Use of Geospatial Techniques to Improve Bee Farming and Bee Health Across Four Main Agroecological Zones in Kenya

David Masereti Makori

(214585752)

A thesis submitted in fulfilment of the academic requirements of

Doctor of Philosophy in Geography

School of Agricultural, Earth & Environmental Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

February 2023

Abstract

Amid augmented climate change and anthropogenic influence on natural environments and agricultural systems, the global socioeconomic and environmental value of bees is undisputed. Bee products such as honey, pollen, nectar, royal jelly and to a lesser extent bee venom are important supplemental sources of income generation especially in the underdeveloped rural African areas. Moreover, bee farming is an important incentive for forest conservation, biodiversity and ecosystem services in terms of pollination services. Bee pollination services play a vital role in crop production, hence directly contribute to food and nutritional security for African smallholder farmers. Nevertheless, bee farming and bee health in general are under threat from climate change, agricultural intensification and associated habitat alteration, agrochemicals intensification, bee pests and diseases. Therefore, there is need to establish spatial distribution of bees, their food substrates, floral cycle and biotic and abiotic threats, especially bee pests. Bee pests devastate bee colonies through physical injury and as vectors of pathogens, hence causing a considerable reduction in bee colony productivity. Thus, this study sought to establish geospatial techniques that could be used to improve bee farming and bee health in Kenya. Firstly, this study aimed to determine the spatial and temporal distribution of stingless bees in Kenya using six machine learning ecological niche approaches and non-conflating variables from both bioclimatic, vegetation phenology and topographic features. All machine learning algorithms used herein performed at an 'excellent' level with a true skills statistics (TSS) score of up to 0.91. Secondly, the study assessed the suitability of resampled multispectral data for mapping melliferous (flowering plants that produce substance used by bees to produce honey) plants in Kenya. Bi-temporal AISA Eagle hyperspectral images, resampled to four sensors' (i.e., WorldView-2, RapidEye, Spot-6 and Sentinel-2) spatial and spectral resolutions, and a RF classifier were used to map melliferous plants. Melliferous plants were successfully mapped with up to 93.33%

overall accuracy using WorldView-2. Furthermore, the study predicted the distribution of four main bee pests (*Aethina tumida*, *Galleria mellonella*, *Oplostomus haroldi* and *Varroa destructor*) in Kenya using the maximum entropy (MaxEnt) model and random forest (RF) classifier. The effect of seasonality on the abundance of bee pests was apparent, as indicated by the Wilcoxon rank sum test, with up to 6.35 times more pests in the wet than the dry season. Furthermore, bioclimatic variables especially precipitation contributed the most (up to 77.8%) to all bee pest predictions, while vegetation phenology provided vital information needed to sharpen the prediction models at grain level due to their higher spatial resolution and seasonal and phenological features. Moreover, topography had a moderate influence (14.3%) on the distribution of bee pests. Also, there was a positive correlation between bee pests' abundance, habitat suitability and high altitude. Anthropogenic influence (as depicted by human footprint data) on the distribution of bee pests was relatively low (1.2%) due to the availability of a variety of bee food substrate from the mixed land use/land cover (LULC) classes, especially farmlands. Using the Pearson correlation coefficient, the prediction models for all bee pests scored at an excellent level (0.84), except for the *G. mellonella* prediction model, which was ranked 'fair' (0.55). Due to the relatively high accuracy for models developed herein to map stingless bees' distribution, melliferous plants and bee pests' occurrence and abundance, this study concluded that the models developed could reliably be used to indicate high suitability areas for bee farming. They could also be used to predict high bee pests risk areas for mitigation and management purposes, hence improving bee health and hive productivity.

Keywords:

Bee health; stingless bees; honeybee pests; melliferous plants; climate change; vegetation phenology; remote sensing; hyperspectral data; multispectral data; machine learning (ML); ecological niche modelling; ecosystem services

PREFACE

This research was carried out in the School of Agricultural, Earth and Environmental Sciences (SAEES), University of KwaZulu-Natal (UKZN), Pietermaritzburg Campus, under the supervision of Prof. Onisimo Mutanga, Prof. John Odindi and Dr. Elfatih M. Abdel-Rahman.

I would like to declare that the research reported in this thesis has never been submitted in any form to any other university and represents my original work, except where due acknowledgments are made.

David Masereti Makori

Signed: [redacted] Date: 10th February, 2023

As the candidate's supervisor, we certify the above statement and have approved this thesis for submission.

Prof. Onisimo Mutanga

Signed: [redacted] Date: 10th February, 2023

Prof. John Odindi

Signed: [redacted] Date: 10th February, 2023

Dr. Elfatih M. Abdel-Rahman

Signed: [redacted] Date: 10th February, 2023

DECLARATION 1: Plagiarism

I, **David Masereti Makori**, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research,
2. This thesis has not been submitted for any degree or examination at any other university,
3. This thesis does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons,
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written, but the general information attributed to them has been referenced;
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced;
5. This thesis does not contain text, graphics, or tables copied and pasted from the internet, unless specifically acknowledged and the source detailed in the thesis and in the references section.

David Masereti Makori

Signed: 

Date: 10th February, 2023

DECLARATION 2: Publications

1. **Makori,** D. M., Fombong, A. T., Abdel-Rahman, E. M., Kiatoko, N., Ongus, J., Irungu, J., Mosomtai, G., Makau, S., Mutanga, O., Odindi, J., Raina, S. and Landmann, T. 2017. Predicting spatial distribution of key honeybee pests in Kenya using remotely sensed and bioclimatic variables: key honeybee pests distribution models, International Journal of Geo-Information, 6(66), https://www.mdpi.com/2220-9964/6/3/66
2. **Makori**, D. M., Abdel-Rahman E. M., Landmann, T., Mutanga, O., Odindi, J., Nguku, E., Tonnang, H. E., and Raina, S. 2020. Suitability of resampled multispectral datasets for mapping flowering plants in the Kenyan savanna, PloSONE, 15(9), e0232313, https://doi.org/10.1371/journal.pone.0232313
3. **Makori**, D. M., Abdel-Rahman, E. M., Ndungu, N., Odindi, J., Mutanga, O., Landmann, T., Tonnang H. E. Z. and Kiatoko, N. 2022. The use of multisource spatial data for determining the proliferation of stingless bees in Kenya, GIScience and Remote Sensing, 59(1), 648–669, https://doi.org/10.1080/15481603.2022.2049536
4. **Makori**, D. M., Abdel-Rahman, E. M., Odindi, J., Mutanga, O., Landmann, T. and Tonnang H. E. Z. 2023. Multi-pronged abundance prediction of bee pests' spatial distribution in Kenya, In review.

In the above manuscripts, conceptualisation, designing of experimental protocol, data collation, collection and analysis, and manuscript preparation was achieved by the candidate, **David Masereti Makori**. This was done under the supervision of Prof. Onisimo Mutanga, Prof. John Odindi and Dr. Elfatih M. Abdel-Rahman.

David Masereti Makori

Signed: 

Date: 10th February, 2023

ACKNOWLEDGEMENTS

My sincere gratitude goes to Prof. Onisimo Mutanga, Prof. John Odindi, Prof. Elfatih M. Abdel-Rahman, Dr. Tobias Landmann and Dr. Henri E. Z. Tonanng. The patience, energy and determination extended to me by these great geospatial scientists during conceptualization, data collection, manuscript preparation and corrections is immense. I am a better and sharper scientist because of your collective effort. I would also like to acknowledge Prof. Suresh Raina, Dr. Evelyne Nguku and Dr. Mercy Ojoyi for helping me out during the initial stages of my PhD.

I would like to express my heartfelt gratitude to bee farmers in Kenya who provided valued information for this research, without which this research may not have been successful. My gratitude goes to Mutemi and Munywoki, farmers in Kasanga area in Mwingi for dedicating their time to help me carry out my field work. I would also like to thank the personnel who participated in field sampling for response variables and information gathering. In addition, I am grateful for the support granted by Darrel Makori, Lloyd Magara, Darlene Moraa, Sidi Mangi, Lydia Nthenya, Felix Sereti, Fr. Gabriel Dolan, Meg Ryan, Tara Madden, Salome Nduta, Simphiwe Sidu, Dr. Nkoba Kiatoko, Dr. Nelly Ndungu, Ms. Emily Kimathi and Dr. Bester Mudereri for data collation, processing, analysis and support. I would also like to acknowledge the effort of Sebit Diyar, Saum Hassan, Henry Oindo, George Gesaka and Davis Okoth for providing me with the necessary assistance during data processing and manuscript preparation. I am very thankful to Dolorosa Osogo and other colleagues in *icipe* for proof reading and giving technical comments to various manuscripts that formed part of this thesis.

I would like to appreciate the European Union and NORAD for providing resources utilized in the field sampling exercises. I am grateful to the Almighty God for giving me the impetus to soldier on despite many challenges I faced in the course of my PhD. Finally, to my family and friends who stood by me when the tides were at their strongest; may the Almighty God bless you and yours abundantly!

To whom much is given, much is expected. The title 'Doc' demands a lot all round. To this end, I am now ready to wear that title with honour. I hereby permit you all to call me **Doc**!

DEDICATION

To my parents

Ms. Priscah Moraa Makori and the Late Leonard Makori.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

1 General Introduction

A honeybee (*Apis mellifera*) on a pollination spree. Image by Madrigal (2022); source https://www.naturettl.com/how-to-photograph-bees/

1.1 Background

The African savanna is characterized by unreliable rainfall with low and dispersed pockets of natural vegetation. Unreliable rainfall is to a large extent unable to sufficiently support rainfed agriculture, necessitating for diversification of income to supplement the unpredictable crop yield (Makori et al., 2017; Raina et al., 2011). Apiculture and related ecosystem services such as pollinator activities boost local economies, food and nutritional security and biodiversity. Hence it is a valuable and sustainable socioecological practice in the African savanna (Kiatoko et al., 2014; Klein et al., 2007).

Commercially beneficial insects such as bees have been demonstrated as important conservation incentives especially to communities living near forests (Raina et al., 2011; Warui et al., 2018a). This is because they are important as income sources due to products such as honey, wax, dyes and silk on one hand (Adolkar et al., 2007; Kasina et al., 2009), and pollination services in agricultural and forests ecosystems on the other hand. Furthermore, they have been shown to play a vital role in conservation of degraded forests and dilapidated land (Blay et al., 2008; Raina et al., 2011). An estimated 60% to 90% of plant species are dependent on both bees, birds and other animals for pollination (Potts et al., 2010; Raina et al., 2011). For instance, pollination services are estimated to contribute to about $14 billion of agricultural economy every year in the USA alone (Hines and Hendrix, 2005; Klein et al., 2007). However unequivocal the contribution of pollination might be in Africa, its economic importance is yet to be fully established and documented. Moreover, the abundance and spatial distribution of stingless bees, honey bees, bee pests and their food substrate, especially flowers is yet to be well understood.

1.1.1 Stingless bees

Also known as meliponine bees, stingless bees belong to the Meliponini tribe. They are eusocial (with high level of social organization) in nature and comprise of hundreds of species that are highly abundant and diverse in the African continent (Iraheta et al., 2015; Michener, 2013). There are over 800 meliponine species documented worldwide with distinctions in body morphology, behaviour and colony size. However, there are only six genera that have been documented on the African continent, that is *Hypotrigona, Cleptotrigona, Liotrigona, Plebeina, Dactylurina* and *Meliponula* (Eardley and Kwapong, 2013; Michener, 2013). There are over 20 species endemic to the African continent, of which 12 have been documented in Kenya (Eardley, 2004; Michener, 2000). The moliponine bees are characterised by a generalist pollination (polylactic) nature, hence beneficial to numerous crops and wild plants

(Njoya, 2010; Slaa et al., 2006). In this regard, they supplement honeybees (*Apis mellifera*), which have since been the major pollinators of crops and wild plants (Garantonakis et al., 2016; Giannini et al., 2015; Winfree et al., 2007). Honeybees have in the past been affected by among others climate change, intensification of agriculture and pesticides, pests and diseases (Makori et al., 2017; Muli et al., 2014a; Pirk et al., 2015). This has led to colony collapse hence threatening bee population and the ecosystem services they offer. Therefore, there is need to study on alternative pollinators such as meliponine bees that may supplement honeybees in provision of products and ecosystem services such as pollination. Indeed, studies such as Kiatoko et al. (2022) and Slaa et al. (2006) have demonstrated improved fruit quality and seed set in crops pollinated by meliponine bees. In effect, this has substantially increased interest in meliponiculture (meliponine bee farming) around the African continent.

Furthermore, production of stingless bee honey and other by products has improved livelihoods of rural communities in Africa (Cortopassi-Laurino et al., 2006; Kiatoko et al., 2014). Meliponine honey and their by-products are used as a source of income, food and medicine due to its anti-carcinogenic and antibiotic properties, which are associated with their unique enzymes (Bijlsma et al., 2006; Libério et al., 2009; Mokaya et al., 2021). This translates to higher market prices compared to regular honey (Kumar et al., 2012). Additionally, by products such as propolis, cerumen and pollen have properties associated with immunomodulatory effects, antioxidants, antibacterial and anti-inflammatory effect, which are desirable for medicinal purposes (Araujo et al., 2012; Libério et al., 2009).

Nevertheless, the spatial and temporal distribution of stingless bees on the African continent and Kenya in particular has not been well explored. The anthropogenic factors and environmental conditions that influence the distribution of meliponine bees remain largely undocumented (MM Fierro et al., 2012; Iraheta et al., 2015). Therefore, efforts to exploit technological advancement in the fields of machine learning, artificial intelligence and ecological niche modelling is highly desirable.

These techniques provide statistical pathways aimed at establishing spatial distribution of meliponine bees in a spatiotemporal spectrum (Fernández and Hamilton, 2015; Kearney and Porter, 2009).

1.1.2 Flower mapping

In the African Savanna, communities adjacent to the forests either directly or indirectly depend on various forest resources and ecosystem services (Kiatoko et al., 2014; Klein et al., 2007). This is because forest habitats are vital to different stages of various bee species (Allsopp et al., 2008; Kremen et al., 2007). Therefore, apiculture could be incentivised to improve conservation in such areas. Practices such as placement of apiaries in close proximity to forest stands has been demonstrated to double hive productivity as compared to those placed further from such ecosystems (Sande et al., 2009). Furthermore, identification and utilization of the right types and amounts of nectar and pollen by bee farmers has a correlation with the quality of hive products, hive productivity, and ability and agility of bees to fend off enemies such as pests and diseases (Pirk et al., 2015; Sande et al., 2009; Warui et al., 2018a).

Climate change and habitat alteration from natural land uses to disturbed ones such as farms and unvegetated land, reduces pollen and nectar available to pollinators (Ricketts et al., 2008). Furthermore, a reduction or elimination of certain pollen and nectar sources could have a negative impact to source-specific pollinators which rely on flower colour, pollen type, flower morphology or physiology (Ushimaru et al., 2007). This is because insects rely on visual signals such as colour, shape and size for the choice of flowers to visit (Cooley et al., 2008; Holzschuh et al., 2007; Pohl et al., 2011). For instance, hummingbirds prefer red coloured flowers, flies like pale colours, while butterflies and bees prefer brightly coloured flowers (Cooley et al., 2008; Holzschuh et al., 2007; Lunau and Maier, 1995; Pickering and Stock, 2003; Pohl et al., 2011). This provides a basis on which conservation measures for flowering plants around vulnerable communities who depend on both agriculture and apiculture for

livelihood are anchored. Therefore, identification of the type, location, intensity and health of flowering plants is essential. Techniques, methods and tools that enables this objective to be met are highly desirable. Studies such as Landmann et al. (2015) mapped flowering plants in Kenya using hyperspectral and hyperspatial datasets. The cost of acquiring these datasets were however inhibitory and the technical skills and software required to process the data are not easily available. Before this research was conducted, techniques developed around readily and freely available datasets, and less technical skills and software had not been developed.

1.1.3 Bee pests

Threats such as habitat alteration, agricultural and agrochemical intensification, climate change, bee diseases and pests negatively affect apiculture (Muli et al., 2014b; Zayed, 2009a). Particularly, bee pests affect bees directly through inflicting physical injury or indirectly as vectors of pathogens (Mumoki et al., 2014; Ongus et al., 2017). The ability of bee pests to adversely affect bee colonies and their ability to traverse and thrive across diverse spatial extents and agroecologies, has attracted considerable scientific interest (Fombong et al., 2012; Muli et al., 2014b; Ongus et al., 2017). The common and economically important bee pests are the varroa mites (*Varroa distructor*), large hive beetles (*Oplostomus haroldi*), small hive beetles (*Aethina tumida*) and wax moths (*Galleria mellonella*) (Fombong et al., 2012; Muli et al., 2014b; Pirk et al., 2016; Torto et al., 2010; Zayed, 2009a). Studies such as Fombong et al. (2012), Muli et al. (2014b) and Torto et al. (2010) have established that these pests are able to thrive in different agroecological gradients. However, the spatio-temporal distribution of these pests in Kenya and Africa has not been adequately established. This research was cognizant that this knowledge lacuna could be addressed by geospatial and earth observation techniques and tools, by using vegetation phenology, topological and bioclimatic datasets, and efficient and cutting-edge AI, ML and EN modelling techniques.

1.2 Aim and objectives

The study aimed at developing geospatial techniques that could be used to improve bee farming and bee health in Kenya.

Specifically, the study sought to;

i. Determine the spatial proliferation of stingless bees in Kenya using multi-source spatial datasets,
ii. Map melliferous plants in Kenya using resampled multispectral data datasets,
iii. Predict the spatial distribution of bee pests in Kenya using bioclimatic and vegetation phenological datasets, and
iv. Predict the spatial distribution of bee pests' abundance (population) using bioclimatic and land use/ land cover datasets

1.3 Scope of the study

The study sought to develop geospatial techniques that could be used to improve bee farming and bee health in Kenya. Firstly, the study established the spatial distribution and distribution of stingless bees in four main agroecological zones, covering 40 administrative counties in Kenya using bioclimatic, vegetation phenology and derived topographical variables. Occurrence data for stingless bees was collected from 19 counties within the modelling region. The study also limited itself to six machine learning algorithms and an ensemble modelling approach to develop prediction models for determining the spatial distribution of stingless bees. Additionally, simulation models were developed to establish spatial distribution of stingless bees in a future epoch (2055).

Secondly the study assessed the utility of satellite-based multispectral datasets with different spatial resolution and RF classifier for mapping flowering plants in the

Kasanga region in Kitui County, Kenya. The study region covered 7.88 km^2 and used the AISA Eagle imagery with high spatial and spectral resolution. The hyperspectral datasets had a spectral range from 400 to 1,000 nm with 64 spectral bands and a spatial resolution of 10 cm. These images were captured in two flowering periods (i.e. onset and peak flowering periods). They were further resampled to the specifications of four multispectral sensors (WorldView-2, RapidEye, Spot-6 and Sentinel-2). Due to the nature of the data sets used in this study, and owing to the cost of capturing the AISA Eagle hyperspectral images used to simulate multispectral sensor datasets used in this study, and further due to the cost of tasking the WorldView-2 satellite to collect the actual 2014 image used to calibrate the flowering models developed in this study, a smaller study site covering 7.88 km^2 and located within Kitui County was selected. The site had a range of differently coloured melliferous plants that could offer a wide range of spectral mix. In addition, this site had well established apiculture practices more than any other site within the region. Moreover, this site is within one of the counties covered by other objectives in this thesis.

Thirdly, the spatial distributions of four species of bee pests were predicted using the maximum entropy (MaxEnt) model (Phillips et al., 2006). Both the response and predictor variables were collected from four main agroecological zones within the Coastal, Mwingi, Kakamega and Mount Kenya regions. The models were also restricted to these spatial extents. Remotely sensed vegetation datasets were used to derive a number of phenological variables. The vegetation phenological variables improved the predictive ability of all models used. Fourthly, bee pests' abundance data, bioclimatic datasets, vegetation phenological data and human footprint data used as a proxy for anthropogenic influence, were used to improve bee pests' prediction techniques previously developed in Makori et al. (2017). The geographical scope used in Makori et al. (2017) was also adopted herein. The RF algorithm was used to predict the spatial abundance of the four species of bee pests, which was then projected to the year 2055.

1.4 Thesis outline

The thesis is comprised of six chapters. The first chapter consists of the general introduction of the study while the last chapter comprises of a synthesis. Chapter two through to five are semi-autonomous comprising of stand-alone research papers. These chapters are made up of the introduction, methodology, results, discussion and conclusion of the individual research papers. Three of the papers were published in reputable peer-reviewed journals, while one paper (Chapter 5) has been submitted for review. The three stand-alone chapters (two through five) have been arranged chronologically to depict the beneficial insects (stingless bees in Chapter 2) which form the basis for beehealth, their food substrate (flowering plants in Chapter 3) and their threats (pests in Chapters 4 and 5).

Chapter one gives the highlights of the study and the **general introduction**. It consists of the background of the study, aims and objectives of the study and the thesis structure.

Chapter two establishes and models the spatial distribution of **stingless bees** in Kenya using spatial modelling techniques, and both bioclimatic and remotely sensed vegetation phenological variables. It aims at establishing suitable bee farming regions and predicting the effect of climate change on stingless bee farming suitability.

Chapter three highlights the suitability of resampled multispectral datasets in **mapping melliferous plants** in Kenya. It aims at developing scalable flower mapping techniques that are accessible to farmers in Kenya for improved farming management.

Chapter four presents methodologies for predicting the spatial distribution of **bee pests** in Kenya using the MaxEnt models, bioclimatic and remotely sensed vegetation phenological datasets. It aims at establishing high bee pest risk regions in which management efforts could be directed to diminish their impact on reduction of beehive productivity.

Chapter five highlights a multi-pronged approach used to predict the **spatial abundance (population) of bee pests** using bioclimatic and land use/land cover (LULC) data as a proxy for anthropogenic influence. It aims at providing bee pests prediction models that are more accurate and realistic, and which capture bee pest population dynamics and the effect human activities have on the distribution of bee pests.

Chapter six on the other hand provides an **overall synthesis** of the thesis, including study conclusions and recommendations.

All the references used in this study, including those cited in the four main research papers have been combined and listed in the references section at the tail end of this thesis.

CHAPTER TWO

2 The Use of Multisource Spatial Data for Determining the Proliferation of Stingless Bees in Kenya

Stingless bee searching for pollen and nectar. Image by Gomollón-Bel (2020); source https://www.chemistryworld.com/news/rare-trehalulose-sugar-found-in-stingless-bees-honey/4012235.article

This chapter is based on:

Makori, D. M., Abdel-Rahman, E. M., Ndungu, N., Odindi, J., Mutanga, O., Landmann, T., Tonnang H. E. Z. and Kiatoko, N. 2022. The use of multisource spatial data for determining the distribution of stingless bees in Kenya, GIScience and Remote Sensing, 59(1), 648–669, https://doi.org/10.1080/15481603.2022.2049536

Abstract

Stingless/meliponine bees are eusocial insects whose polylactic nature enables interaction with a wide variety of wild plants and crops that enhance pollination, hence support ecosystem services. However, their true potential regarding pollination services and honey production is yet to be fully recognized. Worldwide, there are over 800 species of meliponine bees, with over 20 species documented on the African continent. Out of these, only 12 species have been well documented in Kenya. Moreover, interest on meliponine bees has increased amid climate change, agricultural intensification, and other anthropogenic effects. Generally, stingless bees are under researched with no previous documented evidence of their ecological niche (EN) distribution in most African countries. Hence, this study sought to establish the influence of bioclimatic, topographic and vegetation phenology on their spatial distribution and change patterns. Stingless response variables from 490 sample points were collected and used in conjunction with 11 non-conflating features to build stingless ecological niche models. Six machine learning based EN models were used to predict the distribution of seven stingless bees' species combined. The results from the EN models showed that annual precipitation was the most influential variable to stingless bee distribution (contributing 43.09% of the modelling information), while potential evapotranspiration and temperature seasonality contributed 21.18% of the information needed to predict the spatial distribution of stingless bees. Vegetation phenology (21.36%) and topography (14.36%) had moderate effect on stingless bees' distribution. On the other hand, high seasonality in precipitation and temperature indicated high stingless niche variability in the future (i.e., 2055). The performance of six EN algorithms used to predict distribution of stingless bees were found to be "excellent" for random forest (true skills statistics (TSS) = 0.91), and ranger (TSS = 0.90), and "good" for generalized additive models (TSS = 0.87), multivariate adaptive regression spline (TSS = 0.80), and boosted regression trees (TSS = 0.80), while they were "fair" for recursive portioning and regression trees (TSS = 0.79). These EN

models could be utilized to inform stingless bee farming and insects pollinated crops, by highlighting regions that provide highly suitable conditions for stingless bees. Additionally, the findings could be harnessed to increase both bee and agricultural productivity, and forest conservation efforts through supplementary pollination services.

Keywords:

Stingless bees, climate change, vegetation phenology, artificial intelligence, machine learning, species distribution modelling (SDM).

2.1 Introduction

Stingless bees, also known as meliponine bees, are eusocial insects that comprise of hundreds of highly diverse and abundant species (Iraheta et al., 2015; Michener, 2013). They are widely spread, with over 800 documented species worldwide that differ in their body morphology, colony size and behavior. There are six genera of stingless bees found in Africa; *Hypotrigona, Cleptotrigona, Liotrigona, Plebeina, Dactylurina* and *Meliponula* (Eardley and Kwapong, 2013; Michener, 2013), with over 20 species documented and endemic to the African continent, and 12 species documented in Kenya (Eardley, 2004; Michener, 2000). The polylactic nature of meliponine bees is of particular importance with regard to pollination services they provide to numerous crops and wild plants (Njoya, 2010; Slaa et al., 2006). They are known to supplement honeybees (*Apis mellifera)*, which have over the years been the major pollinators of wild plants and crops (Garantonakis et al., 2016; Giannini et al., 2015; Winfree et al., 2007). Generally, honeybee colonies are on the decline due to colony collapse disorder occasioned by the effects of climate change, agricultural intensification, use of pesticides and bee diseases and pests (Makori et al., 2017; Muli et al., 2014a; Pirk et al., 2015). This necessitates for inclusion of other pollinators such as stingless bees, to complement honeybee pollination services. Indeed, it has been demonstrated that stingless bees improve the quality of fruits and seed set in crops (Kiatoko et al., 2022, 2014; Slaa et al., 2006), hence the interest in meliponiculture has substantially grown around the African continent.

Furthermore, livelihoods of many rural African communities have improved due to the production of high quality meliponine bee honey and other bee keeping products (Cortopassi-Laurino et al., 2006; Kiatoko et al., 2014) used as food, medicine and a source of income. Stingless bee honey has antibiotic and anti-carcinogenic properties (Libério et al., 2009) due to the enzymes associated with these properties (Bijlsma et al., 2006; Mokaya et al., 2021). The medicinal value of meliponine bee honey translates into higher price in the market compared to regular honey (Kumar et al., 2012). Also

stingless bee cerumen, propolis and pollen contain anti-oxidants, immunomodulatory effect, antibacterial and anti-inflammatory properties that are important for medical purposes (Araujo et al., 2012; Libério et al., 2009).

Generally, the spatial and temporal distribution of stingless bees in Kenya and the African continent at large has not been well explored. The environmental and anthropogenic factors and conditions necessary for the distribution of stingless bees remain largely undocumented (MM Fierro et al., 2012; Iraheta et al., 2015). These environmental conditions include ambient temperature, precipitation, humidity, altitude, soil wetness and solar irradiance and other factors such as shading (Ackerly et al., 2010; Iraheta et al., 2015). Furthermore, the seasonal pattern of vegetation growth influences stingless bees population dynamics (Miller et al., 2009; Williams et al., 2012), as their forage circles are within short distances (250-500 m) of their nesting sites (Grundel et al., 2010; Torné-Noguera et al., 2014). Since food resources significantly influence their spatial distribution, stingless bees are more sensitive to changes in local environmental conditions compared to honey and bumble bees (Torné-Noguera et al., 2014). Additionally, stingless bee species respond differently to food resources whose distribution is based on local environmental conditions (MM Fierro et al., 2012; Roulston and Goodell, 2011). Specifically, floral community composition and distribution play a major role in spatial dispersion of stingless bees than other factors, such as nesting substrates (Torné-Noguera et al., 2014).

To understand the role of biotic and abiotic factors variations on stingless bees distribution, artificial intelligence (AI), machine learning (ML) and ecological niche (EN) models are commonly used to provide a statistical pathways that influence their distribution in space and time (Fernández and Hamilton, 2015; Kearney and Porter, 2009). On the other hand, accurate data on location and temporal distribution of stingless bee species can provide an important preview for describing their habitats using reliable EN and species-specific predictions modelling approaches. Therefore, response and predictor variables used in these models should be carefully selected to

increase prediction accuracy (Peterson, 2011; Phillips and Dudík, 2008). In this context overfitting of EN models should be avoided to achieve reliable prediction of species' spread (Cord et al., 2014a). Also, remotely sensed predictor variables, such as vegetation phenology should be incorporated into prediction models to minimize probable biasness brought about by generalist nature of bioclimatic variables. This is because bioclimatic predictor variables are commonly interpolated from discrete observations while vegetation seasonality variables are continuous observations whose resolutions are higher both in the temporal and spatial standpoints (Cord et al., 2014a; Pau et al., 2012).

Although many studies (e.g., Njoya, 2010; Fierro *et al.*, 2012; Kumar *et al.*, 2012; Nkoba *et al.*, 2012; Kiatoko *et al.*, 2014; Torné-Noguera *et al.*, 2014; Iraheta *et al.*, 2015; Ndungu *et al.*, 2019) have documented the importance of biology, nesting patterns and distribution of stingless bees, exploitation of vegetation phenology and its interactions with other biotic and abiotic factors to determine the distribution of stingless bees in Kenya is largely unavailable. Furthermore, vegetation phenology provide detailed vegetation information that is derived in narrow and accurate patterns which could be selected using variable selection procedures in EN models to reduce parameterization and overfitting in prediction, improving model accuracy (Cord et al., 2014a; Naimi and Araújo, 2016). In addition, since future bioclimatic conditions are already modelled and readily available, EN models can offer tools to predict possible change scenarios, hence providing informed decisions to avert unwarranted future climatic effects on stingless bee distribution and biodiversity loss. Therefore, this study sought to establish the influence of bioclimatic, topographic and vegetation seasonality variables on the spatial distribution of stingless bees in Kenya. This study also modelled the future distribution scenarios and patterns of stingless bees using simulated bioclimatic data.

2.2 Methods

2.2.1 Study site

The study area spans four agroecological zones with a representative climatic gradient comprising of 40 administrative counties in Kenya. The area covers the coastal, eastern, Mount Kenya, Nyanza, Rift Valley, and western regions of the country, totalling 256,638.49 square kilometres (Figure 1). The stingless bee occurrence data were collected from nineteen (19) counties, i.e., Kilifi, Mombasa, Taita Taveta and Tana River (in the coastal region), Kitui, Machakos and Makueni (in the eastern region), Baringo, Bomet, Nandi, Narok and Elgeyo Marakwet (in the Rift Valley region), Kisii and Nyamira (in the Nyanza region), Meru (in Mount Kenya region), and Kakamega, Bungoma, Vihiga and Busia (in the western region). The western region, including Nandi Hills (in Rift Valley region), Kisii (in Nyanza region) and Mount (Mt.) Kenya regions are characterized by high humidity, with relatively low temperature and high altitude (Githui et al., 2009). They comprise major natural forest and water towers such as the Mau Complex, Cherangani Hills, Kakamega Forest and Kisii Highlands (Kinyanjui, 2011; Odawa and Seo, 2019). The coastal region, a low altitude area, experiences high humidity and temperature, with one major indigenous forest (Arabuko Sokoke) which is part of the coastal Kenya and Tanzanian forests arc (Schürmann et al., 2020). Natural vegetation in the coastal region is under threat from agricultural intensification and expansion of human settlement (Habel et al., 2017). The eastern region, Tana River (in the coastal region) and Baringo (in the Rift Valley region) receive relatively lower rainfall than other regions, with relatively high temperature during the day. Vegetation is predominantly shrubs with isolated stands of acacia trees (Speranza et al., 2009). Elevation of the study site ranges from 2 meters above sea level at the coastal region to 2,045 meters above sea level in the Nandi Hills (in the Rift Valley region).

Mt. Kenya, Kisii highlands (in Nyanza region) and Western regions are predominantly agricultural areas with relatively high maize productivity, a staple

food crop in Kenya. In addition, stingless bees are used to improve crop productivity both in the field and controlled greenhouse environments (Kiatoko et al., 2022). Small scale subsistence farming is practiced in the coastal and eastern regions due to the relatively low precipitation (Githui et al., 2009; Schürmann et al., 2020).

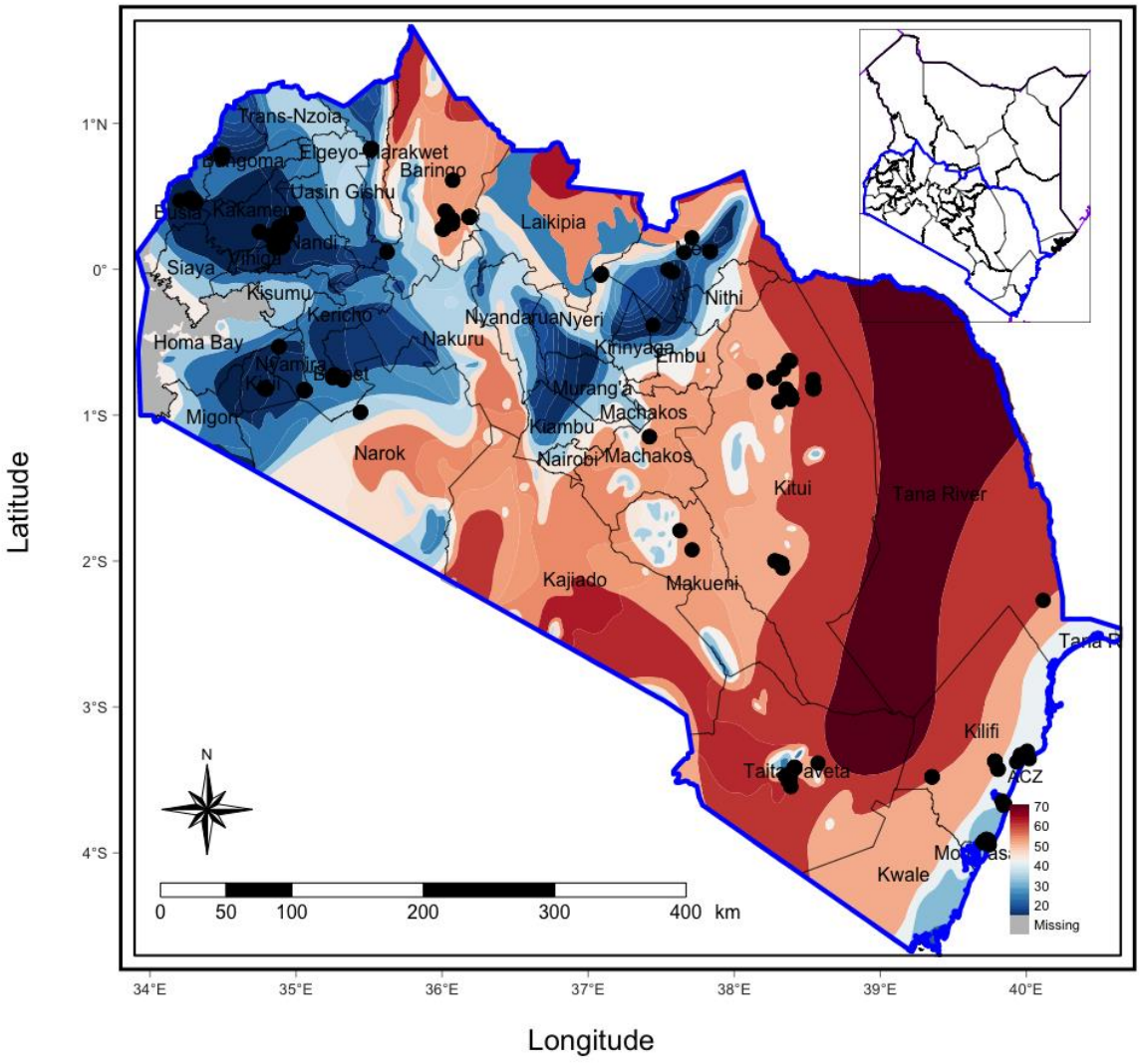


Figure 1: The study area indicating extent of the modelled region (blue outline), the national boundary (in the inset map) in an agroclimatic zone (ACZ) backdrop, ranging from hot and dry (acz id 50 to 70) to cold and wet (acz id 20 to 40) regions and stingless occurrence data (black points).

2.2.2 Data Collection and Pre-Processing

The datasets used in this study were divided into response (stingless bee occurrence) and predictor (bioclimatic, topographic and vegetation phenology) variables (Table 1). The occurrence data were obtained in a point vector format while the predictor variables were satellite raster images. All the predictor variables were clipped to the boundaries of the study site. Since the predictor variables were in different spatial resolutions, hence the bioclimatic and topographic data were resampled to a pixel resolution of 250 m using the Quantum GIS software (QGIS Development Team, 2021). The warp option under the raster processing tool in QGIS was used to edit the resolution of the bioclimatic and topographic variables to fit the resolution of the vegetation phenology variables with intermediate resolution in the predictor variables. This is based on the recommendation by Mudereri et al (2020) that predictor variables of different spatial resolution should be harmonized and resampled to the same resolution, preferably to match the resolution of one of the predictors.

2.2.2.1 Response variables

As aforementioned, stingless bees' occurrence data were collected from nineteen (19) counties in Kenya during the dry (December to February and June to August) and wet (March to May and September to November) seasons, between January 2016 and December 2020 (Figure 1). In each county, meliponaries and colonies observed with the help of the local community within the study area were randomly sampled, with their locations and all species of the stingless bees enumerated. Unidentified stingless bee species were also recorded. Stingless bee presence only data were used for this study since it was logistically impractical to collect 'actual absence' data from the field. Furthermore, lack of sighting during this study was not guarantee of absence (Jiménez-Valverde et al., 2011; Yackulic et al., 2013). It was assumed that the stingless bees had the same probability of occurrence across the landscape, hence every pixel had the same probability of being tagged as "background" pseudo-absence across the

environmental and geographic space of the study area (Naimi and Araújo, 2016). Moreover, the species diversity modelling (sdm) package (Naimi and Araújo, 2016) in R software (R Core Team, 2021) used in this study generates and uses the "pseudo-absence" observations data to compensate for lack of "actual absence" data (Guan et al., 2020).

Data were collected from a total of 490 sample points distributed along the coastal, eastern, Kisii, Mt. Kenya, Rift Valley and the Western regions of the country (Figure 1). Specifically, the counties where stingless bee data was collected span across four main agroecological zones, which had a representative climatic gradient. Data on six stingless bee species was randomly collected whereby a meliponary with more than one domesticated colony placed in one location or a wild nest with a colony used as an observation/sample point. These were *Hypotrigona gribodoi* (n=64), *Meliponula togoensis* (n=24), *Meliponula bocandei* (n=23), *Meliponula ferruginea* (n=55), *Plebiana armata* (n=44) and unidentified species (n=280 sample points). Locations of the stingless bees were recorded using a handheld global positioning system (GPS) with an accuracy of 3 meters ($\pm$3).

2.2.2.2 Predictor variables

2.2.2.2.1 Vegetation phenology

The 16-day Enhanced Vegetation Index (EVI) observations acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) at 250-metre spatial resolution, was used to derive vegetation phenological variables (Table 1). EVI over 20-years period (from 2001 to 2020) was processed using TIMESAT software (Eklundha and Jönsson, 2017) to derive 13 vegetation phenological variables (Table 1) for the two known rainy seasons. To achieve best fitting vegetation phenological parameters, TIMESAT settings were configured and set following Makori et al. (2017) recommendation.

A total of 13 vegetation phenological variables (Table 1) were produced for each of the two rainy seasons in the study area. However, since the second season in Kenya is not stable and vegetation phenological indicators are not consistent throughout the years, only phenological parameters from the first season were used in the present study. Start of season (*start_t*) was calculated from the left edge minimum using a 20% EVI deviation, while time for the end of season (*end_t*) was calculated using a 20% EVI deviation threshold from the minimum EVI, detected towards the latter part of the season. The length of season (*length*) measures the time from the start to the end of the growing season and the base level value (*base*) describes the average value of the right and left minimum values of the vegetation in the season. On the other hand, middle of the season (*mid*) is the mean of the time of the season after removing 20% of the minimum from both the right and left vegetation productivity indicated by EVI, and the maximum value (*max*) is the largest vegetation (EVI) value within the season. Amplitude (*amplitude*) gives the difference between the base value and the largest vegetation (EVI) value in the season, left derivative (*left_d*) shows how fast the season is increasing from the left, right derivative (*right_d*) measures the rate of decrease from the right and large integral (*large_i*) describes the season from the start to the end. The small integral (*small_i*) variable is used to describe the difference between the season end and base level, whereas the value of the function at the start of season (*start_v*) and the value of the function at end of season (*end_v*) indicate the vegetation (EVI) values at the start and end of the season respectively.

