from BIOCLIM ENM with the three IPCC future climate projections have been summarized in Figure 4.26 below. Similar ecological niches are denoted with the same color for each climate scenario. The total area for the six ecological niches from the current climate was found to vary slightly from the total area obtained in each future climate scenario. This can be attributed to the fact that ecological zones overlap with grid duplication thus variation in area summation as observed in the light blue bars.

Interesting to note is the trend in which the viable malaria vector ecologies are spreading from the current climate to projected climate conditions. The non-suitable ecological niche has diminished in all the predictions with IPCC projected climate as compared to the current climate prediction. There will be gain of area in the five suitable ecological niches cumulatively. These observations strongly demonstrated that climate change plays a major role in altering the suitable ecologies for malaria vector to thrive in new areas in Kenya. The generated spatial models should be incorporated in any strategic plans to increase malaria surveillance and combat malaria epidemics.

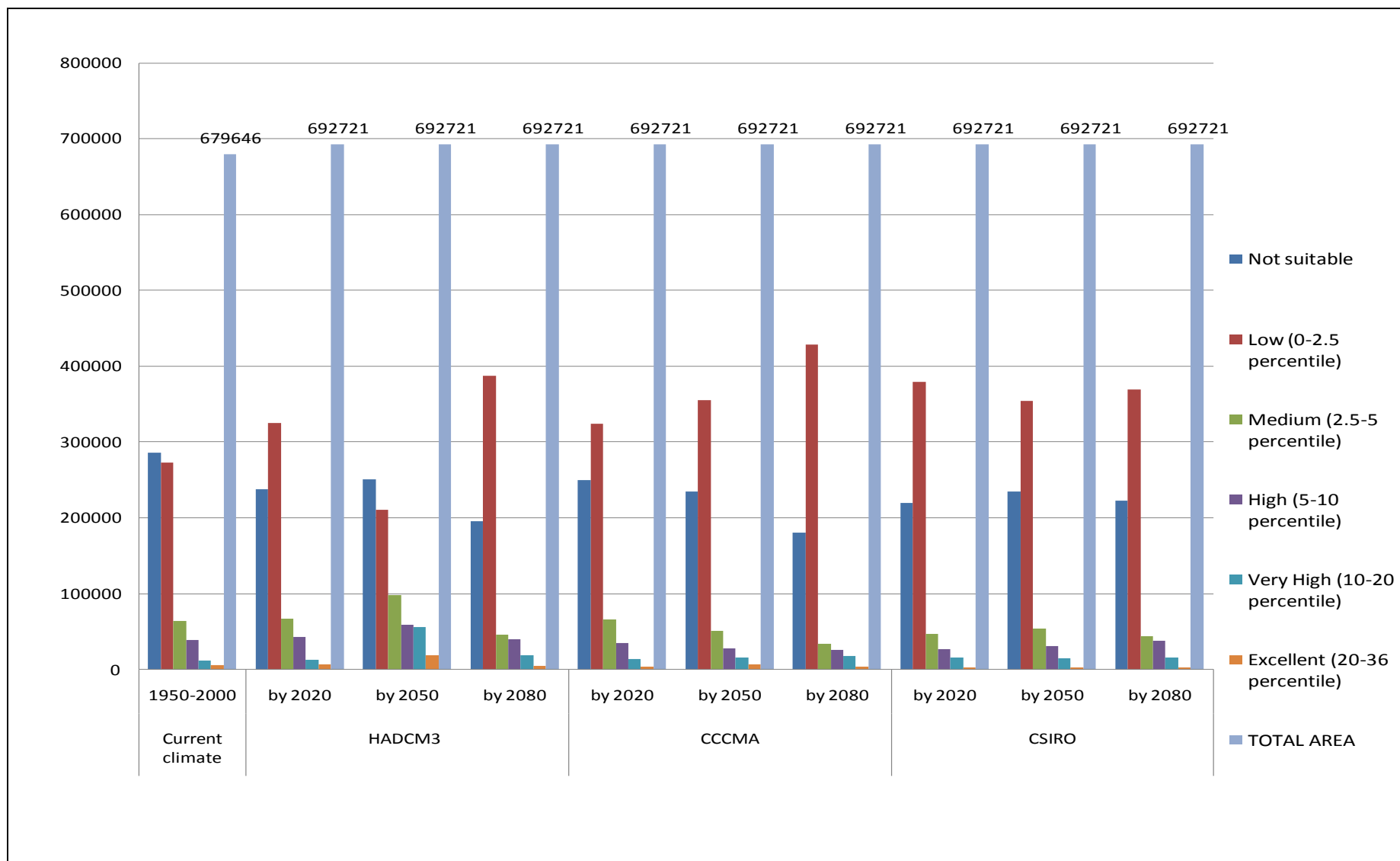


Figure 4. 26: Area Comparison of BIOCLIM predicted Ecological Niches from Current Climate (1950-2000), HADCM3, CCCMA and CSIRO

4.2.2 Analysis and Discussion of BIOCLIM True or False Model Prediction Results

Ecological Niche Modeling using BIOCLIM True or False model resulted into two classes of either suitable or non suitable niches as shown compared in Figure 4.27 below. The highest suitable ecology was 227, 092 km² obtained when prediction was done with HADCM3 future climate by the year 2050. The least suitable ecology was 80, 060 km² which was predicted with CCCMA projection by the year 2050. Ecological niche prediction from HADCM3 and CSIRO showed a similar trend although at different magnitudes. The predicted suitable ecology will increase from by the year 2020 to the year 2050 at higher magnitude in HADCM3 than in CSIRO projection, followed by further decrease of suitable ecology for malaria to survive by the year 2080 in both projections. However, the predicted ecological niche that will be suitable by 2080 will be lower than the area in the year 2020 as predicted with HADCM3 projection while the reverse is true for prediction with CSIRO projection.

CCCMA future climate prediction deviated from the trend observed in the two other climate projections. The area of suitable malaria ecology will decline from by the year 2020 prediction, through the year 2050 up to by the year 2080. Some currently suitable malaria habitants will become unsuitable, as new suitable ecologies emerge. These emerging ecologies will extend from the current niches, but new isolated hot spots will emerge. For instance, prediction from HADCM3 by the year 2050 has shown wide spread of malaria in counties like Narok, Kajiado, Kitui, Makueni, Machakos, Meru, Marsabit, Isiolo, Samburu, Baringo, West Pokot. Turkana county and Mandera among a few others will have some emerging isolated malaria hot spots. ENM prediction with HADCM3 future climate showed that Laikipia County will become unsuitable malaria ecology by the year 2050 and the case remains the same by the year 2080.

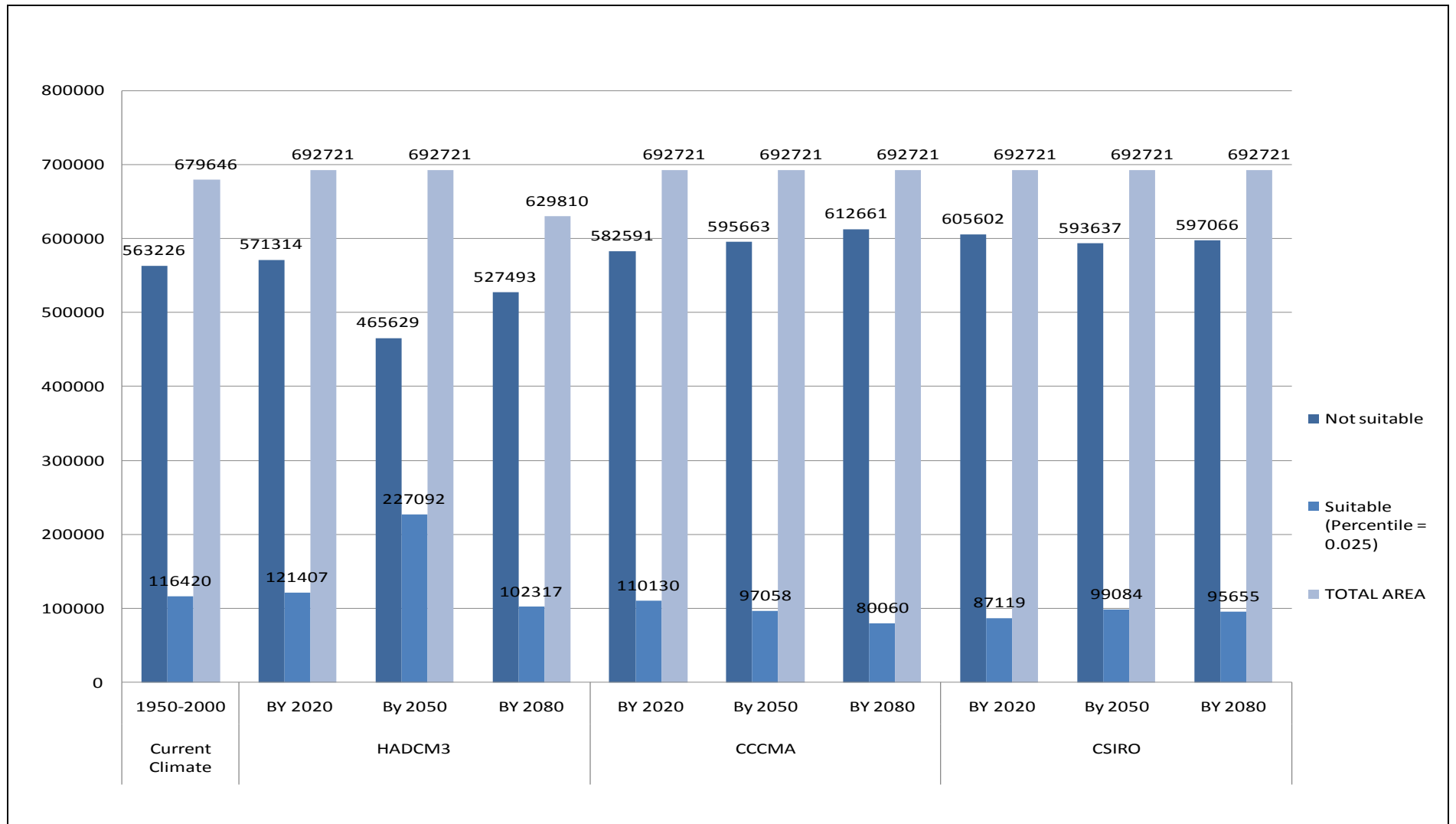


Figure 4. 27: Area Comparison of BIOCLIM True or False predicted Ecological Niches from current Climate (1950-2000), HADCM3, CCCMA and CSIRO

4.2.3 Analysis and Discussion of DOMAIN Model Prediction Results

DOMAIN Ecological Niche Modeling with the three different IPCC projected climate (HADCM3, CCCMA and CSIRO) resulted into six suitability ecological zones for malaria vectors in Kenya. The classification was done based on the computed Gower distances and domain similarity statistics multiplied by 100. The most excellent ecology is the one denoted as 100. Ecological class denoted as 98 - 99 harbors very high suitability while 96-97 has relatively low suitability to malaria vectors. All the classes below 95 are not very suitable as malaria vector habitats. These results are displayed in Figure 4.28 below.

The highest area of excellent suitability was 202, 216 km² resulted from ENM prediction with HADCM3 climate projection by 2050. Prediction with HADCM3 projection by the year 2020 produced the highest ecology of very high suitability (98-99) which covered a total 566, 713 km². The highest ecology classified as high suitability (96-97) was 55, 947 km² yielded by prediction with HADCM3 by the year 2080. None of the future climate projections produced areas of non-suitability and low suitability except that HADCM3 projection by 2050 yielded a total area of 86 km² of low suitability (51-90). The ecology classified as medium suitability (91-95) produced small areas less than 100 km² in all predictions except HADCM3 by 2050 which produced the highest area in this class of 1, 858 km² and CSIRO projections by 2020, 2050 and 2080 with areas of 289 km², 103 km² and 571 km² respectively. Therefore, the model prediction portrays future malaria prevalence in Kenya to be climate driven.

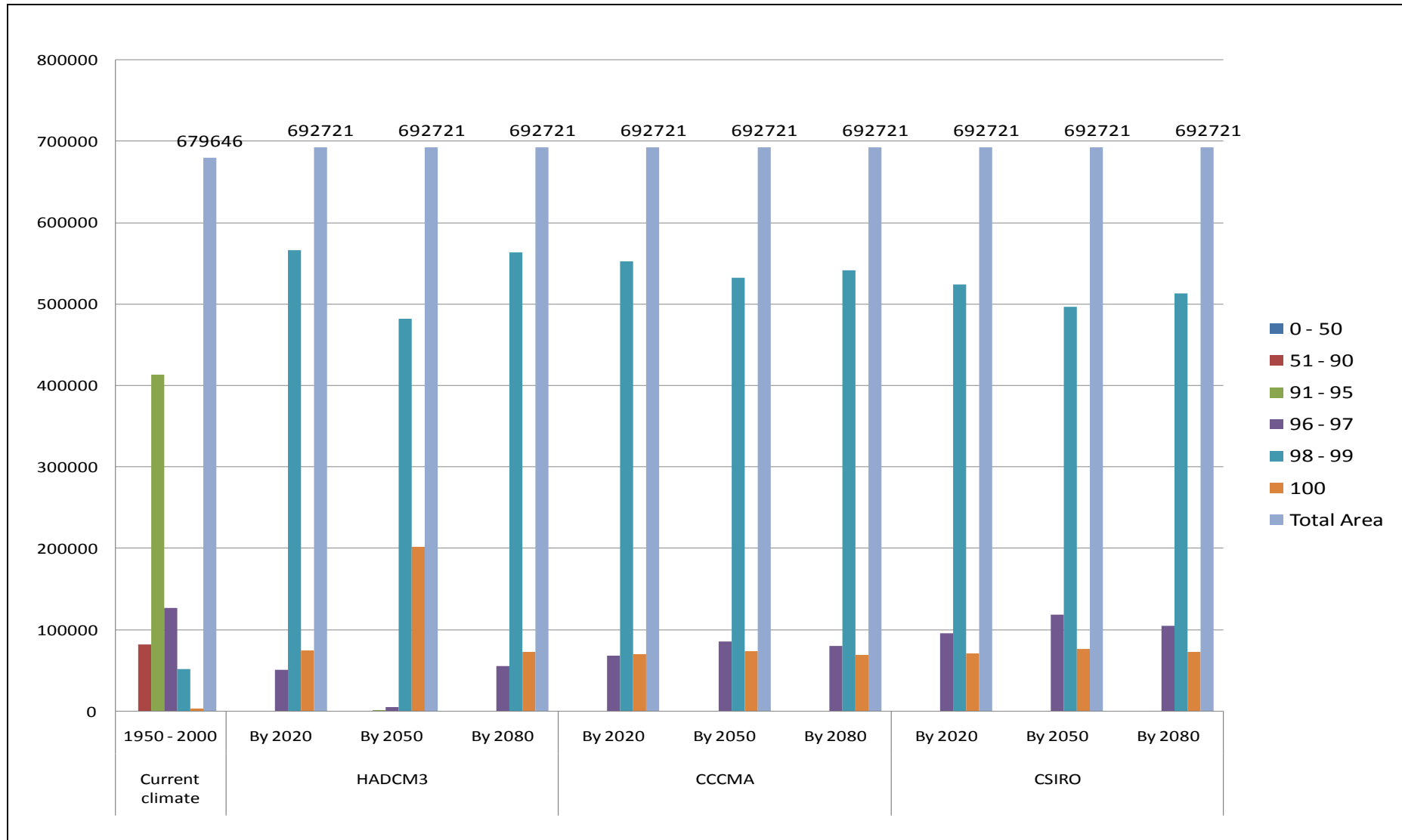


Figure 4.28: Area Comparison of DOMAIN predicted Ecological Niches from current Climate (1950-2000), HADCM3, CCCMA and CSIRO

4.3 Evaluation of Prediction Model Performance

Data was randomly divided into training and testing subsets, with 75% of the records used in model training while 25% of the records were used to test the model. Random absence data was included from the mask since the validation requires an input of presence-absence records. Point values were extracted from the generated bioclimatic grids and Receiver Operating Characteristic (ROC) files from presence – absence distribution of malaria vectors generated for the entire mask. Area under curve (AUC) and Kappa statistic values were computed using the ROC file for each grid. These were the values used to validate the accuracy performance and significance of the ecological niche models.

The malaria vector ecological niches prediction results generated with IPCC model data from the HADCM3, CCCMA and CSIRO future climate projections were validated to assess the model performance of BIOCLIM, BIOCLIM True or False and DOMAIN prediction models. The generated kappa and AUC statistical graphs for model validation outcomes for the current climate and future climate scenarios A2a_2020, A2a_2050 and A2a_2080 are shown in Appendix B. For all the three cases, BIOCLIM and BIOCLIM True or False validation graphs are shown combined in one figure while the results for DOMAIN model are shown in a separate figure. The validation outcome showed that all the models predictions were acceptable as they all met the set conditions for either good or excellent values given in the AUC interpretations (Table 3.7). The kappa statistics for all the predictions were acceptable when interpreted according to the summary in Table 3.6. Further discussions of the resulting kappa and AUC from the validations were done on section 4.4 below.

4.4 Discussion and Comparison of model validation outcomes

The prediction models were all validated using the kappa statistics and Receiver Operator Characteristics (ROC) area under curve (AUC). Interpretation of the acquired model performance values was done for kappa statistics and AUC values. The results from all the IPCC projection models were summarized for comparison. Visualization and comparison of the model performance for all the prediction models was achieved through generation of bar graphs representing the AUC and kappa values.

From the compared results of model validations, DOMAIN model performance with each future climate projection was found to be excellent as the Kappa and AUC values were the highest when compared to those obtained in both BIOCLIM and BIOCLIM True/False. BIOCLIM True/False model performed worse amongst the three models. For instance, the highest AUC value was 0.954 achieved from DOMAIN model with CCCMA projection by 2020, while the lowest AUC value was 0.714 achieved from BIOCLIM True or False Model with HADCM3 projection by 2020 (Figure 4.29a). The highest Kappa value was 0.909 resulting from DOMAIN model with CCCMA projection by 2020 while the lowest kappa value was 0.427 obtained from BIOCLIM True or False model with HADCM3 projection by 2020 (Figure 4.29b)

However, the overall prediction performance for all the models was found to be within acceptable range as per the interpretation guidelines provided by Kappa and AUC statistics in Tables 3.6 and 3.7 respectively. All the models had errors in prediction as none of them had kappa =1 or AUC=1. Therefore, the prediction error for each model can be obtained by subtraction the obtained kappa and AUC validation value from 1. Figure 4.30 compares both kappa and AUC for all the prediction models under the three IPCC climate projections.