2.2.2.2.2 Topographical variables

To investigate the influence and effect of landscape morphology on the distribution of stingless bees, topographical variables were included in the EN models. Aspect, hillshade, roughness, slope, topographic position index (*TPI*) and the terrain ruggedness index (*TRI*) (Table 1) were derived from a 3 arc sec resolution (~90m pixel resolution) digital elevation model (DEM) data acquired by the Shuttle Radar

Topography Mission (SRTM) instrument (CGIAR-CSI, 2020), and used together with other variables to predict the distribution of stingless bees within the study sites.

2.2.2.2.3 Bioclimatic variables

To determine the contribution of climatic conditions on the distribution of stingless bees, the freely available bioclimatic variables with 1 km spatial resolution (Table 1) from AfriClim (Fick and Hijmans, 2017; Platts et al., 2015) were utilized in this study, i.e., the AfriClim data were downscaled from the WoldClim platform (www.worldclim.org). These bioclimatic variables contain grids of temperature, rainfall and derived climatic summaries. The variables describe a long-term mean of the current conditions (1970 to 2000) and simulated future conditions in 2055. This study utilized simulated climate data under the intermediate CO_2 emissions, representative concentration pathway scenario (RCPs) 4.5 watt/m^2 (Pachauri and Mayer, 2015) as set by the International Panel on Climate Change (IPCC) using total radioactive forcing values. The simulated future bioclimatic data were calculated as a mean over the years 2041 to 2070 (Guan et al., 2020), and obtained from the fourth version of the community climate system (CCSM4), which has the most reliable climate projections (Mohammadi et al., 2019; Mudereri et al., 2020b). For each of the timesteps, twenty-one (21) bioclimatic variables (Table 1) were utilized. Specifically, ten (10) temperature and eleven (11) precipitation variables were generated for each timestep.

Table 1: Bioclimatic, topographic and vegetation phenology variables used in the prediction of the spatial distribution models of stingless bees.

Variables		Description	Units
Name	**Abbreviation**		
a) Temperature variables (n = 10)			
Bio 1	*bio1*	mean annual temperature	°C
Bio 2	*bio2*	mean diurnal range in temperature isothermality	°C
Bio 3	*bio3*	isothermality	°C
Bio 4	***bio4***	**temperature seasonality**	**°C**
Bio 5	*bio5*	maximum temperature warmest month	°C
Bio 6	*bio6*	minimum temperature coolest month	°C
Bio 7	*bio7*	annual temperature range	°C
Bio 10	*bio10*	mean temperature warmest quarter	°C
Bio 11	*bio11*	mean temperature coolest quarter	°C
Potential evapotranspiration	***pet***	**potential evapotranspiration**	**mm**
b) Precipitation variables (n = 11)			
Bio 12	***bio12***	**mean annual rainfall**	**mm**
Bio 13	*bio13*	rainfall wettest month	mm
Bio 14	*bio14*	rainfall driest month	mm
Bio 15	*bio15*	rainfall seasonality	mm
Bio 16	*bio16*	rainfall wettest quarter	mm
Bio 17	*bio17*	rainfall driest quarter	mm
Moisture index	*mi*	annual moisture index	n/a
Moisture index moist quarter	*mmiq*	moisture index moist quarter	n/a
Moisture index arid quarter	*miaq*	moisture index arid quarter	n/a

Dry months	*dm*	number of dry months	months
Length of the longest dry season	*llds*	length of longest dry season	months
c) Topographical variables (n = 6)			
Slope	***slope***	**angle of inclination to the horizontal**	**% rise**
Aspect	*aspect*	slope direction	**degrees**
Hillshade	*hillshade*	shading effect	n/a
Roughness	***Roughness***	**degree of surface irregularity**	**n/a**
Topographical position index	***TPI***	**position of surface inclination**	**n/a**
Terrain ruggedness index	*TRI*	ground ruggedness	n/a
d) Vegetation phenological variables (n = 13)			
Start of season time	***start_t***	**time for the start of season**	**decades**
End of season time	*end_t*	time for end of season	decades
Length of season	*length*	length of time from start to end of season	decades
Base value	*base*	average minimum EVI value	n/a
Time of middle of season	*mid*	time of middle of season	decade
Maximum value	*max*	maximum value of fitted data	n/a
Amplitude	***amplitude***	**difference between maximum and base level**	**n/a**
Left derivative	***left_d***	**rate of increase of beginning of season**	**%**
Right derivative	*right_d*	rate of decrease of end of season	%
Large integral	***large_i***	**integral from season start to season end**	**n/a**
Small integral	*small_i*	integral difference between season and base level	n/a
Start of season value	*start_v*	value at the start of season	n/a
End of season value	***end_v***	**value at end of season**	**n/a**

2.2.2.3 Variable selection

Multicollinearity among predictor variables is the single major problem that leads to instability and volatility in the model's parameterization and performance (Dormann et al., 2013; Naimi and Araújo, 2016). Before testing for multicollinearity, the recursive feature elimination criteria in the "caret" package in R was employed to assess the possible minimum number of predictor variables that could produce comparable results. This analysis enabled the study to gain insight into the number of uncorrelated predictor variables that could give the most accurate modelling results. Thereafter, a two-stage elimination criteria was performed based on the Pearson correlation coefficient and variance inflation factor (VIF) to select the predictor variables used in the EN models. The aim of the variable selection experiment was to reduce multicollinearity among predictor variables and establish relevant orthogonal features that are most suited for predicting stingless bee distribution using EN models (Dormann et al., 2013; Plant, 2012). To detect collinearity among predictor variables, multiple regression models were utilized. Each predictor variable was regressed against all other variables, and the VIF was computed for each combination (Plant, 2012). The VIF is commonly used to first select important variables among the different categories, then iteratively those with high linear regressions coefficients are eliminated.

This study firstly set the threshold for the Pearson correlation coefficient at $th = 0.7$ ($r \geq 0.7$), which has been demonstrated to yield the best result without any significant loss of information (Araújo et al., 2019; Dormann et al., 2013; Plant, 2012). Secondly, the "vifstep" function available in the "usdm" package in R (Naimi and Araújo, 2016) was used to further assess the collinearity of the selected variables based on a recommendation by Dorman et al. (2013), that VIF values of 10 or higher between variables indicate collinearity. The correlation matrix presented in Figure 2 used blue colour to indicate positive correlation while brown colour denotes negative correlation. Colour intensity and size of the shape indicate level of collinearity. Shape

orientation denotes direction of correlation. This figure shows that some of the essential and relevant variables to the stingless bees' distribution such as isothermality and mean annual rainfall were highly correlated. Therefore, decision on which variable to select was based on their importance and ecological significance in species distribution models using expert-knowledge approach (Dormann et al., 2013; Naimi et al., 2014). The variable selection experiment enabled the study to select eleven (11) relevant and uncorrelated variables (Table 1) out of a total of 40 (21 bioclimatic, 6 topographic and 13 vegetation phenology). Using a good number of relevant response variables would lead to acceptable and reasonable prediction accuracies in EN models (Stockwell and Peterson, 2002; Wisz et al., 2008). It has been established that 90% prediction accuracy of machine learning algorithms can be obtained using as low as 10 response variables, but higher number of variables, from 50 sample points, would ultimately yield better or near maximal accuracy (Araújo et al., 2019; Wisz et al., 2008). Given that the selected response variables consisted of more than 10 sample points for each stingless bee species, it was expected that the EN models developed in this study would yield "acceptably" high prediction accuracies.

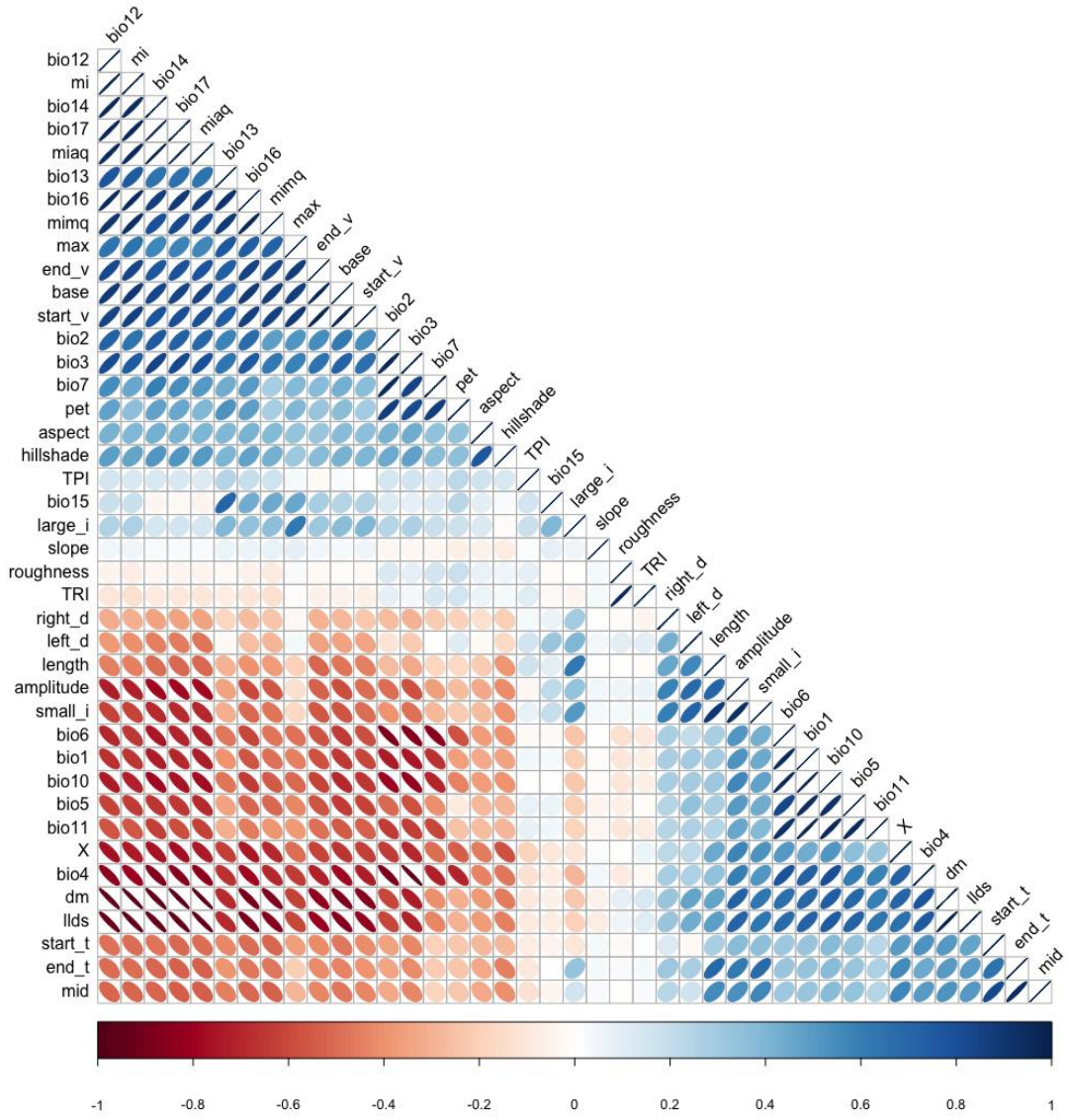


Figure 2: Collinearity matrix of all predictor variables (refer to Table 1) used in the ecological niche (EN) models for the spatial distribution of stingless bees in Kenya.

2.2.3 Model Fitting

Machine learning EN models in the "sdm" package (Naimi and Araújo, 2016) in the R software (R Core Team, 2021) were used to relate the stingless bee presence-only observations to the selected uncorrelated predictor variables. Since this study utilized presence-only observations, 1,000 pseudo-absence points were generated in R environment using the "sdmdata" function in the "sdm" package.

The "sdm" package has a selection of over 22 EN model executions, which are object oriented, reproducible and have an extensible framework in R (Naimi and Araújo, 2016). These machine learning models employ somewhat the same principles of presence-pseudoabsence response variables. In the present study, trial runs were performed for all 22 EN modelling algorithms, from which the best six were selected based on their prediction accuracy (Table 2). The six selected algorithms have been demonstrated to yield high performance accuracies in predicting species distribution in complex environments (Mudereri et al., 2020a; Naimi and Araújo, 2016). An ensemble projection in the "sdm" package inherent in "R" software, was employed to estimate the mean predictions of stingless bee distribution over the six selected EN modelling methods based on the ranking of their predicted accuracy.

Table 2: Selected ecological niche modelling methods used in parallel executions in the "sdm" package in the "R" environment to predict the spatial distribution of stingless bees in Kenya.

S/N	Modelling methods	Abbreviation	Reference
1	Random forest	"RF"	(Breiman, 2001)
2	Ranger	"RANGER"	(Wright et al., 2018)
3	Generalized additive models	"GAM"	(Hastie and Tibshirani, 1990)
4	Multivariate adaptive regression spline	"MARS"	(Friedman, 1991)
5	Boosted regression trees	"BRT"	(Elith et al., 2008)
6	Recursive portioning and regression trees	"RPART"	(Therneau et al., 2010)

It was hypothesized that the model accuracy for six algorithms used in this study could differ in predicting some stingless bee instances, hence an ensemble modelling approach was employed to harmonize the variations among the six EN algorithms. Furthermore, previous studies have demonstrated that ensemble modelling approach usually has a superior predictive power than individual modelling methods (Araújo

et al., 2019; Hao et al., 2019; Makaya et al., 2019; Naimi and Araújo, 2016). An ensemble approach fits and maximizes model accuracy resulting in improved reliability since it utilizes highly ranked models with higher precision. To run the ensemble model, the study utilized the "ensemble" function in the "sdm" package under the "R" environment (Araújo and New, 2007; Naimi and Araújo, 2016). The results of the six stingless bee models were harmonized under the ensemble model using the true skill statistics (TSS) weighted average approach which improves the model predictability (Naimi and Araújo, 2016). However, including individual models of low accuracy in an ensemble experiment would yield a low predictive power. Therefore, the study included only EN models of a TSS of 0.7 in the ensemble modelling experiment (Dormann et al., 2013). A TSS is an accuracy measure that estimates the agreement between the response (e.g., stingless bees presence observations) and predictor (e.g., bioclimatic, topographic and vegetation phenology) variables. It ranges between -1 and 1, with 1 indicating high predictive accuracy (Hao et al., 2019).

Predictions for the distribution of stingless bees in this study area were done under both current and future climatic projections in 2055. For consistency, similar model settings and packages were used for both prediction scenarios to establish the effect of climate change on the distribution of stingless bees in Kenya. For consistency and comparability, the future stingless bee prediction scenario utilized simulated climatic variables, selected to match predictors used in the current climatic scenario. On the other hand, the current vegetation phenological and elevation variables were used in the future model, as projected vegetation phenological scenarios are absent, and elevation is unlikely to considerably change in future. In this study it was assumed that elevation and vegetation phenology will remain unchanged over the projected future period in 2055. The "ensemble" function in the "sdm" package under the "R" environment was used to predict the spatial distribution of stingless bees in future.

2.2.4 Stingless Bees' Distribution Model Validation

This study employed random data splitting method inbuilt in the "sdm" package to assess the accuracy of the stingless bee distribution models developed in the current study. In both the current and future cases, a 10-fold cross validation approach was used (Araújo et al., 2019; Naimi and Araújo, 2016). Automation was used to independently and randomly draw a 70% sample for calibration of the SDMs while the remaining 30% were used to validate their performance. The study utilized both threshold-independent statistics (i.e., area under curve – AUC of the receiver operating curves - ROC) and threshold-dependent statistics (i.e., true skills statistics – TSS, sensitivity and specificity) to assess the performance of the stingless bee EN models (Phillips and Dudík, 2008). The ROC is a graphical representation of the model's fitting of sensitivity (presence data points) and specificity (pseudo-absence data points) metrics The area under such graphical representation (AUC) indicates the overall accuracy of model performance. AUC values close to zero (0) indicate poor ordering or impossible occurrence while those close to one (1) represent perfect prediction by the model or optimal occurrence (Du et al., 2014). The study modified the Swets discriminatory power (Mohammadi et al., 2019; Swets, 1988), which ranks the models' performance in a way that: (i) excellent $\geq$ 0.90, (ii) good = 0.80-0.89, (iii) fair = 0.70-0.79, (iv) poor = 0.60-0.69 and (v) fail $\leq$ 0.59.

This study employed TSS to test the accuracy of the current and future ensemble models. Equations 1 to 3 describe the TSS variables and their parameters (Richard et al., 2018). Optimal or perfect occurrence or perfect ordering of observed stingless bee presence versus its predicted occurrence were presented by TSS values close to +1. Values lower than zero (TSS $\leq$ 0) represented poor or impossible stingless bee occurrence (Guan et al., 2020). The current and future stingless bee distribution maps were used to calculate the predicted area covered by stingless bees and to quantify change in their probability of occurrence. To quantify change in area, the probability of occurrence was converted to presence/absence using the evaluate function in the

"sdm" package in R. The presence/absence observation data were then used to extract probability of occurrence in the "raster" package by thresholding maximum TSS values. This was eventually converted to absolute data frame change estimates whose areas were calculated and quantified in the "sdm" package in R. The change in probability of stingless bees' occurrence was grouped into the following classes: (i) high probability of reduction (≤ -0.41), (ii) moderate probability of reduction (-0.40 to -0.21), (iii) low probability of reduction (-0.20 to -0.06), (iv) no change (-0.05 to 0.05), (v) low probability of proliferation (0.06 to 0.20), (vi) moderate probability of proliferation (0.21 to 0.40), and (vii) high probability of proliferation (≥ 0.41) (Araújo et al., 2019; Rapacciuolo et al., 2014).

$$TSS = (sensitivity + specificity) - 1 \qquad \textbf{\textit{Equation 1}}$$

where *sensitivity* represents the presence observation and *specificity* is absence

$$Sensitivity = \frac{a}{a+b} \qquad \textbf{\textit{Equation 2}}$$

where *a* is true positive presence and *b* is false negative presence

$$Specifity = \frac{d}{c+d} \qquad \textbf{\textit{Equation 3}}$$

where *c* is false positive presence and *d* is true negative presence

2.3 Results

2.3.1 Variable Interaction and EN Model Development

The collinearity model (Figure 2) indicated that 29 out of 40 variables used in the current prediction models were conflating. These variables were eliminated from the EN models based on their importance and individual interactions as indicated in Figure 3. Hence, 11 non-conflating variables (highlighted in bold in Table 1) were deemed fit for developing the final EN models. Furthermore, the recursive feature elimination (RFE) criteria indicated that a few variables (14) could give comparable results before eliminating the colinear ones (Figure 4). In addition, selected bioclimatic variables had the highest relative influence (64.27%) on the EN models developed to predict the stingless bee occurrence, while vegetation phenological and topographical predictor variables contributed to 21.36% and 14.36% information respectively (Figure 5).

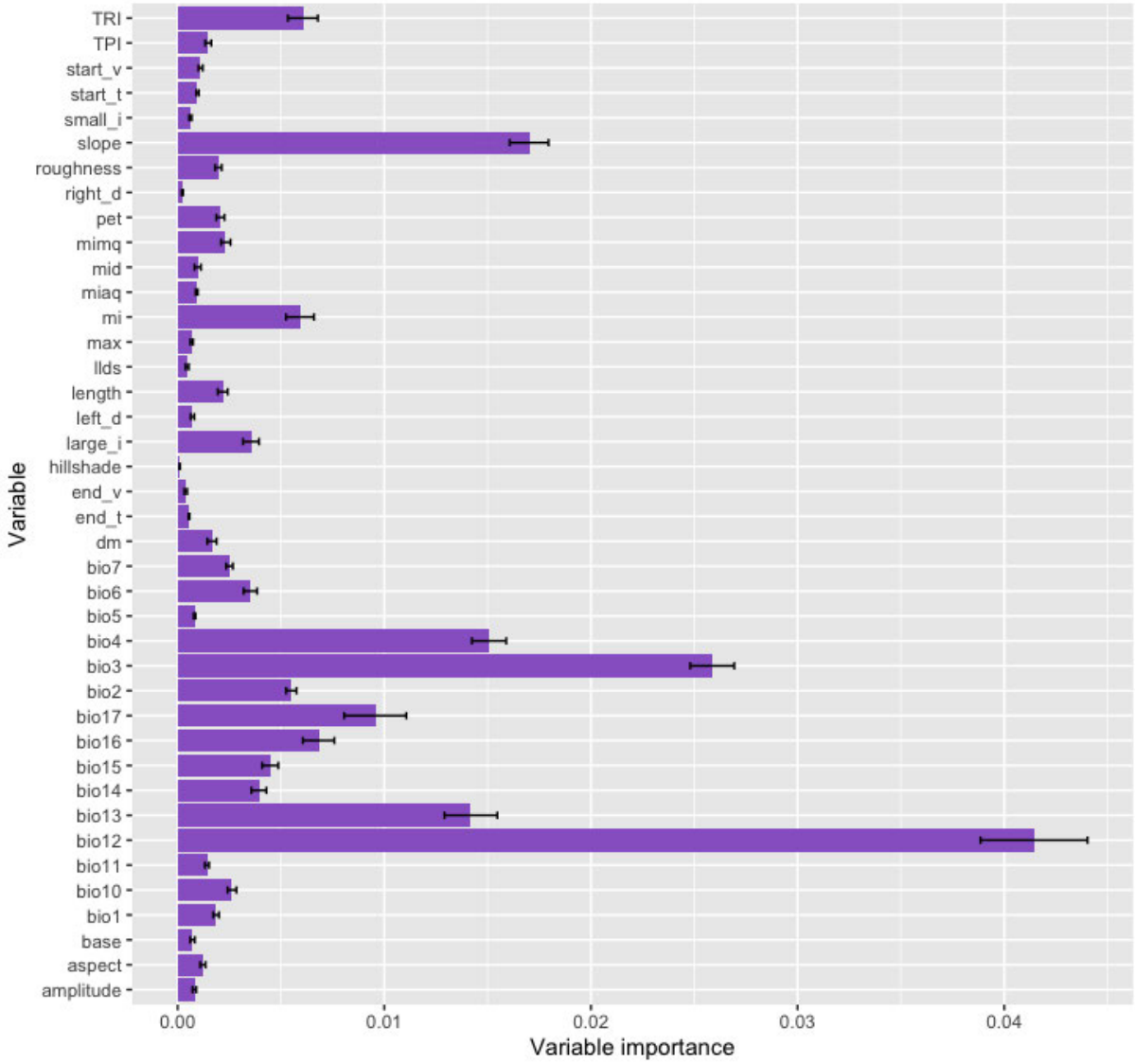


Figure 3: Relative variable importance of all 40 variables (refer to Table 1) used in EN models for predicting the distribution of stingless bees.

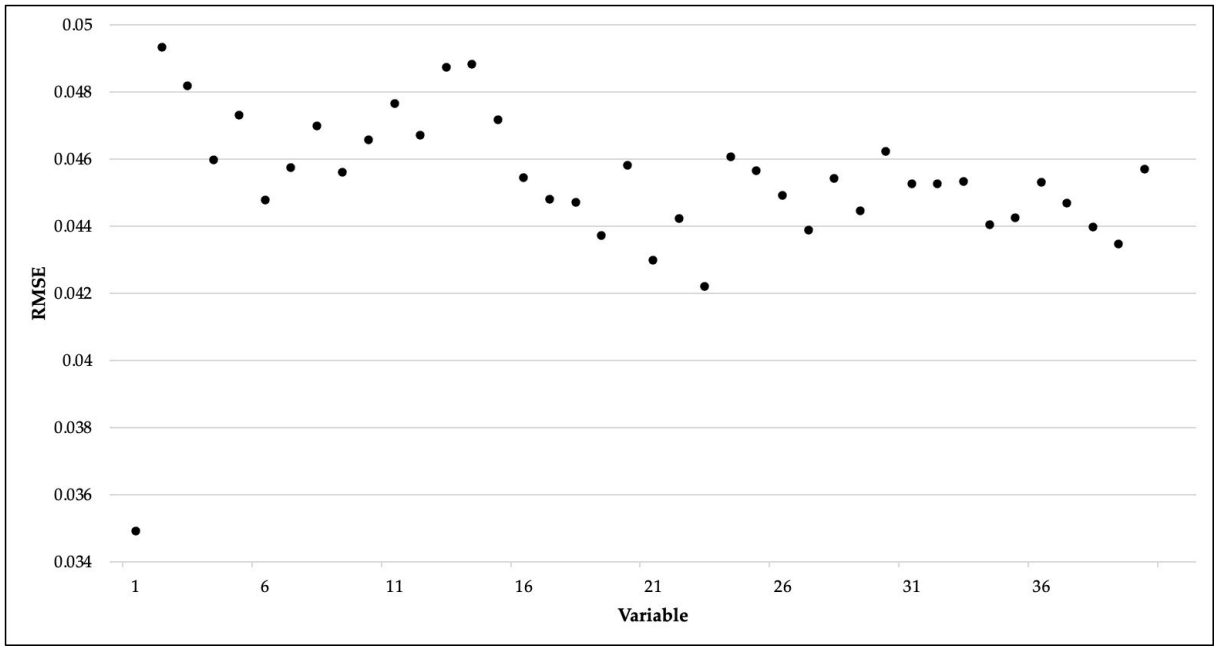


Figure 4: Recursive feature elimination (RFE) model using the root mean square error (RMSE) to indicate variables that could yield comparable prediction model accuracy.

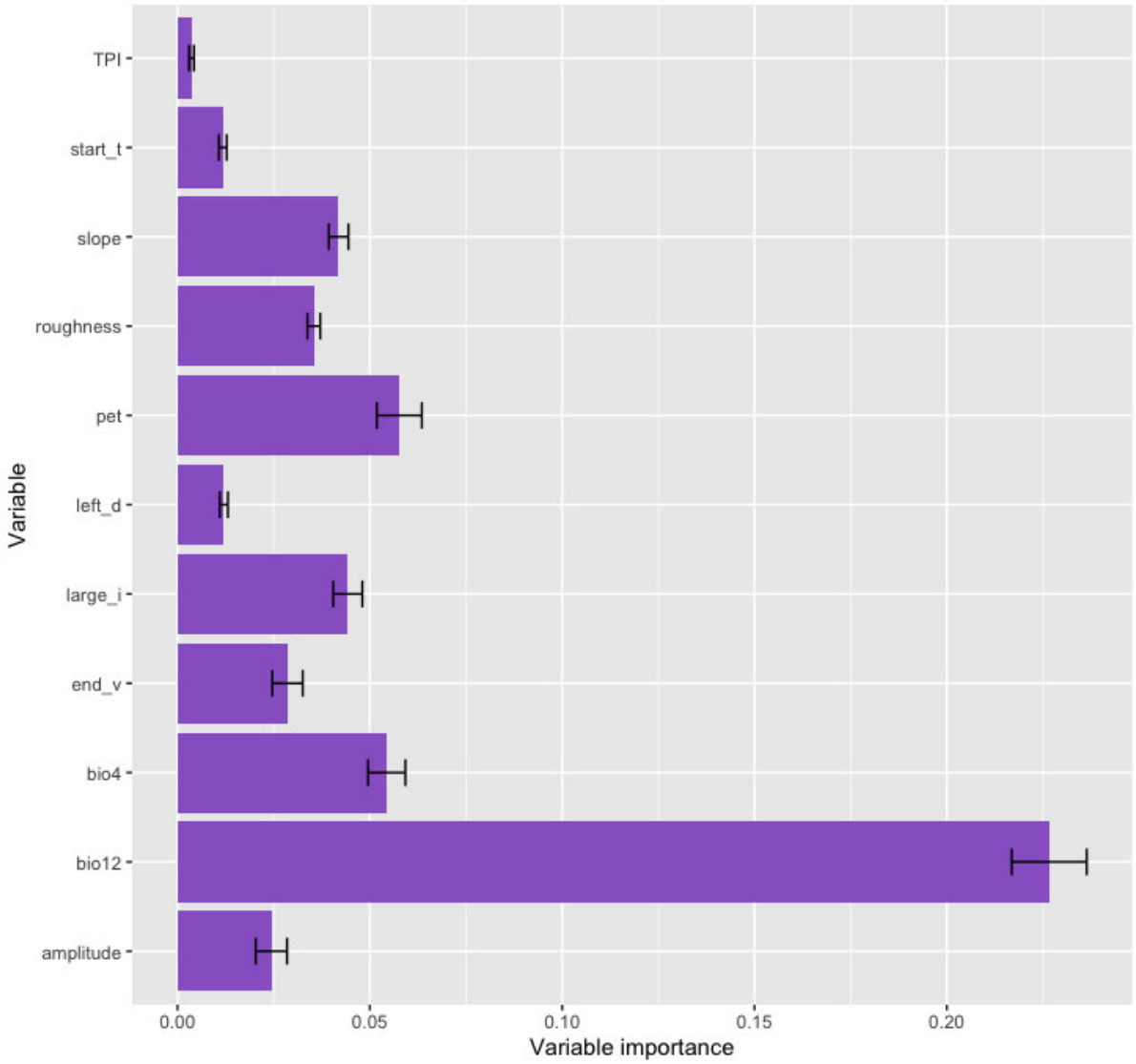


Figure 5: Relative importance of selected non-conflating predictor variables (refer to Table 1) used in the ecological niche models.

Additionally, the results showed that among the three selected bioclimatic variables, mean annual rainfall (*bio12*) was the most important variable in the EN models. It contributed almost half (43.09%) of the information required for distribution of stingless bees (Figure 5). Temperature seasonality (*bio4*) was the second important bioclimatic variable, followed by potential evapotranspiration (*pet*), which together added 21.18% of the information needed to predict stingless bee distribution (Figure 5). On the other hand, vegetation phenology provided the greatest number (5) of

variables viz, integral from season start to end (*large_i*), value at the end of the season (*end_v*), difference between maximum and base level (*amplitude*), rate of increase at the beginning of the season (*left_d*) and time for the start of the season (*start_t*). However, they ranked among the least important variables for predicting the distribution of the stingless bees, except *large_i*, which contributed about 8.98% in the EN models (Figure 5). For topographical variables, slope and roughness were the most important with slope having relatively higher comparative influence (7.18%) to the EN models than the other topology variables. Even though vegetation phenology and topography contributed the least logit (modelling information) when combined (35.73%), they both helped to sharpen the EN models further by providing more prediction information at grain level, hence improving the model's accuracy.

2.3.2 Stingless Bees' Prediction Model and Validation

The AUC threshold-independent statistics indicated that all the EN models developed in this study were "excellent" in their performance according to the modified Swets discriminatory power (Table 3). On the other hand, the TSS threshold-dependent statistics denoted that the EN models achieved varied results ranging from "fair" to "excellent" ranking. In both cases, random forest and ranger algorithms achieved "excellent" results for both the AUC (0.98) and TSS (0.91). Additionally, the differences between attained AUC and TSS model accuracies ranged between 7% (0.07) for random forest to 15% (0.15) for MARS and BRT (Table 3). This was an indication that the latter two EN algorithms had higher disparities in threshold dependent and independent evaluation criteria than the rest. The ROC curves (Figure 6) used the red and blue curves to indicated smoothed mean area under curve (AUC) of the training data and test data, while cyan curves indicated runs for individual models. The black dotted line showed the one-to-one mid-point interaction between sensitivity and specificity for each model. All model algorithms demonstrated a good fit between training and test datasets for smoothed means, with little variations

between individual 10-fold replicated runs. Random forest and RANGER displayed better fitting for 10-fold replicates than other algorithms. This ndicated model stability and relatively consistent predictions within replicates. RPART had the highest dispersions of AUC for individual replicates, suggesting relative model instability before mean averaging. Additionally, some individual replicate runs were demonstrated to be closer to the one-to-one midpoint ordering line between sensitivity and specificity, in line with the low attained model accuracy (Table 3). Ultimately, there was no algorithm skewness beyond the one-to-one midpoint of ordering for specificity and sensitivity. The ordering of presence and absence data (Figure 7) also indicated robustness of the response variables used to develop prediction models. Furthermore, presence-absence data of response variables presented limited dispersion from the mean. This was an indication of good separability between classes and high integrity within classes.

Table 3: Area under curve (AUC) and true skills statistics (TSS) of six EN modelling methods used to predict the distribution of stingless bees in Kenya.

Methods	**AUC**	**TSS**	**Difference**	**Swets (TSS)**
RF	0.98	0.91	0.07	Excellent
RANGER	0.98	0.90	0.08	Excellent
GAM	0.96	0.87	0.09	Good
MARS	0.95	0.80	0.15	Good
BRT	0.95	0.80	0.15	Good
RPART	0.91	0.79	0.12	Fair

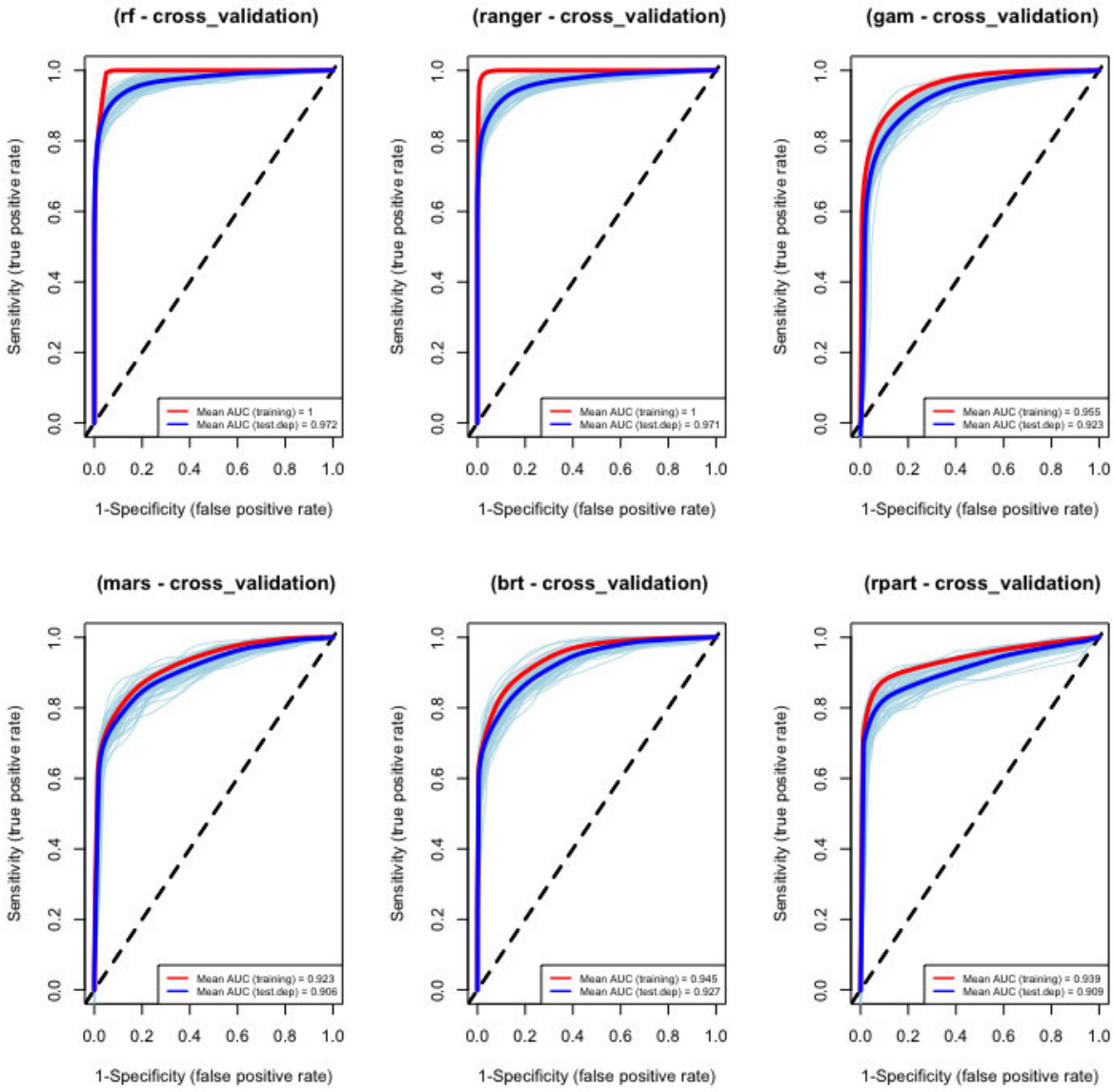


Figure 6: Receiver operating curves (ROC) of EN modelling methods used to predict distribution of stingless bees.

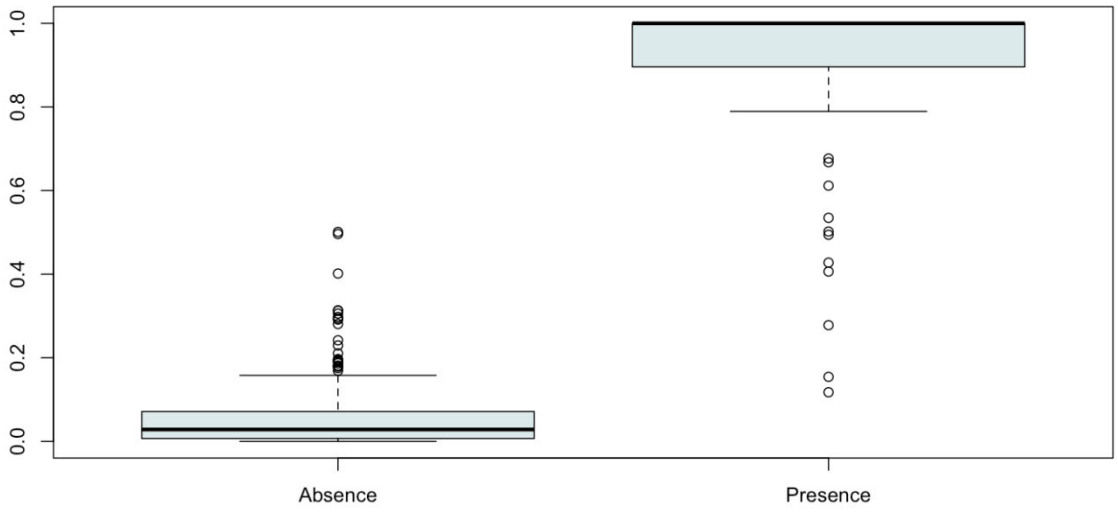


Figure 7: Ordering of response variables used in training and evaluation of the various models.

2.3.3 Stingless Bees' Ecological Niche Prediction

The EN models predicted high stingless bee suitability in the western, Kisii (in Nyanza region), Bomet (in Rift Valley region), coastal, eastern and Mt. Kenya regions of Kenya for both current (a) and future (b) scenarios (Figure 8). Specifically, Mombasa, Taita hills, and Kilifi at the coast, Kasanga in Mwingi and Machakos in Eastern region, Kakamega and Busia in western and Mt. Elgon and Nandi hills in Rift Valley regions were predicted to have relatively high habitat suitability. Kwale and larger Taita-Taveta in coastal region, and Rift Valley regions had varied predicted probability of occurrence, while large parts of Tana River (in the coastal region) and Narok (in the Rift Valley region) counties had low predicted probability (Figure 8). These regions have varied temperature and precipitation gradients as well as seasonality. Additionally, their characteristic climatic and agroecological gradients are diverse (Camberlin et al., 2014; Ochieng et al., 2016). The coastal region has high temperature and humidity, while the eastern parts of the country have high temperature and relatively lower precipitation. Kakamega (in the western region), Kisii in the Nyanza

region), Nandi hills (in the Rift Valley region) and the Mt. Kenya regions experience cooler temperatures with elevated levels of precipitation. Indeed, mean annual rainfall and temperature seasonality (*bio12* and *bio4*) contributed significantly to the EN distribution models. Furthermore, topography of parts predicted with high suitability was similar, with high elevation values that are consistent with low temperature and high precipitation. Mombasa and Kilifi (in the coastal region) and Kitui (in the eastern region) counties are lower altitude regions, while Kakamega and Busia (in the western region), Kisii (in the Nyanza region), Nandi (in the Rift Valley region) and Taita hills (in the coastal region) are largely hilly. The future map (b) displays increased suitability for stingless bees' in some areas, while low and moderate suitability was predicted for areas such as Nandi hills (in the Rift Valley region), Mt. Kenya and the eastern regions of Kenya. Conversely, most regions predicted with high stingless bee habitat suitability had reduced in 2055. There was an exception in the Kisii highlands (in the Nyanza region), which indicated increased suitability from moderate to high suitability of proliferation. The western and eastern (specifically Kasanga in Mwingi) regions had reduced from high to moderate suitability of proliferation.

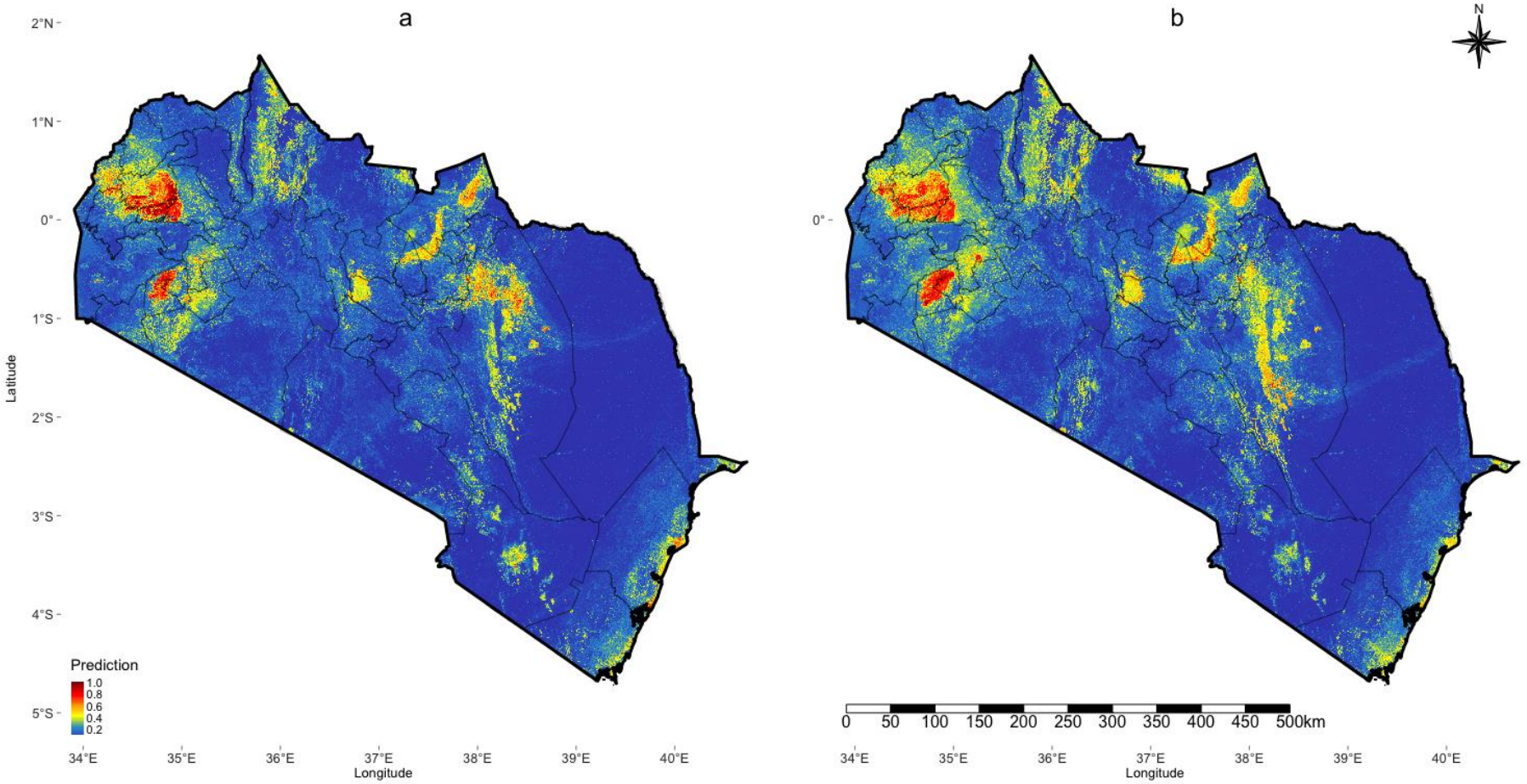


Figure 8: Predicted stingless bee distribution in Kenya for the current (a) and future (b) developed using ensemble of six machine learning EN models.

The change map based on probability of stingless bee occurrence (Figure 9) revealed that many areas especially Kisii (in Nyanza region), Mt. Elgon and Nandi hills (in the Rift Valley region), Mt. Kenya, and some parts of Kitui (in the eastern region), had increased suitability while Kakamega and Busia (in the western region), Bomet (in Rift Valley region), Mwingi (in the eastern region), and Taita Taveta, Mombasa and Kilifi counties (in the coastal region) had moderate to high decrease in probability of stingless bee occurrence. Moreover, the categorized change classes indicated that probability of stingless bees' distribution increased in regions totalling 35,958.68 km^2, while a total of 99,363.70 km^2 was predicted to have diminishing probability of occurrence (Table 4) with unaltered predicted future climate change pathways. Most interclass variations were centred between moderate and low diminishing probability to low and moderate probability of proliferation. Overall, there was more probability of diminishing than proliferation by a total of 63,405.02 km^2 in the study area.

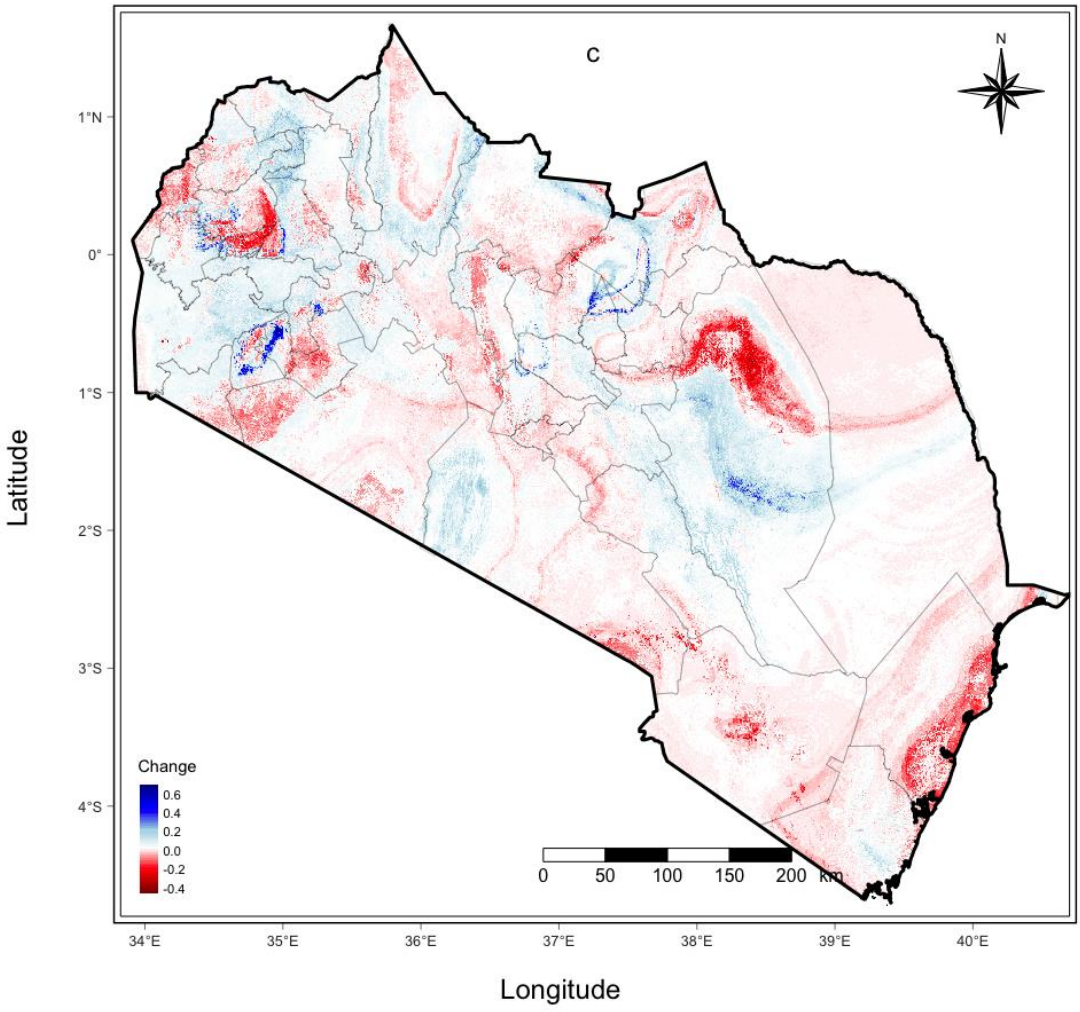


Figure 9: Stingless bees distribution change map (c) indicating difference between the current and future probability of occurrence.

Table 4: Probability of change for stingless bees' distribution maps between the current and future predicted areas.

S/No	Class	Area (km²)
1	High probability of reduction	931.25
2	Moderate probability of reduction	14,449.04
3	Low probability of reduction	83,983.40
4	No change	121,316.12
5	Low probability of proliferation	33,501.14
6	Moderate probability of proliferation	2,234.14
7	High probability of proliferation	223.40
	Total	**256,638.49**

2.4 Discussion

Stingless bees are important in fragmented landscapes due to their polylactic nature. Specifically, their wide range of flower preference make them valuable pollinators, especially in areas with low plant density and diversity. Amid climate change, the influence of climate on spatial distribution of stingless bees is an important element in the determination of their future spatial distribution.

2.4.1 Variable Interaction and EN Model Development

Despite stingless bees diverse and generalist foraging ability, climate and vegetation phenology influence their distribution (Njoya, 2010; Slaa et al., 2006). The EN models established that precipitation, particularly mean annual rainfall, was the most influential variable determining the distribution and proliferation of stingless bees. This variable provided almost half (43%) of the information required by the models to predict their spatial distribution. The models predicted more stingless bee suitability in regions with high annual precipitation across the country. Furthermore, it was established that variability in precipitation affected stingless bees' distribution across temporal and spatial dimensions. This is consistent with literature (MM Fierro et al., 2012; Lovett, 2015; Makori et al., 2017; Mwalusepo et al., 2015; Platts et al., 2015) that note a positive correlation between precipitation and the distribution of stingless bees.

The recursive feature elimination (RFE) model (Darst et al., 2018; Granitto et al., 2006; Yan and Zhang, 2015) indicated that more than half of the predictor variables initially poised for the stingless bee distribution models were redundant. The recursive feature elimination (RFE) experiment established that a few selected features would otherwise yield same or higher prediction accuracy as opposed to using all the 40 predictor variables. The RFE, collinearity and variable importance procedures were useful in eliminating redundant features and enhanced the stingless bee EN model, reducing instability, volatility and parameterization (Dormann et al., 2013; Naimi and Araújo, 2016). On the other hand, vegetation phenological variables contributed lower

permutation on the prediction of stingless bees' distribution compared to bioclimatic variables, but were palpably greatly valued, as demonstrated by the number of features tagged important (Figure 5). It was speculated that the availability of pollen and nectar determine niche colonization of stingless bees and facilitate their distribution in vegetated areas that could be less suitable in terms of climate. This is more important for providing foraged bee bread and nectar products that are indicators of colony strength and trigger of brooding (Pau et al., 2012; Torto et al., 2010). Indeed, vegetation's large integral (*large_i*) and end of season values (*end_v*), which indicate the amount of pollen and nectar available for stingless bees within and end of the season respectively, were tagged as the most important in this category. Higher large integral values (*large_i*) indicate abundance of forage within the season and longer foraging periods which encourage stingless bee colony growth and distribution. On the other hand, a large value at the end of the season (*end_v*) demonstrates availability of enough foraging materials for bees within the season. The coastal, Mt. Kenya, Kisii (in the Nyanza region), Nandi hills (in the Rift Valley region) and western regions had high vegetation's large integral values within the season, because vegetation seasons start earlier and the vegetation phenology values at the end of the season were more than other regions. Consequently, these regions were predicted as highly suitable to stingless bees due to these conducive vegetation conditions.

Slope and terrain roughness topographic variables contributed substantially (13.64% of the total logit) to the performance of the stingless bee EN models compared to the total contribution (14.36%) of the selected topographical variables (Figure 5). Specifically, suitable ENs spanned through regions with high altitudinal gradients across the study site. Previous studies (Birrell et al., 2020; Jackson et al., 2018; Kimathi et al., 2020; Mani, 2013) indicated that most insects have diverse altitudinal regimes and could thrive in regions with varying topography. In Kenya, the coastal, Baringo (in the Rift Valley region) and eastern regions are characterized by low altitude, while

Taita hills (in the coastal region), Nandi hills (in the Rift Valley region), Kisii highlands (in the Nyanza region), Mt. Kenya, Mt. Elgon (in the Rift Valley region) and Kakamega (in the western region) are more elevated. Therefore, the impact of varying altitudes on the probability of stingless bee occurrence could not be conclusively deciphered. Nevertheless, such an impact could influence other parameters such as food resources and climate regimes, which could impair or encourage distribution of the bees. Therefore, topographical variables in seclusion may not be reliable in developing EN models but should be coupled with other features such as vegetation phenology and climate.

2.4.2 Stingless Bees' Prediction Models and Validation

After eliminating the least accurate stingless bee prediction models, scores ranged from "fair" to "excellent" based on the modified Swets discriminatory power. It is reported that predictive model precision and accuracy in space depend on careful and accurate selection of predictor features viz a vis response variables (Aranda and Lobo, 2011; Araújo et al., 2019; Naimi and Araújo, 2016). Also, they are largely dependent on the selection of important non-conflating variables while circumventing redundancy.

Although all predictor models were subjected to same response variables and model conditions, model disparities were observed especially in threshold dependent and independent accuracy measurements for MARS and BRT. Models with low prediction accuracy exhibited less accurate ordering of sensitivity and specificity, closer to the one-to-one mid-point interaction line on the ROC models. Individual model replicates used to average model means indicated increased dispersion from both test and training averaged mean curves. In addition, dispersions increased as prediction accuracy reduced, pulling closer to the mid-point hence lowering the averaged mean curves. Observed model instability indicate decreased model dependability for niche prediction of species, especially where high precision and accuracy are desirable (Farley, 2017; Felton et al., 2021). However, individual model replicate dispersion

observed in this study did not pass the one-to-one midpoint ordering of specificity and sensitivity. Even though prediction accuracy did not reach the excellent level for at least four of the selected models, prediction skewness was not significant since all models attained acceptable ordering of sensitivity and specificity (Gao and Tian, 2021; Kajita et al., 2017). Notwithstanding, this study suggests that robust and rigorous approaches should be employed to ensure accurate collection and collation of response variables, thorough selection of predictor features, and good choice of artificial and deep learning prediction algorithms for improved stingless bee distribution predictability power.

2.4.3 Stingless Bees' Ecological Niche Prediction

Studies have shown that increased variability in both temperature and precipitation decrease the population of beneficial insects such as stingless, honey and bumble bees (Roulston and Goodell, 2011; Torné-Noguera et al., 2014). Indeed, the EN models indicated that there was a general decline of suitability of stingless bee niches across the study area by 63,405.02 km^2, more than proliferation. Additionally, most regions reduced in stingless bee niche suitability from high and moderate suitability to lower classes such as Kakamega (in the western region), Nandi hills and Bomet (in the Rift Valley region), Mwingi (in the eastern region), Taita hills, and other coastal regions. Besides, these regions were demonstrated to have reduced annual rainfall with increased potential evapotranspiration and temperature seasonality (Fick and Hijmans, 2017; Platts et al., 2015). The interaction of these climate variables could have shifted the primal climatic conditions for stingless bees in these regions and affected their phenological patterns. However, the stingless bee distribution EN models indicated that there was a slight increase in spatial suitability of stingless bees, by 4,197.53 km^2, in many regions across the study site. This could be attributed to the stingless bees' polylactic nature, since they forage on a wide range of plant species including weeds, which are predicted to increase with climate change, hence keeping

their colonies strong. On the other hand, improved colony strength would enable them to fight off diseases and pests, whose increase has been demonstrated to cause bee population decrease and colony collapse across the world (Dainat et al., 2012; Williams and Tarpy, 2010).

The generalist behaviour of stingless bees suggests that over time, they have adapted to shifting vegetation density and diversity (Njoya, 2010; Slaa et al., 2006). This adaptation could inform dispersion and distribution of their suitable niches across different agroecological and agroclimatic zones, with subtle differences within given zones. For instance, the coastal region of Kenya displayed varied suitability of stingless bee occurrence, even though temperature and precipitation regimes are almost similar. Also, the vegetation's phenological composition varies from Kwale through Mombasa to Kilifi County in the region. The latter county has more vegetation diversity and density, making it more suitable for stingless bees. On the other hand, regions neighbouring Arabuko Sokoke Forest in Kilifi County were predicted to be more suitable with high prediction accuracy. Besides, their colonization rates were demonstrated to be higher than the neighbouring Kwale County.

Furthermore, inclusion of vegetation phenological information in species diversity models improve their prediction power and sharpens their predictability. Vegetation phenology variables contain information extracted from MODIS EVI data on each of the growing seasons, hence reducing the residual signal noise and data dimensionality in unprocessed EVI time series data (Hinton and Salakhutdinov, 2006; Shen et al., 2013). This could have introduced bias into the stingless bee distribution prediction models. Furthermore, vegetation phenology variables reduce overfitting of species distribution prediction models, since bioclimatic variables are coarser at the grain level (1 km spatial resolution) and only offer interpolated climatic information, which depict homogeneity at large spatial extents. On the other hand, vegetation phenology data is finer at grain level offering higher spatial resolution (250 m), hence have more

power to incorporate environmental heterogeneity on the ground (Saatchi et al., 2008). In this regard, remotely sensed vegetation phenology data captures adverse environmental conditions, which limit the spatial distribution of stingless bees within homogenous climatic pockets. Also, remotely sensed vegetation phenology data captures subtle differences in vegetation variability on the ground, which determine availability of pollen and nectar, the primary sources of food and energy for the stingless bees. Moreover, remotely sensed vegetation phenology data depicts heterogenous ground conditions at near real-time basis capturing short term changes in small spatial extents, contrary to bioclimatic data that is interpolated over large homogenous regions and over extended temporal periods. In this regard, incorporating remotely sensed vegetation phenology data into the stingless bee spatial distribution models made them more realistic and dependable.

On the other hand, the prediction models revealed that topography had little influence on the distribution of stingless bees (Birrell et al., 2020; Jackson et al., 2018). The predicted maps showed high habitat suitability across altitudinal gradients with no regard of ground elevation. Nonetheless, direct impact of topography on the stingless bees could not be straightforwardly established and its influence on microclimate and vegetation density and diversity could only be inferred and deduced to have similar effects on stingless bees' distribution as phenology and climate.

2.5 Conclusions

This study developed the first ever predictive EN models for the spatial distribution of stingless bees in Kenya using an ensemble modelling approach of six machine learning EN algorithms. The polylactic nature of stingless bees is particularly important due to non-selective pollination of a wide range of both food and non-food plants, in mixed cropping systems and virgin land. Careful selection and preparation of response and predictor variables is however paramount in developing accurate, dependable, and robust predictive models. Relative variable importance, recursive

feature elimination (RFE), collinearity determination procedures and selection of non-conflating features are equally important. Based on TSS accuracy, six EN modelling algorithms proved most accurate in predicting the occurrence and spatial distribution of stingless bees. Specifically, random forest (TSS = 0.91), ranger (TSS = 0.90), generalized additive models (TSS = 0.87), multivariate adaptive regression spline (TSS = 0.80), boosted regression trees (TSS = 0.80), and recursive portioning and regression trees (TSS = 0.79) were the most accurate models that were included in the ensemble for predicting the EN of stingless bees in the study area. Ensembling these six EN algorithms enhanced their strengths, while diminishing their weakness, hence providing a more reliable and accurate stingless bee distribution mapping method. The bioclimatic data, especially mean annual rainfall, provided the most useful information needed for predicting the spatial distribution of stingless bees. However, vegetation phenological variables substantiated most features towards the stingless EN models. Although topography was significant to the models, its contribution could not be directly established, hence only inferred through other features such as vegetation phenology and microclimate. Projection of the EN models into the future (2055) based on hypothesized and modelled climatic pathways by the IPCC, indicated that there was more spatial diminishing of stingless bee suitable habitats than distribution. This was occasioned by high temperature and precipitation seasonality in 2055, which affected their ecological niches the most. In effect, the stingless EN models could be utilized to provide important information to beekeepers and farmers practicing insects pollinated mono and mixed cropping systems for increased productivity. On the other hand, the results for this study could provide insights for conservationists grappling with effects of climate change, increased anthropogenic disturbance and habitat fragmentation towards natural vegetation stands.

Acknowledgements

The authors express their heartfelt gratitude to bee farmers in Kenya who provided information towards this research, without which this research may not have been successful. We would like to thank the personnel who participated in field sampling for response variables and information gathering. In addition, we are grateful for the support granted by Sidi Mangi, Felix Sereti, Fr. Gabriel Dolan, Meg Ryan, Tara Madden, Salome Nduta, Simphiwe Sidu and Dr. Bester Mudereri for data collation, processing, analysis and support. We would also like to appreciate the European Union for providing resources utilized in the field sampling exercise (Project number; DCI-FOOD-2011/023-520), Bayer AG Crop Sciences (Project number: NC20662450) and NORAD (Project number: RAF-3058 KEN-18/0005). We gratefully acknowledge the financial support for this research by the following organizations and agencies: UK's Foreign, Commonwealth & Development Office (FCDO); the Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. The views expressed herein do not necessarily reflect the official opinion of the people and agencies that supported the authors.

CHAPTER THREE

3 Suitability of Resampled Multispectral Datasets for Mapping Flowering Plants in the Kenyan Savanna

Terminalia brownie flowering on the landscape of Kasanga in Mwingi, Kenya. Photo taken by the candidate in January 2014.

This chapter is based on:

Makori, D. M., Abdel-Rahman E. M., Landmann, T., Mutanga, O., Odindi, J., Nguku, E., Tonnang, H. E., and Raina, S. 2020. Suitability of resampled multispectral datasets for mapping flowering plants in the Kenyan savanna, PloSONE, 15(9), e0232313, https://doi.org/10.1371/journal.pone.0232313

Abstract

Pollination services and honeybee health in general are important in the African savannas particularly to farmers who often rely on honeybee products as a supplementary source of income. Therefore, it is imperative to understand the floral cycle, abundance and spatial distribution of melliferous plants in the African savanna landscapes. Furthermore, placement of apiaries in the landscapes could benefit from information on spatiotemporal patterns of flowering plants, by optimising honeybees' foraging behaviours, which could improve apiary productivity. This study sought to assess the suitability of simulated multispectral data for mapping melliferous (flowering) plants in the African savannas. Bi-temporal AISA Eagle hyperspectral images, resampled to four sensors (i.e., WorldView-2, RapidEye, Spot-6 and Sentinel-2) spatial and spectral resolutions, and a 10-cm ultra-high spatial resolution aerial imagery coinciding with onset and peak flowering periods were used in this study. Ground reference data was collected at the time of imagery capture. The advanced machine learning random forest (RF) classifier was used to map the flowering plants at a landscape scale and a classification accuracy validated using 30% independent test samples. The results showed that 93.33%, 69.43%, 67.52% and 82.18% overall accuracies could be achieved using WorldView-2, RapidEye, Spot-6 and Sentinel-2 data sets respectively, at the peak flowering period. Our study provides a basis for the development of operational and cost-effective approaches for mapping flowering plants in an African semiarid agroecological landscape. Specifically, such mapping approaches are valuable in providing timely and reliable advisory tools for guiding the implementation of beekeeping systems at a landscape scale.

Keywords: Melliferous plants, bee health, AISA Eagle, hyperspectral data, resampled multispectral data sets, random forest.

3.1 Introduction

African savannas are characterized by unreliable and erratic rainfall with low and dispersed forest pockets. They do not efficiently support rainfed agriculture, necessitating for alternative sources of income to supplement the unpredictable crop yield (Makori et al., 2017; Raina et al., 2011). Apiculture and related ecosystem services such as pollinator activities boosts local economies, food and nutritional security and improve biodiversity, hence valuable and sustainable socioecological practices in the African savanna (Kiatoko et al., 2014; Klein et al., 2007).

Beneficial insects such as honeybees, stingless bees, wasps, butterflies, mulberry and wild silk moths (Raina et al., 2011; Warui et al., 2018a), are important income sources to local communities and living adjacent to the forest, paramount in pollination and pivotal incentives in forest conservation. These insects are valued for among others production of honey, wax, dyes and silk on one hand (Adolkar et al., 2007; Kasina et al., 2009), and pollination of agricultural and forest ecosystems, as well as conservation of derelict land and degraded forest (Blay et al., 2008; Raina et al., 2011). It is estimated that between 60% to 90% of plant species depend on insects and other animals such as birds for pollination (Potts et al., 2010; Raina et al., 2011). In the USA for instance, pollination is estimated to contribute more than $14 billions a year to the agricultural economy alone (Hines and Hendrix, 2005; Klein et al., 2007). However, whereas the contribution of pollination in Africa is unequivocal, its economic importance is yet to be fully documented.

As human pressure on land increases, communities living within four kilometres radius from the forest increases. These communities directly or indirectly depend on a range of forest resources and ecosystem services (Kiatoko et al., 2014; Klein et al., 2007). For instance, placement of apiaries in close proximity of a forest (less than one kilometre) doubles the production of honey than when placed out of a three kilometre radius (Sande et al., 2009). Additionally, proximity to the right types and amounts of pollen and nectar improves hive productivity, honey quality and the agility of bees to

fight off pests and diseases (Pirk et al., 2015; Sande et al., 2009; Warui et al., 2018a). Forest habitats are important to the various life cycles of many beneficial insect species (Allsopp et al., 2008; Kremen et al., 2007). As aforementioned, these insects are useful for among others pollination, improving biodiversity, diversification of livelihood options and as natural agents of pest control (Buchmann and Nabhan, 2012; Büchs, 2003). Disturbance of forest stands and savanna vegetation leads to reduction in pollen and nectar sources and ultimately pollinators, which are susceptible to habitat alteration and changes in climate (Ricketts et al., 2008). A decline or elimination of some plant species leads to a reduction of certain type of pollinators which are specific to either pollen type, flower colour and morphology or physiology (Ushimaru et al., 2007). Most insects rely on visual signals in the choice of flowers to visit. Colour, shape and size influence insects in flower preference (Cooley et al., 2008; Holzschuh et al., 2007; Pohl et al., 2011). For instance, hummingbirds prefer red coloured flowers, flies like pale colours, while butterflies and bees prefer brightly coloured flowers (Cooley et al., 2008; Holzschuh et al., 2007; Lunau and Maier, 1995; Pickering and Stock, 2003; Pohl et al., 2011). Since most of the crops are pollinated by social pollinators such as honeybees, agricultural production in Africa is predicted to reduce as their numbers decline (Winfree et al., 2008, 2007). In this regard, measures aimed at locating, conserving and improving cover of relevant flowering vegetation around vulnerable communities that depend on agriculture for their livelihoods are necessary.

Geoinformation and earth observation tools are increasingly being used to establish, locate and secure sources of nectar and pollen for enhanced hive productivity and ecosystem services (Evans and Schwarz, 2011; Pirk et al., 2015; Zayed, 2009b). Specifically, plant species mapping techniques adopting remotely sensed data assumes that each plant species has a unique spectral niche that is defined by the species biophysical and biochemical make-up (Asner and Martin, 2012; Cho et al., 2012; Clark et al., 2010). Therefore, it is possible to identify and separate every tree species using their spectral features. However, commonly, operational mapping of

tree species using remote sensing systems is hampered by the low spectral resolution in multispectral images and high acquisition cost of hyperspectral images. The improved division of the electromagnetic spectrum in hyperspectral data for instance gives narrow band data the ability to resolve subtle spectral canopy features associated with carotenoid, chlorophyll content and foliar nutrient content (Cho et al., 2012; Mitchell et al., 2012; Thulin et al., 2012). However, prohibitive cost, high dimensionality and multicollinearity, especially when using conventional parametric classification and regression procedures often make the use of hyperspectral data unfeasible. Specifically, the Hughes effect and the high redundancy rates of some bands in models developed using hyperspectral data impede landscape classification (Colgan et al., 2012; Rodriguez-Galiano et al., 2012; Sarhrouni et al., 2012; Yin et al., 2012). In this regard, it is paramount to explore the utility of multispectral images with fewer broad bands for optimal discrimination of functional flowering groups (Abdel-Rahman et al., 2015; Cho et al., 2012; Landmann et al., 2015; Vanbrabant et al., 2020). However, whereas broadband multispectral data of lower spectral and medium spatial resolution such as the Landsat series have become popular in landscape mapping, they could mask out specific spectral features of functional flowering groups, resulting in very low mapping accuracy. The newly launched relatively improved spectral and/or spatial resolution sensors such as WorldView-2, RapidEye, Spot-6 and Sentinel-2 offer great potential in detecting different colours of functional flowering groups (Hedley et al., 2012; Immitzer et al., 2016, 2012). Such sensors are specifically designed to capture spectral properties at additional wavebands such as red-edge and yellow spectrum that mimic over 90% of plant biophysiological information (Marshall et al., 2012; Zhang et al., 2009; Zheng et al., 2018). However, the acquisition cost of some of the commercial multispectral data could limit their operational mapping applications. Hence, there is need first to explore the utility of simulated image data of such multispectral sensors and compare their usefulness with freely available ones for flower mapping (Becker et al., 2007; Schmidt and Skidmore, 2003).

In this study, we explored the utility of four simulated multispectral data (i.e., WorldView-2, RapidEye, Spot-6 and Sentinel-2) for detecting and mapping functional flowering groups in the African savannas during the beginning and peak flowering seasons.

3.2 Methods

3.2.1 Study area

The study was carried in Kasanga (0.770°S and 38.143°E and approximately 933 metres above sea level), about 17 km north of Mwingi town, in Mwingi Central Sub-county, Kitui County, Kenya (Figure 10). The study area covers about 7.88 km^2 in a semi-arid agroecological zone with relatively high temperatures (ranging from 15°C to 31°C). The lowest temperatures are experienced between the months of July to August, while higher temperatures are experienced twice a year, from February to March and September to October. Rainfall in Mwingi is relatively low with typically two peaks in April and November (mean annual precipitation of between 147 to 270) (Landmann et al., 2015). Vegetation in Mwingi varies from woody plants, shrubs to crops. They include *Azandirachta indica, Melia volkensii, Markhamia lutea, Zizyphus abyssinica, Albizia gummifera* and *Vachellia* spp. Flowering plants in Mwingi include *Terminalia brownie, Cassia diambotia, Aspilia mossambicensis, Solanum incanum, Cassia semea, Grewia* spp, *Boscia* spp, and *Vachellia* spp (Raina and Kimbu, 2005). Most of these plants start flowering from December to May, with peak flowering season in February. The flowering season is triggered by the onset of rainfall in November, while the March-April rainfall extends the flowering through to May.

In Mwingi, traditional agricultural practices such as tilling, bush clearance and charcoal burning lead to deforestation of natural forest patches, including melliferous plants. This in turn leads to reduction in honeybee products, pollination services and biodiversity (Delaplane, 2010; Landmann et al., 2015; Williams et al., 2010).

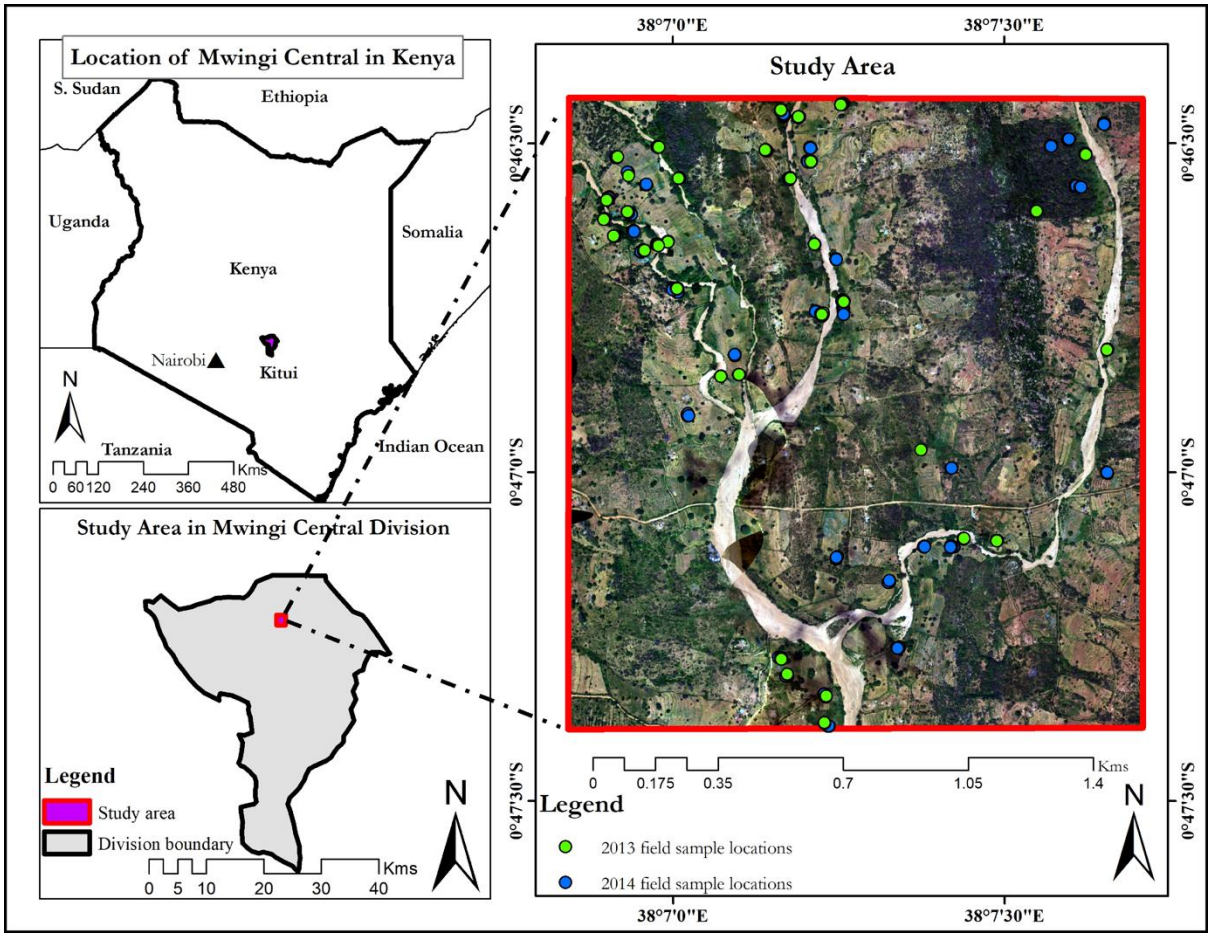


Figure 10: The study area indicating field data sample points on theAISA Eagle image background captured in February 2013 over the Mwingi study site.

3.2.2 Image acquisition and pre-processing

3.2.2.1 AISA Eagle hyperspectral images

The airborne AISA Eagle hyperspectral images were obtained during the onset and peak of flowering periods, i.e., 11th January 2014 and 14 February 2013, respectively. AISA Eagle has a pushbroom scanning sensor with 0.037° instantaneous field of view and 36.04° and 969 pixels across the spatial axis (Fauvel et al., 2008). To produce an optimal number of bands with high signal-to-noise-ratio (SNR), the sensor was set on eight times spectral binning at full width at half maximum (FWHM) of 8 – 10.5 nm in the 400 to 1000 nm spectral range. Hence, the product had 64 bands and a 0.6m spatial resolution after geo-referencing. Since this study sought to re-sample AISA Eagle

hyperspectral images to a range of multispectral sensor specifications, all AISA Eagle image bands were utilized to accommodate a range of sensor characteristics.

A digital elevation model (DEM) with 90m resolution, interpolated to match the 0.6m resolution of the images (Rabus et al., 2003), together with a 2m WorldView-2 image (37S UTM projection) captured over Mwingi in April 2014 were used to geo-reference the AISA Eagle images. The AISA Eagle raw digital values were also converted to at-sensor spectral radiance using the CaliGeoPro atmospheric correction tool (Specim Limited, Oulu, Finland).

3.2.2.2 Resampling of multispectral data from AISA Eagle

Both AISA Eagle hyperspectral images captured in January 2014 and February 2013 were resampled to WorldView-2, RapidEye, Spot-6 and Sentinel-2 multispectral spatial and spectral sensor specifications using the spectral resampling tool in Environment for Visualizing Images (ENVI) 5.3 software (Exelis, 2015). ENVI uses the full-width-half-maximum wavelength information to spectrally resample images. Since the spectral responses of these multispectral images were predefined in ENVI, the pre-defined filter function was used as a resampling method under the spectral resampling tool. These four multispectral image data were used to test the possibility of mapping flowering plants using 10 m or less spatial resolution. Also, we selected multispectral sensors that have infrared and red edge bands that have been established to be effective for detecting vegetation and flowers spectral signals (Abdel-Rahman et al., 2015; Jung-Rothenhäusler et al., 2007). Resampled multispectral data were used to avoid image acquisition cost before testing their suitability for mapping flowering plants. Further, the Sentinel-2 sensor was not yet launched when the study was conducted. Therefore, we obtained AISA Eagle data from previous studies (Abdel-Rahman et al., 2015; Landmann et al., 2015) that looked at the possibility of mapping flowering plants using hyperspectral image data.

Four multispectral sensors (WorldView-2, RapidEye, Spot-6 and Sentinel-2) were considered for this study. These images were resampled from their respective January 2014 and February 2013 AISA Eagle images. Additionally, a WorldView-2 image captured over the study site in April 2014 was used to compare the accuracy of the classification achieved using the resampled WorldView-2 image and provide insight on the reliability for our image resampling procedure.

3.2.2.2.1 WorldView-2

WorldVeiw-2 is an eight (8) waveband multispectral image sensor that was launched in October 2009 with a 1.84 m spatial resolution and an additional 0.46 metre panchromatic band (Anderson and Marchisio, 2012; Immitzer et al., 2012; Latif et al., 2012). The WorldView-2 sensor operates at an altitude of 770 kilometres and images at coastal, blue, green, yellow, red, red edge, near infrared 1, and near infrared 2 regions of the electromagnetic spectrum (EMS). These wavebands have distinct spectral separation (Table 5), hence potential for differentiating vegetation communities, health and other spectral features (Cho et al., 2012; Colgan et al., 2012; Immitzer et al., 2012; Latif et al., 2012). The cost of acquiring a tasked WorldView-2 image is $58 per km^2, with a 100 square kilometre and 5-kilometre minimum order width.

Table 5: Spectral responses of WorldView-2 indicating the lower and upper wavelength with the specific resolution of each band.

	Waveband	Lower λ (nm)	Upper λ (nm)	Spatial resolution (m)
1	Panchromatic	450	800	0.46
2	Coastal	400	450	1.84
3	Blue	450	510	1.84
4	Green	510	580	1.84
5	Yellow	585	625	1.84
6	Red	630	690	1.84
7	Red edge	705	745	1.84
8	Near infrared 1	770	895	1.84
9	Near infrared 2	860	1040	1.84

3.2.2.2.2 Rapid Eye

RapidEye is a 5-waveband multispectral sensor with 5 m spatial resolution (Table 6) that was launched in August 2008 (Anderson et al., 2014; Naughton et al., 2011). In addition to other multispectral bands, RapidEye is notable for its red edge band which has potential for mapping flower blooms (Abdel-Rahman et al., 2015). The cost of RapidEye image is $1.90 per square kilometre, with a 500 square kilometre and 10-kilometre minimum order width.

Table 6: RapidEye wavebands indicating the lower and centre wavelength with the specific resolution of each band.

	Waveband	Lower λ (nm)	Centre λ (nm)	Spatial resolution (m)
1	Blue	440	475	5
2	Green	520	555	5
3	Red	630	657.5	5
4	Red edge	690	710	5
5	Near infrared	760	805	5

3.2.2.2.3 Spot-6

Spot-6 is a multispectral sensor launched in September 2012. It is characterised by four wavebands with 6 m spatial resolution and a panchromatic band with 1.5 metres spatial resolution (Wang et al., 2015). Table 7 further details the sensor spectral and spatial information. The cost of Spot-6 image is $5.75 per square kilometre, with a 500 square kilometre and 20-kilometre order width.

Table 7: Spot-6 wavebands showing the lower, upper and centre wavelength with the resolution of specific image bands.

	Waveband	Lower λ (nm)	Centre λ (nm)	Spatial resolution (m)
1	Panchromatic	450	597.5	1.5
2	Blue	450	487.5	6
3	Green	530	560	6
4	Red	625	660	6
5	NIR	760	825	6

3.2.2.2.4 Sentinel-2

Sentinel-2 is multispectral sensor that was launched in 2015. It is characterised by 13 bands in the spectral ranges of visible/near infrared (VNIR) and shortwave infrared (SWIR) (Table 8), with spatial resolution ranging from 10 to 60 metres (Clasen et al., 2015; Hedley et al., 2012; Stratoulias et al., 2015, p. 2). Sentinel-2 data are freely available. When this study was conducted, the sensor was not yet launched hence the need to resample this dataset from AISA Eagle images.