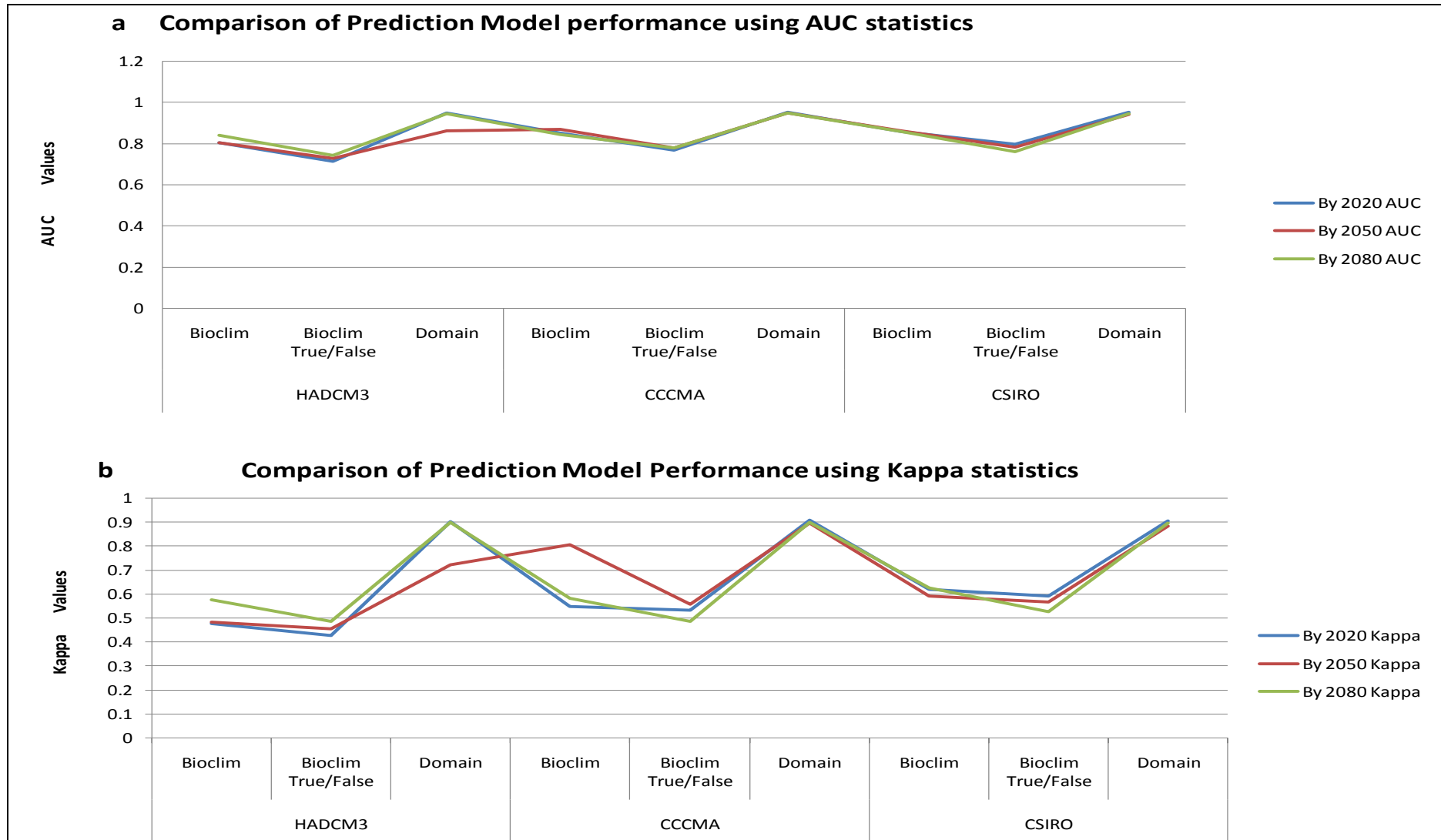


Figure 4. 29: Comparison of prediction model performance using (a) AUC and (b) Kappa statistics

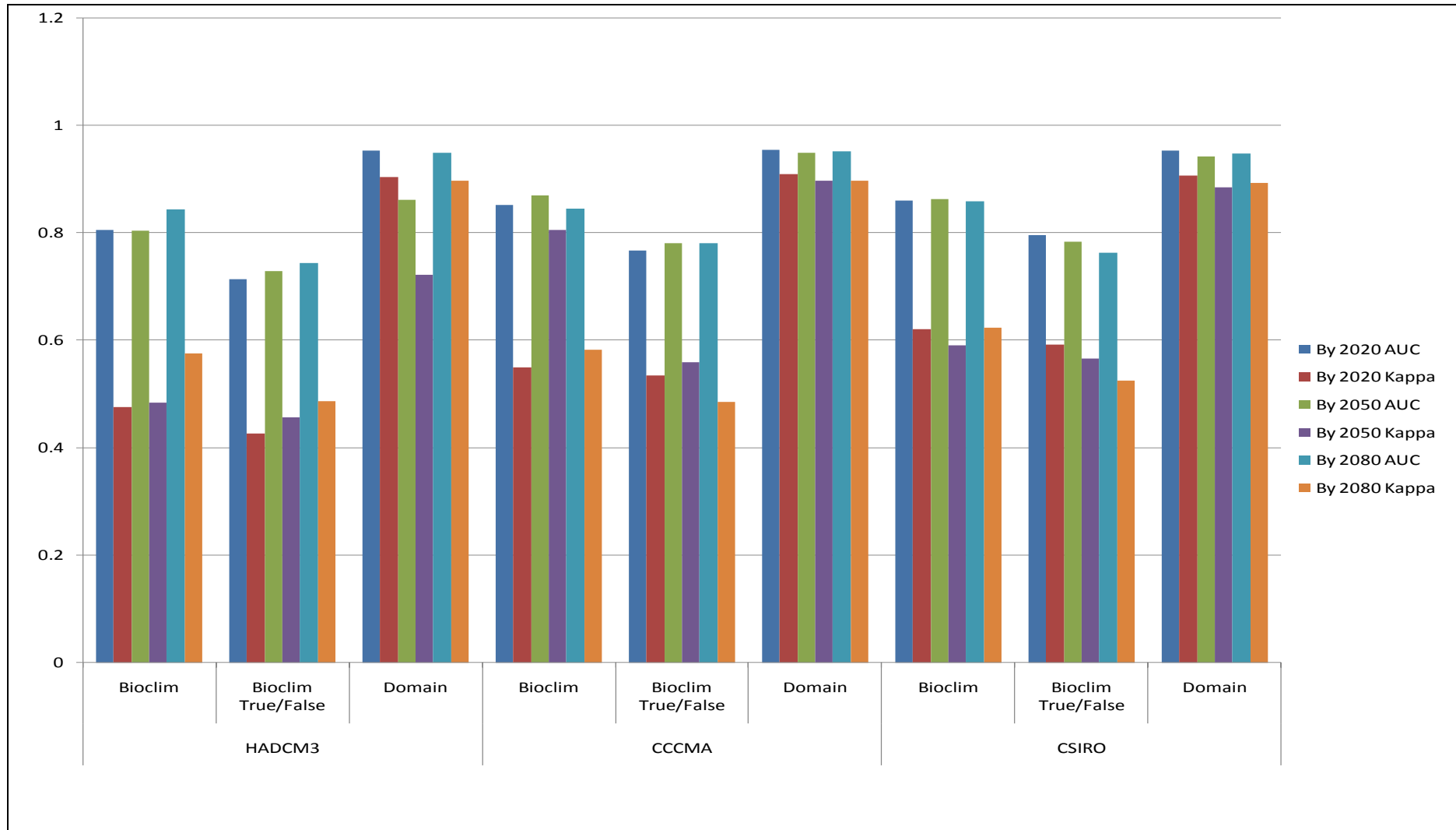


Figure 4. 30: Model Prediction Validation Comparison for BIOCLIM, BIOCLIM True or False and DOMAIN with HADCM3, CCCMA and CSIRO future climate projections

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

This research was set to correlate the relationship between climate change and the distribution of the main malaria vectors in Kenya. Ecological Niche Modeling was applied to spatially establish the effect of climate as an explanatory variable in main malaria vector distribution, hence malaria prevalence in new hotspots. Predictive modeling was done to investigate the spatial-temporal vector distribution under different IPCC future climate projections. The period from 1950 – 2000 was treated as the current climate scenario to explain where malaria vectors are existing as per vector spatial presence data and the prediction of the ecological niches where the vectors would also be found to thrive.

BIOCLIM and DOMAIN models were applied to determine the appropriate prevalence ecological niches. Future prediction for the prevalence and distribution of the malaria vectors was done with model data from HADCM3, CCCMA and SCIRO IPCC future climatic projections under A2a scenario in Kenya. Spatial models were generated for the current climate and projections done to depict how the vectors will be distributed by the years 2020, 2050 and 2080. Validation for the model performance was achieved through analyses of the Receiver Operating Characteristic (ROC) Area under Curve (AUC) and Kappa statistics.

5.1 Conclusions

Ecological Niche Modeling (ENM) was done using IPCC projections for HADCM3, CCCMA and CSIRO models. Climate change has been shown to be a major driving force in the spread of suitable malaria vector ecologies. Qualitative investigation has shown that by use of data from HADCM3, CCCMA and SCIRO models of IPCC projected future climate under the A2a scenario, there is correlation between climate change and malaria vector

spread to new regions. Excellent ecologies for malaria vector to thrive in new environment will expand to continuous spatial areas, with some isolated new hotspots. This was demonstrated through visualization of the generated spatial-temporal models for malaria vector distribution.

The predictions showed that by the year 2020, the suitability areas for malaria vectors in Kenya will start to change from the current ecological suitability. Most areas where the malaria vectors are thriving currently will still remain suitable ecologies. New suitability zones will emerge ranging from low to very high suitability as shown by the predictions. By the year 2050, areas of suitability will expand at an alarming extent. The year 2080 has been predicted to show that the suitable ecologies will start to revert to the original areas of suitability as in the current climate. Therefore, climate change in Kenya will adversely affect the environment at an alarming rate by 2050, but beyond that there will be a level of stabilization, where further change will trigger reversal to the past climate.

The emerging new suitable areas can be clearly distinguished from the BIOCLIM True or False predictions. BIOCLIM True or False prediction with CCCMA future climate prediction showed that the area of suitable malaria ecology will decline from by the year 2020 prediction, through the year 2050 up to by the year 2080. Some currently suitable malaria habitats will become unsuitable, as new suitable ecologies emerge. These emerging ecologies will extend from the current niches, but new isolated hot spots will emerge. For instance, prediction from HADCM3 by the year 2050 has shown wide spread of malaria in counties like Narok, Kajiado, Kitui, Makueni, Machakos, Meru, Marsabit, Isiolo, Samburu, Baringo, West Pokot. Turkana county and Mandera among a few others will have some emerging isolated malaria hot spots. Prediction with HADCM3 future climate showed that

Laikipia County will become unsuitable malaria ecology by the year 2050 and the case remains the same by the year 2080.

Validation accuracies for the three models were graphically displayed. The validation outcomes were tabulated and compared in bar graphs. When the DOMAIN model was applied, the highest validation values were achieved as compared to the BIOCLIM Model. Kappa statistics and AUC values were generated for the three projections in order to investigate the predictive accuracy in determining the malaria vector prevalence with different IPCC projections. The highest kappa ($k = 0.909$) and Area Under ROC Curve (AUC=0.954) values were achieved from DOMAIN model with CCCMA projection by the year 2020. The lowest model performance values of $k = 0.427$ and AUC = 0.714 were obtained from BIOCLIM True or False model with HADCM3 projections by the year 2020. Validation results for prediction model performance showed that all the models used had errors in prediction as none of them had kappa =1 or AUC=1. However, the overall prediction performance for all the models was found to be within acceptable range as per the interpretation guidelines provided by Kappa and AUC statistics.

Intervention measures for malaria catastrophes in Kenya should be well planned to capture the future suitable ecologies in all the newly predicted regions, rather than the witnessed overemphasis in the current epidemic zones. Adaptation strategies should be well formulated and communicated to the unsuspecting populations residing in the currently malaria free zones found to harbor the potential for future epidemics. Climate change has been changing at an alarming rate and the only way to safeguard the population is by creating awareness for preparedness and early risk management. The recent decision by the health sector to collect actual laboratory tested data on malaria incidences in hospitals, coupled with the ecological

niche predictions, will suffice for strategic adaptation for early malaria epidemic risk reduction in Kenya.

The following conclusions were drawn from the Ecological Niche Modeling done using BIOCLIM, BIOCLIM True or False and DOMAIM prediction models: There is correlation between climate change as an explanatory variable and the distribution of main malaria vectors in Kenya. The spatial-temporal distribution of the main malaria vectors in Kenya varies under different IPCC future climate projections which are HADCM3, CCCMA and CSIRO. The future ecological niches for malaria vector occurrence in Kenya will extend from the current niches in most endemic areas, new hotspots will emerge and some suitable ecology will become unsuitable, resulting in varying areas from current climate predictions to projections by the year 2020, 2050 and 2080 under IPCC A2a scenario. Intervention strategies such as indoor or outdoor residual spraying, distribution of insecticide-treated mosquito nets (ITNs) and long-lasting insecticide-treated nets (LLINs) should be diversified in new emerging areas for disaster risk reduction and increase adaptive capacity and resilience among local communities.

Therefore, this research made a breakthrough in establishing the relationship between climate change and the occurrence of main malaria vectors in Kenya. Suitable Climate Envelope Models (CEMs) that use the current geographic distribution of species to infer its environmental requirements and to predict malaria vector species' geographic distribution for the current, or for future climates were investigated and validated. Thus, Ecological Niche Modeling (ENM) has been applied to correlate climate change and the emerging hotspots of malaria vector geographical distribution in Kenya. Predictive modeling showed that the future climate projections under different scenarios will impact malaria vector distribution

negatively, thus the prevalence of malaria in areas which are not currently threatened by the epidemic.

5.2 Recommendations

The accuracy of predicted malaria vector ecological niches, hence malaria prevalence is dependent on the accuracy of projected climate data. Therefore, the author of this research would recommend that local climate data should be collected at finer resolution, so as to localize the trend of malaria vector distribution. The results should be validated with laboratory tested actual county malaria incidence data for a period not less than thirty years. Such data is not currently available as the recorded hospital data on malaria cases in the past categorized most fever incidences as malaria, before the recent undertaking by the health sector to perform clinical laboratory test for any suspected malaria patient.

The predictions in this research employed correlative Ecological Niche Modeling rather than mechanistic approach. The author recommends further research that can adopt mechanistic approach to model the suitable ecologies for malaria vector distribution. The physiologically limiting mechanisms in malaria vectors' tolerance to environmental conditions should be incorporated in mechanistic models to predict the future geographic space. Results from the two different methods can be compared.

The climate change data used in this research has been projected by IPCC by use of different scenarios that are estimates of future dynamics. However, there are uncertainties as many processes are not fully described in the mathematical models, such as turbulence in the atmosphere and ocean, precipitation growth in clouds, cumulus convection, radiation transfer in and around clouds, and CO₂ transfer processes in heterogeneous biosphere canopies. IPCC

should consider modeling climate with simple one-dimensional or volume-integrated models of the atmosphere with prescribed ocean, cryosphere, biosphere, and land surface specifications are necessary. Further, interactive studies of the ocean processes using simple diffusion, radiation and convection, and energy balance relationships are important. This approach can increase plausibility to obtain qualitative simulations of climate change which can be used modeling ecological niches.

5.3 Areas for Further Research

Further to the research findings elaborated in this study, there is need to carry on the following investigations:

- If climate change dictates the distribution of malaria vectors in Kenya, then surveillance studies should be done to investigate the climate change impacts at county level. Analyses techniques that can provide differencing to extract models showing the new hotspots should be explored. This would be important as the boundary of ecological niches does not follow any stipulated county boundaries.
- From the generated 19 bioclimatic variables used in Ecological Niche Modeling, what combination would impact the distribution most significantly? This will assist in determining the climatic variables that can be most attributed to affect the distribution of malaria vectors in Kenya.
- Apart from BIOCLIM and DOMAIN models, other integrated techniques that can model the correlation between climate change and vector distribution should be investigated. The prediction performance for these techniques should be compared to what has been achieved in this research.

- Modeling should be done with the results of Representative Concentration Pathways (RCP) from the Fifth Assessment Report of the IPCC with an aim to review the findings in this research.

REFERENCES

- Afrane, Y.A., Lawson, B.W., Githeko, A.K. and Yan, G., 2005. Effects of micro-climatic caused by land use and land cover on duration of gonotrophic cycles of *Anopheles gambiae* (Diptera: Culicidae) in western Kenya highlands. *Medical Journal of Entomology*, (42), pp. 974-980.
- Alonso, D., Bouma, M.J. and Pascual, M., 2011. Epidemic malaria and warmer temperatures in recent decades in an East African highland. *Proc Biol Sci*, 278(1712), pp. 1661-9.
- Altman D.G., 1991. Practical statistics for medical research. London: Chapman and Hall.
- Benson Owour Ochieng, 2009. Effective communication of science and climate change information to policy makers, report of the ccaa/3rd EAHSC Symposium held during the 3rd East Africa health and scientific conference on climate change, environment and health. Kenyatta International Conference Centre (KICC), Nairobi, 27th March, 2009.
- Berkhout, F., 2005. Rationales for adaptation in EU climate change policies. Climate Policy, special issue: climate policy options Post-2012: European Strategy, technology and adaptation after Kyoto. Retrieved from http://www.hm-treasury.gov.uk/media/8AA/9C/stern_review_supporting_technical_material_frans_berkhout_231006.pdf
- Boko, M., Niang, I., Nyong, A., Vogel, C., Githeko, A., Medany, M., Osman-Elasha, B., Tabo R. and Yanda, P., 2007: Africa. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., *Cambridge University Press, Cambridge UK*, pp. 433-467.

- Bouwer, L., and Aerts, J., 2006. Financing climate change adaptation. *Diasters*, (30), pp. 49-63.
- Brad Lobitz, Louisa Beck, Anwar Huq, Byron Wood, George Fuchs, A. S. G. Faruque, and Rita Colwell, 2000. Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *PNA*, 97(4), pp. 1438 – 1443.
- Brohan, P., et al., 2006. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.*, (111), D12106, doi:10.1029/2005JD006548.
- Chaves, L.F., Satake, A., Hashizume, M. and Minakawa, N., 2012. Indian Ocean Dipole and Rainfall Drive a Moran Effect in East Africa Malaria Transmission. *Journal of Infectious Diseases*, 205(12), 1885-1891.
- Chen, H., Githeko, A.K., Zhou, G., Githure, J.I. and Yan, G., 2006. New records of *Anopheles arabiensis* breeding on the Mount Kenya highlands indicate indigenous malaria transmission. *Malaria Journal*, (5), pp. 17.
- Cohen, J., 1960. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, (20), pp. 37–46.
- Confalonieri, U., Menne, B., Akhtar, R., Ebi, K.L., Hauengue, M., Kovats, R.S., Revich, B. and Woodward, A., IPCC, 2007. Human health. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, *Cambridge, UK: Eds., Cambridge University Press*. pp. 391-431.
- Connor S, Thomson M, Molyneux D., 1999. Forecasting and prevention of epidemic malaria: new perspectives on an old problem. *Parassitologia*, (41), pp. 439–448.

- Douville, H., et al., 2002. Sensitivity of the hydrological cycle to increasing amounts of greenhouse gases and aerosols. *Clim. Dyn.*, (20), pp. 45–68.
- Douville, H., D. Salas-Méla, and S. Tyteca, 2005. On the tropical origin of uncertainties in the global land precipitation response to global warming. *Clim. Dyn.*, (26), pp. 367–385.
- EPA, 2007. Retrieved from HYPERLINK "<http://www.epa.gov/nonco2/econ-inv/table.html>" <http://www.epa.gov/nonco2/econ-inv/table.html> (accessed 29, May2012).
- Fielding, A.H. & Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, (24), pp. 38–49.
- Francis Harvey, 2008. A premier of fundamental Geographic and Cartographic concepts. *The Guilford Press – A Division of Guilford Publications, Inc. 72 Spring Street, New York, NY 10012.*
- Geary, R.C.; 1954: The contiguity ratio and statistical mapping. *Incorporated Statistician*, 5, 115-141.
- Gill C.A.; 1923: The prediction of malaria epidemics: with special reference to an actual forecast in 1921. *Indian J Med Res*, (10), pp. 1136–1143.
- Githeko, A.K., Lindsay, S.W., Confalonieri, U.E. and Patz, J.A., 2000. Climate change and vector-borne diseases: a regional analysis. *Bull World Health Organ*, 78(9), pp.1136-1147.
- Githeko, A.K. and Ndegwa, W., 2001. Predicting malaria epidemics in the Kenyan Highlands using climate data: a tool for decision makers. *Global Change Human Health*, (2), pp.54 - 63.