Table 8: Sentinel-2 wavebands showing the central wavelength, bandwidth with the resolution of specific image bands.

	Waveband	Central λ (nm)	Bandwidth (nm)	Spatial resolution (m)
1	Coastal aerosol	442.7	21	60
2	Blue	492.4	66	10
3	Green	559.8	36	10
4	Red	664.6	31	10
5	Vegetation red edge	704.1	15	20
6	Vegetation red edge	740.5	15	20
7	Vegetation red edge	782.8	20	20
8	Near infrared	832.8	106	10
8A	Narrow near infrared	864.7	21	20
9	Water vapour	945.1	20	60
10	Shortwave infrared – Cirrus	1373.5	31	60
11	Shortwave infrared	1613.7	91	20
12	Shortwave infrared	2202.4	175	20

3.2.2.3 Field data collection

A stratified random sampling approach was used to collect field data on white and yellow flowering plants, flowering fobs, shrubs, green trees, senesced trees and crops (maize and sorghum) within three days of the AISA Eagle and WorldView-2 image data acquisition. Data on crops were only collected in January 2014 as they had already been harvested in February 2013. Additionally, trees that had chlorophyll-

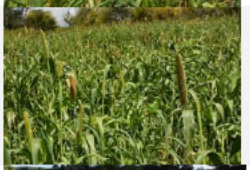
inactive leaves were identified, collected and tagged as 'senesced trees'. The Geospatial Modelling Environment (GME) tool was used to randomly generate 156 ground control points (GCPs) within the study site (Figure 10), from which the field data were collected. Table 9 details the number of samples on each sensor data, depending on their spatial resolution (pixel size) during the flowering seasons (i.e., January 2014; onset of flowering, February 2013; peak flowering and April 2014; end of flowering). The variations in the number of samples on each image data is due to differences in sensor pixel sizes.

Table 9: Field reference data used in the classification of the various flowering and other plant classes from different image datasets.

Class	AISA Eagle	WORLDVIEW-2	RAPIDEYE	SPOT-6	SENTINEL-2	
	February 2013					
Flowering fobs	320	89	115	34	31	
Green trees	241	36	54	21	15	
Senesced trees	450	85	211	66	56	
Shrubs	639	109	149	53	38	
Soil	1,369	432	422	142	111	
White flowers	472	84	161	54	44	
Yellow flowers	319	77	184	53	51	
	January 2014					WorldView-2 (April 2014)
Flowering fobs	1,659	155	22	27	35	148
Green trees	2,503	226	36	52	44	224
Senesced trees	1,178	107	19	43	13	101
Shrubs	4,304	392	65	112	83	389
Soil	8,740	787	126	236	161	783
White flowers	980	93	18	41	14	87
Yellow flowers	893	85	15	42	16	76

To locate the randomly generated GCPs, the data were loaded into a global positioning system (GPS) device with a 3-metre accuracy that was used to locate identified trees on the ground. Trees with three (3) metre canopy sizes (or more) were tagged and measurements taken to ensure all GCP readings were within the tree crowns. Other GCP's were collected from a 10 cm spatial resolution Nikon D3X digital camera image that was captured together with the AISA Eagle image data. Photos of the various functional groups taken by the lead author during the field surveys curried out in January 2014 and February 2013 are presented in Table 10 below;

Table 10: Photos of representative plants in the various flowering functional groups used for generating the flower maps in the Mwingi study site.

FUNCTIONAL GROUP	SPECIES	PHOTO
WHITE FLOWERS 1	*Terminalia brownie*	
WHITE FLOWERS 2	*Vachellia tortilis*	
YELLOW FLOWERS	*Vachellia nilotica*	
CROPS 1	*Zea mays*	
CROPS 2	Millet	
FLOWERING FOBS	*Ipomea vatke*	

SENESCED TREES	Senesced trees	

* crops 1 and crops 2 both belong to the functional group *'crops'* while white flowers 1 and white flowers 2 both belong to the functional group *'white flowers'*.

3.2.2.4 Random forest classifier

The supervised random forest (RF) ensemble algorithm (Breiman, 2001; Lawrence et al., 2006; Liaw and Wiener, 2002) with recursive partitioning was used to classify the different multispectral datasets. The ensemble is a robust machine learning algorithm that allows for growing of many regression trees (*ntree*) from bootstrap samples with replacement from the original data. It uses a majority voting procedure to assign classes to the reference datasets. Each tree uses two thirds (67%) of the randomly and independently selected dataset (*mtry*) for training the algorithm and one third (33%) of the remaining dataset for testing its accuracy (Breiman, 2001) using the out-of-bag (OOB) instances. All the bands in the resampled multispectral images were used as variables in the prediction while being optimized on the OOB error rate (Liaw and Wiener, 2002) using grid search and a 10-fold cross-validation method (Waske et al., 2009). RF has a high level of randomness in bagging (selection of datasets) with low sensitivity towards noise and overtraining, making it more suitable for simultaneous classification and variable selection (Breiman, 2001; Liaw and Wiener, 2002; Peters et al., 2007). The randomForest (Breiman, 2001) library in 'R statistical software' version 3.6.1 (Liaw and Wiener, 2002) was used for this study.

3.2.2.5 Accuracy assessment

An independent 30% random sample of the reference data points was used to test the classification accuracy of the RF classification model. To establish the versatility of the models, the following matrices were calculated; overall accuracy (OA), users' accuracy (UA), producer's accuracy (PA), quantity disagreement (QD) and allocation disagreement (AD) (Cohen, 1960; Congalton and Green, 2008; Pontius and Millones,

2011). QD and AD are absolute values that are used to evaluate the difference among predictions and reference data while comparing the percentage of observations that do not have optimal spatial locations as opposed to the reference samples (Pontius and Millones, 2011). It gives a better indication of the portion of the predictions that are of a good fit with the prediction data (Warrens, 2015a, 2015b).

3.3 Results

3.3.1 Flower compaction and spread

The classification maps for the different flowering seasons; onset of flowering season (January 2014), maximum flowering season (February 2013) and end of the flowering season (April 2014), are presented in Figures 11, 12, and 13, respectively.

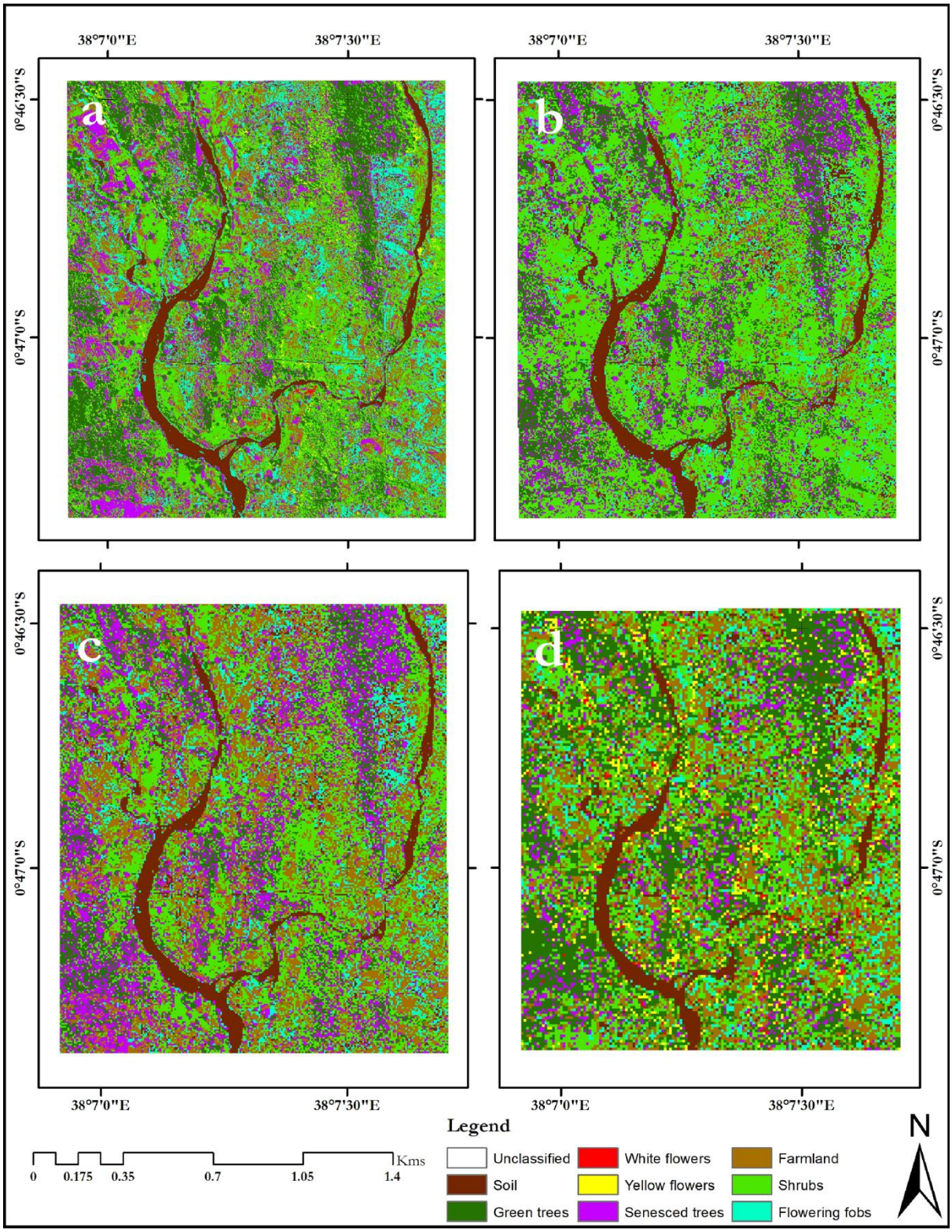


Figure 11: January 2014 classification maps for resampled simulated WorldView-2 (a), RapidEye (b), Spot-6 (c), Sentinel-2 (d) images.

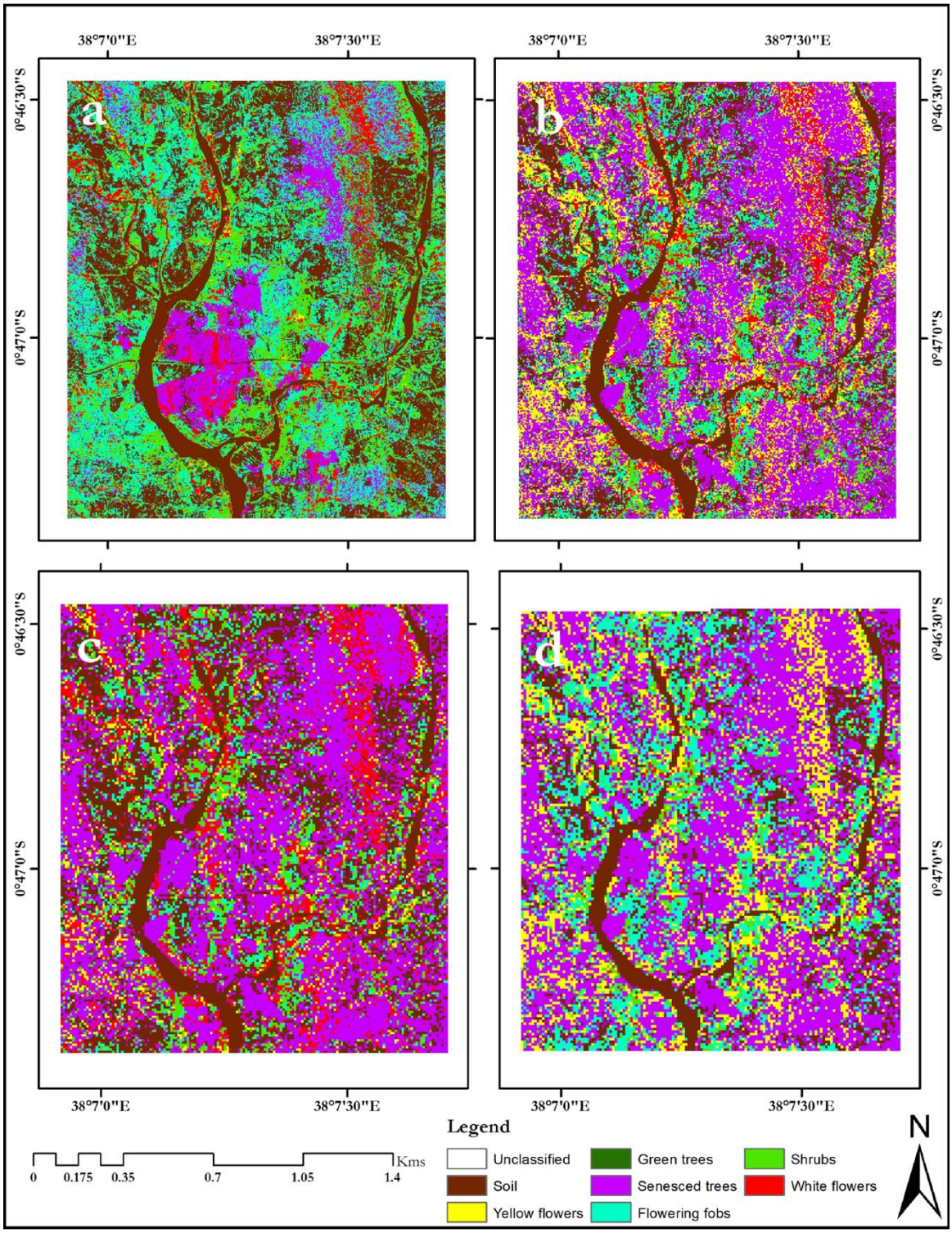


Figure 12: February 2013 classification maps for resampled simulated WorldView-2 (a), RapidEye (b), Spot-6 (c), Sentinel-2 (d) images.

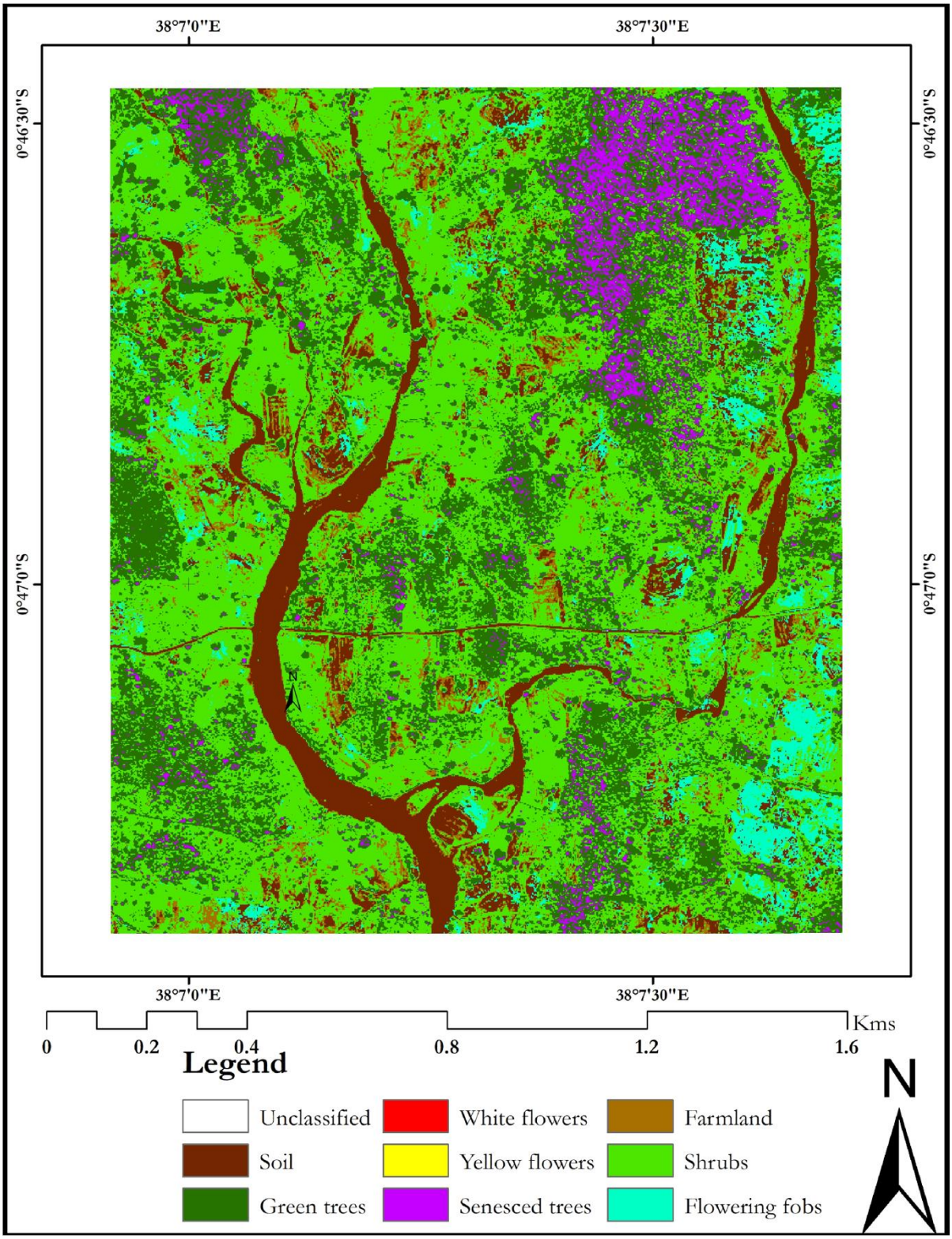


Figure 13: April 2014 classification map obtained using random forest classifier and WorldView-2 image.

Generally, flower classes were mapped with lower individual class accuracies (up to 87.36%) during the onset of the flowering season (Figure 11 and Table 11) as opposed to the maximum flowering season (up to 93.33%, Figure 12 and Table 12), using different multispectral sensor data. Specifically, both yellow and white flowers were reliably delineated with accuracies of up to 94.44% and 90.00% respectively (Table 11). These accuracies were higher than mapping accuracies of other non-flowering plants (Table 11). There was more flower compaction in February (peak flowering season) than January (onset of flowering season). Furthermore, there was more flowering fobs spread across the Mwingi study site in the peak flowering season compared to the onset of the flowering season, but less towards the end of the flowering season (Figure 13 and Table 12). Flowering fobs had the highest classification accuracy compared to the other flower classes (Table 11). The spread of the flowering fobs reduced as the peak flowering season came to an end towards April (Tables 12 and 13). Typically, April is the end flowering season in Mwingi, hence the reduced flowering. On the other hand, flowering crops reduced towards the peak flowering period as flowering in other melliferous plants increased. Results also show that crops were mapped with relatively higher accuracy (79.31%) at the start than later in the flowering season (65.08%).

The delineation of flowering plant species differed with sensor spatial and spectral resolutions, with a decrease in spatial and spectral resolution, leading to a decline in classification accuracies (Tables 11, 12 and 13). WorldView-2 generated the best overall classification results (87.36% at the onset, 93.33% at the peak and 76.40% at the end of the flowering seasons, Tables 11 and 12). Using the WorldView-2 data, subtle flowering differences were identified between the three capture periods; i.e., January 2014, February 2013 and April 2014.

Visual observation of the classification maps indicated that Sentinel-2 had the poorest presentation of the different flowering plants under consideration (Figures 11 and 12). The image and classes were more pixilated with more 'salt and pepper' than

classification maps from other sensors. However, the mapping accuracy presented in Tables 11 and 12 indicated that Sentinel-2 data mapped the flowering plants more accurately than RapidEye and Spot-6 data, even though the latter two had better spatial resolution. Spectrally, RapidEye, Spot-6, WorldView-2 and Sentinel-2 have five, six, eight and twelve spectral bands respectively, a clear indication that spectral resolution influence mapping of melliferous plant species. On the other hand, our maps show over prediction of senesced tree class, particularly during the peak flowering season (February 2013).

Table 11: Confusion matrix for the classification of flowering vegetation communities obtained using resampled simulated WorldView-2, RapidEye, Spot-6 and Sentinel-2 datasets.

	January 2014									February 2013								
	WorldView-2																	
	FB	GT	ST	SB	SL	WF	YF	Totals	UA	FB	GT	ST	SB	SL	WF	YF	Totals	UA
FB	38	0	0	3	2	0	0	43	88.4	25	0	1	1	0	1	0	28	89.3
GT	0	62	1	8	0	2	0	73	84.9	0	10	0	0	0	1	0	11	90.9
ST	0	0	24	1	0	0	0	25	96	1	0	24	0	0	0	0	25	96
SB	3	5	5	92	0	15	14	134	68.7	0	0	0	29	0	5	5	39	74.4
SL	2	0	0	0	234	0	0	236	99.2	0	0	0	0	129	0	0	129	100
WF	0	0	1	2	0	10	1	14	71.4	0	0	0	1	0	18	1	20	90
YF	0	0	1	2	0	0	10	13	76.9	0	0	0	1	0	0	17	18	94.4
Totals	43	67	32	108	236	27	25	538		26	10	25	32	129	25	23	270	
PA (%)	88.4	92.5	75	85.2	99.2	37	40			96.2	100	96	90.6	100	72	73.9		
OA (%)									87.4									93.3
Kappa									0.811									0.908
AD (%)									6.69									2.96
QD (%)									5.95									3.70
	RapidEye																	
FB	3	0	0	1	0	0	0	4	75	16	2	2	4	0	2	4	30	53.3
GT	0	6	2	3	0	1	2	14	42.9	0	4	0	0	0	0	1	5	80
ST	0	1	3	1	0	1	0	6	50	7	1	48	3	4	2	3	68	70.6
SB	2	2	0	10	0	2	2	18	55.6	5	2	6	25	0	7	4	49	51
SL	0	0	0	0	37	0	0	37	100	2	0	4	0	122	2	1	131	93.1
WF	0	0	0	0	0	0	0	0	0	2	4	0	6	0	23	12	47	48.9
YF	0	1	0	0	0	1	0	2	0	2	3	3	6	0	12	30	56	53.6
Totals	5	10	5	15	37	5	4	81		34	16	63	44	126	48	55	386	

PA (%)	60	60	60	66.7	100	0	0			47.1	25	76.2	56.8	96.8	47.9	54.6		
OA (%)									72.8									69.4
Kappa									0.551									0.619
AD (%)									17.28									26.42
QD (%)									9.88									4.15
									Spot-6									
FB	2	0	0	1	2	0	0	5	40	9	0	2	4	0	2	3	20	45
GT	0	9	1	2	0	1	2	15	60	0	4	0	0	0	3	3	10	40
ST	0	3	4	0	0	2	3	12	33.3	3	2	40	3	4	0	3	55	72.7
SB	0	1	3	24	0	4	3	35	68.6	6	0	1	19	1	9	4	40	47.5
SL	2	0	0	0	68	0	0	70	97.1	2	0	3	1	83	1	3	93	89.3
WF	0	1	0	2	0	2	0	5	40	3	3	1	3	0	15	7	32	46.9
YF	0	1	0	1	0	1	2	5	40	0	3	0	0	0	6	15	24	62.5
Totals	4	15	8	30	70	10	10	147		23	12	47	30	88	36	38	274	
PA (%)	50	60	50	80	97.1	20	20			39.1	33.3	85.1	63.3	94.3	41.7	39.5		
OA (%)									75.5									67.5
Kappa									0.620									0.597
AD (%)									17.69									24.09
QD (%)									6.80									8.39
									Sentinel-2									
FB	5	0	0	2	0	0	0	7	71.4	6	0	0	1	0	0	0	7	85.7
GT	0	11	1	0	0	1	0	13	84.6	0	3	0	0	0	0	0	3	100
ST	0	1	2	0	0	0	0	3	66.7	0	0	14	0	0	0	0	14	100
SB	5	1	0	19	1	2	1	29	65.5	1	0	0	8	0	0	2	11	72.7
SL	0	0	0	1	47	0	0	48	97.9	0	0	2	0	33	0	1	36	91.7
WF	0	0	0	0	0	0	1	1	0	0	1	0	1	0	8	1	11	72.7
YF	0	0	0	0	0	1	2	3	66.7	2	0	0	1	0	5	11	19	57.9
Totals	10	13	3	22	48	4	4	104		9	4	16	11	33	13	15	101	

PA (%)	50	84.6	66.7	86.4	97.9	0	50		66.7	75	87.5	72.7	100	61.5	73.3	
OA (%)								82.7								82.2
Kappa								0.699								0.777
AD (%)								10.58								10.89
QD (%)								6.73								6.93

Key:

FB - Flowering fobs
GT – Green trees
ST - Senesced trees
SB - Shrubs
SL - Soil
WF - White flowers
YF - Yellow flowers
PA - Predictor accuracy
UA - User accuracy
OA - Overall accuracy
AD - Allocation disagreement
QD - Quantity disagreement

Table 12: Confusion matrix for the classification of flowering vegetation communities obtained the end of the flowering season (April 2014) using the WorldView-2 image.

	Ground truth								
Prediction	FB	GT	ST	SB	SL	WF	YF	Total	UA (%)
FB	23	2	0	10	2	3	0	40	57.5
GT	1	40	12	6	0	6	8	73	54.8
ST	0	7	16	0	0	3	0	26	61.5
SB	8	13	0	83	6	7	7	124	66.9
SL	3	1	0	4	224	1	0	233	96.1
WF	1	1	2	2	0	5	2	13	38.5
YF	0	2	0	1	0	1	4	8	50
Totals	36	66	30	106	232	26	21	517	
PA (%)	63.9	60.6	53.3	78.3	96.6	19.2	19.1		
OA (%)									76.4
Kappa									0.695
AD (%)									17.79
QD (%)									5.80

Key:

FB – Flowering fobs
GT – Green trees
ST - Senesced trees
SB - Shrubs
SL - Soil
WF - White flowers
YF - Yellow flowers
PA - Producers' accuracy
UA - Users' accuracy
OA - Overall accuracy
AD - Allocation disagreement
QD - Quantity disagreement

3.3.2 Accuracy Assessment

Results presented in Tables 11 and 12 showed consistent differences between the January 2014 and February 2013 classification maps and between the different sensor datasets. WorldView-2 data had superior classification accuracy (87.36% in January 2014 and 93.33% in February 2013) than all the other sensor datasets. Spot-6 had the least overall accuracy (67.52%) in February 2013 while RapidEye had the least overall accuracy (72.84%) in January 2014. Both the users' and producers' accuracies reduced as the spatial resolution decreased (Table 12). On the other hand, the users' and producers' accuracies of Sentinel-2 were higher than those of both RapidEye and Spot-6.

Moreover, the maps had low quantity disagreement (QD) (3.70 to 9.88%) compared to the allocation disagreement (AD) (2.96 to 26.42%). Generally, there was more disagreement both in allocation and quantity as the spatial resolution decreased (Table 9). WorldView-2 had the best agreement fit in the two dimensions (as low as 2.96% AD), while RapidEye and Spot 6 had the highest disagreements (as high as 26.42% and 24.09% AD, respectively). On the contrary, Sentinel-2 with the least spatial resolution had better agreement (6.73% QD) than RapidEye and Spot-6.

Table 13: Classification accuracy of flowering vegetation communities during the start of the flowering season (January 2014) and the peak of the flowering season (February 2013).

Sensor	*Overall Accuracy*		*Allocation Disagreement*		*Quantity Disagreement*	
	2014	**2013**	**2014**	**2013**	**2014**	**2013**
WorldView-2	87.36	93.33	6.69	2.96	5.95	3.70
WorldView-2 (April 2014)	76.40	N/A	17.79	N/A	5.80	N/A
RapidEye	72.84	69.43	17.28	26.42	9.88	4.15
Spot-6	75.51	67.52	17.69	24.09	6.80	8.39
Sentinel-2	82.69	82.18	10.58	10.89	6.73	6.93

Figure 14 compares deviation in means between different classes in the two flowering periods. Results showed that flowering plants had less deviation from the mean than non-flowering plants. There was lower deviation at the peak flowering season i.e., February 2013 than at the onset of flowering i.e., January 2014. Contrary to the others, yellow flowers and flowering fobs signatures deviated more from the mean during the peak flowering period than at the onset of flowering. In the study site, yellow flowers start flowering at the onset of October rains while other flowering plants have their peak in February. Generally, flowering fobs were confused with soil and other background features in January 2014 than February 2013.

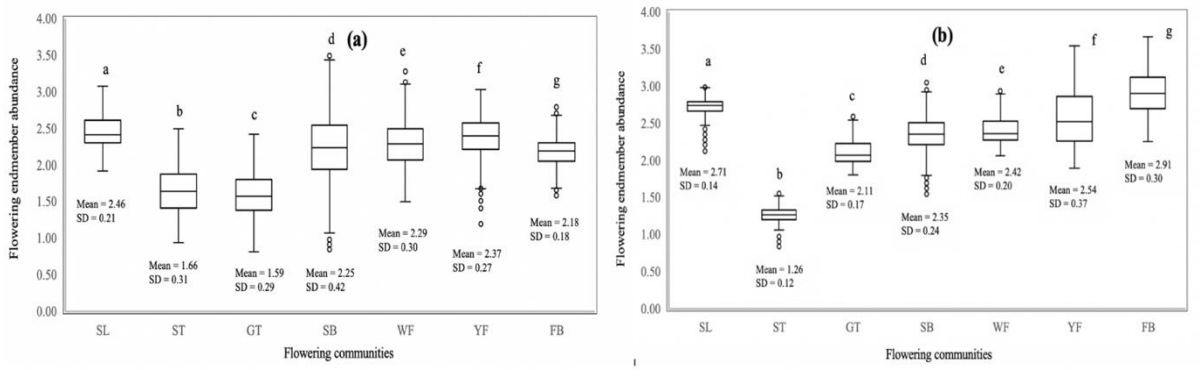


Figure 14: Box plots distribution of the flowering endmember abundancy of the flowering vegetation communities at the onset (a) and peak (b) of flowering seasons.

3.4 Discussion

Results in this study showed that resampled multispectral data of different spectral and spatial resolutions can be used to reliably map flowering plants in a heterogenous landscape in the African Savanna at different flowering seasons. Multispectral images provide affordable or freely available data useful for generating flower maps. Such maps could be operationally valuable in guiding the placement of honeybee apiaries at a landscape scale. Plants of different flower colour (i.e., white and yellow) were mapped with relatively higher accuracies. Specifically, we grouped the flowering plants in the study site into different functional flowering groups of different colours

as honeybees foraging behaviour is highly influenced by among others, flower colour (Hempel de Ibarra et al., 2014; Lunau and Maier, 1995; Morawetz et al., 2013; Pernal and Currie, 2001). Our study shows that large canopy white flowered acacia trees (predominantly *Vachellia tortilis* and *Vachellia brevispica*), white flowered *Terminalia brownie* and yellow flowered acacia (predominantly *Vachellia nilitica*) were mainly observed on low altitude areas within the study area. These three *Vachellia* spp and *Terminalia brownie* were observed to be the most foraged plants by honeybees in the Mwingi region (Raina et al., 2011). Therefore, apiaries in such areas could be placed within the flight radius for honeybees to capitalize on such flowering plants for optimised productivity.

The high flower mapping accuracy achieved in our study could be attributed to the relatively large flowering plant canopies, which ranged between 8 and 42 feet. These flowering plants were also characterized by higher blooming density and compaction (Landmann et al., 2015), particularly during the peak flowering season in February. A large plant canopy size with higher flower density enhance the floral spectral signal detected by the sensor and reduce the background heterogeneity on flowering signal, leading to improved flower detection accuracies (Shen et al., 2010).

Our results showed a better delineation of flowering plants using multispectral sensors with better spectral resolution. For instance, Sentinel-2, with relatively coarser spatial resolution (10m), mapped flowering plants with better accuracy than RapidEye (with 5 m spatial resolution) and Spot 6 (with 6 m spatial resolution). This is consistent with Mumby and Edwards (2002) who demonstrated better mapping accuracy of marine environments using higher spectral resolution images as opposed to images with fewer number of bands. Mumby and Edwards (2002) and Stratoulias et al., (2015) noted that in discriminating vegetation communities, spectral resolution could be more influential than spatial resolution.

Whereas other studies have argued that spatial resolution is more influential in vegetation mapping than spectral resolution (Becker et al., 2007; Schmidt and

Skidmore, 2003; Stratoulias et al., 2015), Abdel-Rahman et al. (2015) and Landmann et al. (2015) noted that spectral resolution improves the ability of an image to identify finer vegetation features like flower colour and compaction. However, we acknowledge that spatial resolution is complimentary to spectral resolution when mapping features such as flowering since flower blooms are more often smaller in size compared to other background features (Dalponte et al., 2013; Govender et al., 2008). It would therefore be expected that an image with higher spatial resolution would yield higher accuracy when coupled with higher spectral resolution.

The classification of the January 2014 image revealed that the onset of the flowering season was dominated by green trees and shrubs as compared to the peak flowering season (February 2013) image. Some of the trees, such as *Combretum* spp and *Terminalia brownie*, that were identified as green at the onset of flowering season (January 2014 image) were observed to be flowering at the peak of flowering (February 2013 image). The flowering of these species is triggered by the onset of the October and April rains (Muok et al., 2000). Although *Terminalia brownie* and *Combretum* spp scored lower than the *Vachellia* spp in honeybee foraging preferences (Raina and Kimbu, 2005), their adaptation at higher altitudes provide a viable foraging option for honeybee colonies from nearby apiaries.

The flowering fobs were also observed at higher and lower altitudes. These flowering fobs (predominantly *Ipomoea kituiensis vatke*) had the highest classification accuracies compared to the other flower classes studied. This finding is consistent with Landmann et al. (2015) who established that flowering fobs were the most accurately mapped class in the study area using hyperspectral data and spectral unmixing approach. The flowering fobs were more spread across the landscape as compared to other flowers that were mainly restricted to farms and vast open grounds. The flowering fobs had higher flower compaction, which could have supressed the background interferences from adjoining green leaves and soil. In this regard,

identification of their specific spectral signature was more accurate, hence the higher classification accuracy.

Flowering fobs, being ephemeral, are characterized by very short growing periods, especially on the higher grounds of the study site. The flowering fobs respond to the change in rainfall both in growth vigour and flowering intensity. The flowering fobs sprout out at the onset of the rains but lack in sustaining the rapid growth with low rainfall, hence rapidly die off. These flowering fobs however bridge the pollen and nectar gaps in low rainfall periods for improved bee diversity and honey production (Raina et al., 2011).

Additionally, there were more yellow flowers at the onset of the flowering season than at the peak of the flowering season at the study site. It was also noted that *Vachellia nilotica* (with yellow flowers) bloomed earlier than the other flowering trees. These tree species (*Vachellia nilitica* and *Combretum* spp) and flowering crops blossom as early as October, hence could provide foraging options to honeybees before the peak flowering periods. It's worth noting that our results showed over prediction of senesced trees during the peak flowering period, especially in images with low spatial and spectral resolutions (Spot-6 and RapidEye). This could be due to the confusion with the signature of some white flowering plants, soil, and trash, especially from bare farmlands.

Yellow flowers (predominantly *Vachellia nilotica*), were generally better mapped across all the data sets than white flowers. On the other hand, white flowering trees were mapped with the least accuracy as compared to other large sized flower blooms. This is consistent with Abdel-Rahman et al. (2015), who reported that white flowers signal was confused with other classes, especially soil, leading to their lower mapping accuracy than the yellow flowers. This could explain the low deviation in January 2014 as the yellow flowers were more compact during this period as compared to the February 2013 period. Landmann et al. (2015) also established that the user accuracy of mapping white flowering plants was lower during the peak than the onset of

flowering period due to confusion with dry white sand at the study site. The elevated water content in soil at the start of the flowering period helps distinguish the soil signature from white flowering plants. As opposed to the onset of flowering at the study site, peak flowering period is characterized by low rainfall and increased temperature. Therefore, the soil has bright characteristics that are similar to those of white flowering plants.

3.5 Conclusions

This study demonstrated the possibility of using multispectral images to map flowering plants in a semi-arid African savanna. The multispectral images are easily accessible, less expensive with less complex flower map production methods. Overall, the multispectral images tested produced acceptable classification accuracies (over 67%) which improved with both spatial and spectral resolutions. WorldView-2 produced maps with the highest classification accuracy while Spot-6 had the least classification accuracy. It was also evident that spectral resolution was paramount in mapping flowering plants. Increasing spectral resolution resulted in better classification accuracies. Furthermore, this study poised the freely available medium resolution Sentinel-2 as a valuable dataset for mapping flowering patterns in the African Savannas. Additionally, the methods used herein are more practical and available to practitioners with limited remote sensing knowledge, skills and resources, hence could be used to generate more information to farmers in the semi-arid African savannas. Results from this study could therefore be used to improve farmers' access to 'educated' information on the optimal locations for setting up apiaries in the African savannas to maximize honeybee products output and ecological services such as pollination. Even though this research considered images from three different flowering periods, this methodology could be more applicable to images from other vegetation periods within the. This could aid in phasing the optimal times for plant specific mapping and feature selection, which could be upscaled to wider regions.

3.6 Acknowledgements

We are indebted to the farmers of Mwingi for their contribution towards data collection and information gathering used in this manuscript. We give our gratitude to Mutemi and Munywoki for dedicating their valuable time to take the team round even to the very remote areas of the Kasanga area in Mwingi, and for being the invaluable link with some of the local people. Also, we would like to thank Betty Sidi, Sebit Diyar, Saum Hassan, Henry Oindo, Lesley Karwitha and Davis Okoth for providing ample working conditions and ensuring all that was needed for this research was availed. The views expressed herein do not necessarily reflect the official opinion of the people and agencies that supported the authors.

CHAPTER FOUR

4 Predicting Spatial Distribution of Key Honeybee Pests in Kenya Using Remotely Sensed and Bioclimatic Variables: Key Honeybee Pests Distribution Models

Small hive beetles (*Aethina tumida*) in a honey bee colony. Image by Giusti (2017); source https://agronotizie.imagelinenetwork.com/zootecnia/2017/06/09/aethina-tumida-nuovo-focolaio-ampliata-la-zona-rossa/54432

This chapter is based on:

Makori, D. M., Fombong, A. T., Abdel-Rahman, E. M., Kiatoko, N., Ongus, J., Irungu, J., Mosomtai, G., Makau, S., Mutanga, O., Odindi, J., Raina, S. and Landmann, T. 2017. Predicting spatial distribution of key honeybee pests in Kenya using remotely sensed and bioclimatic variables: key honeybee pests distribution models, International Journal of Geo-Information, 6(66), https://www.mdpi.com/2220-9964/6/3/66

Abstract

Bee keeping is indispensable to global food production. It is an alternate income source, especially in rural underdeveloped African settlements, and an important forest conservation incentive. However, dwindling honeybee colonies around the world are attributed to pests and diseases whose spatial distribution and influences are not well established. In this study, we used remotely sensed data to improve the reliability of pest ecological niche (EN) models to attain reliable pest distribution maps. Occurrence data on four pests (*Aethina tumida, Galleria mellonella, Oplostomus haroldi* and *Varroa destructor*) were collected from apiaries within four main agroecological regions responsible for over 80% of Kenya's bee keeping. Africlim bioclimatic and derived normalized difference vegetation index (NDVI) variables were used to model their ecological niches using Maximum Entropy (MaxEnt). Combined precipitation variables had a high positive logit (model information) influence on all remotely sensed and biotic models' performance. Remotely sensed vegetation variables had a substantial effect on the model, contributing up to 40.8% for *G. mellonella* and regions with high rainfall seasonality were predicted to be high-risk areas. Projections (to 2055) indicated that, with the current climate change trend, these regions will experience increased honeybee pest risk. We conclude that honeybee pests could be modelled using bioclimatic data and remotely sensed variables in MaxEnt. Although the bioclimatic data were most relevant in all model results, incorporating vegetation seasonality variables to improve mapping the 'actual' habitat of key honeybee pests and to identify risk and containment zones needs to be further investigated.

Keywords: honeybee pests; bioclimatic variables; remotely sensed variables; phenology; ecological niche modelling

4.1 Introduction

Bee keeping is an important economic activity globally. Its pollinator services are particularly indispensable to global food production and nutritional security (Klein et al., 2007; Nkoba et al., 2012). Additionally, bee keeping aids natural resource conservation, especially in communities living around forests. It diversifies households' incomes in African savannas often dominated by erratic and unreliable rainfall unable to sufficiently support rain-fed agriculture (Raina et al., 2011). The supplemental income comes from the sale of hive products such as honey, wax, propolis, royal jelly and to a lesser extent bee venom. However, honeybee health, pollination services and associated livelihood benefits are threatened by climate change, habitat alteration (fragmentation and loss), agriculture intensification, over dependence on agrochemicals, and increasingly by pathogens, pest and diseases (Muli et al., 2014c; Zayed, 2009a).

Pests are particularly considered the most economically important due to their spatial coverage and ability to inflict both direct (through physical injury) and indirect (as pathogen and disease vectors) damage. Despite their role in colony destruction, occurrence and distribution, honeybee pests remain understudied within the Africa tropics, with investigations restricted to Kenya, Malawi and South Africa (Pirk et al., 2016). Furthermore, data on in-country variations from one agroecological zone to another are scarce. In Africa, small hive beetle (*Aethina tumida*), large hive beetles (*Oplostomus haroldi* and *Oplostomus fuligineus*), wax moths (*Galleria mellonella* and *Achroia grisella*) and Varroa mite (*Varroa destructor*), are important pests capable of causing substantial economic damage (Fombong et al., 2012; Muli et al., 2014c; Pirk et al., 2016; Torto et al., 2010). Among these, the Varroa mite remains the single most devastating pest to honeybees due to its ability to inflict direct damage via its hematophagous (feeding on host) trophic habit and indirectly through the active transmission of honeybee viruses (Muli et al., 2014c; Mumoki et al., 2014; Pirk et al., 2016).

Little is known of the spatial and temporal distribution of these pests on the continent. The occurrence and distribution of these pests is influenced by various biotic and abiotic variables. Like other pests, honeybee pests can survive in certain optimal bioclimatic conditions. For instance, the optimal temperature, humidity, precipitation, altitude and biomass/net primary productivity ranges for different honeybee pests can vary significantly (Peterson and Nakazawa, 2008). Studies have shown that the reproductive ability of honeybee pests can be limited by the prevailing dry conditions and enhanced by hot and humid conditions (Neumann and Ellis, 2008; Torto et al., 2010). Vegetation phenology can also influence honeybee pest population dynamics and their species richness. Vegetation phenology determines the availability of food resources, which their hosts (honeybees) use to make hive products such as honey and beebread on which the pests thrive on. In addition, plant resins collected by honeybees are useful for hive construction and defence against foreign organisms (social encapsulation) and self-medication (Neumann et al., 2001; Simone-Finstrom and Spivak, 2012). Therefore, vegetation phenology plays a central role in colony health both directly and indirectly by affecting the life stages of pests such as the hive beetles, which occur outside the hive environment (Pau et al., 2012; Pirk et al., 2016).

Ecological niche (EN) modelling approaches can provide a pathway that statistically link the spatial variations in the bioclimatic variables to the distribution of a particular species (including honeybee pests). EN models rely on the correlations between habitats where different organisms (e.g., honeybee pests) thrive and environmental variables such as bioclimatic conditions, terrain and vegetation (Fernández and Hamilton, 2015; Kearney and Porter, 2009). However, the accuracy of the locations and temporal distribution of the honeybee pests and their habitats are important factors for deriving accurate models (Champetier et al., 2014; Peterson et al., 2002). Appropriate EN models that could accurately predict the spatial distribution of honeybee pest species rely on representative and accurate pest presence data together with carefully selected ecological and climatic predictor variables (Phillips and Dudík, 2008).

A fine and realistic geographic distribution of honeybee pest species is attained using remotely sensed variables such as vegetation 'greenness' dynamics (Cord et al., 2014a). This is because remotely sensed variables are continuous observations with high spatial and temporal resolution compared to bioclimatic variables that are modelled or interpolated (Pau et al., 2012). Well processed remotely sensed data also help in reducing over-fitting (over predicting) of the EN models as such improving the accuracy and value of species and pest distribution estimates (Cord et al., 2014a). Although a number of studies have elucidated the effects of various biological, climatic, vegetation and edaphic factors on some honeybee pests, such as small hive beetles (Neumann et al., 2016) and large hive beetles (Fombong et al., 2012; Torto et al., 2010), the contribution of vegetation phenology and bioclimatic interactions on their distribution and abundance in Kenya remains minimally exploited. Therefore, mapping the spatial distribution of the honeybee pests is valuable for risk assessment and zonation of honeybee pest risk areas to establish containment zones and pathways of source transmission. State of the art remote sensing techniques can be used to derive phenological variables (such as start and end of the season) from remotely sensed data that offer fine scaled and specific information for prediction of the spread of honeybee pests. A model that links remotely sensed and bioclimatic variables provides an operational system to monitor long term changes related to the spread of pests (Cord et al., 2014a). It also offers a tool that can model futuristic scenarios of the honeybee pest distribution since projected climate and phenological responses can be modelled. In this study, we aimed to identify key ecological/remotely sensed and bioclimatic variables associated with the distribution of honeybee pests in Kenya. We also sought to integrate remotely sensed variables for more coherent distribution maps using derived vegetation phenology parameters. We alluded to future scenarios of the distribution of these pests using modelled predictive data.

4.2 Methods

4.2.1 Study site

Occurrence data used in this study were collected from four main regions (agroecological zones) in Kenya; Coast, Mount Kenya, Mwingi and Kakamega (Figure 15), which are categorized by a representative climate gradient. An area covering 251,322.66 square kilometres, spanning from the coastal region through the central and eastern part of Kenya to Lake Victoria, was used to model the ecological niches of these bee pests (Figure 15). This study site consists of 32 administrative counties. Temperature and humidity are highest in the coastal region while the Mwingi site is located in a semi-arid savanna region which is hot and dry (Speranza et al., 2009). Mount Kenya and Kakamega are high altitudes regions that exhibit cooler and wetter climatic conditions (Githui et al., 2009).

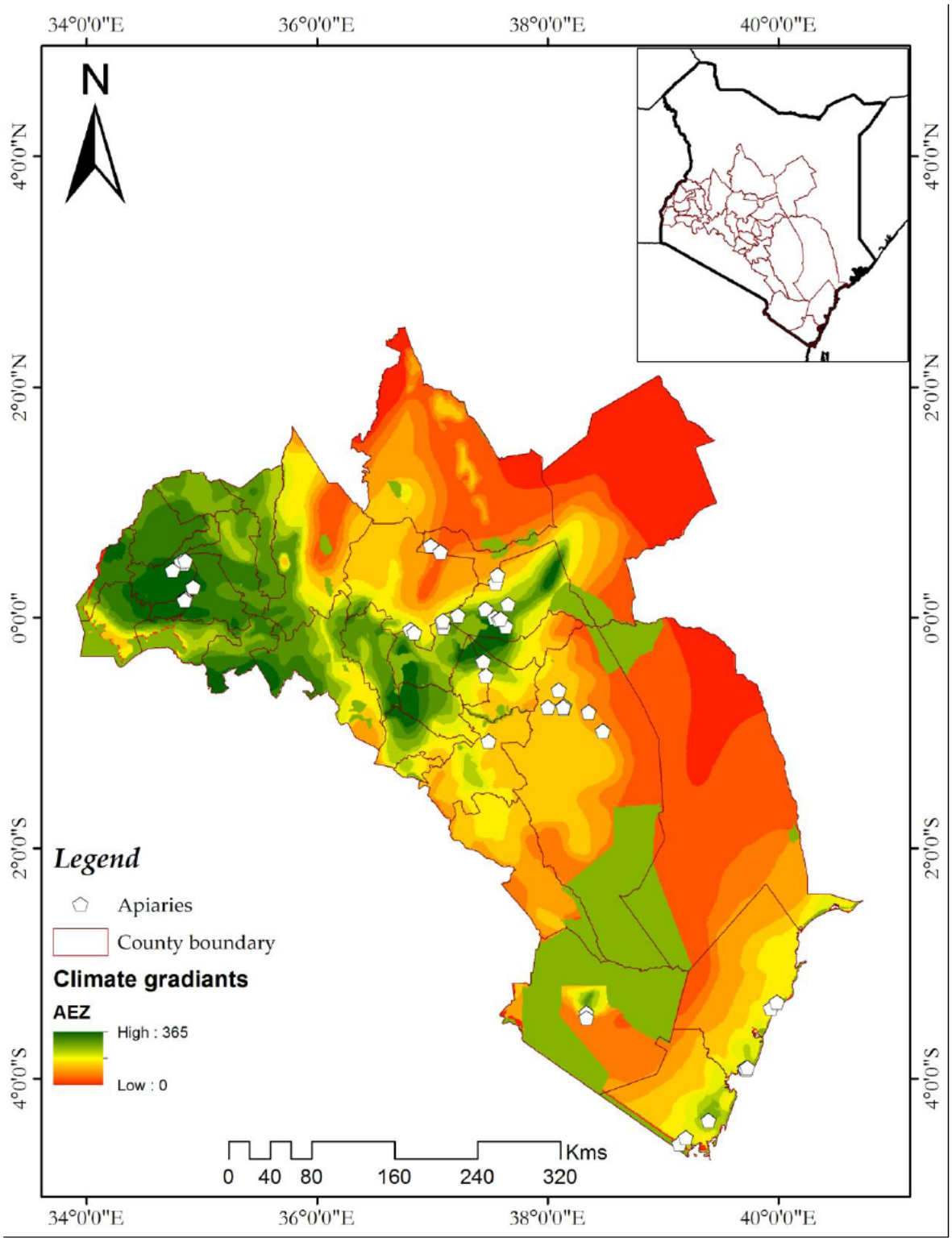


Figure 15. Distribution of apiaries (white points) from where occurrence data were collected in four regions in the study area on an agroecological zone backdrop (acz id 0 to 365, series 2).

4.2.2 Occurrence Data

Honeybee pest species occurrence data were collected during the wet season (March to April 2014) and dry season (June 2015). An apiary with 5–10 colonies in one location was treated as one sample point. Four honeybee pest species (*A. tumida, G. mellonella, O. haroldi* and *V. destructor*) were selected for this study. These four pests are the most frequently encountered and considered the most economically important with widespread prevalence as invasive pests globally (Fombong et al., 2012; Torto et al., 2010). A census of pests in five colonies in each apiary was done and their presence recorded using standard methods as described by Dietemann et al. (2013) for *A. tumida, G. mellonella* and *V. destructor*, and by Torto et al. (2010) for *O. haroldi*. We sampled 37 apiaries which had *A. tumida* present, 38 had *G. mellonella*, 24 had *O. haroldi*, and 45 had *V. destructor* in all the regions included in this study. These were within the sample sizes acceptable in the MaxEnt modelling environment (Amirpour Haredasht et al., 2013; Peterson and Nakazawa, 2008).

We assumed that individuals of these species were uniformly distributed across space. Presence data were collected from different sites independent from the environmental variables according to unknown probability of distribution. Therefore, apiaries were sampled randomly according to the requirements of estimation density, which obliges that individuals be sampled randomly across landscape in proportion to population density (Merow et al., 2013).

To test the influence of seasonality on the honeybee pest abundance, Mann-Whitney U-test was performed to test the significance ($p \leq 0.05$) differences between wet and dry seasons following the observation of heterogeneous variance.

4.2.3 Preparation of Data for Analysis

Prediction analyses were based on various geographical data sets, which were raster and vector, clipped to the boundary of the study site. These data sets were divided into two main groups; occurrence and predictor variables. Predictor variables were categorized into remotely sensed and bioclimatic variables (Table 14).

Table 14: Variables used in the ecological niche model clustered to remotely sensed biotic, bioclimatic and topographical variables.

Variable Name	Abbreviation	Description	Units	Year
A) Remotely sensed variables				
	1. Biotic Variables			
Start of the season	***seas_start***	Time for the start of the season	decades	2001–2014
End of the season	***seas_end***	Time for the end of the season	decades	2001–2014
Length of the season	***seas_length***	Length of the season from start to end	decades	2001–2014
Mid of the season	***seas_mid***	Mid of the season	decades	2001–2014
Base level	***base_level***	Average minimum NDVI value	n/a	2001–2014
Maximum NDVI	***max_ndvi***	Largest NDVI value in the season	n/a	2001–2014
Amplitude	***amplitude***	Difference between maximum and base level	n/a	2001–2014
Left derivative	***left_der***	Rate of increase at the beginning of season	%	2001–2014
Right derivative	***right_der***	Rate of decrease at the end of season	%	2001–2014
Large integral	***large_int***	Large season integral	n/a	2001–2014
Small integral	***small_int***	Small season integral	n/a	2001–2014
Number of seasons	***num_seas***	Number of seasons within the year	number	2001–2014
	2. Topographical variables			
Slope	***slope***	Steepness of the ground	% rise	
Aspect	***aspect***	Slope direction	degrees	
Hillshade	***hillshade***	Shading effect	n/a	
Elevation	***elevation***	Ground height	m	
B) Bioclimatic data				
	1. Temperature variables			
Bio 1	***bio1***	Mean annual temperature	°C	1961–1990
Bio 2	***bio2***	Mean diurnal range in temperature	°C	1961–1990

Bio 3	***bio3***	Isothermality	°C	1961–1990
Bio 4	***bio4***	Temperature seasonality	°C	1961–1990
Bio 5	***bio5***	Max temp warmest month	°C	1961–1990
Bio 6	***bio6***	Min temp coolest month	°C	1961–1990
Bio 7	***bio7***	Annual temp range	°C	1961–1990
Bio 10	***bio10***	Mean temp warmest quarter	°C	1961–1990
Bio 11	***bio11***	Mean temp coolest quarter	°C	1961–1990
Potential evapotranspiration	***pet***	Potential evapotranspiration	mm	1961–1990
	2.	**Precipitation variables**		
Bio 12	***bio12***	Mean annual rainfall	mm	1961–1990
Bio 13	***bio13***	Rainfall wettest month	mm	1961–1990
Bio 14	***bio14***	Rainfall driest month	mm	1961–1990
Bio 15	***bio15***	Rainfall seasonality	mm	1961–1990
Bio 16	***bio16***	Rainfall wettest quarter	mm	1961–1990
Bio 17	***bio17***	Rainfall driest quarter	mm	1961–1990
Moisture index	***mi***	Annual moisture index	n/a	1961–1990
Moisture index moist quarter	***mimq***	Moisture index moist quarter	n/a	1961–1990
Moisture index arid quarter	***miaq***	Moisture index arid quarter	n/a	1961–1990
Dry months	***dm***	Number of dry months	month	1961–1990
Length of longest dry season	***llds***	Length of longest dry season	month	1961–1990

4.2.4 Remotely Sensed Data Processing

4.2.4.1 Biotic Variables

Remotely sensed variables on vegetation productivity dynamics from Moderate Resolution Imaging Spectroradiometer (MODIS) time series data, at 250-meter spatial resolution, were processed and used as ecological predictor variables in the EN modelling (Table 1). Corrected MODIS 16-day NDVI composites were derived for the years 2001 to 2014 (14-year observation period). The corrected time series data sets were further processed in TIMESAT software (Eklundha and Jönsson, 2017; Jönsson and Eklundh, 2004) to derive vegetation seasonality variables that formed part of the input environmental variables in the Maximum Entropy modelling algorithm (Table 14).

TIMESAT analyses phenological signals found in time series satellite data by fitting local functions to the time series data points, then combines them into a global model (Clark et al., 2010; Jamali et al., 2015). Thereafter, a smooth model function is used to extract phenological variables for each growing season, which in turn reduces the influence of residual signal noise in the NDVI time series data (Foi et al., 2008; Shen et al., 2013) and data dimensionality (Fu and Wang, 2003; Hinton and Salakhutdinov, 2006). Function-fitting parameters used in TIMESAT for this study were: a Savitzky-Golay filter procedure, 3- and 4-point window over 2 fitting steps, adaptation strength of 3.0, no spike or amplitude cutoffs, season cutoff of 0.0, and begin and end of season threshold of 20%.

In total, 12 remotely sensed variables were extracted for each growing season; time for the start of the season (*seas_start*) which was set at 20% from the left edge minimum, time for the end of the season (*seas_end*) set at 20% from the right minimum, length of the season or time from the start to the end of the growing season (*seas_length*), base level, which was calculated by averaging the left and right minimum values (*base_level*), time for the mid of the season (*seas_mid*), largest data value for the fitted function during the season (*max*), seasonal amplitude (*amplitude*), rate of increase at

the beginning of the season (*left_der*), rate of decrease at the end of the season (*right_der*), large seasonal integral (*large_int*), small seasonal integral (*small_int*) and number of seasons in a calendar year (*num_seas*) (Table 14). Only variables for the first season were used in this study since data from the second season were not consistent throughout all the years.

4.2.4.2 Topographical Variables

Various topographical variables; elevation, slope, hillshade and aspect in radiation degrees (Table 14) were included in the EN model to predict occurrence of the four honeybee pests species. The topographical variables were derived from a filled 90 m digital elevation model (DEM) data from the Shuttle Radar Topography Mission (SRTM) instruments (Jarvis et al., 2008). The DEM data were resampled to fit the 250 m pixel size and the spatial extent of the MODIS 16-day NDVI datasets.

4.2.5 Bioclimatic Data

For simulation of the honeybee pest species distribution, current climatic conditions at one kilometre grid resolution from the AfriClim data set were used (Lovett, 2015; Mwalusepo et al., 2015; Platts et al., 2015). This data set contains grids of temperature, rainfall and derived bioclimatic summary variables. For prediction of distribution under future climatic condition simulations, downscaled data of the Representative Concentration Pathways Scenarios, Fifth Assessment Report (RCPs-AR5) (Pachauri and Mayer, 2015) future year 2055 (mean over 2041–2070) were used.

Bioclimatic data (Table 14) were divided into temperature and moisture variables. The temperature variables included mean annual temperature (*bio1*), mean diurnal temperature (*bio2*), isothermality (*bio3*), temperature seasonality (*bio4*), maximum temperature for the warmest month (*bio5*), minimum temperature of the coolest month (*bio6*), annual temperature range (*bio7*), mean temperature of the warmest quarter (*bio10*), mean temperature of coolest quarter (*bio11*), and potential

evapotranspiration (*pet*). Moisture variables were; mean annual rainfall (*bio12*), rainfall of the wettest month (*bio13*), rainfall of the driest month (*bio14*), rainfall seasonality (*bio15*), rainfall of the wettest quarter (*bio16*), rainfall of the driest quarter (*bio17*), annual moisture index (*mi*), moisture index of the moist quarter (*mimq*), moisture index of the arid quarter (*miaq*), number of dry months (*dm*), and length of the longest dry season (*llds*). All the 21 climatic variables were resampled to 250 m cell size and to the spatial extent of the MODIS 16-day NDVI datasets.

4.2.6 Ecological Niche Modelling

NMaxEnt tool package version 3.3.3k (Phillips et al., 2006). Probability of estimation for honeybee pest species occurrence was fitted to a set of pixels of features to maximize entropy of estimation density under given constraints of unknown variable phenomena. The values of these pixels were then used to estimate the distribution probability of these pest species. Automated random sampling of pseudo-absence/background samples from a set of pixels within the boundaries of Kenya (Rodríguez-Castañeda et al., 2012), where these species have not been detected (Peterson et al., 2007) was used to maximize predictivity. Pseudoabsence of the probability of presence which would otherwise be confined to presence only (Ward et al., 2009). Each species was assumed to have the same probability of being anywhere on the landscape, hence every pixel or environment on the landscape had the same probability of being tagged as "background" in geographic and environmental space.

4.2.7 Variable Selection

To minimize multicollinearity among predictor variables (Merow et al., 2013), we performed a Pearson correlation test between all the predictors presented in Table 14 to get orthogonal features suitable for the EN model. A correlation coefficient of $|r| > 0.7$ (Dormann et al., 2013) was set as a collinearity indicator for variables that would severely affect our EN model. Variables that met this criterion were eliminated based

on the variable importance analysis, which was performed to identify those with high explanatory power. Test runs were carried out using all bioclimatic and remotely sensed variables in Table 1 to analyse their contribution (Figure 16) to each pest species before building their niche models using the jackknife test (Peterson and Cohoon, 1999).

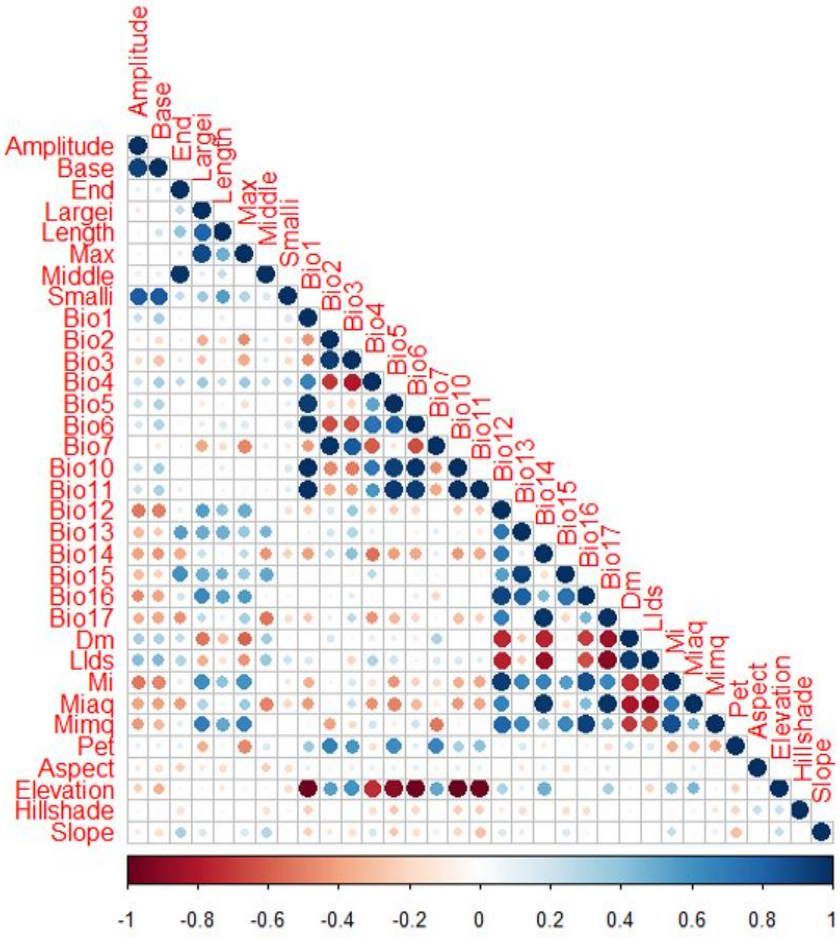


Figure 16. Collinearity matrix for ecological niche models predictor variables. Darker shades of blue and red colour indicate high variable collinearity while lighter shades indicate low collinearity.

A summary of the methodology is illustrated in a flowchart (Figure 17). The interactions of the various variables, occurrence data and selection of the suitable model are graphically represented.

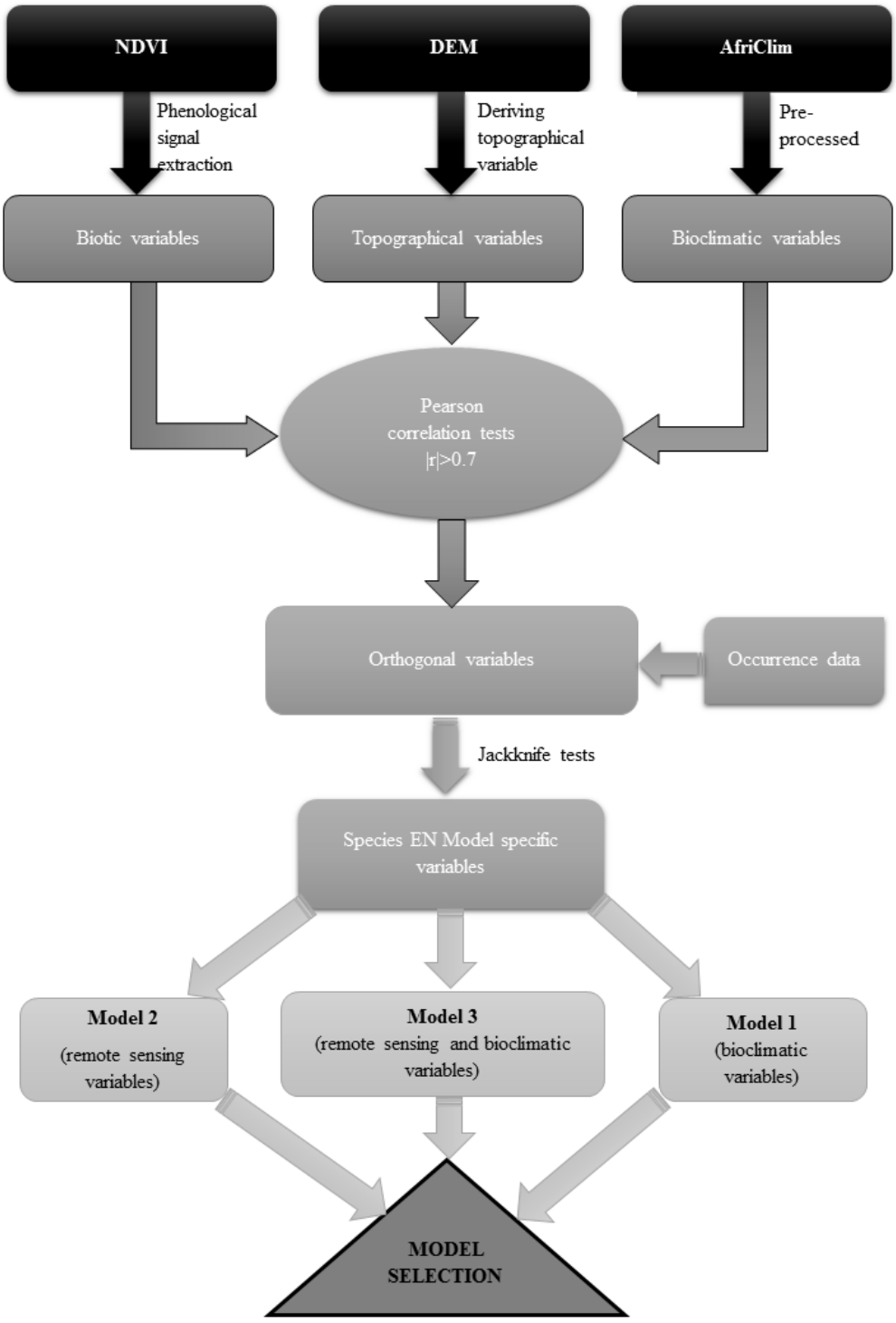


Figure 17. Graphical summary of the methodology illustrating the interaction between different variables used to build the three bee pest models and the selected model.

4.2.8 Model Setting

We used default MaxEnt settings with a few alterations aimed at attaining precise and accurate predictions (Merow et al., 2013; Phillips and Dudík, 2008). Twenty percent of occurrence records were withheld from each model to be used as independent test data. However, the default settings penalized our model by overfitting. Therefore, we opted to use the regularization multiplier of three to spread out the predictions and set the replicate runs at ten to obtain a robust model. Due to increased replication, random seed was used to select different random test/run partition for each run. We used the '10 percentile training presence threshold' which predicts the 10% most extreme presence observation as absent (Cord et al., 2014a) to eliminate outliers from the model.

Modelling was performed under current bioclimatic, vegetation and topographical conditions. Specifically, models were developed using only bioclimatic predictors (model 1), only remotely sensed variables (model 2) and both bioclimatic and remotely sensed variables (model 3) combined. These models were projected into the future (year 2055). We used future climate simulations from Africlim to project future scenario and due to absence of vegetation projection we assumed vegetation and topography to be unchanging over the projection period.

4.2.9 Model Evaluation

We used threshold independent area under curve (AUC) of receiver operating characteristic (ROC) analysis model (Phillips and Dudík, 2008) to assess model performance. The area under a ROC curve indicates the probability that presence (sensitivity) versus absence (specificity) or background points (Yackulic et al., 2013) were ordered correctly by a classifier. AUC values of zero (0) indicate impossible occurrence area while one (1) indicate optimal occurrence area or perfect ordering (Du et al., 2014). We used the Swets (Swets, 1988) (discriminatory power to rank the models as; i) excellent = 0.91–1.00, ii) good = 0.81–0.90, iii) fair = 0.71-0.80, iv) poor =

0.61–0.70, and v) fail = 0.51–0.60. To assess the utility of remotely sensed variables in improving the model performance, the null model (Raes and ter Steege, 2007) was used. Thirty-six randomly generated points from the observed presence cells were used to extract AUC values from each of the three models (model 3, model 1 and model 2). Mann-Whitney U-test was used to evaluate the significance ($p \leq 0.05$) differences between the three models. We further calculated the omission error of the various models developed, since our models used presence only data and calculating the commission error requires absence data (Hernandez et al., 2006).

4.3 Results

4.3.1 Honeybee Pest Abundance

Table 15 shows the abundance data of the four honeybee pests combined across the four regions (Mount Kenya, Mwingi, Kakamega and coastal regions) collected during the wet and dry seasons. Generally, there were more pests during the wet season compared to the dry season. There was a significant difference in pest abundance between seasons. Furthermore, *A. tumida* and *V. destructor* appeared to be more sensitive to precipitation changes as their abundance increased by almost six folds.

Table 15: Mean pest abundance recorded during the surveillance period in the wet and dry season in Kenya.

	A. tumida	***G. mellonella***	***O. haroldi***	***V. destructor***
Wet season	272a	20a	31a	476a
Dry season	58b	21a	15b	75b

4.3.2 EN Models

Table 16 shows the threshold-independent AUC obtained from the ROC analysis model, which indicated 'good' performance for all model 3 and all model 1 except *G. mellonella's*, and fair model 2 for all honeybee pests. Model 1 achieved higher AUC values for all pests than model 2. However, model 3 had higher AUC scores for *A. tumida* (0.87), *O. haroldi* (0.87) and *V. destructor* (0.88) than model 1 or model 2. The AUC values for *G. mellonella* model 3 (0.80) were the same as model 1 (0.80). However, model 3 prediction maps showed lower overfitting with low standard deviation in all honeybee pest species models (Table 16).

The null model analysis showed that there was significant difference between model 2 and models 1 and 3 for all species except for *O. haroldi.* However, there was no significant difference between model 1 and model 3. Therefore, model 2 was eliminated from further model comparison.

Table 16: Area under curve (AUC) and standard deviation (SD) of model 1, 2 and 3 developed for *A. tumida, G. mellonella, O. haroldi* and *V. destructor*.

	Model 1		Model 2		Model 3	
	AUC	**SD**	**AUC**	**SD**	**AUC**	**SD**
A. tumida	0.85	± 0.08	0.80	± 0.07	0.87	± 0.07
G. mellonella	0.80	± 0.11	0.75	± 0.09	0.80	± 0.08
O. haroldi	0.85	± 0.10	0.76	± 0.16	0.87	± 0.09
V. destructor	0.83	± 0.08	0.76	± 0.09	0.88	± 0.07

The pest models had low omission errors for model 3 than the other models for all pests (listed as a percentage for model 3, model 1 and model 2 respectively, *A. tumida* 16.6, 19.4 and 25.0, *G. mellonella* 27.7, 33.3 and 38.9, *O. haroldi* 30.5, 33.3 and 38.9 and *V. destractor* 19.4, 25.0 and 30.6). Model 3 performed better for all bee pests than the other models.

4.3.3 Predictor Variable Contribution

In all model 3 for the four honeybee pests, bioclimatic variables combined had the strongest relative influence than remotely sensed variables combined. Bioclimatic variables had a contribution of 88.1% for *O. haroldi,* 83.0% for *A. tumida*, 68.3% for *V. destructor, and* 58.7% for *G.mellonella* (Table 17). Further, in the bioclimatic cluster, precipitation variables had more influence (ranging from 58.0% to 81.0%) than temperature variables (ranging from 0.4% to 7.3%). Remotely sensed variables in model 3 on the other hand (biotic and topographical) had a total contribution of 17.0% for *A. tumida,* 41.3% for *G. mollenela,* 11.9% for *O. haroldi* and 31.7% for *V. destructor.* The combined relative contribution of remotely sensed biotic variables in model 3 ranged from 11.4% to 40.8% (Table 17). In this variable cluster, maximum NDVI in the season had a substantial relative contribution (between 7.7% and 29.4%) to the EN models. Topographical variables had a relatively lower contribution, less than 4.2% and no modelled effect on *V. destructor.*

Table 17: Percentage contribution of various variables to the four ecological niche models using jackknife.

	A. tumida	*G. mellonella*	*O. haroldi*	*V. destructor*
Remotely sensed variables (biotic variables)				
amplitude	4.3	2.0	0.7	6.3
base_level	-	8.9	1.5	6.1
large_int	0.8	0.5	-	-
max_ndvi	7.7	29.4	8.3	19.3
small_int	-	-	0.9	-
RS (biotic) total	12.8	40.8	11.4	31.7
Remotely sensed variables (topographical variables)				
hillshade	1.2	0.2	0.5	-
slope	3.0	0.3	-	-
Topographical total	4.2	0.5	0.5	-
Bioclimatic variables (precipitation and temperature variables)				
bio3	2.0	-	7.3	-
bio7	-	0.7	-	20.5
bio12	1.7	13.7	-	20.5
bio15	43.9	11.9	79.7	28.2
llds	-	-	1.1	0.5
mimq	35.4	32.4	-	18.7
Bioclimatic total	83.0	58.7	88.1	68.3
Grand total	**100**	**100**	**100**	**100**

Based on the jackknife tests, the relative variable importance (both bioclimatic and remotely sensed) to model 3 (Figure 18), showed varied contribution to the various EN models. The first five ranked variables (bio15, mimq, max_ndvi, amplitude and slope) cumulatively contributed to 94.3% in the *A. tumida* model. The first five variables (mimq, max_ndvi, bio12, bio15 and base_level) had 96.3% cumulative contribution to *G. mellonella* model and 92.8% to *V. destructor* model, while bio15, bio3 and max_ndvi had a cumulative contribution of 95.3% to *O. haroldi* model.

Bio15 had the highest gain on *A. tumida* (43.9%), *O. haroldi* (79.7%) and *V. destructor* (28.2%) in model 3. Mimq contributed most of the information used in predicting habitat suitability of *G. mellonella* (32.4%). All these variables had a positive logit on the respective model 3. Topographical variables did not have any influence on *V. destructor* with little influence on *G. mellonella* and *O. haroldi* models. On the other hand, *A. tumida* and *G. mellonella* model 1 indicated that mimq had the highest positive effect on their distribution, while bio12 influenced *V. destructor* most. The *O. haroldi* model 1 was mostly influenced by bio15. Compared to the other remotely sensed variables, max_ndvi exhibited the largest influence in all the four EN models pertaining to the four honeybee pests.

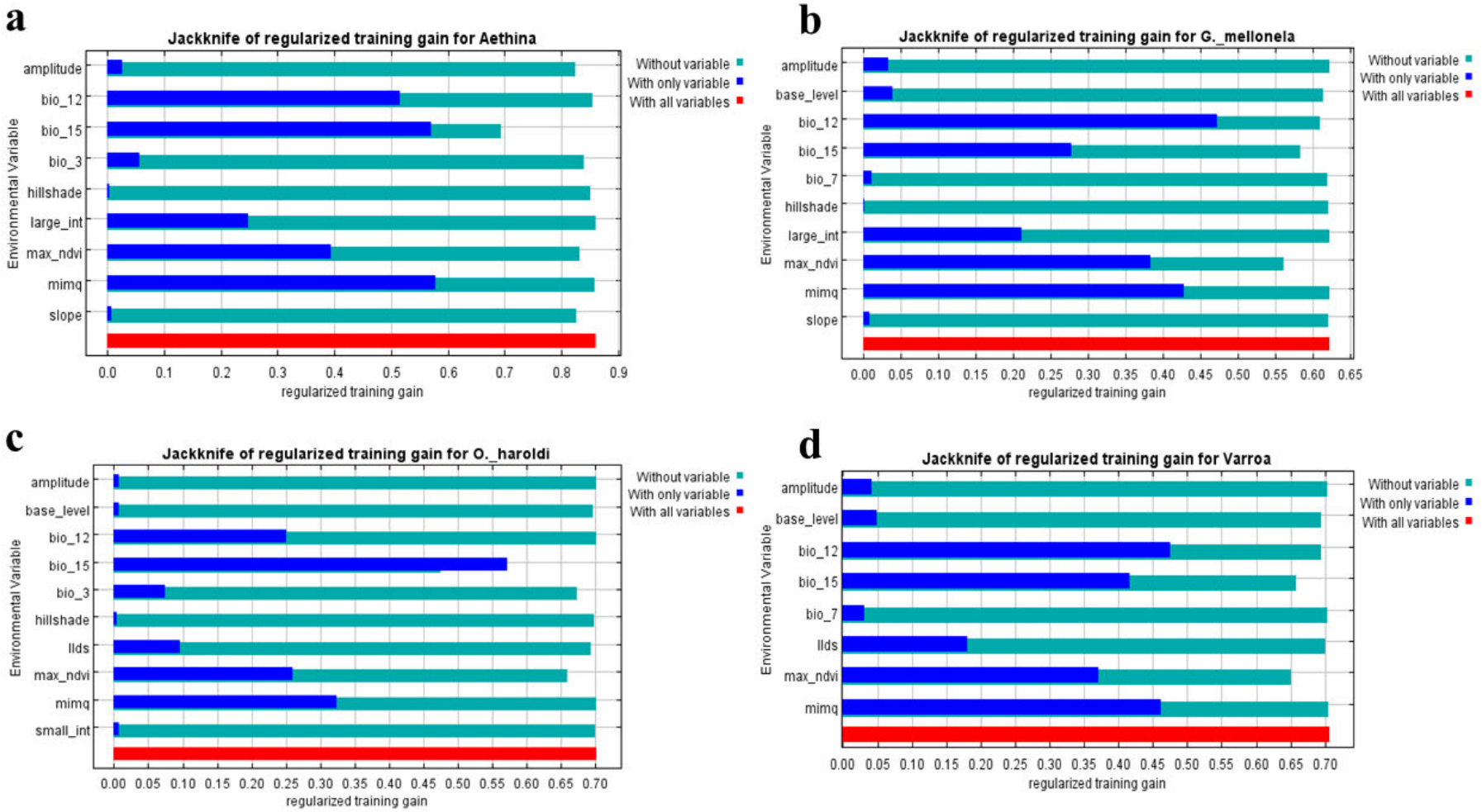


Figure 18. Jackknife variable contribution for (**a**) *A. tumida*, (**b**) *G. mellonella*, (**c**) *O. haroldi* and (**d**) *V. destructor* for remote sensing and bioclimatic model (model 3.

4.3.4 Visualization of Distribution

The honeybee pest models predicted high habitat suitability in the Mount Kenya and the coastal regions for all four honeybee pests (Figure 19). Mwingi region was predicted to have low (*G. mellonella*) to high (*O. haroldi*) habitat suitability. However, Kakamega region had varied likelihood for suitability of all honeybee pests. This area was predicted to have a low *O. haroldi* risk, moderate *V. destructor* to moderately high *A. tumida* and high *G. mellonella* habitat suitability (Figure 20). Mount Kenya, Kakamega and the coastal regions are characterized by high rainfall seasonality. The coastal region however had higher temperature and humidity than Mount Kenya and Kakamega regions. On the other hand, Mwingi had relatively high temperature and lower isothermality than Mount Kenya and Kakamega. Although the coastal and Mount Kenya regions had similar rainfall seasonality, they had different temperature and isothermality gradients. Conversely, the coastal region and Mwingi had higher mean annual temperature and lower isothermality than the other regions but exhibited different honeybee pest risk potentials.

Futuristic scenario models for the four honeybee pests generally indicated that high habitat suitability areas may spread to areas currently classified to have medium to low pest suitability (Figure 20). Modelled bioclimatic data essentially predict an increase in amount of rainfall in the wettest quarter and increased rainfall variability.