- Githeko, A.K., Ogallo, L., Lemnge, M., Okia, M. and Ototo, E.N., 2014. Development and validation of climate and ecosystem-based early malaria epidemic prediction models in East Africa. *Malaria Journal* 2014, **13**:329 doi:10.1186/1475-2875-13-329.
- Gómez-Mendoza, L. and Arriaga, L., 2007. Modeling the Effect of Climate Change on the Distribution of Oak and Pine Species of Mexico. *Conservation Biology*, (21), pp.1545-1555. (Retrieved May 23, 2011).
- Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994. Description of the fifth generation Penn State/NCAR Mesoscale Model (MM5), *Tech. Rep. TN-398+STR, NCAR, Boulder, Colorado*, pp. 121.
- Grell, G., 1993. Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, (121), pp. 764–787.
- Hamann, A. and Wang, T., 2006. Potential Effects of Climate Change on Ecosystem and Tree Species Distribution in British Columbia. *Ecology*, (87), pp.2773-2786.
- Hasselmann, K., 1976. Stochastic Climate Models. *Tellus*, *XXVIII*
- Hay, S.I., Cox, J., Rodgers, D.J., Randolph, S.E., Stem, D.I., Shanks, G.D., Myers, M.F., and Snow, R.W., 2002a. Climate change and the resurgence of malaria in the East African Highland. *Nature*, (415), pp. 905-909.
- Hay, S.I., Cox, J., Rodgers, D.J., Randolph, S.E., Stem, D.I., Shanks, G.D., Myers, M.F., and Snow, R.W., 2002b. Hot topic or hot air? Climate change and the resurgence of malaria in the East African Highland. *Trends in Parasitology*, (18), pp. 530-534.
- Hijmans et al., 2005. Very high resolution interpolated climate surfaces for global land areas. Published online in *Wiley InterScience* (www.interscience.wiley.com). DOI: 10.1002/joc.1276. *Int. J. Climatol.* 25: 1965–1978.

- Hughes, L., Cawsey, E. M. and Westoby, M., 1996. Climatic Range Sizes of Eucalyptus Species in Relation to Future Climate Change. *Global Ecology and Biogeography Letters* (5), pp. 23-29. doi: 10.2307/2997467.
- Hughes, L., Cawsey, E.M. and Westoby M., 1996. Climatic Range Sizes of Eucalyptus Species in Relation to Future Climate Change. *Global Ecology and Biogeography Letters*, (5), pp.23-29. doi: 10.2307/2997467.
- IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.
- IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.
- IPCC, 2007: Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, *Eds.*, Cambridge University Press, Cambridge, UK, 976pp.
- Iverson, L. R., and Prasad, A. M., 2001. Potential Changes in Tree Species Richness and Forest Community Types Following Climate Change. *Ecosystems*, (4), pp. 186-199. Retrieved May 23, 2011.
- Jones, R.G., Noguera, M., Hassell, D.C., Hudson, D., Wilson, S.S., Jenkins, G.J. and Mitchell, J.F.B., 2004. Generating high resolution climate change scenarios using PRECIS, *Met Office Hadley Centre, Exeter, UK*, 40pp, (Accessed April 2004).

- Kelly-Hope, L.A., Hemingway, J. and McKenzie, F.E., 2009. Environmental factors associated with the malaria vectors *Anopheles gambiae* and *Anopheles funestus* in Kenya. *Malar J*, 8(268), pp. 1-8.
- Kenya National Bureau of Statistics (KNBS), 2010. 2009 Kenya Population and Housing Census, Vol. 1A & 1C. Nairobi: KNBS.
- Koenraadt C.J.M., Githeko A.K., Takken W., 2004. The effects of rainfall and evapotranspiration on the temporal dynamics of *Anopheles gambiae* s.s. and *Anopheles arabiensis* in a Kenyan village. *Acta Trop*, (90), pp. 141–153.
- Kovats R.S., Bouma M.J., Hajat S., Worrall E., Haines A., 2003. El Nino and health. *Lancet*, (362), pp. 1481–1489.
- Kovats, R.S., Campbell-Lendrum, D.H., McMichael, A.J. Woodward A. and Cox, J.S., 2001. Early effects of climate change: do they include changes in vector-borne disease? *Philos. T. Roy. Soc. B*, (356), pp.1057 - 1068.
- MARA/ARMA, 1998. Towards an atlas of malaria risk in Africa. *Durban, South Africa*.
- McClellan, C. J., Lovett, J. C., Küper, W., Hannah, L., Sommer, J. H., Barthlott, W., Termansen, M., Smith, G. F., Tokumine, S. and Taplin, J. R. D., 2005. African Plant Diversity and Climate Change. *Annals of the Missouri Botanical Garden*, (92), pp. 139-152. Retrieved May 23, 2011.
- Moran, P.A.P., (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B* (10), pp. 243-251.
- Munga, S., Munakawa N., Zhou, G., Mushinzimana, E., Barrak, O.O.J., Githeko, A.K. and Yan, G.; 2006: Association between land cover and habitat productivity of malaria vectors in western Kenya highlands. *American Journal of Tropical Medicine*, (74), pp. 69-75.

- Murray, C.J.L., Rosenfeld, L.C., Lim, S.S., Andrews, K.G., Foreman, K.J., Haring, D., Fullman, N., Naghavi, M., Lozano, R. and Lopez, A.D., 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet* (379), pp. 413-431.
- NASA GSFC, 2010. TRMM: Tropical Rainfall Measuring Mission. <http://trmm.gsfc.nasa.gov/> (accessed June, 2010).
- NASA, 2009. MODIS. <http://modis.gsfc.nasa.gov> (accessed January, 2009).
- Nathan Y. Chan, Kristie L. Ebi, Fraser Smith, Thomas F. Wilson, and Anne E. Smith, 1999. An Integrated Assessment Framework for Climate Change and Infectious Diseases. *Environmental Health Perspectives*, 5(107), pp. 5.
- National Malaria Strategy, 2009-2017. Towards a malaria-free Kenya. Ministry of Public Health and Sanitation, Division of Malaria Control.
- Okara, et al., 2010. Distribution of main malaria vectors in Kenya. *Malaria Journal*, (9)69, doi:10.1186/1475-2875-9-69.
- Omumbo, J., Lyon, B., Waweru, S., Connor, S. and Thomson, M., 2011. Raised temperatures over the Kericho tea estates: revisiting the climate in the East African highlands malaria debate. *Malaria Journal*, 10(12).
- OpenLearn Labspace website; Communicable Diseases Module: 6. Factors that Affect Malaria Transmission (HYPERLINK "<http://labspace.open.ac.uk/mod/oucontent/>" <http://labspace.open.ac.uk/mod/oucontent/>), (accessed Saturday, 15 June 2013, 13:55).
- Pamela Anderson, et al., 2005. Climate Change Futures, Health, Ecological and Economic Dimensions, published by The *Centre for Health and the Global Environment*, *Havard Medical School*.

- Parmesan, C. and Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, (421), pp. 37–42.
- Pascual, M., Ahumada, J.A., Chaves, L.F., Rodo, X. and Bouma, M., 2006. Malaria resurgence in the East African highlands: Temperature trends revisited. *Pnas*, 103(15), pp. 5829-5834.
- Pascual, M., Athumada, J.A., Chaves, L.F., Rodo, X. and Bouma, M.; 2006: Malaria resurgence in the East African highlands: temperature trends revisited. *P. Natl. Acad. Sci. USA*, (103), pp. 5829-5834.
- Patz, J.A., Graczyk, T.K. and Geller, N. and Vittor A.Y., 2000. Effects of environmental change on emerging parasitic diseases. *Int J Parasitol*, (30), pp.395-405.
- Pearce, J. & Ferrier, S.; 2000: Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, (133), pp. 225–245.
- Pearson, R. G. and Dawson, T.P., 2003. Predicting the Impacts of Climate Change on the Distribution of Species: Are Bioclimate Envelope Models Useful? *Global Ecology and Biogeography*, (12), pp.361-371.
- Pearson, R. G., and Dawson T.P. 2003: Predicting the Impacts of Climate Change on the Distribution of Species: Are Bioclimate Envelope Models Useful? *Global Ecology and Biogeography*, (12), pp. 361-371. Retrieved May 23, 2011.
- Pearson, R.G. 2007: Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. *American Museum of Natural History*. Available at <http://ncep.amnh.org>.
- Pollitzer, R., 1959. Cholera (World Health Organization, Geneva), pp. 11–50.
- Randall, D.A., et al, 2007. Climate Models and Their Evaluation. In: *Climate Change 2007: The physical Science Basis*. Contribution of Working Group I to the Fourth

- Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press, Cambridge, UK and New York, NY, USA.*
- Roll Back Malaria and World Health Organization, 2001. Malaria Early Warning Systems: Concepts, Indicators and Partners.
- Root, T., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C. and Pounds, J.A., 2003. Fingerprints of global warming on wild animals and plants. *Nature* (421), pp. 57–60.
- Rosenberg, N., 1986. Adaptations to Adversity: Agriculture, Climate, and the Great Plains of North America. *Great Plains Quarterly*, (6), pp. 202-217.
- Safi N., Adimi F., Soebiyanto R.P., Kiang R.K., 2010. Towards malaria risk prediction in Afghanistan using remote sensing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Volume XXXVIII, Part 8, Kyoto Japan.
- Sawada, M., 1999. Rookcase: an Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. *Bulletin of the Ecological Society of America*, October 1999, pp. 231-243.
- Schreck, C.J. and Semazzi, F.H.M., 2004. Variability of the recent climate of Eastern Africa. *International Journal of Climatology*, (24), pp. 681-701.
- Stern, D.I., Gething, P.W., Kabaria, C.W., Temperley, W.H., Noor, A.M., Okiro, E.A., Shanks, G.D., Snow, R.W. and Hay, S.I., 2011. Temperature and malaria trends in highland East Africa. *PLoS One*, 6(9), e24524.
- Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (Supplement), pp. 234-240.
- Webster R. and Oliver M.A., 2007. Geostatistics for Environmental Scientists. *John Wiley and Sons Ltd*, Second Edition.
- Wigley, T.M.L., 2005. The climate change commitment. *Science*, (307), pp. 1766–1769.

- Williams, S. E., Bolitho, E. E. and Fox, S., (2003). Climate Change in Australian Tropical Rainforests: An Impending Environmental Catastrophe. Proceedings: *Biological Sciences*, (270), pp. 1887-1892. (Retrieved May 23, 2011).
- World Health Organization, (1993) A Global Strategy for Malaria Control. *WHO Geneva*.
- World Health Organization, (1998) Roll Back Malaria: A Global Partnership. *RBM/Draft/1 WHO-Geneva*.
- World Health Organization, 2003. Climate Change and Human Health: Risks and Responses, Summary. 20 Avenue Appia, 1211 Geneva, 27 Switzerland: Publications of the World Health Organization.
- World Health Organization, 2010. World Malaria Report 2010. World Health Organization, Geneva, pp. 1-93.

APPENDIX A: MALARIA SPATIAL DATA

Table A. 1 Malaria Spatial Data from Okara et. al., 2010

<i>Year</i>	<i>Report</i>	<i>Prov</i>	<i>Dist</i>	<i>Name</i>	<i>Lat</i>	<i>Long</i>	<i>Source</i>	<i>Month</i>	<i>Year</i>	<i>Num peop examined</i>	<i>Positive</i>	<i>Method examination</i>
2006	MoH Report	Western Province	Bungoma	Ack Primary School	0.6081	34.7732	GPS	2	2006	100	2	Microscopy
2005	MoH Report	Western Province	Kakamega	Amalemba Primary School	0.273	34.7529	GPS	6	2005	194	15	Microscopy
2003	Journal	Coast Province	Kwale	Amani	-4.1235	39.2847	GPS	5	1998	100	73	Microscopy
2005	MoH Report	Nyanza Province	Migori	Aringo Primary School	-0.846	34.1839	GPS	1	2005	130	6	Microscopy
2006	MoH Report	Nyanza Province	Homa-bay	Arunda Primary School	-0.5374	34.4655	GPS	5	2005	99	29	Microscopy
2003	Journal	Coast Province	Kilifi	Barani	-3.924	39.7675	GPS	5	1998	100	65	Microscopy
2005	MoH Report	Western Province	Kakamega	Bhulukunya Primary School	0.2701	34.7451	GPS	5	2005	163	25	Microscopy
2007	MoH Report	Nyanza Province	Kisii Central	Bigege & Matieko	-0.5734	34.7129	Other	11	2007	150	16	Microscopy
2002	Report	Nyanza Province	Nyamira	Birongo	-0.7482	34.8902	Other	10	1999	162	32	Microscopy
2000	MoH Report	Eastern Province	Meru Central	Bishop Bessone Primary School	0.03265	37.69389	GPS	2	2000	261	17	Microscopy
2001	MoH Report	Rift Valley Province	Bureti	Bishop Nding Junior Academy	-0.49957	35.09869	GPS	10	2001	67	26	Microscopy
2001	Report	Nyanza Province	Nyamira	Blue line Academy				6	2001	34	25	Microscopy
2006	MoH Report	Nyanza Province	Migori	Bondo Kosiemmo Primary School	-0.8252	34.1966	GPS	2	2006	124	4	Microscopy
2002	MoH Report	Rift Valley Province	Nandi	Bonjoge Primary School	0.1092	34.8936	GPS	7	2002	64	20	Microscopy
2001	MoH Report	Nyanza Province	Nyamira	Bonyaiguba Primary School	-0.514	34.8917	GPS	10	2001	140	88	Microscopy
2005	Report	Western Province	Kakamega	Bukhulunya Primary School	0.27012	34.74514	GPS	5	2005	163	25	Microscopy
2001	MoH	Rift Valley	Bureti	Charera Primary School	-0.50464	35.1155	GPS	10	2001	48	13	Microscopy

2003	Report Journal	Province Coast	Kilifi	Chasimba	-3.7291	39.7012	GPS	5	1998	100	39	Microscopy
2005	MoH Report	Province Western	Bungoma	Chebukaka Primary School	0.7577	34.6018	GPS	5	2005	190	28	Microscopy
2004	MoH Report	Province Rift Valley	Keiyo	Chegilet Primary School	0.8301	35.6063	GPS	4	2004	56	9	Microscopy
2003	Report	Province Rift Valley	Nandi	Chepsui	0.169147	35.042803	Other	8	2002	12	1	Microscopy
2005	MoH Report	Province Western	Bungoma	Chesamisi Primary School	0.8392	34.7902	GPS	10	2005	375	62	Microscopy
2006	Report	Province Western	Bungoma	Chesamisi Primary School	0.8392	34.7902	GPS	3	2006	225	18	Microscopy
2006	Report	Province Western	Bungoma	Chesamisi Primary School	0.8392	34.7902	GPS	10	2006	150	44	Microscopy
2007	MoH Report	Province Nyanza	Nyamira	Chirochiro & Nyaturubo	-0.771	34.8852	Other	11	2007	150	1	Microscopy
2000	MoH Report	Province Eastern	Meru	Chorobia				2	2000	267	74	Microscopy
2005	MoH Report	Province Western	Bungoma	Chwele Primary School	0.7665	34.5506	GPS	5	2005	190	35	Microscopy
2004	Journal	Province Central	Kirinyaga	Ciagini	-0.7497	37.3939	GPS	2	2002	69	6	Microscopy
2003	Journal	Province Coast	Malindi	Dabaso	-3.3397	40.0007	Other	5	1998	99	38	Microscopy
2006	Journal	Province North Eastern	Garissa	Dadaab (refugee camps)	0.0544	40.3081	GPS	4	2002	198	21	Microscopy
2007	Report	Province North Eastern	Garissa	Dertu	0.2726	39.7984	Pers. comm	11	2006	157	1	Microscopy
2003	Journal	Province Coast	Kilifi	Dindiriri	-3.734	39.799	Encarta	5	1998	101	62	Microscopy
2003	Journal	Province Coast	Kwale	Dumbule	-4.1164	39.3695	GPS	5	1998	100	82	Microscopy
2007	MoH Report	Province Nyanza	Kisii	Ekerubo Primary School	-0.7103	34.7375	GPS	7	2007	66	1	Microscopy
2002	Report	Province Coast	Tana River	Elan	-0.2219	39.4837	Other	10	1999	59	3	Microscopy

		Province										
2007	MoH Report	Nyanza Province	Gucha	Emenwa Primary School	-0.8304	34.8123	GPS	7	2007	92	2	Microscopy
2004	Report	Rift Valley Province	Keiyo	Emsea Secondary School	0.44169	35.61437	GPS	6	2004	9	3	Microscopy
2007	MoH Report	Nyanza Province	Nyamira	Eremo Primary School	-0.7346	34.8539	GPS	6	2007	100	5	Microscopy
2007	MoH Report	Nyanza Province	Gucha	Etono Primary School	-0.8844	34.7524	GPS	7	2007	145	4	Microscopy
2006	Journal	Rift Valley Province	Kericho	Fort Ternan	-0.199	35.349	Encarta	10	1999	95	6	Microscopy
2006	Journal	Rift Valley Province	Kericho	Fort Ternan	-0.199	35.349	Encarta	4	2000	68	8	Microscopy
2006	Journal	Rift Valley Province	Kericho	Fort Ternan	-0.199	35.349	Encarta	11	2000	56	3	Microscopy
2006	Journal	Rift Valley Province	Kericho	Fort Ternan	-0.199	35.349	Encarta	5	2001	59	7	Microscopy
2005	MoH Report	Eastern Province	Meru Central	Gaceero Primary School	-0.108	37.7615	GPS	6	2005	65	11	Microscopy
2006	MoH Report	Nyanza Province	Siaya	Gangu Primary School	0.0755	34.1835	GPS	8	2006	173	76	Microscopy
2003	Journal	Coast Province	Malindi	Garithe	-2.9966	40.1953	Other	5	1998	100	79	Microscopy
2005	MoH Report	Eastern Province	Meru Central	Gateway Primary School	0.0019	37.8326	Other	3	2004	61	9	Microscopy
2003	Journal	Coast Province	Kwale	Gazi	-4.427	39.503	Encarta	5	1998	100	60	Microscopy
2007	Report	Nyanza Province	Kisii Central	Genga Primary School	-	34.643327	GPS	9	2007	100	7	Microscopy
2007	MoH Report	Nyanza Province	Kisii Central	Gesabakwa Primary School	0.637399	34.7639	GPS	5	2007	100	7	Microscopy
2001	MoH Report	Nyanza Province	Nyamira	Gesura Primary School	-0.4856	34.9741	GPS	7	2001	97	32	Microscopy
2001	MoH Report	Nyanza Province	Nyamira	Getaari Primary School	-0.4641	35.0122	GPS	8	2001	131	104	Microscopy
2004	MoH Report	Nyanza Province	Bondo	Godwa community	-0.17	34.24	Other	10	2004	87	56	Microscopy
2006	MoH	Nyanza	Migori	Gunga Primary School	-0.8049	34.1275	GPS	2	2006	73	2	Microscopy