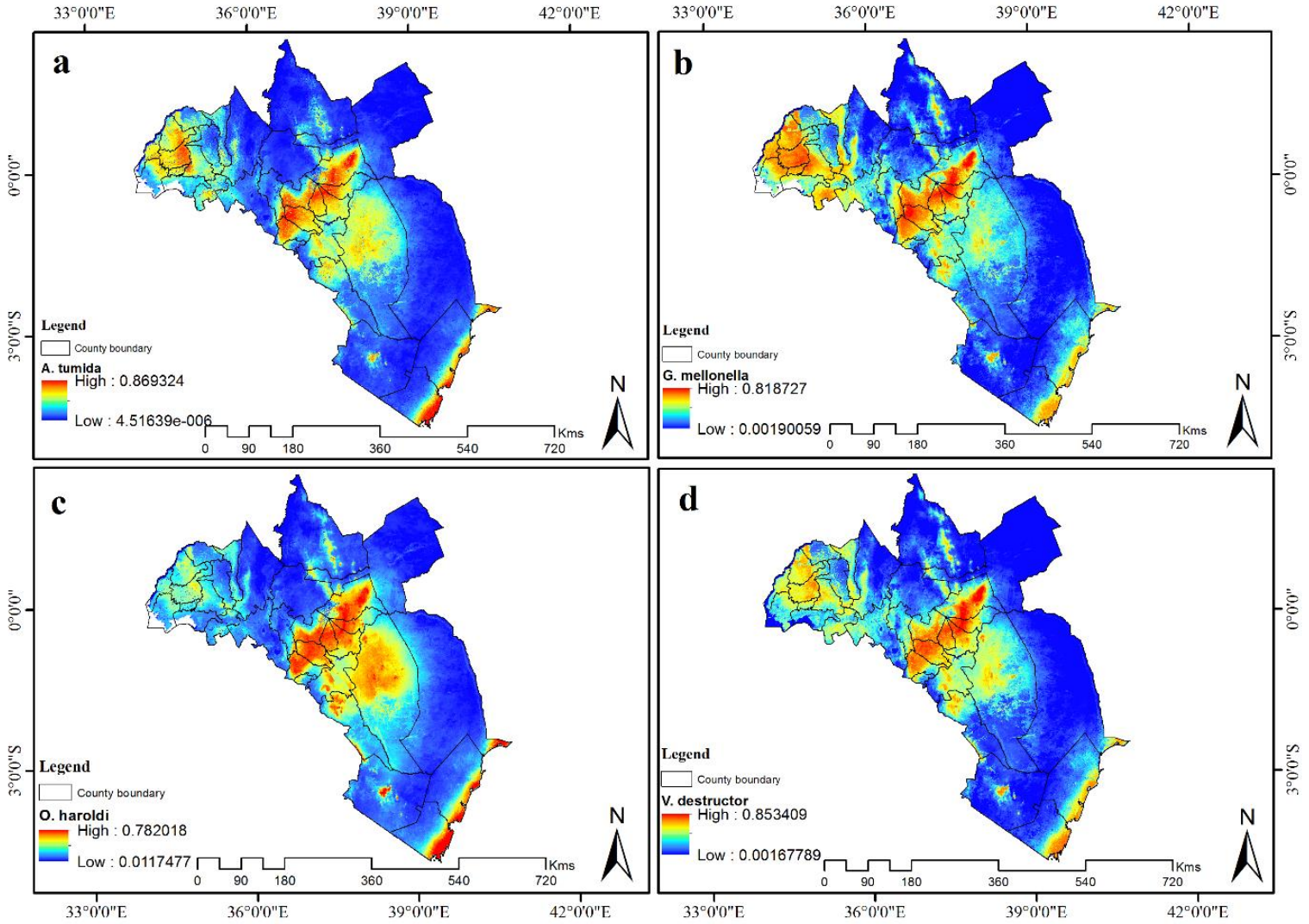


Figure 19. Current habitat suitability for (**a**) *A. tumida*, (**b**) *G. mellonella*, (**c**) *O. haroldi* and (**d**) *V. destructor* in Kenya.

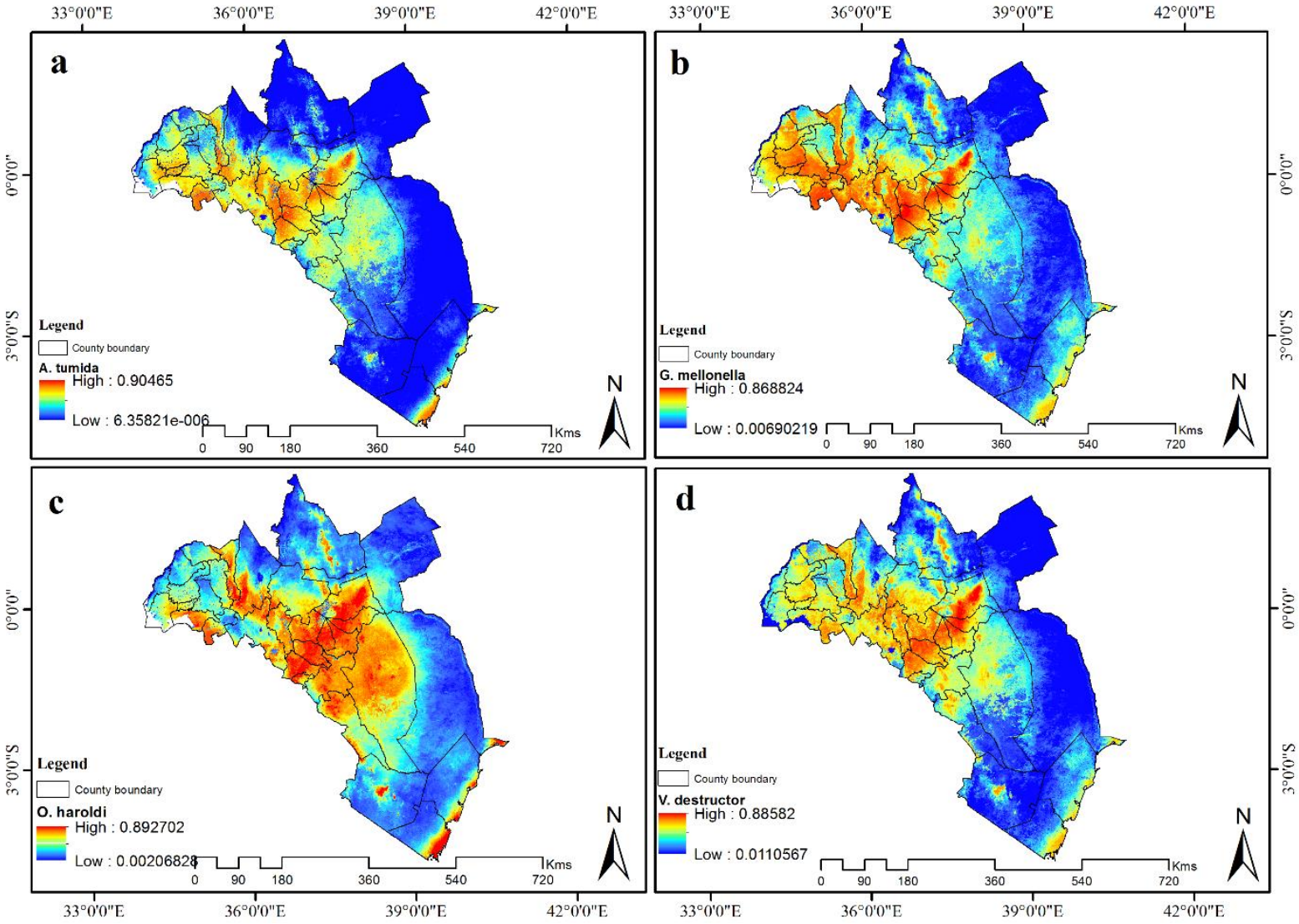


Figure 20. Projected (2055) habitat suitability for (**a**) *A. tumida,* (**b**) *G. mellonela,* (**c**) *O. haroldi* and (**d**) *V. destructor* in Kenya.

4.3.5 Contribution of Remotely Sensed Data

Remotely sensed variables increased the accuracy and precision of the models by reducing overfitting and refining their predictivity power (Table 16). A subset from the *V. destructor* maps showed that model 1 had over predicted distribution compared to model 3, which had remotely sensed variables with smaller pixel size (high resolution) and more environmental heterogeneity incorporated (Figure 21).

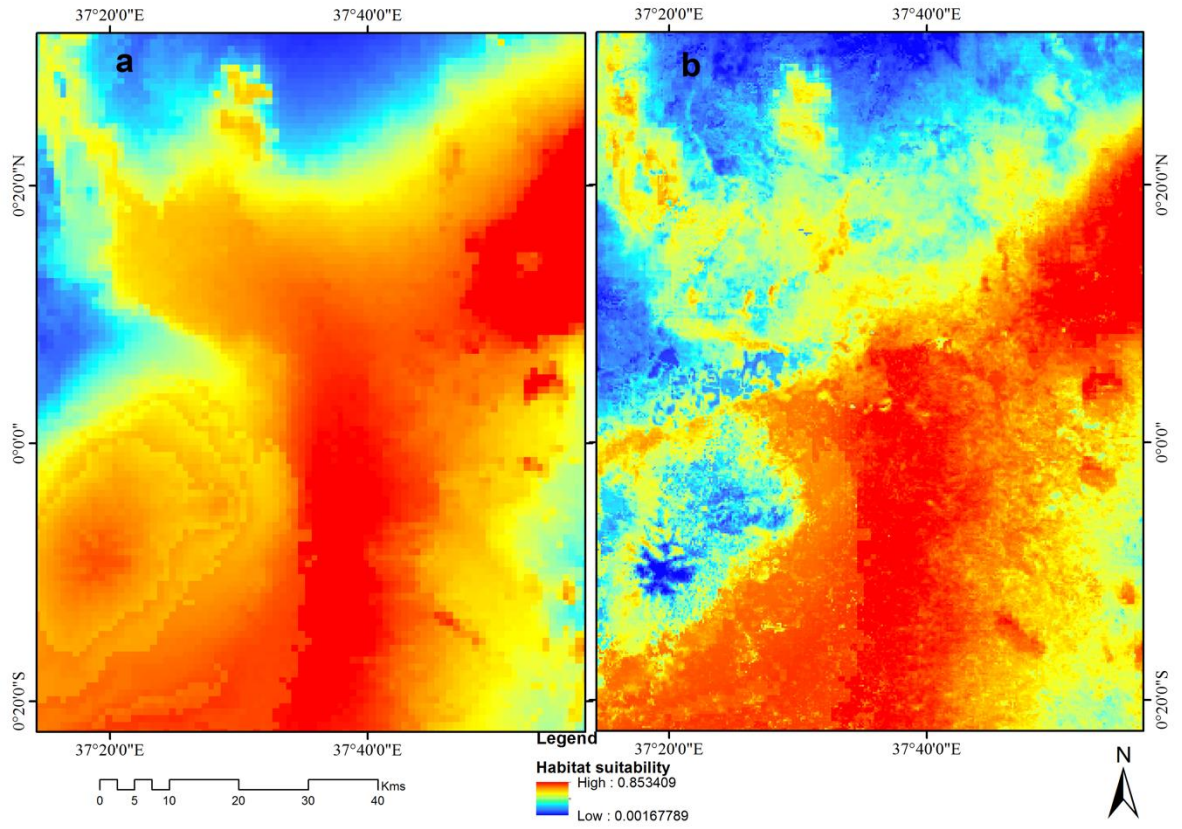


Figure 21: Improved predictivity and reduced overfitting of *V. destructor* observed in model 3 (b) when using remotely sensed variables compared to model 1 (a) .

Model 3 had lower mean of habitat suitability than the model 1 for all the honeybee pests investigated and lower standard deviation for *A. tumida* and *V. destructor* (Figure 22). Model 3 histogram shows that predictivity was consolidated around the mean while bio_model histogram shows skewness to the right. Consolidation around the mean was occasioned by remotely sensed variables, whose measurement at grain size ensured variability was detectable at higher precision limited by spatial resolution.

Moreover, model 3 had higher ranked AUC values and lower standard deviations than model 1 for all models except *G. mellonalla*, which, however, had lower standard deviation than its model 1 (Table 16).

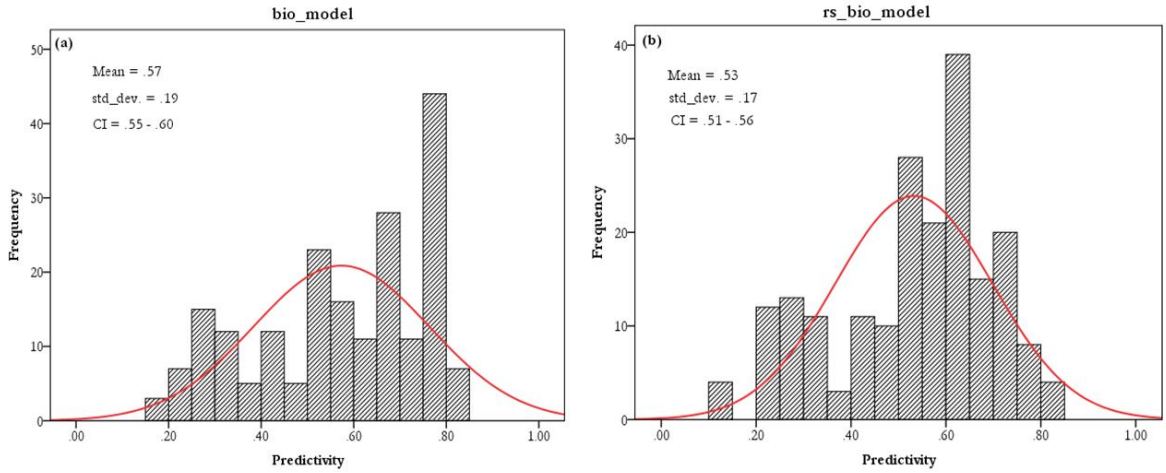


Figure 22. Predictivity frequency of *V. destructor* for model 3 (b) with lower mean and standard deviation (SD) than model 1 (a).

4.4 Discussion

Understanding factors that influence the multiplication and spread of honeybee pests across various spatial extents and the level of risk they pose is important in bee health. Therefore, improved predictability of pest diversity and accurate honeybee pests' distribution maps will provide tools and information necessary for honeybee pest control (Martin et al., 2012). This study sought to identify conditions that encourage increase of honeybee pests in Kenya and developed models that could be used to predict their spread. Projected models served to show the level of the risk posed by the honeybee pest to the beekeepers in future.

Our surveillance data confirmed the presence of variation in abundance and distribution of the four honeybee pest species across the study sites. There were more records observed during the wet than the dry season. These findings were in line with a previous study by Torto et al. (2010) which showed higher hive beetles' abundance

during rainy than the dry seasons in Kenya. However, observed seasonal variation in *Varroa* mite populations is contrary to previous reports of fairly constant mite infestations across *Apis mellifera scutellata* (the savanna honeybee) (Strauss et al., 2013). Additionally, the coastal lowlands are inhabited by *A. m. litorea* and forest highlands by monticolla-like bees, which hybridize with savanna bees along their migratory paths (Raina and Kimbu, 2005). Therefore, there is need to further understand the interaction of these honeybee pests across the different spatial divides.

4.4.1 Predictor Variable Contribution

Although most honeybee pests have wider spatial ranges and can be found in various regions with different climatic conditions (Wiley et al., 2003), there are specific climatic conditions that play a key role in determining their distribution. For instance, hive beetles and wax moth can thrive on alternative food sources while *Varroa* is an obligate ectoparasite. Therefore, the former two are more likely affected by climatic variations than the latter. We therefore identified four key bioclimatic and remote sensing variables to be the most important in determining occurrence of the four honeybee pests. Rainfall seasonality, mean annual rainfall, moisture index in moist quarter and largest NDVI value in the season had the most influence on the model 3 for the four honeybee pests (Table 17).

These variables are predominantly linked to precipitation apart from maximum NDVI in the season. Moreover, field occurrence data (Table 15) showed that there were more pests recorded in the wet season than the dry season. Similar results were presented by Torto et al. (2010) who documented that precipitation's positive effect on *Varroa* was indirectly linked to production of bee forage that influence colony growth and brood increment. This in turn provide ample reproductive sites for the mite (Dietemann et al., 2013). Distribution maps alluded to the influence of precipitation on the occurrence and distribution of the honeybee pest. Occurrence and distribution were high in regions such as Mount Kenya, Kakamega and the coastal strip. Therefore,

these four precipitation variables can be used to successfully predict the seasonal variations of these honeybee pests in various regions within country (Torto et al., 2010). As 'precipitation' is an overly important variable and given the future increases in predicted precipitation amounts and variability (Lovett, 2015; Platts et al., 2015), there may be higher occurrences of the four pests we investigated.

Our EN models showed that topography had little to no influence on the occurrence of the modelled honeybee pests in Kenya. The models predicted both the regions with high altitude (such as Mount Kenya) and low altitude (such as the coast), to have high habitat suitability. This was consistent with Mani's conclusion that most insects (honeybee pests included) had diverse altitudinal gradients (Mani, 2013). Indeed, the honeybee pest we investigated had the same habitat suitability across all altitudinal gradients.

4.4.2 Contribution of Remotely Sensed Data

The inclusion of remotely sensed variables into our EN models improved their ability to predict the distribution of honeybee pest species. Even though bioclimatic variables offered the ability to predict potential distribution of the honeybee pests (Aranda and Lobo, 2011), remotely sensed variables enhanced the predictability of model 3 for the four pests by reducing overfitting as shown in Figure 21. The prediction ability of these EN models improved because remotely sensed variables captured fragmented environmental conditions on the ground. The fine spatial resolution of remotely sensed variables reduced the errors that emanate from the generalization in bioclimatic variables over given spatial extents. Furthermore, the model products (prediction maps) were more realistic in predicting habitat suitability in model 3 and they had reduced overestimation (Cord et al., 2014a; Saatchi et al., 2008).

Remotely sensed variables also capture 'adverse' environmental conditions that limit distribution of these honeybee pests across homogenous bioclimatic conditions. Model 1 would predict similar occurrence within regions with similar bioclimatic

conditions as shown in Figure 21. However, with remotely sensed variables such as maximum NDVI in the season, high honeybee pest potentials were predicted by model 3 in areas with high NDVI values as opposed to those that recorded low values. Torto et al. (2010) demonstrated that honeybee pests' reproduction is influenced by vegetation phenology such as NDVI. Moreover Pau et al. (2012) also indicated that success of honeybee pests is determined by availability of pollen and nectar, on which their hosts depend on. Therefore, remotely sensed variables that capture these changes (i.e., NDVI variables) offer model 3 information that is not available in the model 1. Remotely sensed variables also provide the ability to refine the prediction of honeybee pest distribution as opposed to bioclimatic variables which reflect potential honeybee pest distribution (Cord et al., 2014a). Further, remotely sensed environmental variables capture heterogeneity that is reflective of ground condition on or near real time basis, while bioclimatic data are interpolated at large areas hence depicting homogeneity at large spatial extents (Attorre et al., 2007; Hijmans et al., 2005).

4.4.3 Advantages of Using Integrative EN Models and Applicability

EN models rely only on occurrence to estimate predictions. Availability of absence data improves the model precisions, but such data are not readily available and hence one cannot conclude with certainty that lack of occurrence in an area confirms absence (Engler et al., 2004). Therefore, EN models use pseudoabsence to fill this gap. However, important information such as abundance of a pest in a certain region, which could improve prediction, is not utilized. Various remotely sensed data with different spatial-temporal and spectra resolutions are available, thus widening the EN modelling boundaries (Cord et al., 2014b). Advanced analytical approaches of remotely sensed data also offer the ability to derive and use specific phenological information from raw data to improve detectability.

Although remotely sensed data have been previously used to map the distribution of various plant species (Bradley et al., 2012; Cord et al., 2014b) our models offer tools that could be used to accurately map the distribution of honeybee pests and conceivably other insect species. This is because remotely sensed data incorporated in our models provide timely (recent) information that allow vegetation changes as a result of anthropogenic land use, especially in agricultural areas or unprotected forests under consideration. Consequently, we offer tools that could be used in similar or differently designed methodologies as those documented herein, to improve pest management.

Remotely sensed data however have shortcomings that could negatively impact the prediction models. Since pixel size is an important indicator of the quality of the data, accuracy achieved by prediction models that have used remotely sensed variables are limited to the pixel size. Therefore, EN models that use data with coarse resolution may not give accurate predictions (Saatchi et al., 2008). On the other hand, remotely sensed data with fine spatial and spectral resolution, such as World View-2 and 3, AISA Eagle, are costly and unavailable. Additionally, challenges such as cloud cover greatly degrade the quality of optical remotely sensed data because environmental measurements may not be accurately captured or unavailable when the cloud cover is substantial. This however may be overcome by high temporal resolution which offers options to circumvent data with low quality or using microwave remote sensing technology that is not affected by cloud cover.

EN models require variables with the same pixel size (i.e., spatial resolution) in analysis. Seldom will spatial data from different sources have similar spatial resolution. Resampling the datasets to achieve same pixel size could introduce uncertainties (Clifford and Tarassenko, 2005; Dikshit and Roy, 1996) to the EN models. Such uncertainties and those associated with presence data collection may be misleading (Fielding and Bell, 1997). Therefore, remotely sensed data used to build prediction models should be carefully selected, considering spatial, spectral and temporal resolutions. Moreover, vegetation data projected to future scenarios that

offer EN models the ability to produce more refined prediction maps are not readily available. This leads to model assumption, which may include similar vegetation conditions both in the current and future scenarios that may reduce reliability of projected models in predictions. Additionally, data on flowering conditions that influence bee and pest populations are also not readily available. Flower maps in Kenya are only available in small spatial extents that do not span the whole area under this research (Abdel-Rahman et al., 2015; Landmann et al., 2015). Similarly, even though apiaries were selected from representative agroecological sites in Kenya, the full climatic niches of the studied bee pests might not have indeed been captured, particularly the background points were drawn from the whole extent of Kenya. Therefore, the individual variable contribution to the spread of the bee pests outside their training range should be further investigated.

4.5 Conclusions

EN models developed were 'good', hence deemed appropriate in predicting bee pest habitat suitability. However, it was established that model 1 which used only bioclimatic data had lower AUC values with higher standard deviations than model 3 that used both bioclimatic and remotely sensed data. Although remotely sensed data improved the models' precision, bioclimatic variables influenced the models more than remotely sensed variables. It was also established from modelled bioclimatic data that bioclimatic change will have an impact on the distribution of these honeybee pests. Their spatial distribution would increase to areas currently classified to have low occurrence. Projected pest predictions could however be improved if appropriate pest management measures are put in place. Prediction maps could be used to identify high risk areas for quarantine to limit the spread of the pests. Therefore, these models would be important tools to decision makers in mitigating effects of destructive pests. Furthermore, containment zones need further investigation to establish conditions that prohibit the distribution of these pests. Targeted interventions should focus on zones where future pest outbreaks are expected.

4.6 Acknowledgements

The authors' hearty thanks go to George Makau, Fiona Mumoki, Isabella Nyamoita and Ada A. for their assistance with the field sampling exercise. The authors appreciate the various beekeepers who donated their apiaries for this study. Lastly, our sincere gratitude goes to the European Union for funding of this project (Project number; DCI-FOOD-2011/023-520). The views expressed herein do not necessarily reflect the official opinion of the donors. We are also very thankful to Dolorosa Osogo and other colleagues in ICIPE for proof reading and technical comments.

CHAPTER FIVE

5 Multi-pronged Abundance Prediction of Bee Pests' Spatial Distribution in Kenya

Varroa mites (*Varroa destractor*) on the back of a worker bee. Image by Entine (2020); source https://geneticliteracyproject.org/wp-content/uploads/2017/11/Screen-Shot-2017-08-09-at-9.38.45-AM-1024x623.png

This chapter is based on:

Makori, D. M., Abdel-Rahman, E. M., Odindi, J., Mutanga, O., Landmann, T. and Tonnang H. E. Z. In preparation. Multi-pronged abundance prediction of bee pests' spatial distribution in Kenya.

Abstract

Bee farming and bee health in general are threatened by climate change, agricultural intensification and associated habitat alteration, agrochemicals intensification and bee pests and diseases. Among these threats, bee pest has particularly been identified as a major obstacle to bee farming and bee health. Although previous studies have endeavoured to establish the spatial distribution of bee pests, their abundance (population) and intensification remains poorly understood. Hence, this study sought to determine factors that influence the abundance and spatial distribution of bee pests in Kenya. Abundance data on *Varroa destractor, Oplostomus haroldi, Galleria mellonella* and *Aethina tumida* were collected from apiaries in Kenya's main agroecological regions during the wet and dry seasons. The abundance data were fitted to non-conflating bioclimatic variables, vegetation phenological data, topographical variables and land use/ land cover (LULC) data, as a proxy for anthropogenic influence, using random forest algorithm. The results indicated a significant ($p \leq 0.05$) seasonal influence on bee pests' abundance, while precipitation-related variables contributed the most (up to 77.8%) on all bee pests' abundance prediction models. Topographic and vegetation phenological influence had a moderate influence (at 14.3% and 6.7% respectively) while anthropogenic influence, as denoted by the LULC, had a low (1.2%) contribution on bee pests' abundance models. High seasonality in bioclimatic variables increased the spatial and abundance of projected future (2055) bee pests' risk levels across the study site. Three out of four prediction models developed herein ranked 'excellent' in terms of their performance, while the model for predicting *G. mellonella* was ranked fair. Due to the high precision of prediction models developed herein, this study concludes that the results could reliably be used to identify bee pests' high-risk areas for management and mitigation purposes.

Keywords:

Bee health, food security, ecosystem services, climate change, vegetation phenology, anthropogenic influence, machine learning, random forest

5.1 Introduction

Bee pollination services are vital for biodiversity conservation and crop production, that play a critical role in improving food and nutritional security in many parts of the world (Aizen and Harder, 2009; Kiatoko et al., 2014; Klein et al., 2007). Additionally, bee farmers directly benefit from honey and other hive by-products. The semi-arid African agroecology zones, which are typified by low, erratic and unreliable rainfall, are unable to support rainfed agriculture and bee farming (Raina et al., 2011). Bee farming provides supplemental income from hive products such as honey, royal jelly, propolis, wax and to a smaller extent bee venom (Oladimeji, 2018; Rai et al., 2021). In return, conservation efforts are enhanced in regions bordering forests where bee farming is commonly practiced (Chiawo et al., 2011; Potts et al., 2006; Raina et al., 2011; Warui et al., 2018b). Nevertheless, bee farming and bee health are threatened by climate change, agricultural intensification and associated habitat alterations, agrochemical intensification, bee pests and diseases (Muli et al., 2014b; Zayed, 2009a).

In particular, bee pests are considerably devastating bee health through colony collapse as a result of direct physical injury or indirectly as vectors of pathogens that transmit diseases (Mumoki et al., 2014; Ongus et al., 2017). The impact caused by the pests on bee colonies and their ability to traverse vast spatial extents within varying agroclimatic and agroecological zones has recently attracted profound scientific interest. For instance, various studies have sought to determine the effect of pests on bee health and performance (Fombong et al., 2012; Muli et al., 2014b; Ongus et al., 2017), their spatial extents and distribution patterns (Boncristiani et al., 2021; Makori et al., 2017), and their economic impact on bee health (Boncristiani et al., 2021; Wilkins et al., 2007). The most common and economically important bee pests are the varroa mites (*Varroa distructor*), large hive beetles (*Oplostomus haroldi*), small hive beetles (*Aethina tumida*) and wax moths (*Galleria mellonella*) (Fombong et al., 2012; Muli et al., 2014b; Pirk et al., 2016; Torto et al., 2010; Zayed, 2009a). Although these pests are

known to proliferate across agroecological gradients, their spatial and temporal distribution in Kenya has not been adequately established.

Bee pests have different optimal bioclimatic conditions for their distribution and survival. However, studies have established significant variations in optimal conditions regarding temperature, precipitation, net productivity and altitudinal range among pests (Mumoki et al., 2014; Neumann and Ellis, 2008; Peterson and Nakazawa, 2008; Torto et al., 2010). The habitats' biotic conditions such as net productivity, measured by vegetation productivity and phenology, affect bee pests' population dynamics and richness. Specifically, biotic conditions affect pests hosts' (bees) agility, vigour and ability to produce and accumulate hive products such as bee bread, on which the pests thrive (Neumann et al., 2001; Pau et al., 2012; Pirk et al., 2016; Simone-Finstrom and Spivak, 2012). On the other hand, bioclimatic conditions such as prevailing dry, humid, hot and cold conditions could limit or enhance the distribution and reproductive ability of bee pests (Neumann and Ellis, 2008; Torto et al., 2010). Additionally, anthropogenic effects on bee pests and their hosts influence their spatial distribution and proliferation (Winfree et al., 2009; Zulian et al., 2013). For instance, landscape fragmentation and agricultural intensification have been demonstrated to adversely influence landscape continuity, which is important for flower diversity. Flower intensity and diversity are valuable traits for bees (the bee pests' hosts) food, hence sustaining the pests' distribution (Boncristiani et al., 2021; Cameron and Sadd, 2020). Conversely, human settlements impact bee distribution by reducing their nesting sites, habitats and eventually their ability to fight off pests' invasion (Fombong et al., 2012; Makori et al., 2017; Ongus et al., 2017). Therefore, information that provides various aspects of human effects on the distribution of both bees and their pests are desirable.

To understand and establish the spatial distribution and proliferation of bee pests, ecological niche (EN) modelling, artificial intelligence (AI), and machine learning (ML) algorithms have been adopted (Dormann, 2020; Strebel et al., 2022). In general,

these algorithms provide statistical pathways for linking a feature(s) of interest (response) to reliable predictor variables such as biotic, bioclimatic and anthropogenic factors in space and time (Fernández and Hamilton, 2015; Kearney and Porter, 2009). Moreover, prediction algorithms provide the most relevant predictor variables with important insights on accurate location-specific suitability, from which linkages with response variables can be spatially and temporally established (Allouche et al., 2006; Wisz et al., 2008). To achieve accurate and realistic bee pests' predictions, the predictor variables should be carefully and meticulously selected (Peterson, 2011; Phillips et al., 2009). For example, the currently available bioclimatic variables are somehow generalist in nature as they are interpolated from discrete observations over large spatial expanses (Platts et al., 2015). On the other hand, satellite-based variables like vegetation phenology are continuous observations with better spatial and temporal resolutions that could capture the 'actual' landscape patterns (Cord et al., 2014b; Pau et al., 2012). Response variables with the ability to provide extra information other than 'location-only' are necessary. For instance, bee pests' abundance (population) data can provide both location and population information, which enable relatively reliable information for understanding the bee pests' propagation and developing the best management options (Hallman and Robinson, 2020; Strebel et al., 2022; Waldock et al., 2022). Additionally, for reliable species distribution predictions, ML algorithms should be carefully selected, parameterized and optimized to avoid overfitting (Cord et al., 2014b).

Studies like Makori et al. (2017) have predicted the spatial distribution of bee pests in Kenya using presence-only data as a response variable, and climate and vegetation features as predictor variables. As previously mentioned, a number of studies have shown the usefulness of insect abundance data in providing reliable information (Acevedo et al., 2017; Baldridge et al., 2016; Bradley, 2016; Dallas and Hastings, 2018; DeMarche et al., 2019; Hallman and Robinson, 2020; Strebel et al., 2022; Waldock et al., 2022). Furthermore, Makori et al. (2017) used the maximum entropy (MaxEnt) model

to predict bee pests' distribution. However studies have shown that other cutting-edge ML algorithms such as random forest (RF) might perform relatively better (Abdel-Rahman et al., 2013; Rodriguez-Galiano et al., 2012). Further, the MaxEnt used only utilized presence-only datasets, while other algorithms that could handle quantitative observations such as abundance data should be explored. Furthermore, to the best of our knowledge, studies have not yet included the influence of land use/ land cover (LULC) on the abundance distribution of bee pests in Kenya. The LULC could adequately explain anthropogenic influence on the distribution and proliferation of bee pests in Kenya (Chaudhary et al., 2015; Maney et al., 2022; Newbold, 2018; Newbold et al., 2015). Therefore, this study sought to establish the influence of bioclimatic, vegetation phenology, topography and LULC (as a proxy for anthropogenic influence) variables on spatial distribution and abundance of bee pests in Kenya. This study also predicted the future (i.e., 2055) scenarios of the bee pests' distribution using the above-mentioned predictor variables.

5.2 Methods

5.2.1 Study site

The study was conducted across 38 counties in Kenya encompassing four agroecological zones. The area covers 259,004.34 km^2 and spans more than half of the country that include the Nyanza, Western, Rift Valley, Mount (Mt.) Kenya, Eastern and Coastal regions (Figure 23). Abundance data were collected from 14 counties comprising of Vihiga and Kakamega (in western region), Nandi and Baringo, (in Rift Valley region), Embu, Nyeri, Laikipia, Meru and Isiolo (in Mt. Kenya region), Machakos and Kitui (in Eastern region) and Taita Taveta, Kilifi and Kwale (in coastal region). The study sites varied in elevation from 2 metres above sea level at the coastal region to 2,045 metres above sea level at the Nandi Hills.

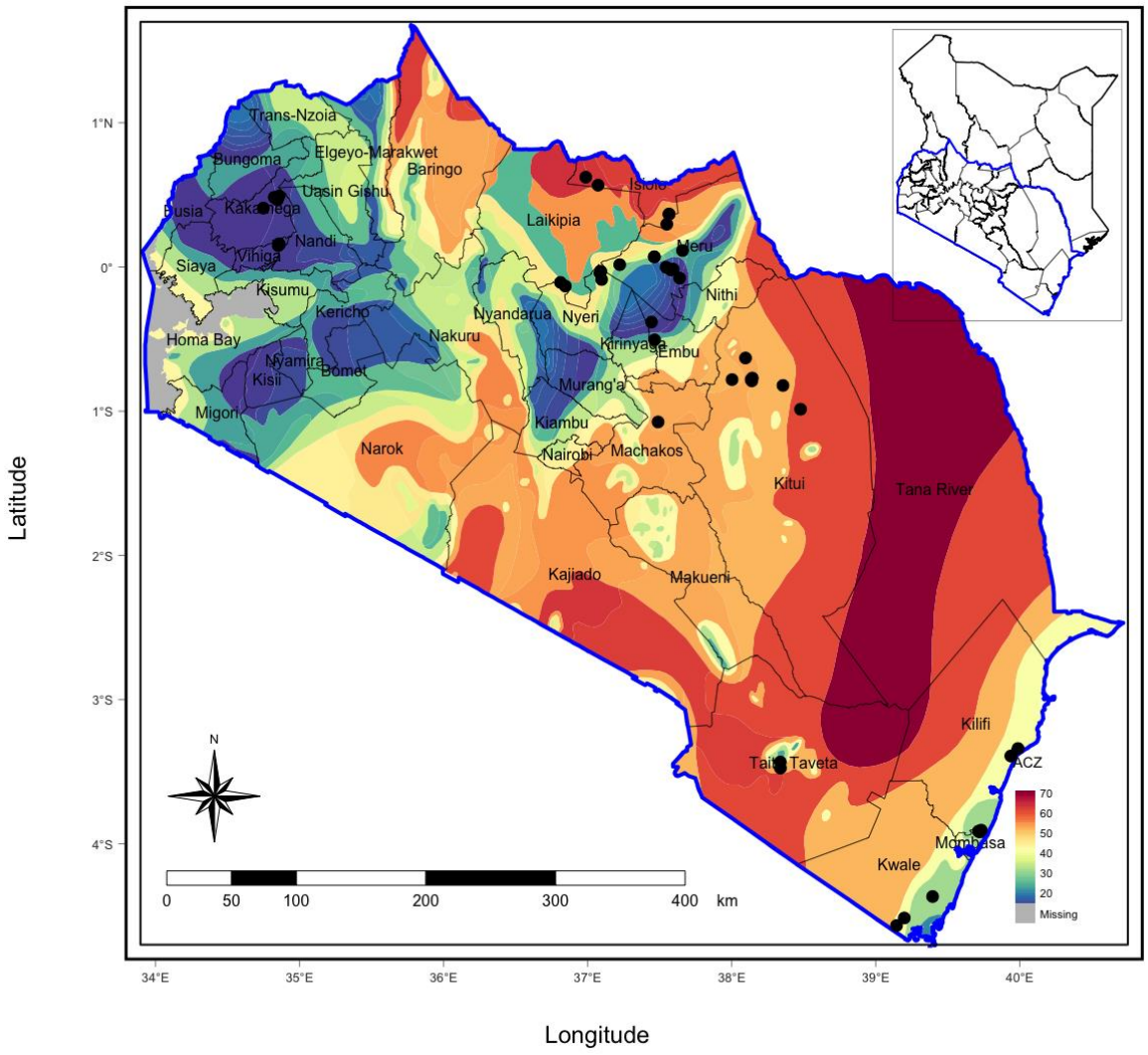


Figure 23: Study area on agroclimatic zones (acz id 20 to 70) backdrop,bee pests' abundance and regions used to predict bee pest distribution in Kenya.

The coastal region is a low altitude area with high temperature and humidity, with Arabuko Sokoke as the major forest (Schürmann et al., 2020). Human settlement and intensification of agriculture are two major threats to natural vegetation in the region (Habel et al., 2017). Mt. Kenya, Nandi and Western regions are characterized by high altitude, low temperature and high humidity (Githui et al., 2009). Major water towers such as the Mau Complex, Cherangani Hills and Kakamega forest are found in these regions (Kinyanjui, 2011). The Tana River County in the coastal region, Baringo

County in Rift Valley region and the larger Eastern region are characterized by high temperature and low humidity. These regions are predominantly covered by shrubs and few stands of Acacia trees (Speranza et al., 2009). Due to low precipitation in the larger Eastern region and most parts of the Coastal region, small scale subsistence farming is the major socioeconomic activity (Githui et al., 2009; Schürmann et al., 2020). On the other hand, Kakamega, Kisii highlands, Nandi hills and Mt. Kenya regions receive higher precipitation hence higher agricultural productivity. These areas are mainly utilized for maize farming, which is a staple food crop in Kenya (Mohajan, 2014; Schroeder et al., 2013). In addition, bees are used in these regions to improve household income from hive products and crop production through pollination services (Suso et al., 2016; Warui et al., 2018b).

5.2.2 Data collection and pre-processing

Data used for prediction of bee pests' distribution and abundance were categorized into response variables (bee pests' abundance) and predictor variables that comprised bioclimatic, vegetation phenology, topological and LULC (representing anthropogenic influence) variables (Table 18).

5.2.2.1 Response variables

Bee pests' abundance data were collected in point vector format between March and April 2014 (wet season) and June 2015 (dry season). A sample point consisted of an apiary with five or more colonies (i.e., hive) in one location. The apiaries were randomly selected for sampling with help from the local community in the study regions (Acharya et al., 2013; Endo et al., 2015; Jia and Barabási, 2013) according to the requirements of estimation abundance (density) across the landscape in proportion to population density (Merow et al., 2013). An enumeration of all pests in selected apiaries was done and data on four bee pests species with global economic importance

recorded; i.e., *Varroa destructor*, *Oplostomus haroldi*, *Galleria mellonella* and *Aethina tumida* (Fombong et al., 2012; Torto et al., 2010) using standard methods as detailed in Dietemann et al. (2013) and Torto et al. (2010). A total of 45 apiaries had *V. destructor*, 24 had *O. haroldi*, 38 had *G. mellonella* and 37 had *A. tumida* (Refer to Table 20 for pest abundance data). These were within the acceptable sample size ranges for most ML modelling algorithms (Amirpour Haredasht et al., 2013; Peterson and Nakazawa, 2008).

The apiaries considered for all the four bee pests span across four agroecological zones in the study area, with a representative climatic gradient. A handheld global positioning system (GPS) device with an accuracy of 3 ($\pm$3) metres was used to collect and record information on positions of sampled apiaries. The Wilcoxon rank sum test was performed to test the significance influence ($p \leq 0.05$) of seasonality on abundance of bee pests between wet and dry seasons, following heterogeneous variance observations (Fagerland and Sandvik, 2009; Neuhäuser and Ruxton, 2009).

5.2.2.2 Predictor variables

All predictor variables were collected and processed in raster format. The variables in this category were clipped to the boundaries of the study area and resampled to the same spatial resolution. According to Mudereri et al. (2020) and Makori et al. (2022), all predictor variables should be resampled and harmonized to a base resolution. Therefore, warping was used to edit the resolution of LULC, topographical and bioclimatic datasets to that of the vegetation phenology datasets, which had moderate resolution (250 metres).

5.2.2.2.1 Vegetation phenology

Vegetation phenology was derived from enhanced vegetation index (EVI), which was acquired from the 250-metre resolution Moderate Resolution Imaging Spectroradiometer (MODIS) imagery at a 16-day interval. This study used EVI

observations from 2000 to 2021 (21 years) to derive vegetation phenology variables (Table 18) in TIMESAT software environment (Eklundha and Jönsson, 2017). Best fitting was achieved using the TIMESAT fitting as recommended by Makori et al. (2017).

A total of 13 vegetation phenological variables (refer to Table 18 for a complete list) were derived and used in this study. These were start of the season time (*start_t*), end of season time (*end_t*), length of season (*length*), base value (*base*), time for middle of season (*mid*), maximum value (*max*), amplitude (*amplitude*), left derivative (*left_d*), right derivative (*right_d*), large integral (*large_i*), small integral (*small_i*), start of season value (*start_v*) and end of season value (*end_v*).

5.2.2.2.2 Topographical variables

Topographical variables derived from a 90-metre pixel resolution (3 arcsec resolution) digital elevation model (DEM), were used to model the influence of land morphology on the distribution and abundance of bee pests in Kenya. The DEM was acquired by a Shuttle Radar Topography Mission (SRTM) instrument (CGIAR-CSI, 2020). Derived topographical variables included topographical position index (TPI), terrain ruggedness index (TRI), roughness, aspect and hillshade (Table 18).