2006	Report	Province											
2006	Report	Eastern Province	Isiolo	Halgani Primary School	0.325	38.20072	GPS	5	2006	25	4	Microscopy	
2006	MoH Report	Nyanza Province	Siaya	Hawinga Primary School	0.0874	34.159	GPS	8	2006	119	61	Microscopy	
2005	Journal	Western Province	Kakamega	Iguhu	0.167	34.7483	GPS	7	2002	163	90	Microscopy	
2005	Journal	Western Province	Kakamega	Iguhu	0.167	34.7483	GPS	6	2003	309	136	Microscopy	
2007	MoH Report	Nyanza Province	Kisii	Iranda Primary School	-0.6278	34.7849	GPS	5	2007	100	8	Microscopy	
2007	MoH Report	Nyanza Province	Central Gucha	Itumbe	-0.7578	34.7757	GPS	10	2007	100	1	Microscopy	
2007	MoH Report	Nyanza Province	Gucha	Itumbe Primary School	-0.8418	34.8298	GPS	7	2007	100	0	Microscopy	
2006	MoH Report	Nyanza Province	Migori	Jangoe Primary School	-0.8416	34.2603	GPS	2	2006	358	24	Microscopy	
2003	Journal	Coast Province	Kilifi	Jaribuni	-3.633	39.733	Encarta	5	1998	98	53	Microscopy	
2004	Thesis	Coast Province	Malindi	Jilore	-	39.903548	GPS	2	2003	99	36	Microscopy	
2003	Report	Rift Valley Province	Nandi	Kabaskei	3.188313	0.159086	34.950867	Other	7	2002	37	28	Microscopy
2006	Report	Nyanza Province	Homa-bay	Kabunde	-	0.610105	34.651789	Other	6	2006	26	16	Microscopy
2004	Journal	Central Province	Kirinyaga	Kagio	-0.617	37.249	Encarta	2	2002	53	9	Microscopy	
2003	Journal	Coast Province	Kilifi	Kagombani	-3.5385	39.6096	GPS	5	1998	100	72	Microscopy	
2005	MoH Report	Western Province	Kakamega	Kakamega Primary School	0.2823	34.754	GPS	7	2005	238	31	Microscopy	
2005	Journal	Rift Valley Province	Turkana	Kakuma	3.717	34.865	Encarta	3	2001	194	13	Microscopy	
2006	MoH Report	Nyanza Province	Kisumu	Kaloleni Primary School	-0.0963	34.7675	GPS	9	2005	161	27	Microscopy	
2003	Report	Rift Valley Province	Nandi	Kamoiywo	0.401687	35.09186	Other	7	2002	38	11	Microscopy	
2005	MoH	Western	Bungoma	Kamukuywa Primary	0.7783	34.7878	GPS	2	2005	200	21	Microscopy	

2005	Report	Province		School									
2005	MoH Report	Western Province	Bungoma	Kamusinde Rc Primary School	0.8249	34.7422	GPS	10	2005	205	43	Microscopy	
2005	MoH Report	Western Province	Bungoma	Kamusinga Primary School	0.8046	34.7073	GPS	3	2005	123	14	Microscopy	
2002	Report	Nyanza Province	Nyando	Kandiang'a	-0.3312	34.8687	Other	10	1999	434	322	Microscopy	
2005	Journal	Nyanza Province	Kuria	Kanyawegi sublocation	-1.2366	34.6462	GPS	5	2003	134	122	Microscopy	
2003	Report	Rift Valley Province	Baringo	Kapkelewa Primary School	0.37685	35.72099	GPS	9	2002	35	0	Microscopy	
2003	Report	Rift Valley Province	Nandi	Kapkinduywa	-	34.91202	Other	7	2002	48	18	Microscopy	
2003	Report	Rift Valley Province	Nandi	Kapkobis Primary School	0.002065	35.0833	GPS	8	2002	35	11	Microscopy	
2004	MoH Report	Rift Valley Province	Baringo	Kapkuikui Primary School	0.3689	36.0165	GPS	5	2004	128	4	Microscopy	
2002	MoH Report	Rift Valley Province	Nandi	Kaptabongeni Primary School	0.2154	35.0371	GPS	5	2002	67	24	Microscopy	
2004	MoH Report	Rift Valley Province	Baringo	Kaptombes Primary School	0.4109	36.0244	GPS	5	2004	80	2	Microscopy	
2001	MoH Report	Rift Valley Province	Bureti	Kapurus Primary School	-	35.106407	GPS	10	2001	83	21	Microscopy	
2003	Report	Rift Valley Province	Nandi	Kapwareng	0.558952	34.912848	Other	7	2002	22	12	Microscopy	
2005	MoH Report	Eastern Province	Meru	Karirwara Primary School	0.037	37.7506	GPS	3	2004	82	10	Microscopy	
2002	MoH Report	Rift Valley Province	Baringo	Katibel Primary School	0.6315	35.6667	GPS	9	2002	70	0	Microscopy	
2005	MoH Report	Eastern Province	Meru	Kaurone Primary School	-0.1101	37.7956	GPS	6	2005	66	29	Microscopy	
2001	MoH Report	Nyanza Province	Nyamira	Kemasare Primary School	-0.5248	34.9184	GPS	10	2001	112	64	Microscopy	
2007	MoH Report	Nyanza Province	Gucha	Kemoreko Primary School	-0.8844	34.77	GPS	10	2007	71	0	Microscopy	
2007	MoH Report	Nyanza Province	Nyamira	Kiamokama sublocation	-0.8316	34.8832	GPS	5	2007	55	0	Microscopy	
2002	Report	Eastern Province	Meru	Kibaranyaki	-0.0017	37.5706	Other	10	1999	138	13	Microscopy	

		Province	Central									
2005	MoH Report	Western Province	Bungoma	Kibichori Primary School	0.755	34.5847	GPS	5	2005	180	39	Microscopy
2006	Report	Coast Province	Malindi	Kibokoni Primary School	-3.1708	40.11838	GPS	6	2005	99	32	Microscopy
2006	Report	Coast Province	Malindi	Kibokoni Primary School	-3.1708	40.11838	GPS	9	2005	89	10	Microscopy
2002	Report	Eastern Province	Meru	Kibureni	0.0167	37.84	Other	10	1999	156	58	Microscopy
2002	Report	Coast Province	Kwale	Kifyonzo	-4.1861	39.2683	Other	10	1999	166	86	Microscopy
2002	Report	Coast Province	Kwale	Kikoneni	-4.45	39.299	Encarta	10	1999	179	76	Microscopy
2001	MoH Report	Rift Valley Province	Bureti	Kimase Primary School	-0.6928	35.10542	GPS	10	2001	83	26	Microscopy
2005	MoH Report	Western Province	Bungoma	Kimilili Boys School	0.796	34.7104	GPS	10	2005	236	31	Microscopy
2006	Report	Eastern Province	Isiolo	Kinna Primary School	0.32118	38.2107	GPS	5	2006	30	7	Microscopy
2002	MoH Report	Rift Valley Province	Baringo	Kinyach Primary School	0.9461	35.6803	GPS	9	2002	72	2	Microscopy
2002	MoH Report	Rift Valley Province	Nandi	Kiptulya Primary School	0.2027	34.9914	GPS	5	2002	65	32	Microscopy
2003	Report	Rift Valley Province	Nandi	Kiroppet	0.152879	35.090002	Other	7	2002	22	11	Microscopy
2006	MoH Report	Nyanza Province	Kisumu	Kisian Primary School	-0.0732	34.663	GPS	3	2005	116	92	Microscopy
2003	Report	Rift Valley Province	Baringo	Kisok Primary School	0.407	35.70509	GPS	9	2002	32	2	Microscopy
2002	Report	Coast Province	Kwale	Kiteje Primary School	-4.1185	39.6063	GPS	2	2002	146	6	Microscopy
2003	Journal	Coast Province	Kilifi	Kitengwani	-3.5269	39.778	Other	5	1998	99	69	Microscopy
2004	Report	Central Province	Kirinyaga	Kithinti Primary School	-0.6107	37.3248	GPS	5	2004	178	62	Microscopy
2003	Journal	Coast Province	Kilifi	Kitsoeni	-3.7034	39.7322	GPS	5	1998	100	69	Microscopy
2005	Report	Nyanza Province	Homa-bay	Kogelo Kalanya Primary	-	34.471327	GPS	3	2005	93	10	Microscopy

		Province		School	0.532914							
2006	Report	Nyanza Province	Homa-bay	Kogwang Primary School	-	34.523555	GPS	6	2006	46	16	Microscopy
2002	MoH Report	Rift Valley Province	Nandi	Koimet Primary School	0.638507	35.0333	GPS	7	2002	70	20	Microscopy
2006	MoH Report	Nyanza Province	Kisumu	Kotetni Primary School	-0.0706	34.7085	GPS	3	2005	120	83	Microscopy
2001	MoH Report	Nyanza Province	Nyamira	Kowidi Primary School	-0.4423	34.9788	GPS	11	2001	82	48	Microscopy
2006	MoH Report	Nyanza Province	Kisumu	Kudho Primary School	-0.072	34.7611	GPS	9	2005	226	75	Microscopy
2002	MoH Report	Rift Valley Province	Baringo	Kuikui Primary School	0.8008	35.7079	GPS	9	2002	65	1	Microscopy
2002	Report	Eastern Province	Kitui	Kyatune	-1.7762	38.1137	Other	10	1999	153	3	Microscopy
2007	MoH Report	Nyanza Province	Nyando	Lela Primary School & Community	-0.1649	34.8922	GPS	6	2007	133	77	Microscopy
2001	MoH Report	Rift Valley Province	Bureti	Lelach Primary School	-0.65368	35.08937	GPS	10	2001	74	20	Microscopy
2007	Journal	Eastern Province	Makueni	Lower Mangelete & Yumbuni	-2.7	38.133	Encarta	11	2003	1044	146	Microscopy
2007	Journal	Eastern Province	Makueni	Lower Mangelete & Yumbuni	-2.7	38.133	Encarta	2	2004	973	464	Microscopy
2004	MoH Report	Nyanza Province	Siaya	Lwango Kotieno community	-0.3859	34.2846	Other	10	2004	120	82	Microscopy
2007	MoH Report	Nyanza Province	Nyamira	Maburi & Amariba	-0.7509	34.8061	Other	11	2007	150	5	Microscopy
2003	Journal	Coast Province	Kwale	Magaoni	-4.38	39.4724	GPS	5	1998	78	50	Microscopy
2005	MoH Report	Western Province	Kakamega	Mahiakalo Primary School	0.2996	34.7665	GPS	11	2005	211	23	Microscopy
2003	Journal	Coast Province	Kilifi	Majajani	-3.6667	39.75	Other	5	1998	101	76	Microscopy
2003	Journal	Coast Province	Malindi	Majenjeni	-3.1403	40.1405	Other	5	1998	100	58	Microscopy
2006	Report	Coast Province	Malindi	Majivuni Primary School	-	40.088636	Other	6	2005	113	32	Microscopy
2006	Report	Coast Province	Malindi	Majivuni Primary	3.202999	40.088636	Other	9	2005	93	12	Microscopy

		Province		School	3.202999								
2006	Journal	Western Province	Kakamega	Makhokho	0.1758	34.7495	GPS	2	2000	80	32	Microscopy	
2005	MoH Report	Western Province	Bungoma	Makhonge Primary School	0.77	34.5769	GPS	5	2005	200	42	Microscopy	
2006	Report	Western Province	Bungoma	Makhonge Primary School	0.77	34.5769	GPS	5	2006	200	39	Microscopy	
2005	MoH Report	Western Province	Bungoma	Malaha Primary School	0.6386	34.7385	GPS	2	2005	220	28	Microscopy	
2005	Report	Nyanza Province	Homa-bay	Manga Primary School	-0.4892	34.6174	GPS	3	2005	80	30	Microscopy	
2006	MoH Report	Nyanza Province	Homa-bay	Manga Primary School	-0.4892	34.6174	GPS	1	2006	63	32	Microscopy	
2006	MoH Report	Nyanza Province	Gucha	Mangere Primary School	-0.8266	34.717	GPS	10	2006	83	0	Microscopy	
2004	Journal	Nyanza Province	Kisii	Marani	-0.5794	34.7991	GPS	2	2002	114	29	Microscopy	
2004	MoH Report	Nyanza Province	Kisii	Marani	-0.5794	34.7991	GPS	5	2004	75	9	Microscopy	
2003	Report	Rift Valley Province	Baringo	Maregut Primary School	0.868322	35.694798	GPS	9	2002	27	0	Microscopy	
2005	MoH Report	Nyanza Province	Migori	Marienga Primary School	-0.9466	34.5581	GPS	7	2005	185	28	Microscopy	
2003	Journal	Coast Province	Malindi	Masheheni	-3.1281	40.109	Other	5	1998	101	67	Microscopy	
2004	Journal	Nyanza Province	Nyamira	Masimba	-0.8607	34.9403	GPS	2	2002	76	1	Microscopy	
2001	Report	Nyanza Province	Nyamira	Masosa Primary	-	34.89943	GPS	7	2001	195	102	Microscopy	
2007	MoH Report	Eastern Province	Makueni	Matangani Primary School	0.490186	-2.6961	38.1442	GPS	5	2004	95	6	Microscopy
2005	MoH Report	Western Province	Kakamega	Matende Primary School	0.26563	34.75392	GPS	7	2005	222	23	Microscopy	
2005	Report	Western Province	Kakamega	Matende Primary School	0.26563	34.75392	GPS	5	2005	160	20	Microscopy	
2003	Journal	Coast Province	Malindi	Maziwani	-3.2085	40.0699	Other	5	1998	100	48	Microscopy	
2003	Journal	Coast Province	Malindi	Mbarak Chembe	-3.294	40.0843	Other	5	1998	100	59	Microscopy	

2004	Journal	Province Nyanza	Suba	Mbita & Lwanda	-0.418	34.204	GPS	11	2001	276	154	Microscopy
2004	Journal	Province Central	Kirinyaga	Mbui Njeru	-0.6981	37.3446	GPS	2	2002	21	0	Microscopy
2002	Report	Province Coast	Kwale	Mbweka Primary School	-4.1634	39.5873	GPS	2	2002	178	0	Microscopy
2001	MoH Report	Province Nyanza	Nyamira	Menyenya Sda Primary School	-0.7625	35.0194	GPS	1	2002	108	77	Microscopy
2005	MoH Report	Province Western	Bungoma	Miendo Primary School	0.6718	34.6801	GPS	2	2005	295	35	Microscopy
2003	Journal	Province Coast	Malindi	Mijomboni	-3.2599	40.0122	Other	5	1998	100	63	Microscopy
2002	Report	Province Coast	Kwale	Miritini	-4.0107	39.5869	GPS	10	1999	133	35	Microscopy
2001	MoH Report	Province Nyanza	Nyamira	Miruka Sibora Primary School	-0.4988	34.8806	GPS	7	2001	60	34	Microscopy
2001	Report	Province Nyanza	Nyamira	Miruka Sibora Primary School	-0.4988	34.8806	GPS	6	2001	127	53	Microscopy
2002	Report	Province Western	Bungoma	Misikhu	0.7148	34.7568	Other	10	1999	173	100	Microscopy
2005	MoH Report	Province Eastern	Meru	Mitunguu Primary School	-0.1045	37.782	GPS	6	2005	114	17	Microscopy
2003	Journal	Province Central	Malindi	Mjanaheri	-3.068	40.1405	Other	5	1998	100	49	Microscopy
2006	Report	Province Coast	Malindi	Mkao Moto Primary School	-3.25125	40.05228	GPS	6	2005	100	18	Microscopy
2006	Report	Province Coast	Malindi	Mkao Moto Primary School	-3.25125	40.05228	GPS	9	2005	91	6	Microscopy
2002	Report	Province Coast	Kwale	Mkumbi Primary School	-4.0986	39.6084	GPS	2	2002	148	6	Microscopy
2003	Report	Province Rift Valley	Baringo	Mogorwa Primary School	0.31206	35.71901	GPS	9	2002	30	2	Microscopy
2005	Report	Province Western	Bungoma	Moi Girls High School Kamusinga	0.802822	34.70819	GPS	4	2005	173	14	Microscopy
2007	Report	Province Nyanza	Gucha	Mosache Primary	-	34.631764	Other	9	2007	100	5	Microscopy
2007	MoH	Province Nyanza	Gucha	Mosasa Primary School	0.753484 -0.8475	34.819	GPS	7	2007	186	4	Microscopy

2007	Report	Province											
2007	MoH Report	Nyanza Province	Gucha	Mosobeti Primary School	-0.9184	34.712	GPS	10	2007	71	0	Microscopy	
2004	Journal	Nyanza Province	Kisii	Mosocho	-	34.736268	GPS	2	2002	35	11	Microscopy	
2003	Journal	Coast Province	Central Kwale	Moyeni	0.591993								
2003	Journal	Coast Province	Kwale	Moyeni	-4.1424	39.3906	GPS	5	1998	100	78	Microscopy	
2003	Journal	Coast Province	Kilifi	Mtepeni	-3.9001	39.7315	GPS	5	1998	99	82	Microscopy	
2002	Report	Coast Province	Kwale	Mteza Primary School	-4.0824	39.5151	GPS	2	2002	139	0	Microscopy	
2007	Report	Nyanza Province	Kisii	Mugori Primary School	-	34.641485	GPS	9	2007	100	17	Microscopy	
2006	MoH Report	Western Province	Bungoma	Mukhuyu Primary School	0.665629								
2006	MoH Report	Western Province	Bungoma	Mukhuyu Primary School	0.598	34.7662	GPS	2	2006	100	0	Microscopy	
2008	Report	Western Province	Kakamega	Mukumu Boys High School	0.223505	34.763689	GPS	5	2008	15	2	Microscopy	
2004	Journal	Central Province	Kirinyaga	Murinduko	-0.5667	37.45	Encarta	2	2002	63	34	Microscopy	
2005	MoH Report	Western Province	Kakamega	Muslim Primary School	0.2818	34.7503	GPS	11	2005	175	10	Microscopy	
2007	Report	Nyanza Province	Nyamira	Mutembe Primary School	-	34.965569	GPS	9	2007	62	1	Microscopy	
2006	Report	Central Province	Kirinyaga	Mutithi village	0.876729								
2006	Report	Central Province	Kirinyaga	Mutithi village	-0.7255	37.3215	GPS	8	2005	186	0	Microscopy	
2002	Report	Nyanza Province	Kisii	Mwamosioma	-0.6378	34.7852	Other	10	1999	203	88	Microscopy	
2003	Journal	Coast Province	Kwale	Mwaroni	-4.2668	39.5762	GPS	5	1998	100	50	Microscopy	
2007	Report	Nyanza Province	Kisii	Mwata Primary School	-	34.648639	GPS	9	2007	140	20	Microscopy	
2006	Report	Central Province	Kirinyaga	Mwea childrens home	0.678266								
2006	Report	Central Province	Kirinyaga	Mwea childrens home	-0.6643	37.3494	GPS	6	2006	76	0	Microscopy	
2005	MoH Report	Western Province	Lugari	Namagara Primary School	0.5026	34.8714	GPS	5	2005	159	29	Microscopy	
2006	Report	Nyanza Province	Kisumu	Nanga Primary School	-0.1346	34.7384	GPS	5	2005	184	69	Microscopy	
2006	MoH	Nyanza	Kisumu	Nanga Primary School	-0.1346	34.7384	GPS	3	2006	269	180	Microscopy	