5.2.2.2.3 Bioclimatic variables

Bioclimatic variables used in this study were obtained from AfriClim (Fick and Hijmans, 2017; Platts et al., 2015) at 1 km spatial resolution. The AfriClim datasets were downscaled from WorldClim platform (www.worldclim.org) for the African continent. Bioclimatic variables contain derived summaries of rainfall and temperature and describe current (1970 – 2000) and future (2055) conditions. Simulated climatic conditions under intermediate CO_2 emissions, set by the

International Panel on Climate Change (IPCC) at representative concentration pathway scenario (RCP) 4.5 watt/m^2 (Pachauri and Mayer, 2015) using total radioactive forcing values were used. Under this pathway, future climatic conditions are simulated means from 2041 to 2070 (Guan et al., 2020) obtained from the fourth mission of community climate system (CCSM4). This mission has been demonstrated to contain most reliable climatic projections (Mohammadi et al., 2019; Mudereri et al., 2020a). Twenty-one bioclimatic variables were used for current and future timesteps, comprising of 11 precipitation and 10 temperature variables (Table 18).

Table 18: Vegetation phenology, bioclimatic, topographic and land use/ land cover (LULC) used as predictor variables by prediction models for the spatial distribution of bee pests in Kenya.

Variables		Description	Units
Name	**Abbreviation**		
e) Temperature variables (n = 10)			
Bio 1	*bio1*	mean annual temperature	^{o}C
Bio 2	*bio2*	mean diurnal range in temperature isothermality	^{o}C
Bio 3	*bio3*	isothermality	^{o}C
Bio 4	*bio4*	temperature seasonality	^{o}C
Bio 5	*bio5*	maximum temperature warmest month	^{o}C
Bio 6	*bio6*	minimum temperature coolest month	^{o}C
Bio 7	*bio7*	annual temperature range	^{o}C
Bio 10	*bio10*	mean temperature warmest quarter	^{o}C
Bio 11	*bio11*	mean temperature coolest quarter	^{o}C
Potential evapotranspiration	*pet*	potential evapotranspiration	mm
f) Precipitation variables (n = 11)			
Bio 12	*bio12*	mean annual rainfall	mm
Bio 13	*bio13*	rainfall wettest month	mm
Bio 14	*bio14*	rainfall driest month	mm
Bio 15	*bio15*	rainfall seasonality	mm
Bio 16	*bio16*	rainfall wettest quarter	mm
Bio 17	*bio17*	rainfall driest quarter	mm
Moisture index	*mi*	annual moisture index	n/a
Moisture index moist quarter	*mmiq*	moisture index moist quarter	n/a
Moisture index arid quarter	*miaq*	moisture index arid quarter	n/a

Dry months	*dm*	number of dry months	months
Length of the longest dry season	*llds*	length of longest dry season	months
g) Topographical variables (n = 6)			
Slope	*slope*	angle of inclination to the horizontal	% rise
Aspect	*aspect*	slope direction	degrees
Hillshade	*hillshade*	shading effect	n/a
Roughness	*Roughness*	degree of surface irregularity	n/a
Topographical position index	*TPI*	position of surface inclination	n/a
Terrain ruggedness index	*TRI*	ground ruggedness	n/a
h) Vegetation phenological variables (n = 13)			
Start of season time	*start_t*	time for the start of season	decades
End of season time	*end_t*	time for end of season	decades
Length of season	*length*	length of time from start to end of season	decades
Base value	*base*	average minimum EVI value	n/a
Time of middle of season	*mid*	time of middle of season	decade
Maximum value	*max*	maximum value of fitted data	n/a
Amplitude	*amplitude*	difference between maximum and base level	n/a
Left derivative	*left_d*	rate of increase of beginning of season	%
Right derivative	*right_d*	rate of decrease of end of season	%
Large integral	*large_i*	integral from season start to season end	n/a
Small integral	*small_i*	integral difference between season and base level	n/a
Start of season value	*start_v*	value at the start of season	n/a
End of season value	*end_v*	value at end of season	n/a
i) Land use variables (n = 1)			
Landcover	*Landcover*	land cover data indicating land use on the ground	land use class

5.2.2.3 Variable selection

Variables used for model predictions should be carefully selected to avoid multicollinearity that could lead to volatility in model performance (Dormann et al., 2013; Naimi and Araújo, 2016). However, before testing for multicollinearity, the recursive feature elimination (RFE) criteria in the 'caret' package in R was used to provide insight on minimum number of uncorrelated variables that could yield comparable prediction results (Darst et al., 2018; Makori et al., 2022). To select preferred prediction variables, two-stage elimination criteria was performed, firstly using variable inflation factor (VIF) and Pearson correlation coefficient. This was meant to reduce multicollinearity amongst predictor variables while establishing orthogonal variables that were most suited for bee pests prediction models (Dormann et al., 2013; Plant, 2012). Secondly, multiple regression models were utilized to regress each predictor variable against all other variables to detect collinearity while computing VIF for each combination (Plant, 2012). This step was used to select important prediction variables from the pool while iteratively eliminating those with high linear regression coefficients.

The Pearson correlation coefficient was set at $th = 0.7$ ($r \geq 0.7$), as the first threshold for best results (Araújo et al., 2019; Dormann et al., 2013; Plant, 2012). The second threshold was set using the 'vifstep' function in 'usdm' package in R (Naimi and Araújo, 2016). This was used to further assess collinearity among variables from the first step while eliminating collinear ones with more than 10 VIF values (Dormann et al., 2013). The resultant correlation matrix (Figure 24) show that brown and blue colours denote negative and positive correlations respectively. Colour intensity and size of shape denote level of correlation, while shape orientation indicate nature of correlation. It is apparent that some predictor variables tagged important for bee pests' distribution were correlated. For instance, start of season time (*start_t*) and rainfall driest quarter (*bio17*) were highly correlated. Hence, 14 important and uncorrelated predictor variables (Table 18) were selected for prediction of bee pests in Kenya. The

selected variables were sufficient to yield high prediction accuracies when fitted with adequate response variables for ML algorithms used (Araújo et al., 2019; Wisz et al., 2008).

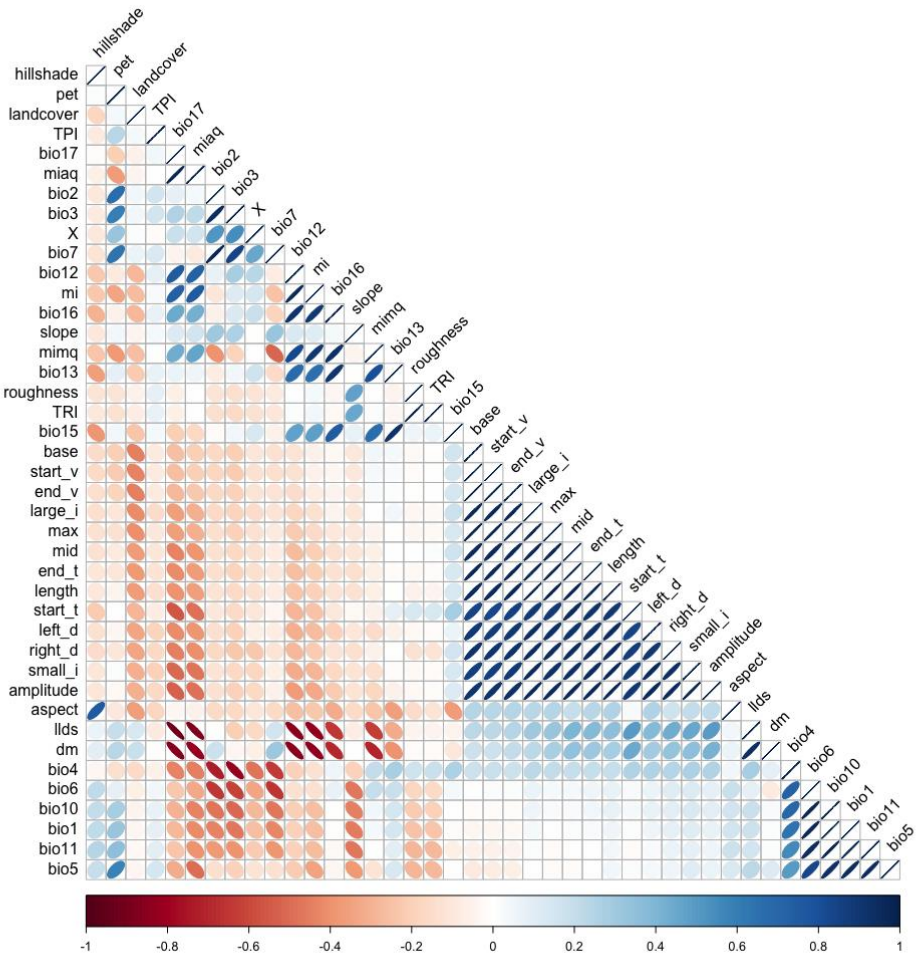


Figure 24: Collinearity matrix indicating correlation interaction of all predictor variables (refer to Table 18) used to predict spatial abundance of bee pests in Kenya.

5.2.3 Fitting modelling environment

Bee pests' abundance observations were related to the selected uncorrelated predictor variables using the RF algorithm. This was done using the RF machine learning

ecological niche modelling environment (Breiman, 2001) in the 'sdm' package (Naimi and Araújo, 2016) in R software (R Core Team, 2021). The RF models were set to ten iterations in each model run and an ensemble approach was utilized to estimate species-specific mean abundance predictions. Hence the variations among the predictions were harmonized (Araújo et al., 2019; Hao et al., 2019). The 'ensemble' function was used in the 'sdm' package to harmonize results of individual bee pests' species prediction replications using true skills statistics (TSS) weighted approach in both the current and future timestamps. A 0.7 TSS cut off was used to select replicate models to include in the ensemble. The TSS is a measure of accuracy that estimates agreement between response and predictor variables. It ranges between -1 and +1, where -1 indicates the lowest or poorest prediction accuracy while +1 indicates the highest prediction accuracy (Hao et al., 2019).

Predictions were done under both current and future (year 2055) projections for all bee pests' species, with simulated climatic and LULC scenarios. Since projected vegetation phenology was unavailable and elevation was assumed to be unchanging in the future epoch, their current datasets were also used for the future projections. Furthermore, similar model settings and packages were used for both current and future epochs for consistency.

5.2.4 Prediction model validation

The RF prediction power of bee pests' was evaluated using the Pearson's correlation coefficient (r) (Equation 4) that was calculated using the R software (de la Fuente et al., 2021). Extracted prediction values were used to calculate power of fit between bee pests' abundance and environmental suitability, which was represented as deviance of fit explained by the developed models. Degree of agreement, sign and significance of the models were used to report suitability and prediction power of each model. Prior to correlation evaluation, data from each model were subjected to covariation

linear test and scatter plot reported for each prediction model. Then model prediction power was ranked for each model from fail to excellent (Table 19).

Pearson correlation coefficient (r) used to validate prediction power of developed bee pests' abundance distribution prediction models **Equation 4**

$$r = \frac{\Sigma(x - m_x)(y - m_y)}{\sqrt{\Sigma}(\mathrm{x} - m_x)^2 \Sigma(y - m_y)^2}$$

where:

i) x is a vector for observed abundance

ii) y is a vector for predicted abundance

iii) m_x and m_y corresponds to means of x and y respectively

Table 19: Model evaluation ranks used to categorise bee pests' abundance distribution prediction models developed for Kenya for *Varroa destractor, Oplostumus haroldi, Galleria mellonella* and *Aethina tumida*.

Pearson's correlation coefficient (r) value	**Rank**
0.20 and below	Fail
0.21 to 0.40	Poor
0.41 to 0.60	Fair
0.61 to 0.80	Good
0.81 and above	Excellent

5.3 Results

5.3.1 Seasonality influence on bee pests' abundance

The Wilcoxon rank sum test with continuity correction performed at 95 % confidence interval ($p \leq 0.05$) indicated that there was a significant difference ($p = 0.047$) between bee pests' abundance observations in the wet and dry seasons across the four agroecological zones in the study area. Moreover, a visual observation showed high seasonal variability for bee pests (Figure 25). Generally, there were more bee pests' abundance in the wet than the dry season (Table 20). Bee pests observations in the dry season were 42.25 pest counts with relatively low dispersions around the mean. On the other hand, bee pests' abundance was more (almost five times) during the wet season with a mean of 199.75 and higher relative dispersions around the mean compared to the dry season. The *V. destractor* had the highest ratio between seasons (6.35 times more in the wet season) while *G. mellonella* had the least seasonal ratio (at 0.95 times).

Table 20: Means abundance of *Varroa destractor, Oplostomus haroldi, Galleria mellonella* and *Aethina tumida* for abundance data collected during the wet and dry seasons in Kenya. The seasonal ratios show difference between seasons.

	Wet season	**Dry season**	**Seasonal ratio**
V. destractor	476.00	75.00	6.35
O. haroldi	31.00	15.00	2.07
G. mellonella	20.00	21.00	0.95
A. tumida	272.00	58.00	4.69
Mean of total	**199.75**	**42.25**	**4.73**

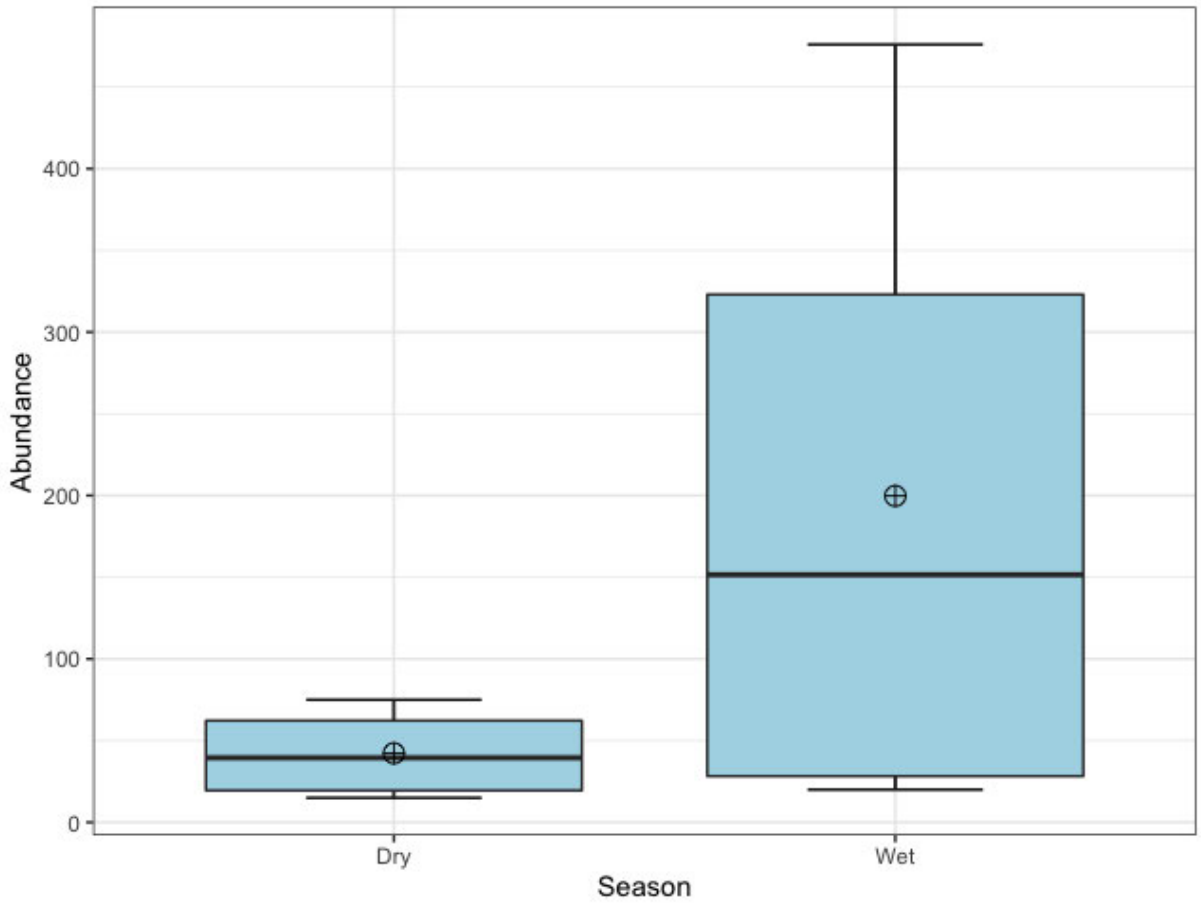


Figure 25: Box plots of bee pests abundance observations for the dry and wet seasons in Kenya, with significant seasonality influence ($p \leq 0.05$).

5.3.2 Predictor variable selectin and bee pests' prediction

Twenty-seven out of 41 predictor variables used in the collinearity model (Figure 24) were conflating. Based on their individual interactions and importance (Figure 26), they were eliminated from further analysis. The 14 variables that were non-conflating and ranked high in variable importance scale were used to develop refined bee pest prediction models. The recursive feature elimination (RFE) model (Figure 27) used to test the least variable interaction before eliminating the collinear ones further indicated that 14 predictor variables would yield acceptably high predictions.

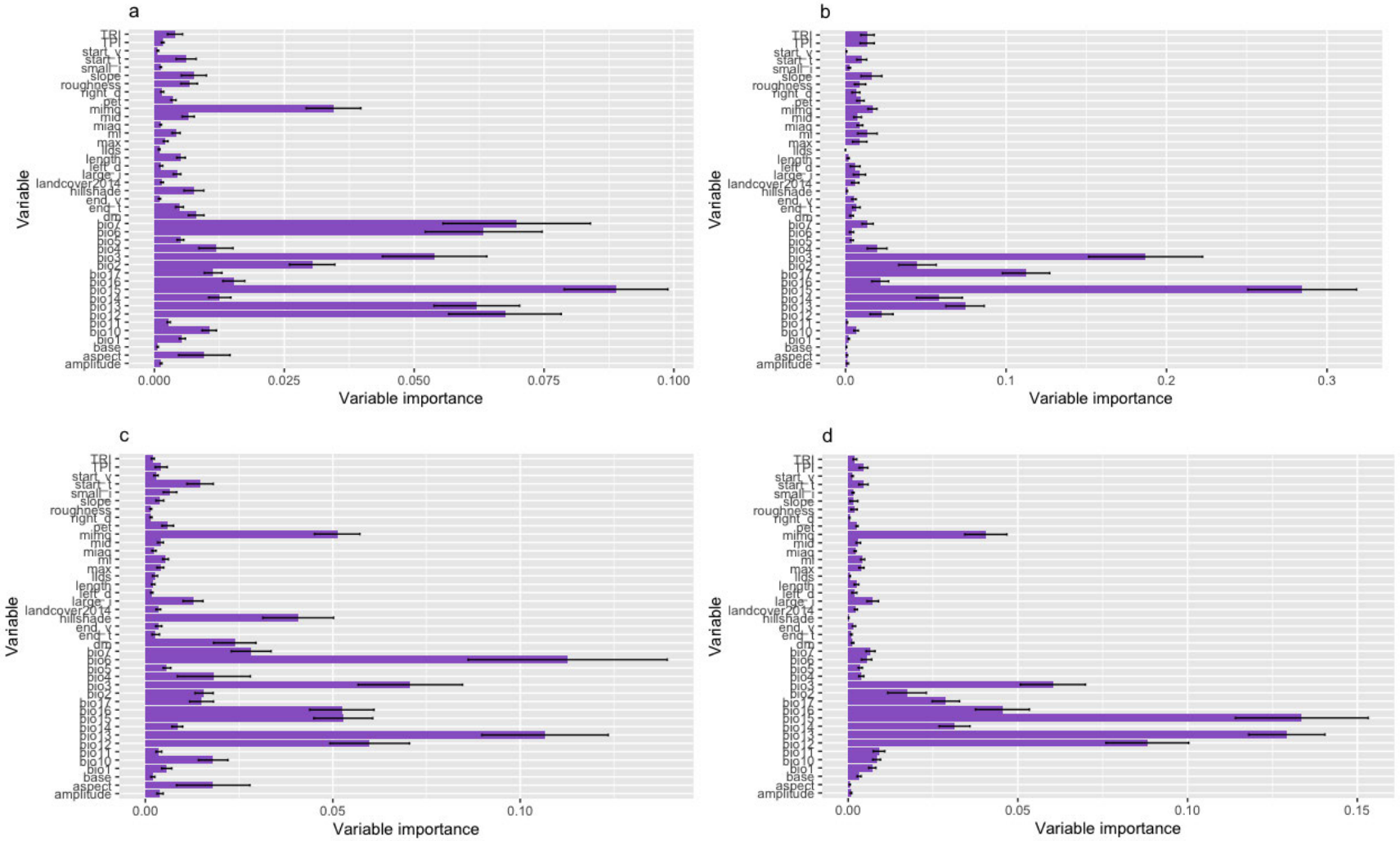


Figure 26: Variable importance of all variables (refer to Table 18) used for predicting *Varroa destractor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d) abundance.

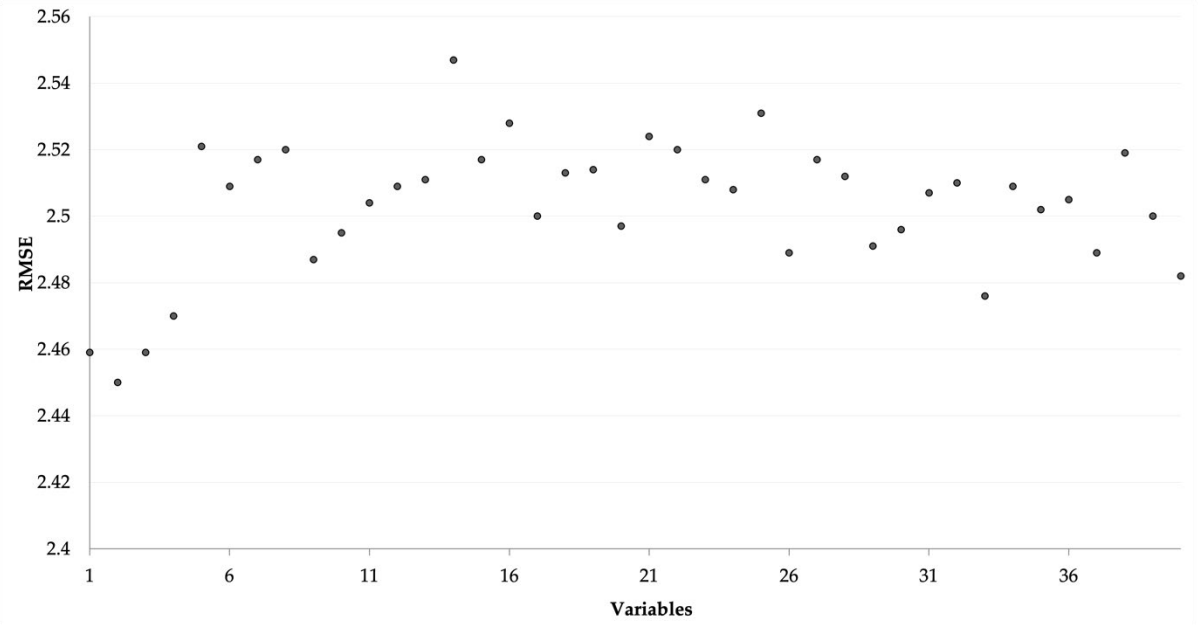


Figure 27: The recursive feature elimination (RFE) model used to indicate independent features (variables) that could yield acceptable prediction accuracy based on root mean square error (RMSE).

Bioclimatic and topographical variable clusters contributed the highest number of variables and the most information on all bee pest-specific models compared to vegetation phenological variables. Furthermore, the bioclimatic variables had the highest influence (77.8%) on *V. destractor* prediction model by more than three quarters, compared to other predictor variables combined (Figure 28). Additionally, rainfall wettest month and annual temperature range alone had almost two thirds (61.0%) contribution on *V. destractor's* prediction models. In general, the rainfall wettest month alone had the highest influence (42.5%), contributing almost half of predictor information needed by the models. Topographical influence on *V. destractor* prediction models was 14.3%, while vegetation phenological variables contributed 6.7% to the models. On the other hand, the LULC had the least contribution (1.2%) on *V. destractor* prediction models.

The bioclimatic and topographical variables contributed 69.9% and 16.3%, respectively (Figure 28) to *O. haroldi* prediction models. Topographical effect was

more substantial (16.3%) than phenological effect (6.8%), suggesting that *O. haroldi* was not very sensitive to vegetative food substrate in its distribution. Moreover, bioclimatic and topographical effect on *G. mellonella* was relatively similar, at 48.5% and 31.6% respectively. The contribution of phenological variables was notably higher (16.6%) to *G. mellonella* prediction models compared to the rest of bee pests' abundance prediction models. Furthermore, bioclimatic variables had comparatively lower contribution on the prediction abundance of *G. mellonella* than the other bee pests. This suggests lower sensitivity of *G. mellonella* to climatic variations, and more responsiveness to vegetative food substrate. On the other hand, *A. tumida*'s abundance prediction models had a much higher contribution for bioclimatic variables (75.1%), indicating high sensitivity of this bee pest to climatic variation. Also, rainfall wettest month and potential evapotranspiration had almost two thirds (64.2%) contribution on the *A. tumida* abundance prediction models, while the topographical and vegetation phonological variables contributed 18.2% and 4.6%, respectively.

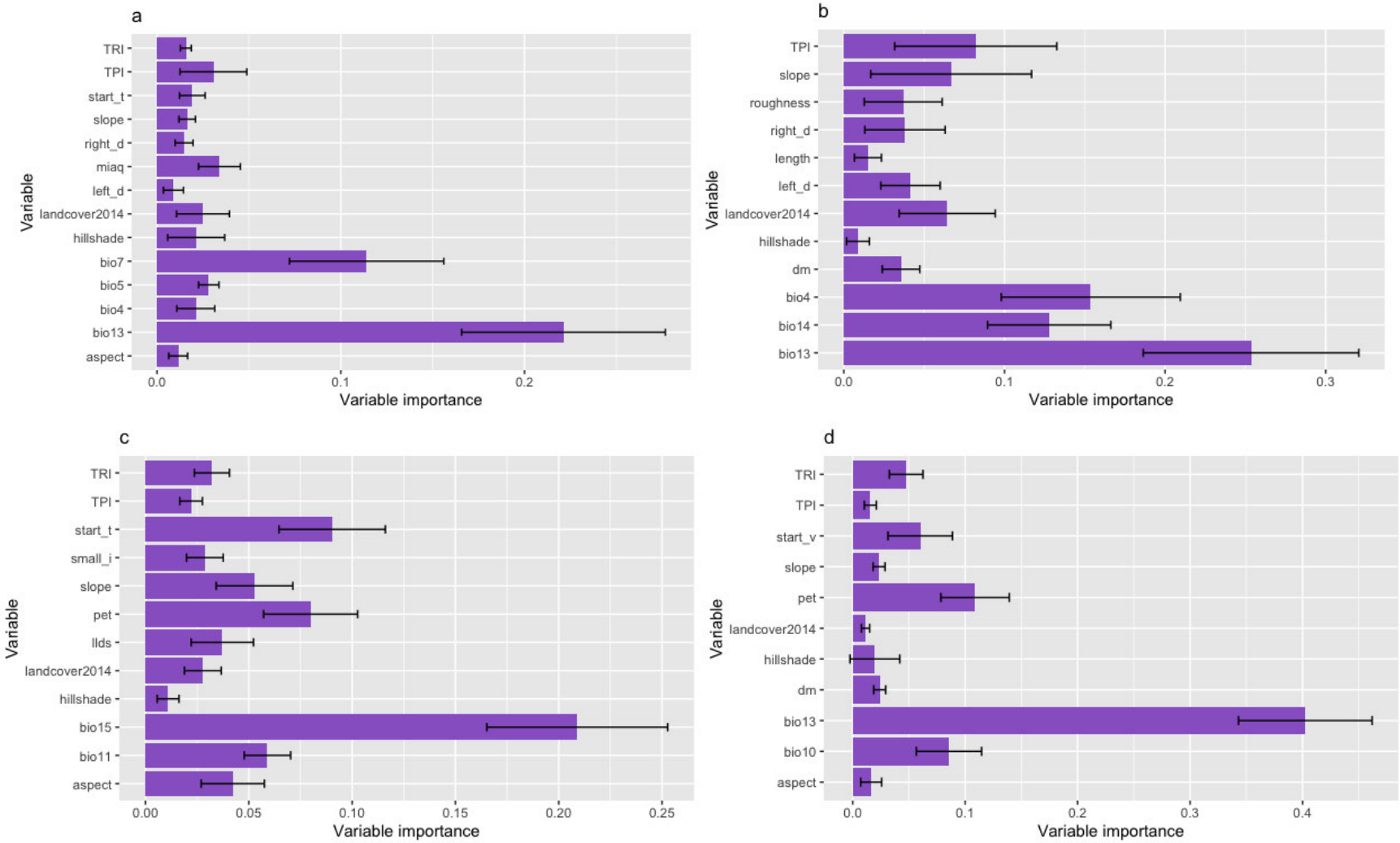


Figure 28: Interaction of selected non-conflating predictor variables for *Varroa destractor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d) prediction models.

5.3.3 Bee pests' abundance prediction validation

The Pearson correlation coefficient indicated that all RF prediction models other than *G. mellonella* had high power of fit between bee pests' abundance and the predictor variables. The *V. destractor, O. haroldi* and *A. tumida* abundance prediction models scored excellent while *G. mellonella* model scored fair (Table 21). Although all *p* values at 95% confidence interval were highly significant ($p \leq 0.05$), the *G. mellonella* model had a much smaller significance score ($p = 0.00045$). Furthermore, dispersion of abundance around line of best fit was much higher in *G. mellonella* prediction model compared to other models developed (Figure 29). This was in line with observations using Pearson's product-moment correlation for the same prediction models, suggesting that *G. mellonella* model was weaker in fitting bee pests' abundance to environmental suitability than the other models developed in this study.

Table 21: Pearson correlation coefficient (*r*) and probability scores (*p* values) of bee pests' prediction abundance models developed to predict spatial distribution of bee pests in Kenya.

Abundance prediction model	***r***	***p***	**Score**
Varroa destructor	0.83	7.17E-12	Excellent
Oplostomus haroldi	0.83	1.63E-06	Excellent
Galleria mellonella	0.55	0.0004452	Fair
Aethina tumida	0.84	1.23E-09	Excellent

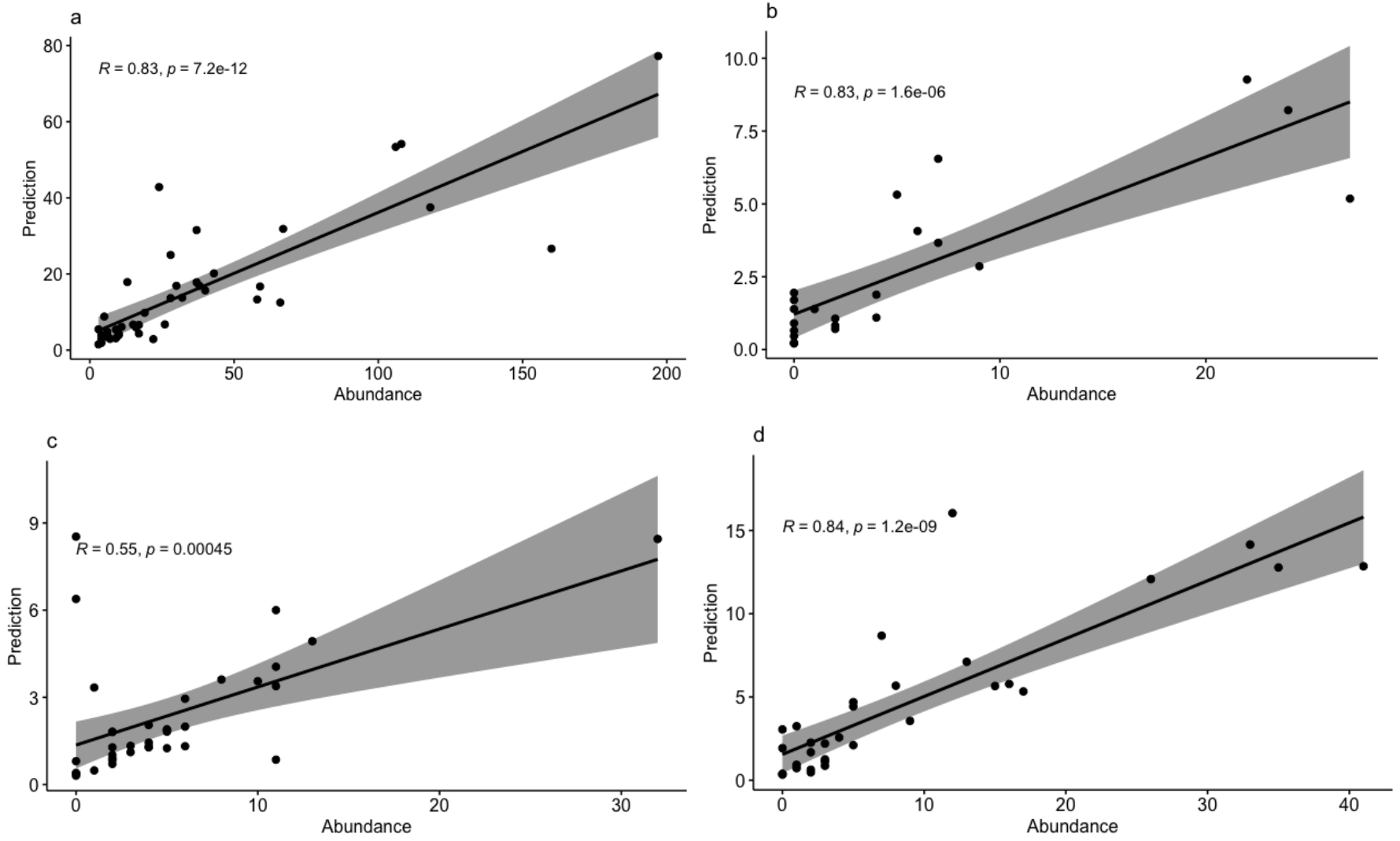


Figure 29: Scatter plots indicating distribution of predicted against observed abundance of *Varroa destractor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d).

5.3.4 Bee pests' distribution and abundance

The RF prediction models indicated varied bee pests' abundance in different regions of the study area. The western Kenya region was predicted with high abundance for *V. destractor* and *A. tumida* and moderate suitability for *G. mellonella.* Besides, these bee pests were predicted to be highly abundant in Mt. Kenya, central, western and coastal regions of Kenya (Figure 30). These regions have relatively high precipitation compared to the rest of the study site. Western, Mt. Kenya and central regions experience low temperatures whereas coastal region experiences high temperature. On the other hand, eastern region had varied predicted bee pests' abundance while Tana River, Kajiado and other semi-arid regions were predicted with low abundance of all bee pests. However, *O. haroldi* was predicted to have low abundance in most regions of the study area in the current epoch apart from Mwingi in eastern region. The latter regions are characterized by relatively dry conditions with low precipitation levels, isothermality and high temperatures. These regions exhibit varied agroecological and climatic conditions whose variations are in line with the predicted abundance. Indeed, rainfall of the wettest month (*bio13*), rainfall seasonality (*bio15*) and annual temperature range (*bio7*) contributed the most to the bee pests' abundance models. Furthermore, high altitudes were predicted to have high abundance of all bee pests. Additionally, high precipitation and low temperature were observed in regions with high altitudes. Regions with low altitude were also characterised by low precipitation levels, isothermality and relatively higher temperature. These regions were consequently predicted to have low bee pests' abundance. They include Narok county, Kajiado, Kilifi, Kwale, Tana River, the larger Taita Taveta county, Makueni and Kitui counties.

The future bee pests' abundance prediction indicates a general increase in potential bee pests' risk in most regions of the study site. There was both increment in abundance and spatial distribution across the study area (Figure 31). The abundance maps showed that most parts of the study area previously predicted to have moderate

bee pests' distribution, increased in abundance at the future epoch. Regions predicted with high proliferation include Mt. Kenya, western Kenya, coast and Mwingi in the eastern region. Notably, *V. destractor* (a) and *A. tumida* (d) had a higher proliferation (8,191.62 and 9,387.89 km^2 respectively) while *O. haroldi* (b) and *G. mellonella* (c) had moderate proliferation rates (2,766.57 and 3,057.60 km^2 respectively) in the future prediction maps (Table 22). Furthermore, bee pests' abundance change maps (Figure 32) indicated that most parts of western, Mt. Kenya, central and coastal regions had high proliferation in the future timestamps while most parts of eastern regions had changed from low to moderate proliferation for *G. mellonella* (c). On the other hand, some parts of the study site had reduced bee pests' abundance distribution (Table 22 and Figure 32) which indicated that these sites became less suitable for bee pests. However, areas with reduced bee pests' abundance were much less compared to areas with increased abundance distribution. Bioclimatic data indicated that rainfall seasonality and rainfall wettest quarter increased in the future epoch compared to current conditions.

Table 22: Predicted abundance change for *Varroa destractor, Oplostumus haroldi, Galleria mellonella* and *Aethina tumida* between current and future timestamps represented in area (km^2).

Class	***Varroa destractor***	***Oplostomus haroldi***	***Galleria Mellonella***	***Aethina tumida***
	Area (km^2)			
Decrease	001,436.54	001,887.18	004,492.30	001,725.62
No change	249,376.17	254,382.18	251,454.44	247,890.83
Increase	008,191.62	002,766.57	003,057.60	009,387.89
Total	**259,004.34**	**259,004.34**	**259,004.34**	**259,004.34**

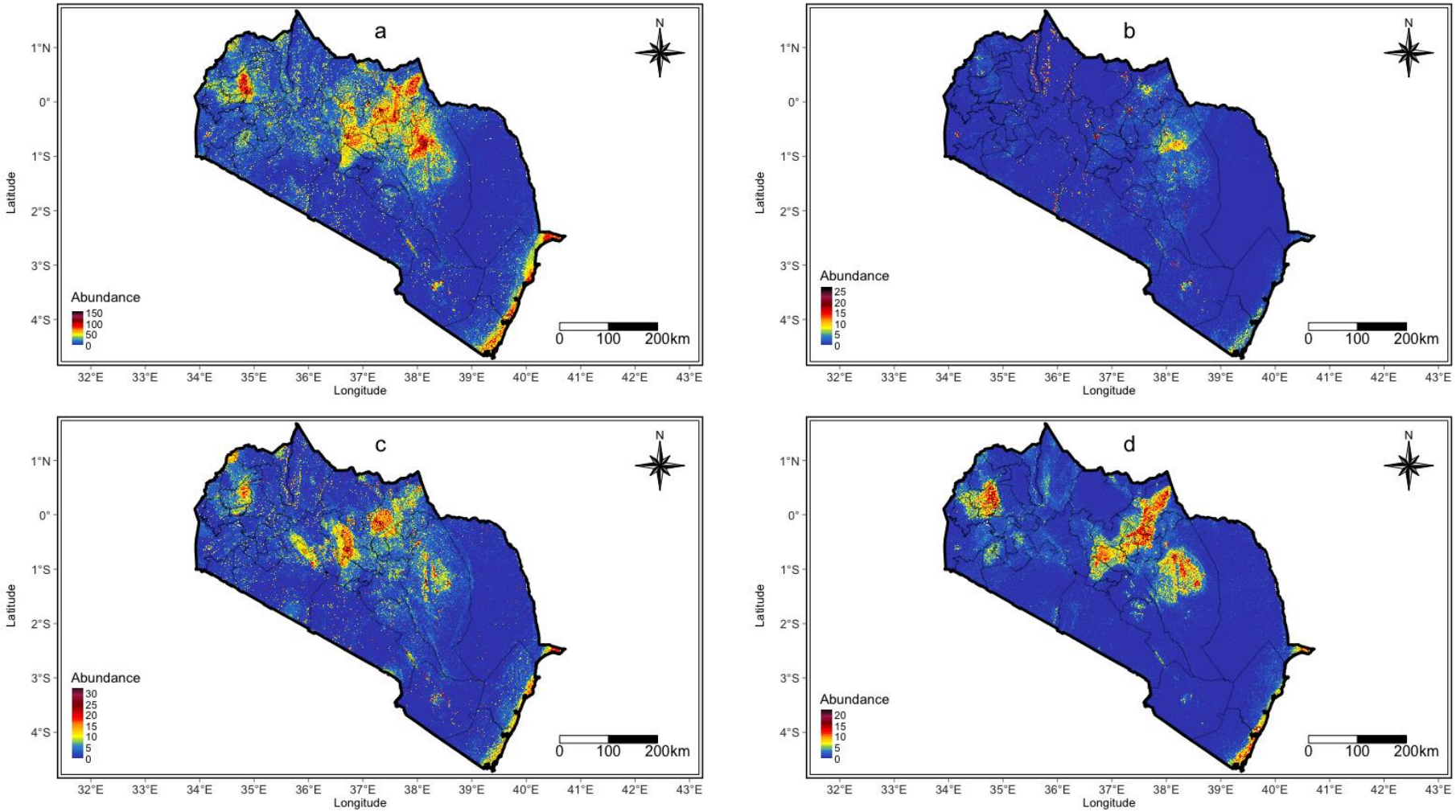


Figure 30: Predicted current abundance of *Varroa destractor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d). Blue, yellow and red colours indicate low, moderate and high predicted abundance, respectively.