2007	Report	Province											
2007	Report	Nyanza Province	Kisii Central	Ndaru Primary School				9	2007	100	2	Microscopy	
2006	Report	Nyanza Province	Homa-bay	Ndiru Primary School	-0.61035	34.51041	GPS	6	2006	22	14	Microscopy	
2004	Report	Central Province	Kirinyaga	Ndomba Primary School	-0.5808	37.3418	GPS	5	2004	179	10	Microscopy	
2002	Report	Eastern Province	Embu	Nga'ratuko	-0.2249	37.3787	Other	10	1999	100	3	Microscopy	
2005	MoH Report	Nyanza Province	Migori	Ngere Primary School	-0.7744	34.5719	GPS	7	2005	121	14	Microscopy	
2002	Report	Coast Province	Kwale	Ningawa Primary School	-4.1256	39.6003	GPS	2	2002	107	0	Microscopy	
2002	Report	Central Province	Nyandarua	Njabini	-0.726	36.6531	Other	10	1999	77	0	Microscopy	
2000	MoH Report	Eastern Province	Meru Central	Nkabune Primary School	0.03445	37.69383	GPS	2	2000	261	35	Microscopy	
2005	MoH Report	Eastern Province	Meru North	Nthamiri Primary School	0.0388	37.7891	GPS	3	2004	102	25	Microscopy	
2004	Journal	Nyanza Province	Kisii Central	Nyagoto	-0.5459	34.8062	GPS	2	2002	65	12	Microscopy	
2001	MoH Report	Nyanza Province	Nyando	Nyakorio community	-0.174	35.0626	Other	8	2001	125	33	Microscopy	
2006	MoH Report	Nyanza Province	Siaya	Nyalaji Primary School	0.0771	34.1717	GPS	8	2006	193	103	Microscopy	
2006	MoH Report	Nyanza Province	Homa-bay	Nyalkinyi Primary School	-0.5199	34.4912	GPS	5	2005	100	6	Microscopy	
2006	Report	Nyanza Province	Kisumu	Nyamasaria Primary School	-0.10475	34.786116	GPS	3	2005	32	12	Microscopy	
2006	MoH Report	Nyanza Province	Gucha	Nyamiobo Primary School	-0.8585	34.699	GPS	10	2006	72	2	Microscopy	
2007	Report	Nyanza Province	Kisii Central	Nyandiwa Primary School	-	34.829368	GPS	9	2007	100	4	Microscopy	
2001	MoH Report	Nyanza Province	Nyamira	Nyaobe Primary School	-0.4407	34.9856	GPS	11	2001	77	60	Microscopy	
2001	MoH Report	Nyanza Province	Nyamira	Nyarichia Primary School	-0.5607	34.9417	GPS	7	2001	126	61	Microscopy	
2007	MoH	Nyanza	Kisii	Nyataro Primary School	-0.7203	34.7635	GPS	6	2007	100	6	Microscopy	

2007	Report	Province	Central										
2007	MoH	Nyanza	Kisii	Nyaura & Inaga	-0.677	34.7718	Other	11	2007	150	18	Microscopy	
2006	Report	Province	Central										
2006	MoH	Nyanza	Migori	Obolo Primary School	-1.0219	34.1448	GPS	2	2006	90	20	Microscopy	
2006	Report	Province											
2006	MoH	Nyanza	Homa-bay	Ogande Primary School	-0.5704	34.4955	GPS	1	2006	71	29	Microscopy	
2001	Report	Province											
2001	Report	Nyanza	Nyamira	Oganga Sibora Academy	-	34.89164	GPS	6	2001	53	39	Microscopy	
2005	Report	Province			0.515042								
2005	MoH	Nyanza	Migori	Ogwedhi Primary School	-1.1097	34.6195	GPS	7	2005	70	4	Microscopy	
2007	Report	Province											
2007	MoH	Nyanza	Kisii	Omwari Primary School	-0.6578	34.6523	GPS	5	2007	100	10	Microscopy	
2006	Report	Province	Central										
2006	MoH	Nyanza	Homa-bay	Ongeti Primary School	-0.4778	34.5434	GPS	5	2005	86	37	Microscopy	
2005	Report	Province											
2005	MoH	Nyanza	Migori	Othochrakuom Primary School	-1.001	34.1784	GPS	4	2005	158	32	Microscopy	
2003	Report	Province											
2003	Journal	Coast	Kilifi	Paziani	-3.5881	39.5266	GPS	5	1998	100	72	Microscopy	
2002	Report	Province											
2002	Report	Coast	Kwale	Pungu Primary School	-4.1181	39.627	GPS	2	2002	168	4	Microscopy	
2006	Report	Province											
2006	Report	Nyanza	Homa-bay	Radiro Primary School	-	34.444089	GPS	6	2006	33	17	Microscopy	
2004	Journal	Province			0.563773								
2004	Journal	Nyanza	Nyamira	Ramasha	-0.8963	34.9748	GPS	2	2002	56	2	Microscopy	
2007	Report	Province											
2007	MoH	Nyanza	Nyando	Ranjira Primary School	-0.1495	34.8629	GPS	7	2007	120	81	Microscopy	
2008	Report	Province											
2008	MoH	Nyanza	Rachuonyo	Rawinji Primary School	-0.429	34.7447	GPS	3	2008	61	20	Microscopy	
2007	Report	Province											
2007	MoH	Nyanza	Gucha	Riamo Primary School	-0.8691	34.7086	GPS	10	2007	100	1	Microscopy	
2004	Report	Province											
2004	Journal	Nyanza	Kisii	Riana	-0.651	34.666	Encarta	2	2002	68	15	Microscopy	
2007	Report	Province	Central										
2007	MoH	Nyanza	Nyamira	Ribate Primary School	-0.8609	34.9385	GPS	6	2007	100	3	Microscopy	
2007	Report	Province											
2007	MoH	Nyanza	Kisii	Rioma & Marani	-0.5334	34.7924	Other	10	2007	150	4	Microscopy	
2006	Report	Province	Central										
2006	Report	Nyanza	Homa-bay	Ruga Primary School	-	34.459306	GPS	6	2006	48	25	Microscopy	

		Province			0.629291							
2002	Report	Central Province	Nyandarua	Rurii	-0.201	36.382	Other	10	1999	117	11	Microscopy
2006	MoH Report	Western Province	Bungoma	S.A. Primary School	0.6163	34.7596	GPS	2	2006	100	1	Microscopy
2006	Report	Coast Province	Malindi	Sabaki Primary School	-3.16368	40.08292	GPS	6	2005	103	51	Microscopy
2006	Report	Coast Province	Malindi	Sabaki Primary School	-3.16368	40.08292	GPS	9	2005	82	32	Microscopy
2007	Report	Nyanza Province	Siaya	Sauri (sublocation)	0.118	34.5126	GPS	4	2005	909	481	Microscopy
2005	MoH Report	Western Province	Kakamega	Sda Satellite Primary School	0.2809	34.7509	GPS	11	2005	241	8	Microscopy
2007	MoH Report	Nyanza Province	Gucha	Sengera Primary School	-0.9277	34.7198	GPS	7	2007	100	3	Microscopy
2006	MoH Report	Nyanza Province	Migori	Senye Primary School	-0.9812	34.0864	GPS	2	2006	148	14	Microscopy
2006	MoH Report	Nyanza Province	Homa-bay	Sero Primary School	-0.5933	34.4932	GPS	1	2006	60	30	Microscopy
2004	Thesis	Coast Province	Malindi	Shakahola	-	39.553492	Other	2	2003	109	19	Microscopy
2003	Journal	Coast Province	Kilifi	Shariani	3.133702	-3.7897	GPS	5	1998	100	54	Microscopy
2005	MoH Report	Western Province	Lugari	Shihome Primary School	0.3977	34.7842	GPS	5	2005	171	27	Microscopy
2007	MoH Report	Western Province	Kakamega	Shisasari And Rosterman Primary Schools	0.2616	34.7147	GPS	7	2007	330	73	Microscopy
2005	MoH Report	Western Province	Kakamega	Shitaho Primary School	0.2713	34.7687	GPS	11	2005	183	28	Microscopy
2006	Journal	Western Province	Kakamega	Sigalagala	0.199	34.749	Encarta	2	2000	80	21	Microscopy
2006	Report	Coast Province	Malindi	Sir Ali Primary School	-3.21838	40.11909	GPS	6	2005	105	1	Microscopy
2006	Report	Coast Province	Malindi	Sir Ali Primary School	-3.21838	40.11909	GPS	9	2005	86	0	Microscopy
2002	Report	Western Province	Busia	Sisenye	0.158	33.994	Encarta	10	1999	171	117	Microscopy

2006	Journal	Western Province	Bungoma	Sitikho	0.4776	34.667	Other	11	1998	300	230	Microscopy
2004	MoH Report	Rift Valley Province	Keiyo	Songeto Primary School	0.5929	35.5601	GPS	4	2004	83	15	Microscopy
2005	MoH Report	Nyanza Province	Migori	Sori Primary School	-0.8456	34.154	GPS	1	2005	110	6	Microscopy
2005	MoH Report	Western Province	Kakamega	St Antony Primary School	0.2066	34.8509	GPS	5	2005	190	43	Microscopy
2005	MoH Report	Western Province	Kakamega	St Josephs Primary School	0.2803	34.7449	GPS	11	2005	123	4	Microscopy
2006	Report	Coast Province	Malindi	St. Andrews Primary School	-3.21874	40.10201	GPS	6	2005	108	7	Microscopy
2006	Report	Coast Province	Malindi	St. Andrews Primary School	-3.21874	40.10201	GPS	9	2005	100	6	Microscopy
2005	Report	Western Province	Bungoma	St. Antony Primary School	0.603587	34.770425	GPS	5	2005	190	42	Microscopy
2005	Report	Western Province	Kakamega	St. Joseph Primary School	0.28357	34.74551	GPS	11	2005	123	4	Microscopy
2003	Journal	Coast Province	Kilifi	Takaungu	-3.683	39.85	Encarta	5	1998	100	52	Microscopy
2006	MoH Report	Nyanza Province	Gucha	Tendere Primary School	-0.8177	34.7207	GPS	10	2006	105	2	Microscopy
2004	Report	Rift Valley Province	Baringo	Tirion village Sunshine ECD	0.38736	36.02541	GPS	5	2004	29	0	Microscopy
2001	MoH Report	Nyanza Province	Nyamira	Tombe Sda Primary School	-0.6528	34.8719	GPS	7	2001	77	58	Microscopy
2005	MoH Report	Nyanza Province	Migori	Tonye Primary School	-0.715	34.6324	GPS	7	2005	100	6	Microscopy
2005	MoH Report	Western Province	Kakamega	Township Primary School	0.2714	34.7585	GPS	6	2005	226	34	Microscopy
2006	MoH Report	Nyanza Province	Siaya	Uhembo Primary School	0.1393	34.2762	GPS	8	2006	187	75	Microscopy
2007	MoH Report	Nyanza Province	Siaya	Usula Primary School	0.0806	34.3679	GPS	10	2007	251	53	Microscopy
2002	Report	Coast Province	Mombasa	Utange	-3.9667	39.7167	Encarta	10	1999	106	36	Microscopy
2003	Journal	Coast Province	Kwale	Vinuni	-4.183	39.533	Encarta	5	1998	100	66	Microscopy

2002	Report	Coast Province	Kwale	Voroni Primary School	-4.1926	39.5618	GPS	2	2002	295	0	Microscopy
2003	Journal	Coast Province	Kwale	Vuga	-4.183	39.499	Encarta	5	1998	100	44	Microscopy
2006	MoH Report	Western Province	Bungoma	Webuye Primary School	0.6109	34.7685	GPS	2	2006	100	0	Microscopy
2006	MoH Report	Nyanza Province	Homa-bay	Wiamen Primary School	-0.6341	34.4658	GPS	1	2006	66	20	Microscopy
2006	MoH Report	Nyanza Province	Homa-bay	Wi-Obiero Primary School	-0.6277	34.4901	GPS	1	2006	52	29	Microscopy
2002	Report	Coast Province	Kwale	Yeje Primary School	-4.1748	39.5876	GPS	2	2002	329	5	Microscopy
2002	Report	Coast Province	Kwale	Zibani Primary School	-4.1277	39.6128	GPS	2	2002	264	15	Microscopy
2003	Journal	Coast Province	Kwale	Ziwani	-4.1553	39.4541	GPS	5	1998	100	59	Microscopy

Table A. 2 Malaria Spatial Data from MARA ARMA, 1998

<i>Country</i>	<i>Locality</i>	<i>Lat</i>	<i>Long</i>	<i>Month</i>	<i>Year</i>	<i>Catch</i>	<i>Species</i>	<i>Iden</i>	<i>Number</i>	<i>Reference</i>
KENYA	AHERO	0	34	5	89,90	HRD	ARABIENSIS	2		GITHEKO SERVICE ETAL 1996, PARASSITOLOGIA 38 481-489
KENYA	AHERO	0	34	5	89,90	HRD	GAMBIAE	2		GITHEKO SERVICE ETAL 1996, PARASSITOLOGIA 38 481-489
KENYA	BARINGO	0.5	36	?	?	HRD	ARABIENSIS	2		MNZAVA ET AL 1994, JAMCA 10:507
KENYA	BUNYALA (IRR. SCHEME)	0.1	34.05	12-Aug	1974	HRD	ARABIENSIS	2	33	HIGHTON (PERS COMM)
KENYA	BUNYALA (IRR. SCHEME)	0.1	34.05	12-Aug	1974	HRD	GAMBIAE	2	2	HIGHTON (PERS COMM)
KENYA	BUNYALA (IRR. SCHEME)	0.1	34.05	3	1974	HRD	ARABIENSIS	1,2	7	R.I.
KENYA	BUNYALA (IRR. SCHEME)	0.1	34.05	3	1974	OSA	ARABIENSIS	1,2	7	R.I.
KENYA	GARASHI	-3.1	40	4	1981	MBO	ARABIENSIS	2	2	MOSHA & SUBRA 1982
KENYA	GARASHI	-3.1	40	4	1981	MBO	GAMBIAE	2	5	MOSHA & SUBRA 1982
KENYA	GARASHI	-0.1	40	4	1981	MBO	MERUS	2	2	MOSHA & SUBRA 1982
KENYA	HOLA AREA	-1.25	39.6333	1	1975	HRD	ARABIENSIS	2	32	HIGHTON (PERS COMM)
KENYA	HOLA AREA	-1.25	39.6333	6	1975	HRD	ARABIENSIS	2	27	HIGHTON (PERS COMM)
KENYA	HOLA AREA	-1.25	39.6333	?	1976	HRD	ARABIENSIS	2	248	HIGHTON (PERS COMM)
KENYA	HOLA AREA	-1.25	39.6333	1	1975	HRD	GAMBIAE	2	1	HIGHTON (PERS COMM)
KENYA	HOLA AREA	-1.25	39.6333	6	1975	HRD	GAMBIAE	2	2	HIGHTON (PERS COMM)
KENYA	HOLA AREA	-1.25	39.6333	?	1976	HRD	GAMBIAE	2	5	HIGHTON (PERS COMM)
KENYA	HOLA AREA	-1.25	39.6333	6	1981	MBO	ARABIENSIS	2	24	MOSHA & SUBRA 1982
KENYA	HOLA AREA	-1.25	39.6333	6	1981	HRD	ARABIENSIS	2	32	MOSHA & SUBRA 1982
KENYA	HOLA AREA	-1.25	39.6333	10	1974	HRD	ARABIENSIS	1,2	2	R.I.
KENYA	HOLA AREA	-1.25	39.6333	1	1975	HRD	ARABIENSIS	1,2	4	R.I.
KENYA	HOLA AREA	-1.25	39.6333	8	1975	HRD	ARABIENSIS	1,2	16	R.I.

KENYA	HOLA AREA	-1.25	39.6333	6	1976	HRD	ARABIENSIS	1,2,3	57	R.I.; MILES 1978
KENYA	JIMBO	-4.6	39.2333	3-Jan	1979	MB?	GAMBIAE	7	16	MOSHA & MUTERO 1982
KENYA	JIMBO	-4.6	39.2333	6-Apr	1979	MB?	GAMBIAE	7	4	MOSHA & MUTERO 1982
KENYA	JIMBO	-4.6	39.2333	12-Oct	1979	MB?	GAMBIAE	7	23	MOSHA & MUTERO 1982
KENYA	JIMBO	-4.6	39.2333	3-Jan	1979	MB?	MERUS	7	71	MOSHA & MUTERO 1982
KENYA	JIMBO	-4.6	39.2333	6-Apr	1979	MB?	MERUS	7	40	MOSHA & MUTERO 1982
KENYA	JIMBO	-4.6	39.2333	12-Oct	1979	MB?	MERUS	7	59	MOSHA & MUTERO 1982
KENYA	JIMBO	-4.6	39.2167	8-May	1981	HRD	ARABIENSIS	2	1	MOSHA & SUBRA 1982; MOSHA & PETRARCA 1983
KENYA	JIMBO	-4.6	39.2167	8-May	1981	HRD	GAMBIAE	2	1	MOSHA & SUBRA 1982; MOSHA & PETRARCA 1983
KENYA	JIMBO	-4.6	39.2167	8-May	1981	HRD	MERUS	2	56	MOSHA & SUBRA 1982; MOSHA & PETRARCA 1983
KENYA	JIMBO	-4.6	39.2167	8-May	1981	MBO	MERUS	2	236	MOSHA & SUBRA 1982; MOSHA & PETRARCA 1983
KENYA	KAPKUIKUI (Baringo DIST)	0.6333	36.2833	6/91-	Aug-92	HRD	ARABIENSIS	2	666	MNZAVA ET AL 1994
KENYA	KAPKUIKUI (BARINGO DIST)	0.6333	36.2833	6/91-	Aug-92	OSA	ARABIENSIS	2	78	MNZAVA ET AL 1994
KENYA	KARIMA	- 0.6667	37.3	6-Dec	86-87	HRD	ARABIENSIS	2	9525	IJUMBA ET AL 1990
KENYA	KARIMA	- 0.6667	37.3	6-Dec	86-87	MBO	ARABIENSIS	2	1278	IJUMBA ET AL 1990
KENYA	KARIMA	- 0.6667	37.3	6-Dec	86-87	OWT	ARABIENSIS	2	454	IJUMBA ET AL 1990
KENYA	KARIMA	- 0.6667	37.3	6-Dec	86-87	ILT	ARABIENSIS	2	242	IJUMBA ET AL 1990
KENYA	KARIMA	- 0.6667	37.3	6-Dec	86-87	OSA	ARABIENSIS	2	3872	IJUMBA ET AL 1990
KENYA	KILIFI (NEAR MOMBASA)	- 3.6167	39.8333	5-4	94-95	MBI, MBO	ARABIENSIS	5		MBOGO ETAL 96, MVE 10:251
KENYA	KILIFI (NEAR MOMBASA)	- 3.6167	39.8333	5-4	94-95	MBI, MBO	GAMBIAE	5		MBOGO ETAL 96, MVE 10:251
KENYA	KILIFI (NEAR	-	39.8333	5-4	94-95	MBI, MBO	MERUS	5		MBOGO ETAL 96, MVE 10:251

	MOMBASA)	3.6167								
KENYA	KIMILILI (ON TANA RIVER)	0.7833	34.6667	6	1980	HRD	GAMBIAE	2	6	R.I.
KENYA	KISIAN	- 0.0667	34.6833	11	1990	HRD	ARABIENSIS	2	144	GITHEKO ET AL 1993
KENYA	KISIAN	- 0.0667	34.6833	4	1991	HRD	ARABIENSIS	2	84	GITHEKO ET AL 1993
KENYA	KISIAN	- 0.0667	34.6833	11	1990	HRD	GAMBIAE	2	188	GITHEKO ET AL 1993
KENYA	KISIAN	- 0.0667	34.6833	4	1991	HRD	GAMBIAE	2	340	GITHEKO ET AL 1993
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	HRD	ARABIENSIS	2	64	MA ET AL 1990
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	ARD	ARABIENSIS	2	93	MA ET AL 1990
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	HRD	GAMBIAE	2	385	MA ET AL 1990
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	ARD	GAMBIAE	2	18	MA ET AL 1990
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	ABN	ARABIENSIS	2	129	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	HRD	ARABIENSIS	2	60	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Sep	1987	ABN	ARABIENSIS	2	61	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Sep	1987	HRD	ARABIENSIS	2	106	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Sep	1987	MBI	ARABIENSIS	2	37	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Sep	1987	MBO	ARABIENSIS	2	1	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Sep	1987	OSA	ARABIENSIS	2	4	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	ABN	GAMBIAE	2	21	PETRARCA ET AL 1991
KENYA	KISIAN	- 0.0667	34.6833	11-Oct	1986	HRD	GAMBIAE	2	417	PETRARCA ET AL 1991
KENYA	KISIAN	-	34.6833	11-Sep	1987	ABN	GAMBIAE	2	8	PETRARCA ET AL 1991

		0.0667								
KENYA	KISIAN	-	34.6833	11-Sep	1987	HRD	GAMBIAE	2	516	PETRARCA ET AL 1991
		0.0667								
KENYA	KISIAN	-	34.6833	11-Sep	1987	MBI	GAMBIAE	2	68	PETRARCA ET AL 1991
		0.0667								
KENYA	KISIAN	-	34.6833	11-Sep	1987	OSA	GAMBIAE	2	9	PETRARCA ET AL 1991
		0.0667								
KENYA	KISUMU	-	34.3833	3	90	HRD	GAMBIAE	5		VULULE ETAL 94, MVE 8:71
		0.1667								
KENYA	KISUMU AREA	-	34.3833	?	1952	HRD	GAMBIAE	1	a	DAVIDSON & JACKSON 1962; MOORES 1953
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Mar	1975	HRD	ARABIENSIS	2	1322	HIGHTON ET AL 1979
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Mar	1975	OSA	ARABIENSIS	2	178	HIGHTON ET AL 1979
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Mar	1975	HRD	GAMBIAE	2	144	HIGHTON ET AL 1979
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Mar	1975	OSA	GAMBIAE	2	25	HIGHTON ET AL 1979
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Jan	1972	HRD	ARABIENSIS	2	450	JOSHI ET AL 1975
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Jan	1972	OSA	ARABIENSIS	2	65	JOSHI ET AL 1975
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Jan	1972	HRD	GAMBIAE	2	1375	JOSHI ET AL 1975
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Jan	1972	OSA	GAMBIAE	2	29	JOSHI ET AL 1975
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Jun	1971	HRD	GAMBIAE	2	32	JOSHI IN SERVICE 1972
		0.1667								
KENYA	KISUMU AREA	-	34.3833	8-Jul	1976	HRD	GAMBIAE	2,3	12	MILES 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7	1981	MBO	GAMBIAE	2	7	MOSHA & SUBRA 1982
		0.1667								
KENYA	KISUMU AREA	-	34.3833	8-Jul	1968	HRD	ARABIENSIS	1	11	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	11-Sep	1968	OSN	ARABIENSIS	1	9	R.I.
		0.1667								

KENYA	KISUMU AREA	-	34.3833	12	1968	MBO	ARABIENSIS	1	1	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12	1968	ARD	ARABIENSIS	1	1	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	3-Dec	1969	OSN	ARABIENSIS	1,2	10	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	3-Jan	1969	HRD	ARABIENSIS	1,2	1	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Apr	1969	HRD	ARABIENSIS	1,2	5	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Apr	1969	OSN	ARABIENSIS	1,2	22	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Apr	1969	OSA	ARABIENSIS	1,2	2	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	2-Dec	1970	ABN	ARABIENSIS	1,2	20	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	3	1974	HRD	ARABIENSIS	1,2	17	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	8-Jul	1968	HRD	GAMBIAE	1	33	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	11-Sep	1968	OSN	GAMBIAE	1	7	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	3-Dec	1969	OSN	GAMBIAE	1,2	2	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	3-Jan	1969	HRD	GAMBIAE	1,2	5	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Apr	1969	HRD	GAMBIAE	1,2	3	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Apr	1969	MBI	GAMBIAE	1,2	1	R.I.
		0.1667								
KENYA	KISUMU AREA	-	34.3833	5-Apr	1976	HRD	ARABIENSIS	1,2,3	31	R.I.; MILES 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1969	L	ARABIENSIS	2	203	SERVICE 1970B
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1969	HRD	ARABIENSIS	2	167	SERVICE 1970B
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1969	OSN	ARABIENSIS	2	187	SERVICE 1970B
		0.1667								

KENYA	KISUMU AREA	-	34.3833	12-Nov	1969	L	GAMBIAE	2	375	SERVICE 1970B
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1969	HRD	GAMBIAE	2	428	SERVICE 1970B
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1969	OSN	GAMBIAE	2	45	SERVICE 1970B
		0.1667								
KENYA	KISUMU AREA	-	34.3833	11	1969	LIGHT TRAP	ARABIENSIS	2	15	SERVICE 1970C
		0.1667								
KENYA	KISUMU AREA	-	34.3833	11	1969	LIGHT TRAP	GAMBIAE	2	38	SERVICE 1970C
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1971	HRD	ARABIENSIS	2	449	SERVICE 1973
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1971	L	ARABIENSIS	2	706	SERVICE 1973
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1971	HRD	GAMBIAE	2	495	SERVICE 1973
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1971	L	GAMBIAE	2	107	SERVICE 1973
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7	1974	L	ARABIENSIS	2	151	SERVICE 1976B
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	L	ARABIENSIS	2	1066	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	HRD	ARABIENSIS	2	297	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	OSN	ARABIENSIS	2	240	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	OSA	ARABIENSIS	2	36	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	MB	ARABIENSIS	2	3	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	OWT	ARABIENSIS	2	1	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	L	GAMBIAE	2	1137	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	HRD	GAMBIAE	2	667	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	OSN	GAMBIAE	2	57	SERVICE ET AL 1978
		0.1667								

KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	OSA	GAMBIAE	2	9	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	MB	GAMBIAE	2	3	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	LIGHT TRAP	GAMBIAE	2	1	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	6-Apr	1975	ABO	GAMBIAE	2	1	SERVICE ET AL 1978
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Jun	1971	HRD	ARABIENSIS	2	27	WHITE & MUNISS 1971
		0.1667								
KENYA	KISUMU AREA	-	34.3833	7-Jun	1971	HRD	GAMBIAE	2	13	WHITE & MUNISS 1971
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1970	HRD	ARABIENSIS	2	405	WHITE 1972
		0.1667								
KENYA	KISUMU AREA	-	34.3833	12-Nov	1970	HRD	GAMBIAE	2	72	WHITE 1972
		0.1667								
KENYA	LAMBWE VALLEY	-	34.25	11	1970	HRD	ARABIENSIS	2	5	WHITE & MUNISS 1970
		0.6667								
KENYA	LAMBWE VALLEY	-	34.25	11	1970	HRD	GAMBIAE	2	2	WHITE & MUNISS 1970
		0.6667								
KENYA	LODWAR	3.1	35.6333	7	1970	HRD	ARABIENSIS	1,2	7	R.I.
KENYA	MAJI MOTO	-	35.7	4	1973	MBI	ARABIENSIS	1,2	4	R.I.
		1.3333								
KENYA	MAJI-NDEGE (Baringo DIST)	0.6333	36.2833	6/91-	Aug-92	HRD	ARABIENSIS	2	600	MNZAVA ET AL 1994
KENYA	MAJI-NDEGE (Baringo DIST)	0.6333	36.2833	6/91-	Aug-92	OSA	ARABIENSIS	2	126	MNZAVA ET AL 1994
KENYA	MALINDI AREA	-3.25	39.9667	5	1973	MBI	GAMBIAE	1,2	1	R.I.
KENYA	MALINDI AREA	-3.25	39.9667	6-May	1975	HRD	GAMBIAE	1	1	R.I.
KENYA	MALINDI AREA	-3.25	39.9667	6-May	1975	HRD	MERUS	1,2,3	14	R.I.; MILES 1978
KENYA	MALINDI AREA	-3.25	39.9667	?	?	?	MERUS	7	a	VAN SOMEREN ET AL 1955
KENYA	MARA GAME PARK	-1.5	35.5	6	1974	HRD	GAMBIAE	1,2	1	R.I.
KENYA	MOMBASA (JARDINI)	-	39.5167	4	1967	MBI	GAMBIAE	2	2	WHITE 1968
		4.0667								
KENYA	MWEA TEBERE	-0.5	37.6833	6	1981	HRD	ARABIENSIS	2	3	MOSHA & SUBRA 1982

KENYA	MWEA TEBERE	-0.5	37.6833	6-May	1970	MBI	ARABIENSIS	1,2	3	R.I.
KENYA	MWEA TEBERE	-0.5	37.6833	6-May	1970	OSN	ARABIENSIS	1,2	10	R.I.
KENYA	MWEA TEBERE	-0.5	37.6833	6-May	1970	HRD	ARABIENSIS	1,2	14	R.I.
KENYA	NYANZA PROV (AHERO)	-0.5	34.5	?	?	MBI	ARABIENSIS	5		WALKER ETAL 96 JAMCA 12:172
KENYA	SARADIDI	-	34	11-Oct	1986	HRD	ARABIENSIS	2	10	MA ET AL 1990
KENYA	SARADIDI	0.0667	34	11-Oct	1986	HRD	GAMBIAE	2	58	MA ET AL 1990
KENYA	SARADIDI	0.0667	34	11-Oct	1986	HRD	ARABIENSIS	2	11	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Sep	1987	ABN	ARABIENSIS	2	10	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Sep	1987	HRD	ARABIENSIS	2	67	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Sep	1987	MBI	ARABIENSIS	2	12	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Oct	1986	HRD	GAMBIAE	2	63	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Sep	1987	HRD	GAMBIAE	2	272	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Sep	1987	MBI	GAMBIAE	2	24	PETRARCA ET AL 1991
KENYA	SARADIDI	0.0667	34	11-Sep	1987	MBO	GAMBIAE	2	2	PETRARCA ET AL 1991
KENYA	TAVETA AREA	3.3833	37.6667	1	1961	HRD	GAMBIAE	1	a	DAVIDSON & JACKSON 1962
KENYA	TAVETA AREA	3.3833	37.6667	5	1981	HRD	ARABIENSIS	2	65	MOSHA & SUBRA 1982
KENYA	TAVETA AREA	3.3833	37.6667	5	1981	MBO	ARABIENSIS	2	28	MOSHA & SUBRA 1982
KENYA	TAVETA AREA	3.3833	37.6667	5	1981	MBO	GAMBIAE	2	1	MOSHA & SUBRA 1982
KENYA	TAVETA AREA	3.3833	37.6667	6	1967	HRD	ARABIENSIS	1	3	R.I.
KENYA	TAVETA AREA	3.3833	37.6667	6	1969	HRD	ARABIENSIS	1,2	6	R.I.

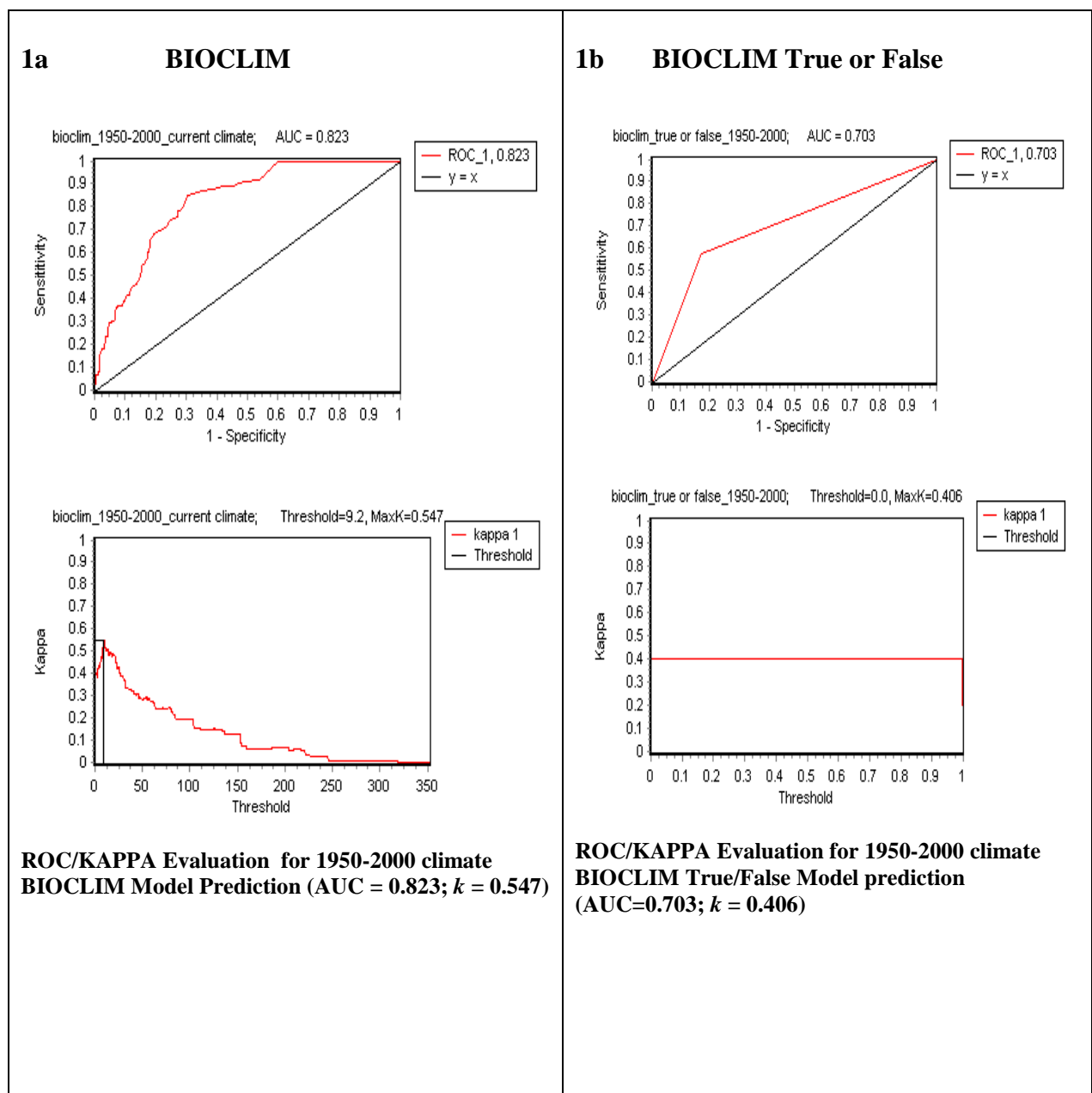
KENYA	TAVETA AREA	-	37.6667	6	1969	OSN	ARABIENSIS	1,2	2	R.I.
		3.3833								
KENYA	TAVETA AREA	-	37.6667	6	1980	HRD	ARABIENSIS	2	5	R.I.
		3.3833								
KENYA	TAVETA AREA	-	37.6667	3	1970	OSA	ARABIENSIS	2	3	WHITE & MUNISS 1970
		3.3833								
KENYA	TAVETA AREA	-	37.6667	3	1970	OSN	ARABIENSIS	2	20	WHITE & MUNISS 1970
		3.3833								
KENYA	TAVETA AREA	-	37.6667	3	1970	HRD	ARABIENSIS	2	247	WHITE & MUNISS 1970
		3.3833								
KENYA	TAVETA AREA	-	37.6667	3	1970	HRD	GAMBIAE	2	31	WHITE & MUNISS 1970
		3.3833								
KENYA	TAVETA AREA	-	37.6667	8	1971	HRD	ARABIENSIS	2	70	WHITE & MUNISS 1971
		3.3833								
KENYA	TAVETA AREA	-	37.6667	8	1971	HRD	GAMBIAE	2	10	WHITE & MUNISS 1971
		3.3833								
KENYA	TURKANA DISTRICT	4	36	6-8	94	HRD, OSN	ARABIENSIS	5		CLARKE, ODIALLA ETAL 1996; TRSTM&H 90 302-304

APPENDIX B: MODEL VALIDATION OUTCOMES

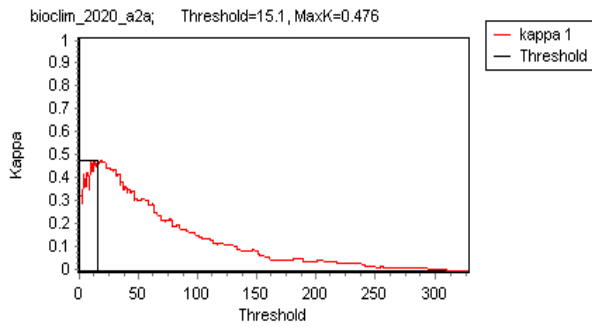
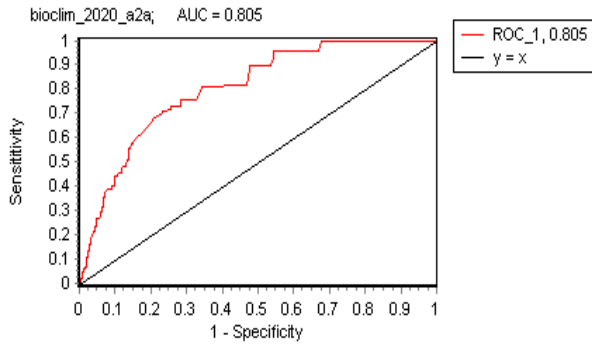
B.1 Validation Results for HADCM3 Predictions

Ecological Niche Modeling was executed with HADCM3 future climate projection resulting in different suitability classes for malaria vectors to thrive. Section B.I.1 shows the graphs for AUC and kappa statistics from BIOCLIM and BIOCLIM True or False predictions. The validation outcomes for DOMAIN model are shown in section B.I.2.

B.1.1 HADCM3 BIOCLIM and BIOCLIM True or False Model Validation Results

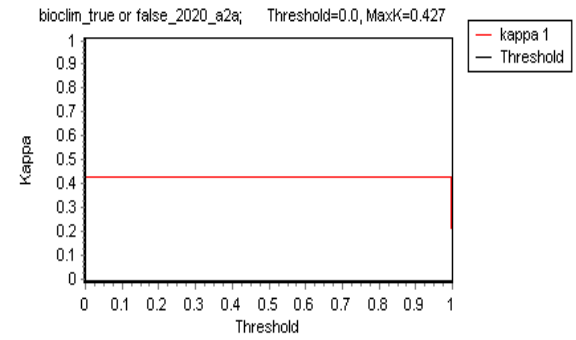
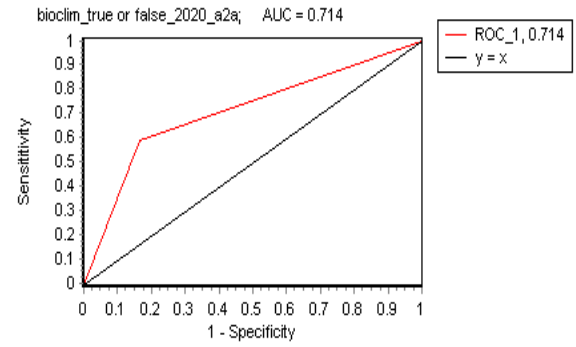


2a BIOCLIM



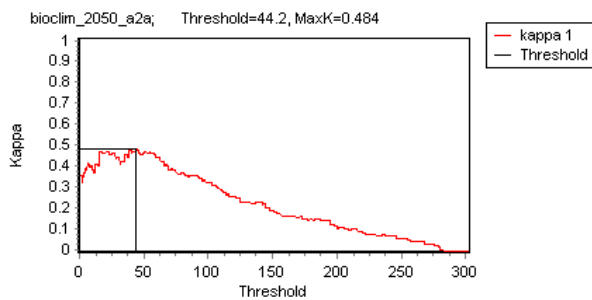
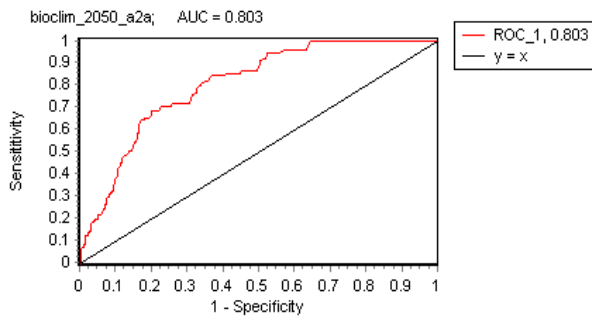
ROC/KAPPA Evaluation for HADCM3_2020_A2a climate BIOCLIM Model Prediction (AUC = 0.805; $k = 0.476$)

2b BIOCLIM True or False



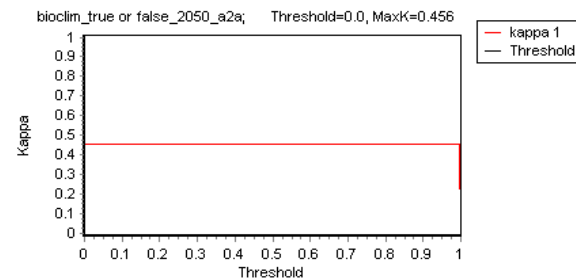
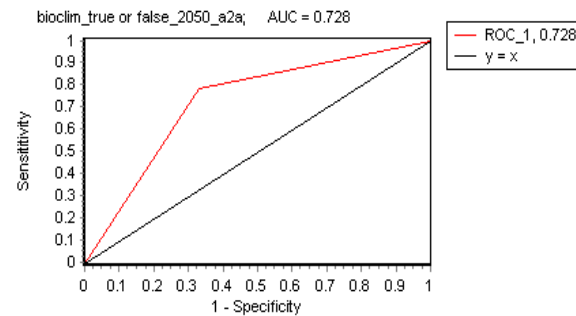
ROC/KAPPA Evaluation for HADCM3_2020_A2a climate BIOCLIM True/False Model prediction (AUC = 0.714; $k = 0.427$)

3a BIOCLIM



ROC/KAPPA Evaluation for HADCM3_2050_A2a climate BIOCLIM Model Prediction (AUC=0.803; $k = 0.484$)

3b BIOCLIM True or False



ROC/KAPPA Evaluation for HADCM3_2050_A2a climate BIOCLIM True/False Model prediction (AUC = 0.728; $k = 0.456$)

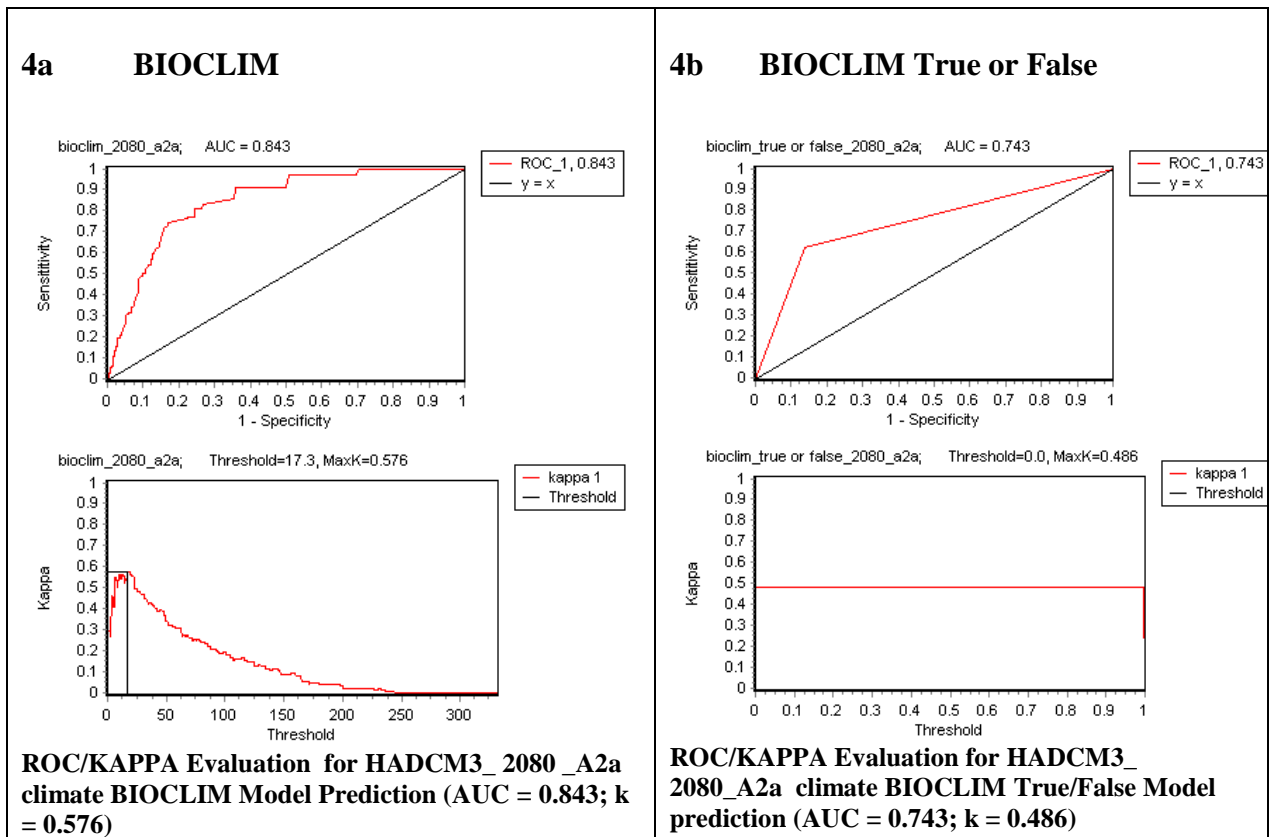
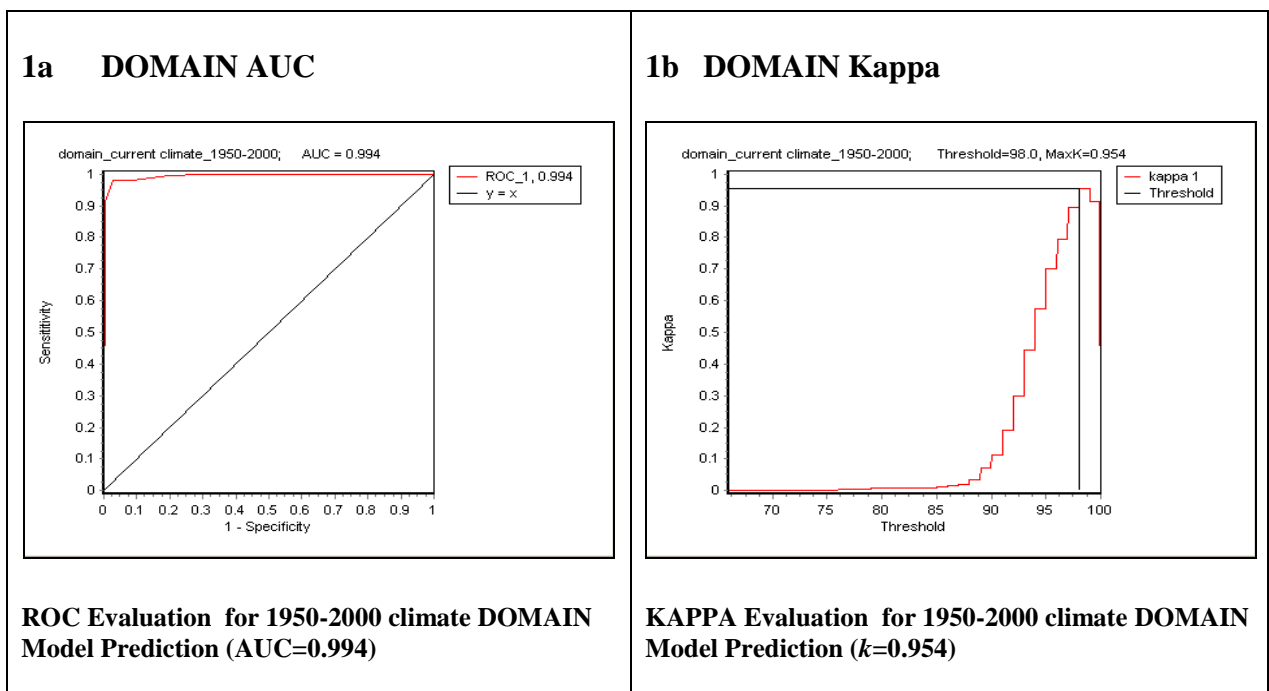


Fig. B. 1: BIOCLIM Model Validation Results for HADCM3 predicted malaria vector distribution cases, Bioclim (a), BIOCLIM True/False (b).

B.1.2 HADCM3 DOMAIN Model Validation Results



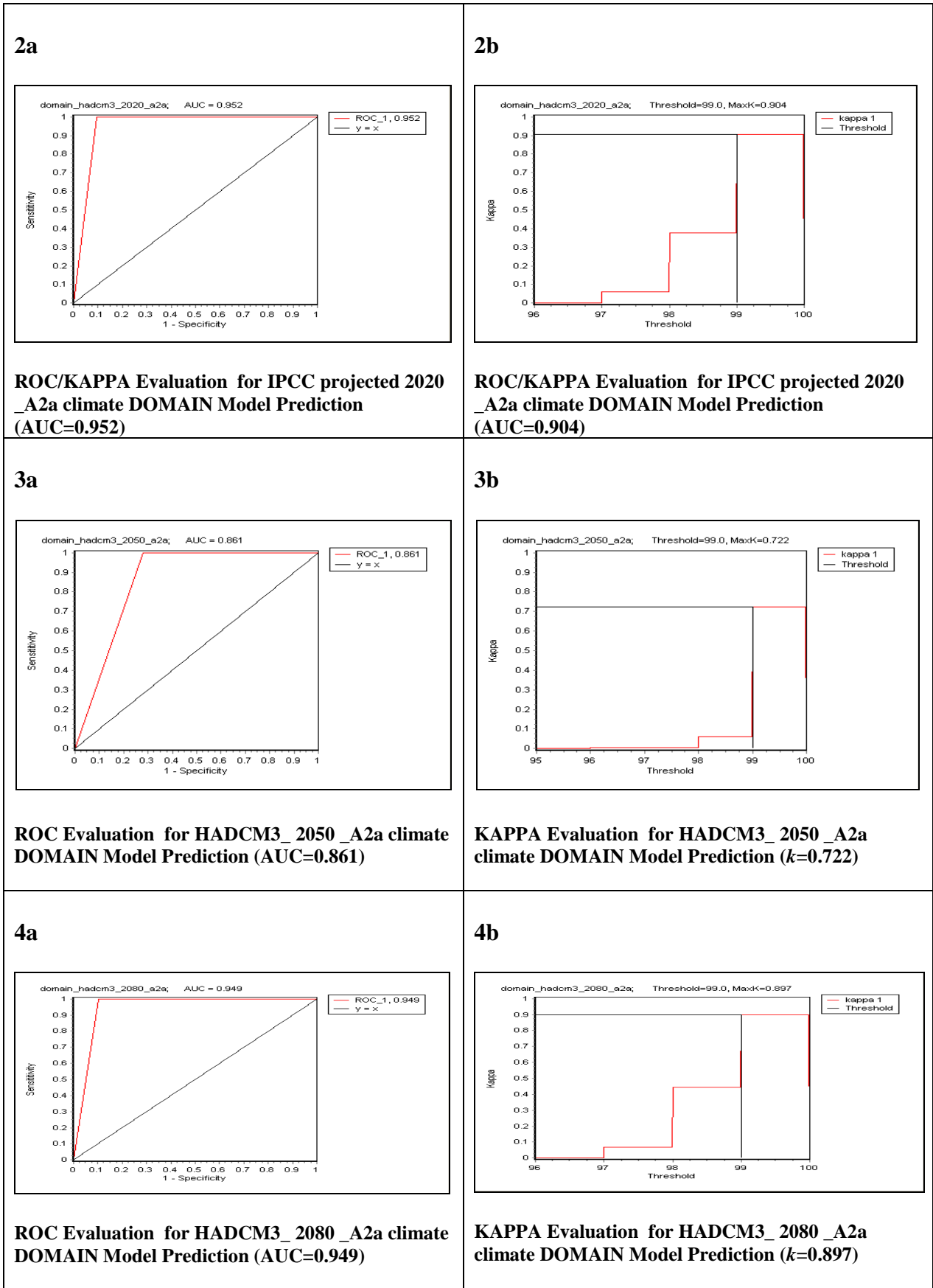
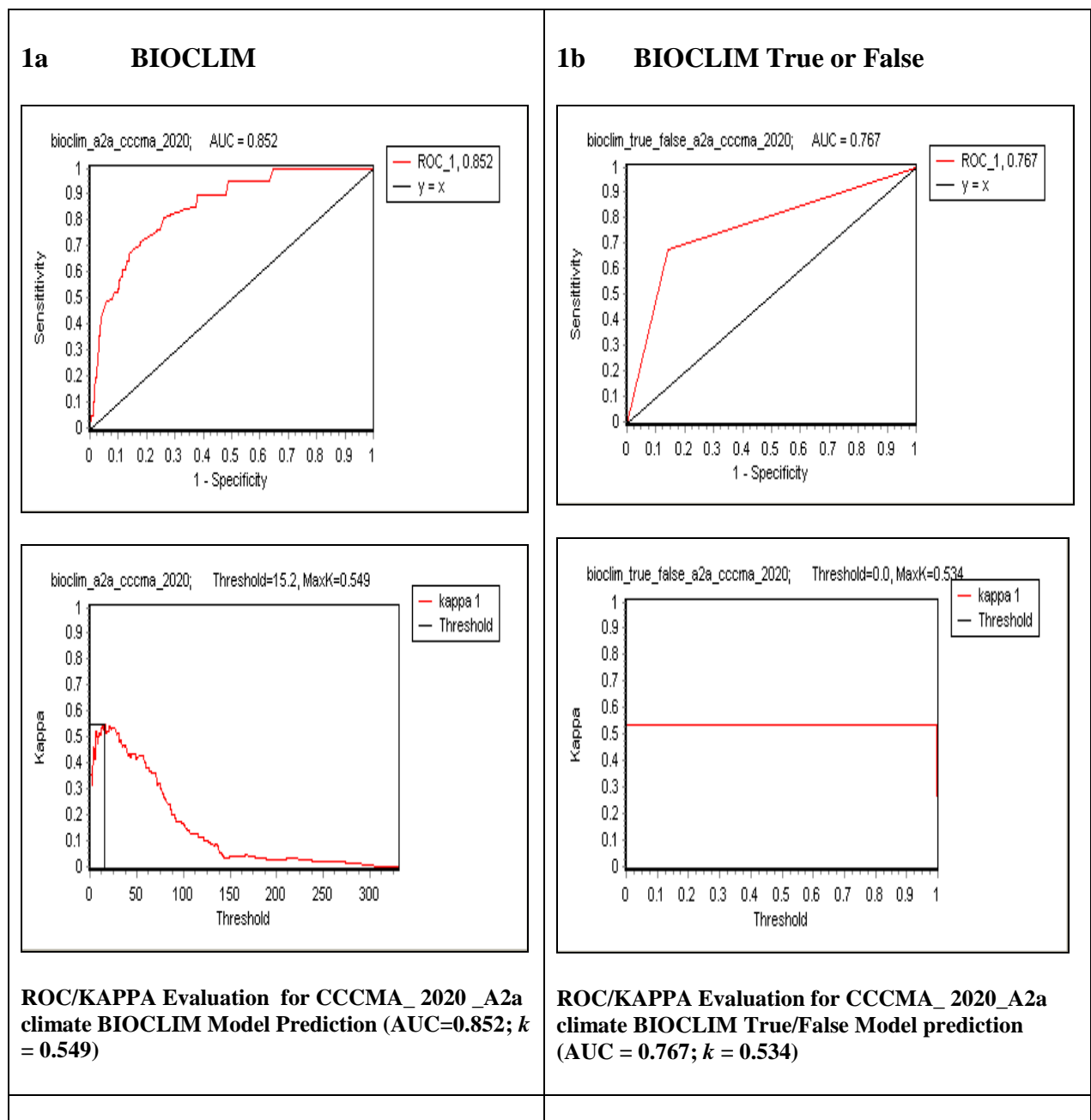


Fig. B. 2: DOMAIN Model Validation Results for HADCM3 predicted malaria vector distribution cases, ROC (1a, 2a, 3a) and Kappa (1b, 2b, 3b)

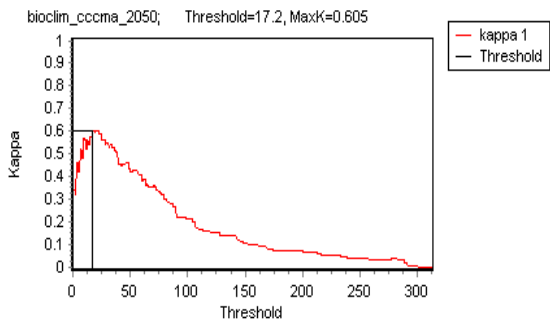
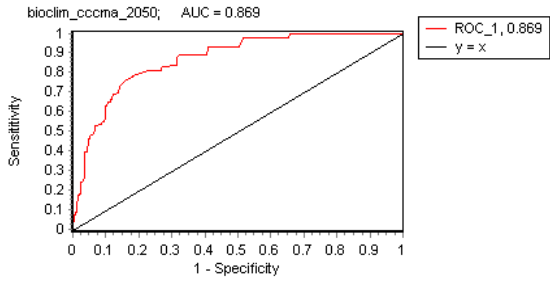
B.2 Validation Results for CCCMA Predictions

When Ecological Niche Modeling was done with CCCMA future climate projection, different suitability classes for malaria vectors to thrive emerged. Section B.2.1 shows the graphs for AUC and kappa statistics from BIOCLIM and BIOCLIM True or False predictions. The validation outcomes for DOMAIN model are shown in section B.2.2.

B.2.1 CCCMA BIOCLIM and BIOCLIM True or False Model Validation Results

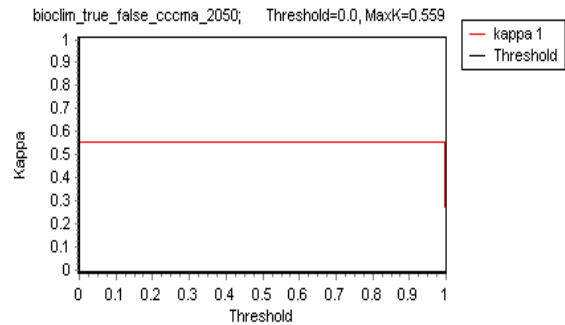
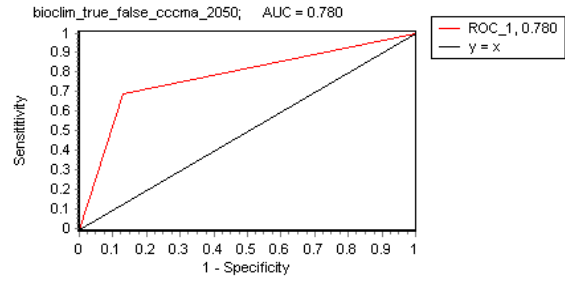


2a BIOCLIM



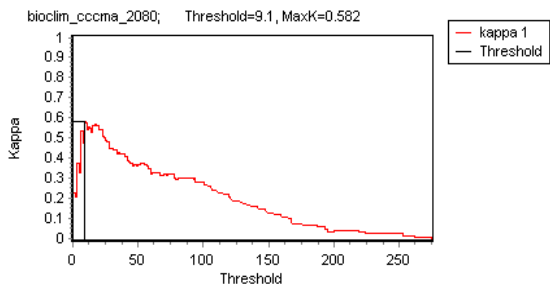
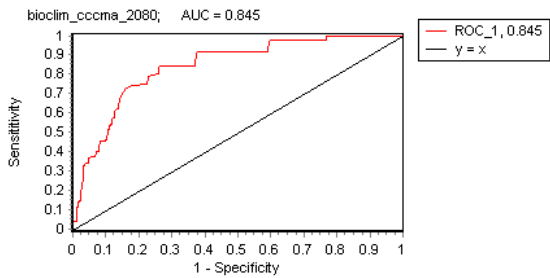
ROC/KAPPA Evaluation for CCCMA_2050_A2a climate BIOCLIM Model Prediction (AUC=0.869; $k = 0.805$)

2b BIOCLIM True or False



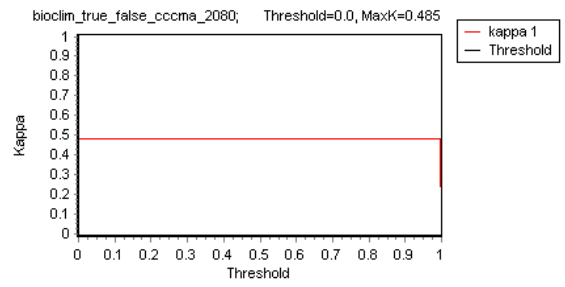
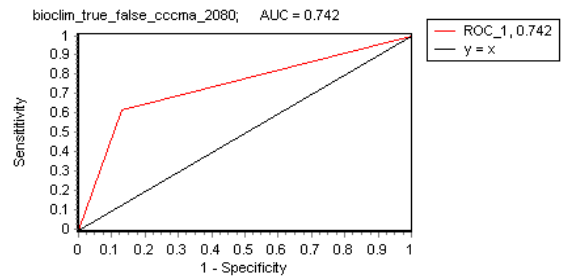
ROC/KAPPA Evaluation for CCCMA_2050_A2a climate BIOCLIM True/False Model prediction (AUC = 0.780; $k = 0.559$)

3a BIOCLIM



ROC/KAPPA Evaluation for CCCMA_2080_A2a climate BIOCLIM Model Prediction (AUC=0.845; $k = 0.582$)

3b BIOCLIM True or False



ROC/KAPPA Evaluation for CCCMA_2050_A2a climate BIOCLIM True/False Model prediction (AUC = 0.780; $k = 0.485$)

Fig. B. 3: BIOCLIM Model Validation Results for CCCMA predicted malaria vector distribution cases, Bioclim (a), BIOCLIM True/False (b)

B.2.2 CCCMA DOMAIN Model Validation Results

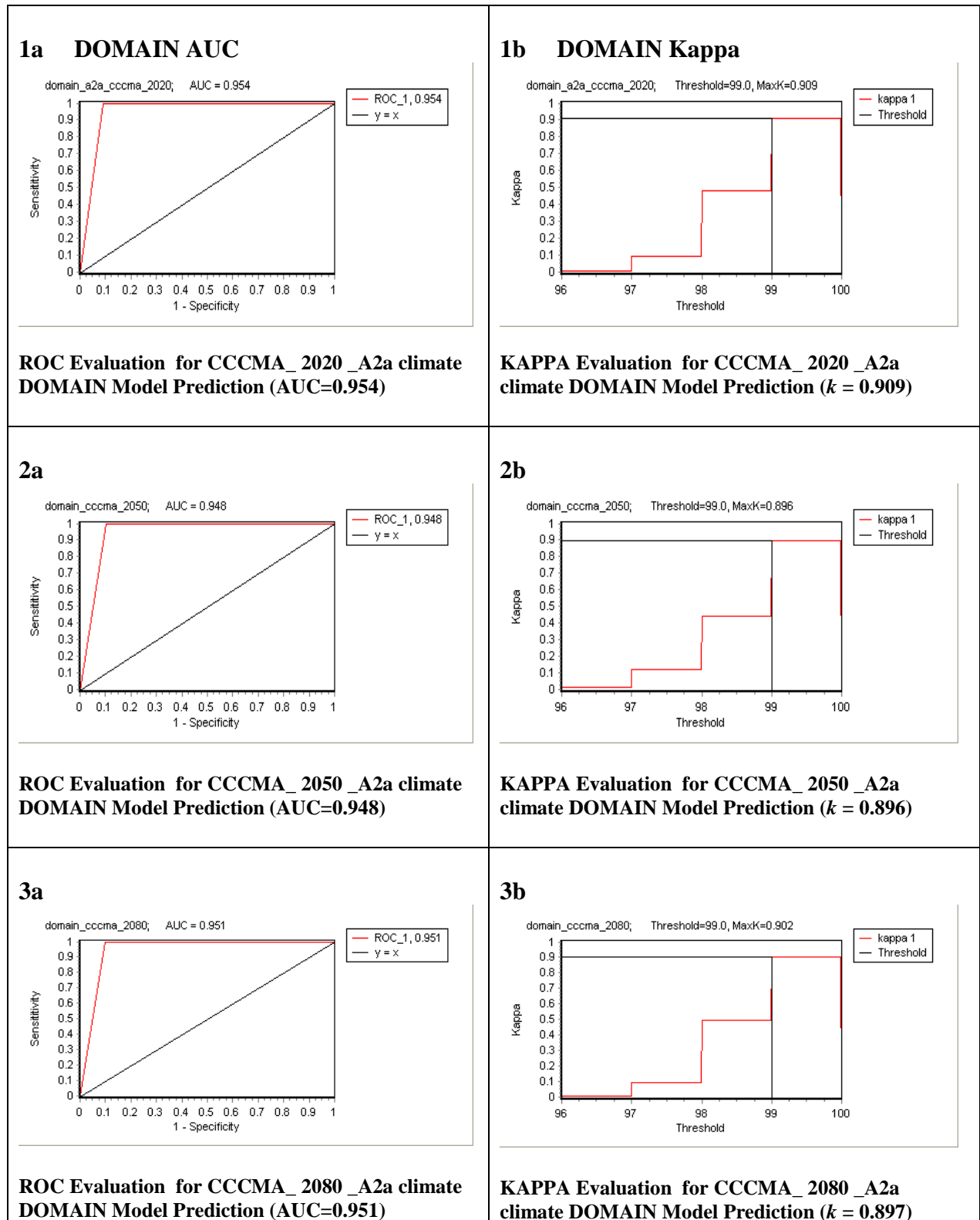
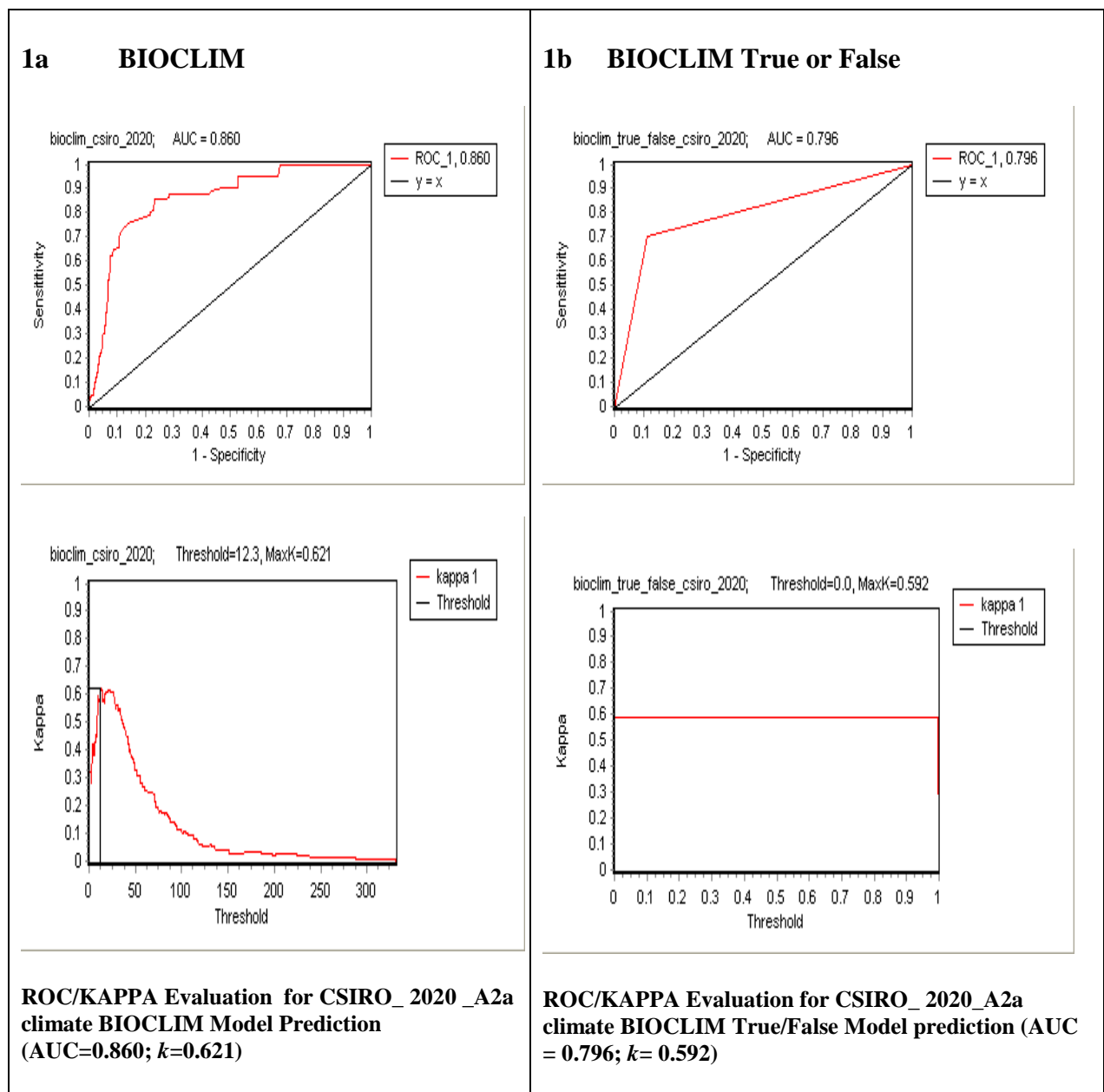


Fig. B. 4: DOMAIN Model Validation Results for CCCMA predicted malaria vector distribution cases, ROC (1a, 2a, 3a) and Kappa(1b, 2b, 3b)

B.3 Validation Results for CSIRO Predictions

Ecological Niche Modeling performed with CSIRO future climate projection, different suitability classes for malaria vectors to thrive emerged. Section B.3.1 shows the graphs for AUC and kappa statistics from BIOCLIM and BIOCLIM True or False predictions. The validation outcomes for DOMAIN model are shown in section B.3.2.

B.3.1 CSIRO BIOCLIM and BIOCLIM True or False Model Validation Results



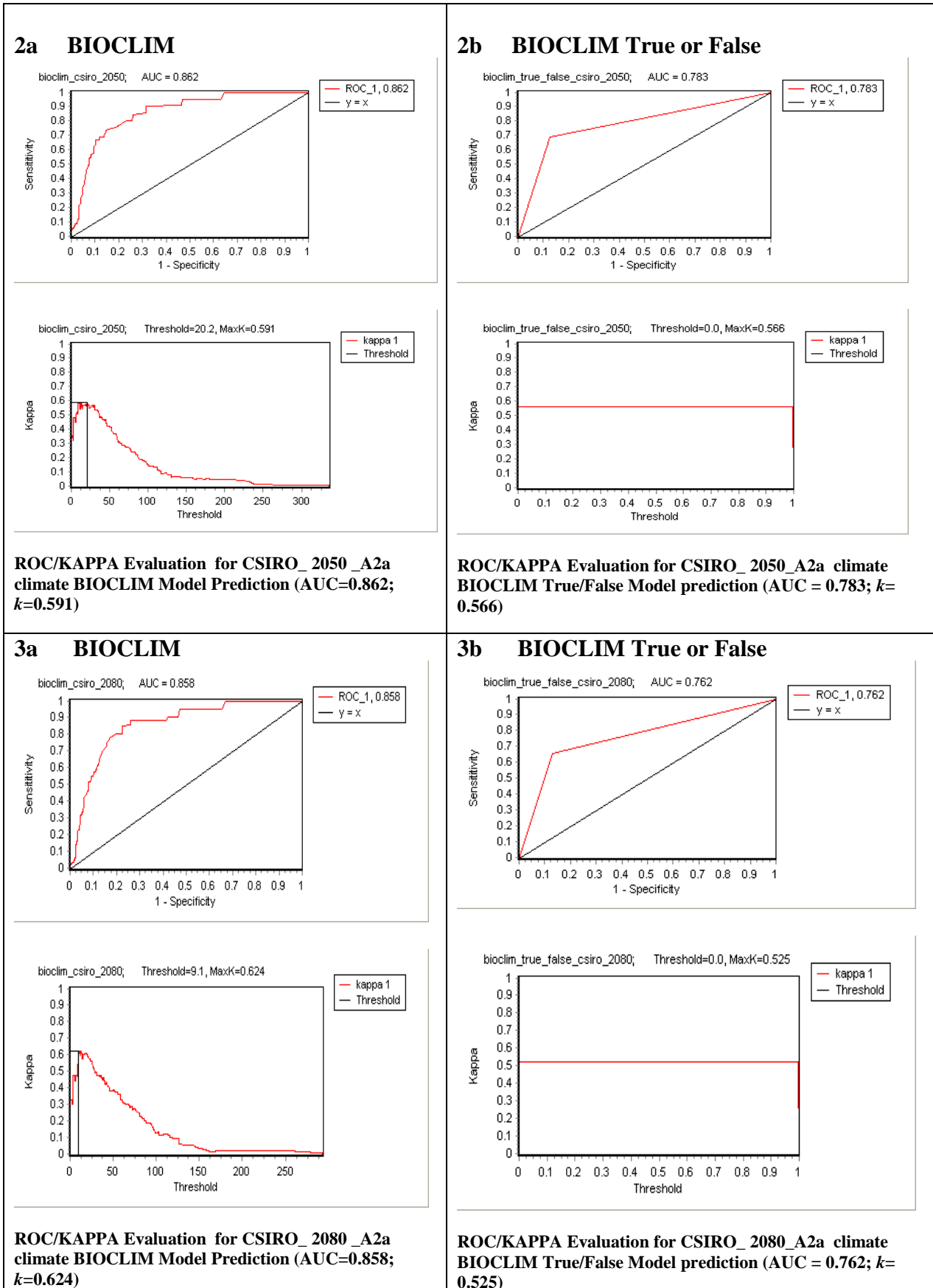


Fig. B. 5: BIOCLIM Model Validation Results for CSIRO predicted malaria vector distribution cases, BIOCLIM (a), BIOCLIM True/False (b)

B.3.2 CSIRO DOMAIN Model Validation Results

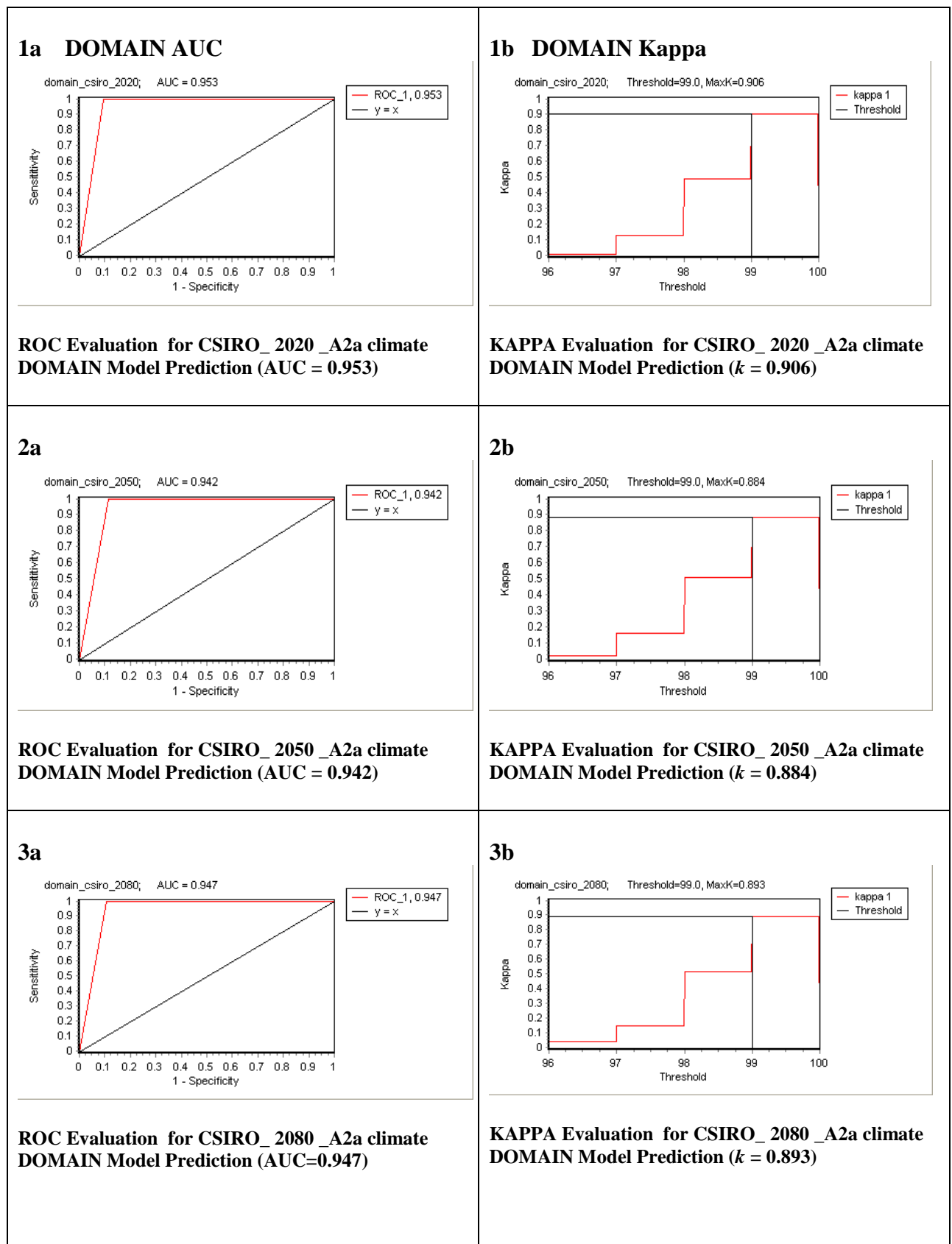


Fig. B. 6: DOMAIN Model Validation Results for CSIRO predicted malaria vector distribution cases, ROC (1a, 2a, 3a) and Kappa (1b, 2b, 3b).