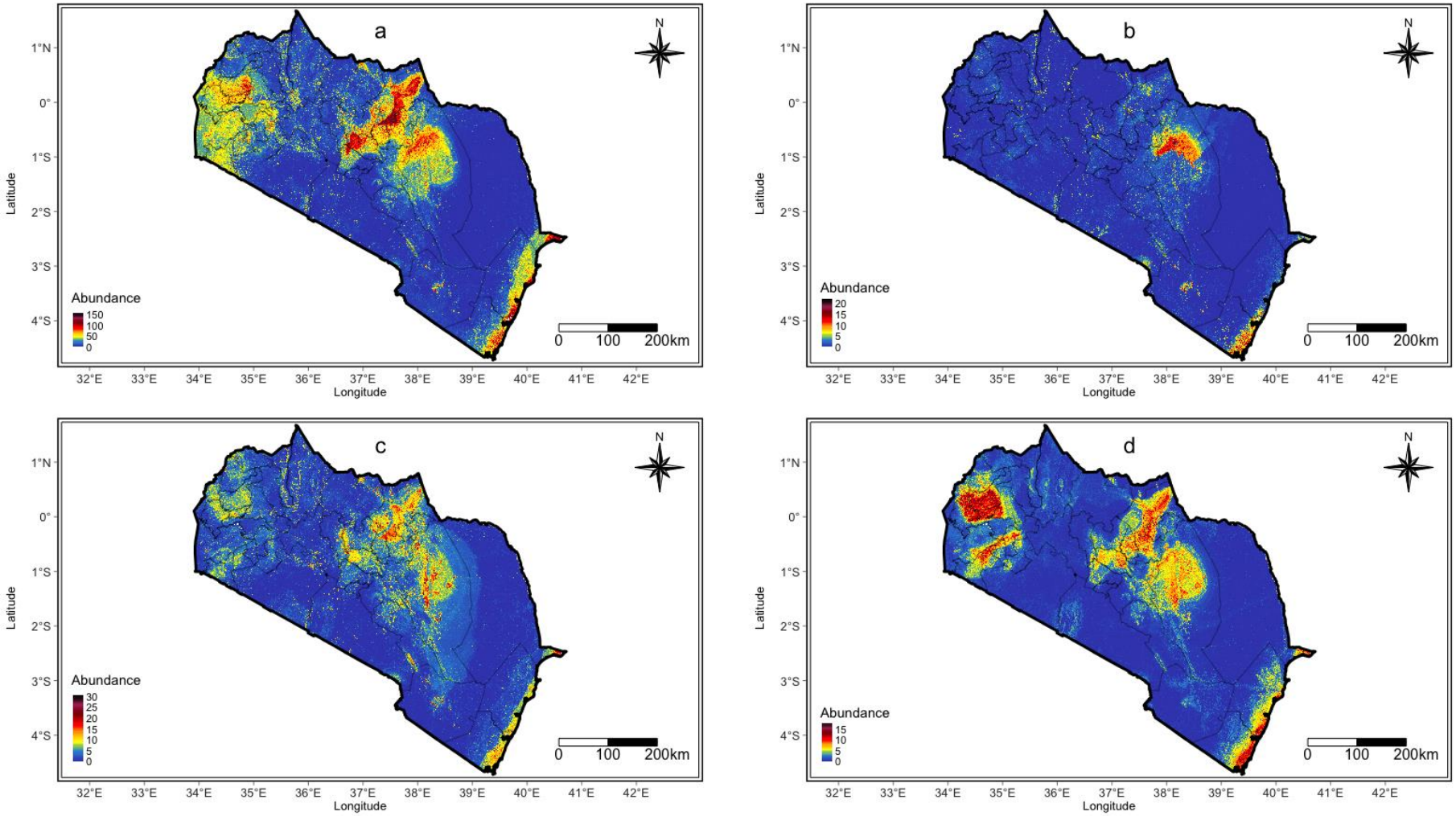


Figure 31: Predicted future (2055) abundance of *Varroa destractor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d). Blue, yellow and red colours indicate low, moderate and high predicted abundance, respectively.

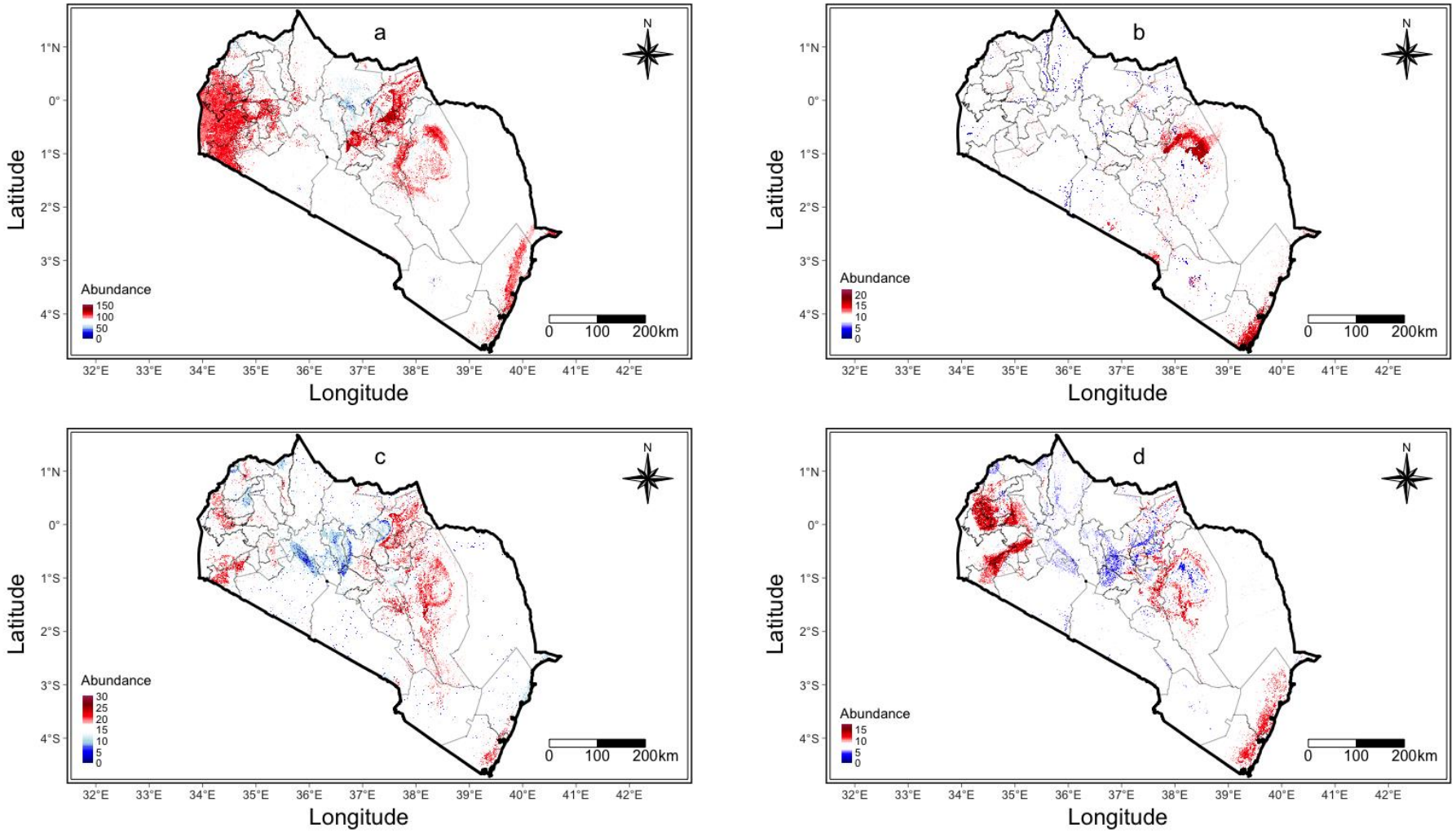


Figure 32: Change maps for abundance of *Varroa destractor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d). Red and blue colours indicate positive and negative abundance change, respectively.

5.4 Discussion

There is an elevated global interest in bee health amid increased climate change, and anthropogenic influence on natural environments and food production. As a supplemental source of income especially in sub-Saharan Africa, and high nutritional value of bee products and pollination services, understanding the spatial distribution and abundance of bee threats is vital. Accurate and reliable predictive tools that establish and demonstrate extent and severity of potential risks posed by bee pests' are important in promoting bee health (Makori et al., 2017; Martin et al., 2012). Establishing natural conditions which encourage distribution of bee pests and developing appropriate predictive models to establish their spatial and temporal risk potentials is vital. Bioclimatic datasets, vegetation phenology, topographical variables and LULC patterns were utilized to develop bee pests' abundance prediction models in this study. Bee pests' abundance data were used as response variables in the prediction models to provide spatially explicit and accurate pest population information for improved predictability as opposed to modelling occurrence data only.

5.4.1 Seasonality influence on bee pests' abundance

The effect of seasonality on bee pests' abundance was apparent. There was an average of almost five times (4.73) more bee pests during the wet season than the dry season. This was much pronounced for *V. destractor* with more than six times (6.35) pests in the wet season compared to the dry season (Table 20). However, previous studies have indicated constant infestation patterns of *V. destractor* across all seasons (Strauss et al., 2013) contrary to findings of this study. Also, all bee pests were significantly more in the wet than the dry season, apart from *G. mellonella* which had slightly more pest counts in the dry season. Similarly, Torto et. al (2010) reported higher incidences of *A. tumida* during the wet than dry season. Seasonality directly and indirectly affects bee pests' abundance and distribution through their hosts (honeybees). Essentially, bee

pests depend on their hosts for food substrate and feed on their brood. On the other hand, bee, as pests hosts, collect nectar and pollen from plants, which are affected by seasonal variations. Moreover, seasonality affects the ability of honeybees to collect and make food substrate on which the pests prey. Vegetation phenology indicate a reduction in raw materials (nectar and pollen) in the dry season and *vice versa*. This is influenced by mean annual rainfall and rainfall seasonality, which were high in the wet season. In addition, both pollen and nectar collected from flowering plants triggered by either onset or end of the rainy season, are available within apiaries preying by bee pests (Fombong et al., 2012).

5.4.2 Predictor variable selection and bee pests' abundance prediction

Honeybee pests proliferate at varying climatic and altitudinal gradients and can be found across varying agroecological and agroclimatic regions (Boncristiani et al., 2021; Fombong et al., 2012; Makori et al., 2017). Their abundance is largely dependent on their hosts distribution and survival, and various climatic and vegetation conditions. Obligate ectoparasites such as *V. destractor* depend fully on their hosts for survival, hence their success and distribution are dependent on the hosts' resilience and vibrancy. Also, generalists' pests such as *A. tumida* and *G. mellonella* depend on other food substrates as well, hence more susceptible to climatic variations. Prediction models developed in this study established that the bioclimatic variables contributed the most (between 48.48% to 77.84%) towards prediction of distribution of bee pests' abundance than vegetation phenology, topography and LULC. Furthermore, precipitation variables were more predominant than temperature variables in this category for all the studied bee pests.

Rainfall wettest month, rainfall seasonality and annual temperature range contributed about half of the information needed by prediction models for spatial abundance distribution of bee pests. The models predicted high bee pests' abundance in regions

with high precipitation levels. Moreover, precipitation variation affected distribution and abundance of bee pests across space and time. There was a positive correlation between precipitation and spatial-temporal distribution of all bee pests as indicated by their current increase in abundance and spatial presence in the future epoch (2055). This was in line with findings by various studies (Mm Fierro et al., 2012; Lovett, 2015; Makori et al., 2022, 2017; Mwalusepo et al., 2015; Platts et al., 2015) that recorded positive correlation between precipitation and insects' distribution. Additionally, response variables indicated that there was a significant difference between abundance observations done in the wet and dry seasons (Table 20 and Figure 25). There was higher bee pests' abundance in the wet than the dry season. This is in agreement with Torto et al. (2010) who indirectly linked distribution of *A. tumida* to precipitation via increased forage that trigger bee colony growth and increase of brood.

Most predictor variables (27 out of 41) that were originally thought could be used for prediction of bee pests' abundance were deemed redundant, hence eliminated from the final prediction models. The recursive feature elimination (RFE) model (Darst et al., 2018; Granitto et al., 2006; Yan and Zhang, 2015) was useful in determining the least number of predictor variables that could be used while providing fundamental information necessary for accurate prediction of bee pests' distribution (Figure 26). In conjunction with collinearity and variable importance elimination criteria, model prediction power was enhanced while minimizing parametrization and volatility (Dormann et al., 2013; Naimi and Araújo, 2016). Although vegetation phenological and topographical variables were not highly palpable towards bee pests' prediction models and contributed comparatively lower permutation, they were important in sharpening the prediction models as indicated by the number of variables from these categories tagged by prediction models (Figure 28). Moreover, vegetation phenology imply availability of both pollen and nectar which are vital to hosts in the production of food substrates and colony strength. Both time and value at the start of the season

(*start_t* and *start_v* respectively), length of the season (*length*) and right derivative (*right_d*) were among vegetation phenological variables tagged important by the bee pests' prediction models. All these variables are important indicators of nectar and pollen availability in a region. Right derivative (*right_d*) on the other hand indicates minimum vegetative matter that is available for foraging in between seasons, hence available food substrate to carry both bees and their pests through the next season. Mt. Kenya, central, western and coastal Kenya regions had high values of these predictor variables indicating a positive correlation with bee pests' abundance, which were also high in these regions.

Slope, terrain roughness index (*TRI*) and topographical position index (*TPI*) were among topographical predictor variables that contributed the most to the prediction models. Combined permutation contribution of topographical predictor variables to bee pests' abundance prediction models ranged from 14.3% to 31.6%. Indeed, most regions predicted with high abundance of bee pests had high altitude such as Mt. Kenya, Cherangani hills and areas around Mt. Elgon in western Kenya region. However, coastal Kenya and eastern regions were predicted with high abundance of bee pests as well. This was in accordance with previous studies (Birrell et al., 2020; Kimathi et al., 2020; Makori et al., 2022) which reported that insects have diverse altitudinal gradients. Although topographical variables were crucial in prediction of bee pests, their direct impact could not be conclusively deciphered. Their impact could however be derived from altitudinal effect on precipitation, temperature regimes and vegetation phenology. In this regard, different categories of predictor variables could not reliably be used in seclusion to predict spatial and abundance distribution of bee pests, hence they should be coupled with other complementary predictor variables.

5.4.3 Bee pests' abundance prediction validation

Model accuracy is dependent on the selection of appropriate response variables and important non-conflating predictive features while circumventing redundancy (Araújo et al., 2019; Naimi and Araújo, 2016). Response variables in this study employed the abundance of four species of bee pests, which were fitted with an initial 41 predictor variables to determine the best environmental fit. Under the same model settings, *V. destractor, O. haroldi* and *A. tumida* abundance prediction models' performance ranked excellent (Table 21), indicating more accurate ordering of specificity and sensitivity. On the other hand, the *G. mellonella* abundance prediction model scored fair, indicating lower accuracy. Ordering of observed versus predicted abundance (Figure 29) indicated low dispersions that pulled closer to the line of best fit in 'excellent' models' performance, with higher dispersion for the 'fair' models' performance with notable outlier abundance scores. Hence, the low scoring model suggested higher instability in abundance and suitability prediction, and thus could not be used with high certainty to predict bee pests' abundance spatial distribution (Farley, 2017; Felton et al., 2021). Despite notable lower prediction accuracy in the *G. mellonella* model (fair accuracy of r = 0.55), it had insignificantly low prediction skewness. Consequently, all models developed in this study could be used to reliably order sensitivity and specificity (Gao and Tian, 2021). Nevertheless, rigorous, and robust acquisition of substantial observed abundance data is necessary to improve predictive power of bee pests' abundance prediction models.

5.4.4 Bee pests' distribution and abundance

Climate simulation under different CO_2 emission pathways suggest an increase in temperature, and localized and cumulative precipitation intensity in future (Fick and Hijmans, 2017; Platts et al., 2015). As a result, the abundance and agility of most beneficial insects such as bees, which hosts some pests, will decrease (Torné-Noguera et al., 2014). This could reduce their ability to defend themselves against enemies

including bee pests. On the other hand, prediction models established a positive correlation between levels of precipitation and bee pests' distribution. Bee pests' abundance increased with an increase in precipitation in the simulated future epoch (2055). Indeed, prediction models demonstrated increased abundance of bee pests in future in some regions in Kenya from moderate to high (Table 22 and Figures 30, 31 and 32). Furthermore, the Wilcoxon rank sum test with continuity correction indicated significant ($p \leq 0.05$) differences in observed bee pests' abundance between the wet and the dry seasons (Table 20 and Figure 25). These differences were more pronounced in *V. destractor* and *A. tumida*. Their prediction models revealed notable differences between predicted bee pests' abundance in current and future epochs. Specifically, there was increased area suitable for *V. destractor* by a total of 8,191.62 km^2 and a decrease of only 1,436.54 km^2. Besides, *A. tumida* suitability increased by 9,387.89 km^2 and decreased by only 1,725.62 km^2. The Mt. Kenya, central, western and coastal Kenya sites were the most affected regions by these changes. These regions are characterized by high precipitation (Githui et al., 2009). In addition, these regions have been demonstrated to have higher rainfall and temperature seasonality with elevated levels of evapotranspiration in future (Fick and Hijmans, 2017; Platts et al., 2015). The primal climatic conditions for bee pests could have shifted with interaction of changing bioclimatic variables in these regions. In addition, these interactions could affect the phonological patterns in habitats of bee pests' habitats and their hosts (i.e., bees). A positive shift of precipitation levels in regions such as Kitui, Narok and Kajiado directly or indirectly triggered an increase of both suitability and abundance of bee pests in the future epoch (Figure 32). Furthermore, such changes affects the hosts negatively (Dainat et al., 2012; Makori et al., 2022; Williams and Tarpy, 2010) making them more susceptible to the pests' invasion. Hence, bee pests threat levels increased in most areas across the study site.

The inclusion of LULC as a proxy for anthropogenic influence and vegetation phenological variables improved the predictive power of bee pests' abundance,

making them more reliable. While LULC data indicate human population effect to the prediction models, vegetation phenological variables provided data on growing seasons at grain level, hence availability of food substrates. Moreover, vegetation phenological variables were processed from MODIS EVI at higher spatial and temporal resolutions (250 metres spatial resolution) as opposed to bioclimatic data (1,000 metres spatial resolution). Therefore, vegetation phenological variables provided prediction models with higher environmental heterogeneity at higher spatial resolution than the bioclimatic variables (Saatchi et al., 2008). In this regard, vegetation variations enabled the prediction models to identify heterogenous pockets on the landscape which either hinder or limit the abundance and distribution of bee pests. In addition, vegetation phenological variables were derived from monthly MODIS EVI datasets, which provide higher temporal resolution (16 days) compared to the bioclimatic variables that are interpolated over longer periods and large homogenous spatial extents. Therefore, vegetation phenological variables provided prediction models with detailed near real-time and actual information that improved their predictive power, hence more credible and dependable.

5.5 Conclusions

Amid elevated global interest in climate change and anthropogenic influence on natural environments and agricultural patterns, bee farming has gained more relevance, especially in African agroecological regions where food and nutritional security are often elusive. However, bee health is threatened by among others bee pests which ravage colonies and even apiaries. This study developed bee pests' abundance and distribution prediction models for four bee pests viz *V. destractor, O. haroldi, G. mellonella* and *A. tumida.* Abundance data were integrated with bioclimatic, vegetation phenological, topographical and LULC (as a proxy for anthropogenic influence) variables to develop precise and reliable bee pests' abundance prediction models. The models ranked excellent for *V. destractor, O. haroldi* and *A. tumida*, and fair for *G. mellonella* using the Pearson correlation scale. Precipitation variables contributed the most to bee pests' abundance prediction models and seasonal variations proved significantly ($p \leq 0.05$) influential in bee pests' abundance as indicated by the Wilcoxon rank sum test with continuity correction. Furthermore, regions under the study area with high rainfall variability and high humidity were predicted with higher threat of bee pests. Furthermore, bee pests' threat levels were predicted to increase both spatially and in intensity (abundance) with climate variations across the study site. The bee pests' abundance prediction models developed in this study were deemed precise, hence could reliably be used to map high risk areas, where management efforts and resources could be employed to curb the spread of bee pests. Therefore, these prediction models could provide decision makers with essential tools to assuage spread of bee pests, hence improving bee health.

5.6 Acknowledgments

The authors would like to appreciate farmers who participated in this research either directly or indirectly by providing information or apiaries where response variables were collected from. In addition, our appreciation goes to personnel who collected field data without which this research may not have been realised. More importantly, our heartfelt gratitude goes to Darrel Makori, Lloyd Magara, Darlene Moraa, George Gesaka, Priscah Moraa, Lydia Nthenya, Sarah Ogutu, Zachary Mburu and Sumaiya Mohamed for support in data collection, processing, and analysis. We would also like to appreciate the European Union for providing resources utilized in field sampling exercise (Project number; DCI-FOOD-2011/023-520), Bayer AG Crop Sciences (Project number: NC20662450), and NORAD (Project number: RAF-3058 KEN-18/0005). We gratefully acknowledge financial support for this research by the following organizations and agencies: UK's Foreign, Commonwealth & Development Office (FCDO); Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Federal Democratic Republic of Ethiopia; and Government of the Republic of Kenya. The views expressed herein do not necessarily reflect the official opinion of people and agencies that supported the authors.

CHAPTER SIX

6 Geospatial techniques in bee farming and bee health: a synthesis

A bee on a flower collecting pollen and nectar. Image by *(Madrigal, 2022)*; source https://cdn.naturettl.com/wp-content/uploads/2021/05/21151609/How-to-Photograph-Bees-13.jpg

6.1 Introduction

Ostensible climate change and anthropogenic influence on natural environments and agricultural systems have heightened the significance of bees on global socioeconomic and environmental aspects. In the rural African savanna, bee products such as honey, pollen, nectar, royal jelly and bee venom are important as supplemental sources of income. These products are particularly important in the underdeveloped rural African savanna. Furthermore, financial benefits associated with apiculture are important incentives to forest conservation, ecosystem services such as pollination of crops and wild plants, and vital to food production, particularly for improving food and nutritional security for smallholder farmers. However, apiculture and bee health are under threat from climate change, agricultural and agrochemical intensification and associated habitat alteration, bee pests and diseases. Therefore, this study sought to establish spatio-temporal distribution of bees (stingless bees), their food substrate (flowering plants) and threats (bee pests) in Kenya using earth observation and geospatial techniques. Firstly, the study predicted the spatial distribution of stingless bees in Kenya using machine learning algorithms based on occurrence data and non-conflating features drawn from bioclimatic, vegetation phenological and topographical variables (Chapter 2). Moreover, the study upscaled flower mapping approaches using hyperspectral datasets captured in both onset and peak flowering periods, to the spectral and spatial specifications of four readily available multispectral sensors (Chapter 3). These datasets were used to map the distribution of flowering plants in Kenya and developed techniques that could be readily available and usable by various stakeholders. Furthermore, the study used both bee pests' occurrence-only datasets, in conjunction with bioclimatic, vegetation phenological, topographical and LULC variables, to predict spatial distribution of bee pests, as biotic threats, using MaxEnt (Chapter 4). The study sought to further refine the prediction of the spatial distribution of bee pest using detailed abundance data and superior machine learning algorithms (Chapter 5). All findings and scientific techniques

developed herein are wrapped up in the synthesis chapter (Chapter 6, this chapter) giving general conclusion and recommendations/way forward based on specific chapter that supersede this one.

6.2 Study implications

This study provides earth observation and geospatial modelling tools, and techniques that could be utilized to improve bee health and apiculture in Kenya, which are also upscalable to other parts of the world. Prediction models and mapping techniques developed in this study ranged from 'fair' to 'excellent' in mapping high suitability areas for meliponiculture. Further, the techniques were utilized for predicting high bee pests risk areas for mitigation purposes, hence improving hive productivity and bee health. The findings in Chapter 2 implied that the stingless EN models developed in this study could reliably be used to predict stingless bee distribution in Kenya, upscalable to a future epoch (2055). The findings indicated that as the effects of climate change and anthropogenic alterations of natural environments increase, the suitability and spatial distribution of meliponine bees decrease. Therefore, these findings are key for both meliponine bee farming and insect pollinated crops for identification of regions with suitable conditions. Additionally, authorities could utilize these data and techniques to improve apicultural and agricultural productivity. In addition, these findings could provide a basis to incentivise apiculture for conservation of natural vegetation by communities adjacent to forests through supplementary pollination services.

Furthermore, in Chapter 3 this study was able to demonstrate methods that could be used to map melliferous plants using readily and freely available datasets. This is vital due to the cost limitations occasioned when acquiring hyperspectral datasets that had only been previously used to map flowering plants. In addition, this study was able to use mapping methods inherent in open source software, as compared to commercially developed software with inhibitory costs. Ultimately, this study

provided the first ever cost-effective approaches for mapping flowering plants in an African semiarid agroecological landscape. These approaches also provide valuable, timely and reliable advisory tools for guiding implementation of beekeeping systems at a landscape scale.

In Chapter 4, this study developed mapping methods that could be used to monitor spatial distribution and proliferation of major bee pest. Under this chapter, the study used simple and straight forward techniques in open source MaxEnt and Timesat software. The novelty of these methods was demonstrated by incorporating vegetation phenology variables derived from freely available MODIS NDVI datasets. These data improved the predictive power of models developed for mapping bee pests' habitat and containment zones.

However, the mapping methods developed in Chapter 4 were limited to predictions that were based on only occurrence data. They did not incorporate important bee pest abundance data, that could not only predict spatial risk zones, but also indicate the severity of the threats. The methods developed under Chapter 5 harnessed this important aspect to provide bee farmers and policy makers with 'excellent' tools that could be utilized to map bee pest hot spots where containment and management measures could be prioritized. To this end, this study has been able to demonstrate the implementation of earth observation and geospatial techniques in improving apiculture and bee health. Therefore, researchers, farmers, managers and policy makers have been provided with tools, methods and techniques that could be used to make apiculture more beneficial. In addition, these techniques could help improve the livelihoods of smallholder farmers, increase their resilience to climate change and improve natural vegetation cover.

6.3 Study limitations

This study sought to develop techniques that could be reliably used to map spatial distribution of stingless bees, bee pests and flowering plants in Kenya. Firstly, it successfully developed models that could predict the spatial distribution of stingless bees across 40 counties in Kenya. To achieve this, it used occurrence-only data from 5 species of stingless bees, collected from 19 out of the 40 administrative counties. Even though the 19 counties covered the main agroecological zones in the country, there could have been subtle micro-climatic pockets that may not be represented. A wider spatial representation in all counties could have been more representative and hence improve model's accuracy and reliability, albeit slightly. On the other hand, even though this study collected species specific field data, it did not develop species-specific distribution models. There could have been species-specific differences which could be easily missed by bundling all species data together. Further, this study used meliponine bees' occurrence-only data to develop spatial distribution models. It did not harness the demonstrated benefits of abundance data to offer the spatial models with more modelling information.

The flower mapping techniques used in this research relied heavily on resampled satellite imagery. The final multispectral image datasets that were developed inherited characteristics of the original AISA Eagle hyperspectral image. Therefore, all faults that could have been associated with acquisition, processing and storage of the hyperspectral datasets could have been inherited by the simulated multispectral datasets. Using 'actual' multispectral dataset might have enabled the study to avert some inaccuracies associated with such errors. In addition, field data was used to train flower mapping models developed herein was collected within three days of acquisition of the imagery data. This is despite alteration of atmospheric conditions within short periods. Therefore, collection of field data within the shortest time possible of the acquisition of image data is recommended. In addition, use of a handheld spectrometer to collect field data could have offered the flower mapping

models developed in this research with more superior training data to improve their accuracies.

The bee pests' models developed in this study were largely excellent but exhibited limitations similar to the models developed for stingless bees. For instance, the spatial distribution of field data used to train models exhibited spatial biasness. Even though they largely covered a representative agroecological zone gradient, there were pockets of the study area that were not covered. In addition, methods developed in Chapter 4 utilized occurrence-only as opposed to abundance data. Furthermore, only MaxEnt software was used as opposed to advanced machine learning (ML) algorithms such as Random Forest (RF). However, these setbacks were largely addressed by methods developed in Chapter 5. Abundance data and ML were used to develop more advanced mapping methods.

Variable multicollinearity was also observed across all spatial prediction models developed in this study. For instance, out of 40 variables presented for modelling distribution of both stingless bees and bee pests, 11 variables were eligible. This was mostly due to both bioclimatic, biotic and topographic variables exhibiting collinearity issues. Even though it was concluded that the excluded variables did not provide the models with substantially extra information and only marginally improved the information base available to the spatial models, they could contain unique information which is lost when they are illuminated from the modelling environment. Therefore, models that could utilize an array of variables while circumventing parameterization could be more desirable.

6.4 Conclusions

Due to the polylactic nature and pollination uniqueness of stingless bees and the economic importance of bee farming in sub-Saharan Africa, the earth observation and geospatial techniques developed herein could improve bee farming and bee health in the region. Specifically, the geospatial techniques studied and developed in this research were demonstrated to be beneficial to bee health in reliably mapping the spatial distribution of stingless bees, bee food substrate (flowering plants) and their biotic threats (bee pests). The study was able to develop the first ever known ecological niche models for mapping of stingless bees in Africa. It harnessed careful selection of both response and predictor variables, and six best machine learning algorithms to develop 'excellent' modelling approaches. Ensembling the best modelling approaches however yielded higher accuracies and more realistic spatial models. It was apparent that bioclimate, especially precipitation variables were most important for stingless bee models. However, vegetation phenology which did not contribute most logit to the prediction models, more legible variables to the models. Indeed, the models were more realistic and accurate with the inclusion of the biotic variables. The effect of topography was not immediately discerned from the contribution of topological variables, but was inferred especially from biotic factors in the models.

Similar conclusions were drawn from bee pest distribution models, especially since bioclimatic variables specifically precipitation proved most important. Mean annual rainfall contributed most of the information to the models. In addition, it was apparent that machine learning algorithms and ensemble methods were the most accurate with more realistic products. Furthermore, careful selection of non-collinear variables and accurately collected and representative response variables improved the accuracy of models developed. Even though occurrence-only EN models could yield products that could give an indication on the pest risk zones, they could not provide much needed information on bee pest severity and intensity. The abundance spatial models

however could indicate bee pest risk and containment zones, severity and changing conditions over time.

Temporal upscalling of the EN models to the year 2055 based on the hypothesized and modelled pathways provided an insight on the effects of climate change and the anthropogenic influence on both bees and their threats. As habitat suitability of stingless bees reduced, the risk levels and spatial distribution of bee pests increased. The same projection epoch indicated high precipitation and temperature seasonality, variables that affected both stingless bees and pests the most. Even though upscaling of vegetation phenology was not available in this study, an increment of the anthropogenic effect from the human footprint data implied a reduction of natural vegetation and in effect both habitat and food sources for bees. As the habitat suitability of stingless bees reduced, the risk intensity and spatial extent of bee pests increased. It was concluded that there was an inverse relationship between the habitat suitability of beneficial insects and their threats.

Food resources/substrates (flowers) were successfully mapped in this study as well using resampled multispectral imagery datasets. The images with higher spatial resolution had higher overall accuracy. In addition, spectral resolution improved the classification accuracy of flowering plants. This study concluded that an improvement in both the spatial and spectral resolutions improves classification accuracy of flowering plants.

6.5 Recommendations

The study demonstrated the importance of earth observation and geospatial techniques in apiculture and bee health. It showed how these techniques can be used to predict the spatial distribution of bees (stingless bees), their food substrate (flowering plants) and their threats (bee pests). The inverse relationship between bees and their pests that has been demonstrated in this study should be an eye opener to researchers, farmers, conservationist, policy makers and authorities concerned with agriculture, environment and forestry. Farmers should embrace friendlier and less invasive practices and reduce intensification of agrochemicals on their farms. On the other hand, they should use the tools that have been developed herein for better informed placement of apiaries to realise improved hive productivity and farm field outputs. Moreover, the tools developed in this study should provide relevant bodies, organization and policy makers with the impetus to improve apiculture within the region and reverse the negative effects occasioned by climate change, habitat alteration, and agricultural and agrochemical intensification. Improved apiculture and agricultural practices could positively impact on the livelihoods of smallholder farmers in the African savanna.

Furthermore, advancement of earth observation and geospatial techniques, and new discoveries around artificial intelligence and climate change, should advance studies around new cutting-edge mapping algorithms including ultra-deep machine learning. Moreover, there is continuous advancement on climate (e.g., wind speed, direction e.t.c.), vegetation (e.g., 3D canopy to height and canopy cover fraction) and anthropogenic (e.g., human footprints) datasets that could be utilized to study their effects on farming and bee health. Therefore, more research should be carried out to incorporate all these advancements in understanding the dynamics of bee farming and bee health. More specifically, future studies should be directed towards optimal exploitation of the following recommendations:

i. Climate change will in a larger part shrink suitable habitats for bees. Therefore, more accurately upscaled datasets available in near real-time and downscaled to regional or country level should be explored and used to develop reliable prediction models.

ii. Bee population datasets (abundance) rather than presence-only datasets should be used to provide more accurate prediction of bee and bee pests' distribution at regional scale.

iii. Species-specific prediction models should be explored in the prediction of spatial distribution of bees since there are subtle variations between bee species which non-species-specific models may be insensitive to.

iv. Mapping of temporal floral patterns should be explored further using high-resolution phenotypic camera and other sensors mounted on unmanned aerial vehicles (i.e., drones and balloons) and artificial intelligence algorithms. In addition, handheld hyperspectral spectrometres should be used to collect floral spectra field training data for improved classification accuracies.

v. Research should be expanded beyond the prediction of spatial distribution of bee pests' species studied in this research. Further research should be carried out on species interaction, especially interaction among the bees as hosts and the pests (i.e., areas that are suitable for different bee and pest species).

References

Abdel-Rahman, E.M., Ahmed, F.B., Ismail, R., 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. Int. J. Remote Sens. 34, 712–728. https://doi.org/10.1080/01431161.2012.713142

Abdel-Rahman, E.M., Makori, D.M., Landmann, T., Piiroinen, R., Gasim, S., Pellikka, P., Raina, S.K., 2015. The Utility of AISA Eagle Hyperspectral Data and Random Forest Classifier for Flower Mapping. Remote Sens. 7, 13298–13318. https://doi.org/10.3390/rs71013298

Acevedo, P., Ferreres, J., Escudero, M.A., Jimenez, J., Boadella, M., Marco, J., 2017. Population dynamics affect the capacity of species distribution models to predict species abundance on a local scale. Divers. Distrib. 23, 1008–1017. https://doi.org/10.1111/ddi.12589

Acharya, A., Prakash, A., Saxena, P., Nigam, A., 2013. Sampling: Why and How of it? Anita S Acharya, Anupam Prakash, Pikee Saxena, Aruna Nigam. Indian J. Med. Specilaities. https://doi.org/10.7713/ijms.2013.0032

Ackerly, D.D., Loarie, S.R., Cornwell, W.K., Weiss, S.B., Hamilton, H., Branciforte, R., Kraft, N.J.B., 2010. The geography of climate change: implications for conservation biogeography. Divers. Distrib. 16, 476–487. https://doi.org/10.1111/j.1472-4642.2010.00654.x

Adolkar, V.V., Raina, S.K., Kimbu, D.M., 2007. Evaluation of various mulberry Morus spp. (Moraceae) cultivars for the rearing of the bivoltine hybrid race Shaanshi BV-333 of the silkworm Bombyx mori (Lepidoptera: Bombycidae). Int. J. Trop. Insect Sci. 27, 6. https://doi.org/10.1017/S174275840774537X

Aizen, M.A., Harder, L.D., 2009. The Global Stock of Domesticated Honey Bees Is Growing Slower Than Agricultural Demand for Pollination. Curr. Biol. 19, 915–918. https://doi.org/10.1016/j.cub.2009.03.071

Allouche, O., Tsoar, A., Kadmon, Ronen, Kadmon, R, 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Allsopp, M.H., Lange, W.J. de, Veldtman, R., 2008. Valuing Insect Pollination Services with Cost of Replacement. PLOS ONE 3, e3128. https://doi.org/10.1371/journal.pone.0003128

Amirpour Haredasht, S., Barrios, M., Farifteh, J., Maes, P., Clement, J., Verstraeten, W.W., Tersago, K., Van Ranst, M., Coppin, P., Berckmans, D., Aerts, J.-M., 2013. Ecological Niche Modelling of Bank Voles in Western Europe. Int. J. Environ. Res. Public. Health 10, 499–514. https://doi.org/10.3390/ijerph10020499

Anderson, C., Brunn, A., Thiele, M., 2014. Absolute Calibration of the RapidEye Constellation. Conf. Charact. Radiom. Calibration Remote Sens. CALCON.

Anderson, N.T., Marchisio, G.B., 2012. WorldView-2 and the evolution of the DigitalGlobe remote sensing satellite constellation, in: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII. Presented at the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, International Society for Optics and Photonics, p. 83900L. https://doi.org/10.1117/12.919756

Aranda, S.C., Lobo, J.M., 2011. How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. Ecography 34, 31–38. https://doi.org/10.1111/j.1600-0587.2010.06134.x

Araujo, M.A.R., Libério, S.A., Guerra, R.N.M., Ribeiro, M.N.S., Nascimento, F.R.F., 2012. Mechanisms of action underlying the anti-inflammatory and immunomodulatory effects of propolis: a brief review. Rev. Bras. Farmacogn. 22, 208–219. https://doi.org/10.1590/S0102-695X2011005000167

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. Sci. Adv. 5, eaat4858. https://doi.org/10.1126/sciadv.aat4858

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Asner, G.P., Martin, R.E., 2012. Contrasting leaf chemical traits in tropical lianas and trees: implications for future forest composition. Ecol. Lett. 15, 1001–1007.

Attorre, F., Alfo', M., De Sanctis, M., Francesconi, F., Bruno, F., 2007. Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. Int. J. Climatol. 27, 1825–1843. https://doi.org/10.1002/joc.1495

Baldridge, E., Harris, D.J., Xiao, X., White, E.P., 2016. An extensive comparison of species-abundance distribution models. PeerJ 4, e2823. https://doi.org/10.7717/peerj.2823

Becker, B.L., Lusch, D.P., Qi, J., 2007. A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery. Remote Sens. Environ. 108, 111–120. https://doi.org/10.1016/j.rse.2006.11.005

Bijlsma, L., Bruijn, L.L.M. de, Martens, E.P., Sommeijer, M.J., 2006. Water content of stingless bee honeys (Apidae, Meliponini): interspecific variation and comparison with honey of Apis mellifera. Apidologie 37, 480–486. https://doi.org/10.1051/apido:2006034

Birrell, J.H., Shah, A.A., Hotaling, S., Giersch, J.J., Williamson, C.E., Jacobsen, D., Woods, H.A., 2020. Insects in high-elevation streams: Life in extreme environments imperiled by climate change. Glob. Change Biol. 26, 6667–6684. https://doi.org/10.1111/gcb.15356

Blay, D., Appiah, M., Damnyag, L., Dwomoh, F.K., Luukkanen, O., Pappinen, A., 2008. Involving local farmers in rehabilitation of degraded tropical forests:

some lessons from Ghana. Environ. Dev. Sustain. 10, 503–518. https://doi.org/10.1007/s10668-006-9077-9
Boncristiani, H., Ellis, J.D., Bustamante, T., Graham, J., Jack, C., Kimmel, C.B., Mortensen, A., Schmehl, D.R., 2021. World Honey Bee Health: The Global Distribution of Western Honey Bee (*Apis mellifera* L.) Pests and Pathogens. Bee World 98, 2–6. https://doi.org/10.1080/0005772X.2020.1800330
Bradley, B.A., 2016. Predicting abundance with presence-only models. Landsc. Ecol. 31, 19–30. https://doi.org/10.1007/s10980-015-0303-4
Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E., Zachmann, L.J., 2012. Species detection vs. habitat suitability: Are we biasing habitat suitability models with remotely sensed data? Ecol. Model. 244, 57–64. https://doi.org/10.1016/j.ecolmodel.2012.06.019
Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324
Buchmann, S.L., Nabhan, G.P., 2012. The Forgotten Pollinators. Island Press.
Büchs, W., 2003. Biotic indicators for biodiversity and sustainable agriculture—introduction and background. Agric. Ecosyst. Environ., Biotic Indicators for Biodiversity and Sustainable Agriculture 98, 1–16. https://doi.org/10.1016/S0167-8809(03)00068-9
Camberlin, P., Boyard-Micheau, J., Philippon, N., Baron, C., Leclerc, C., Mwongera, C., 2014. Climatic gradients along the windward slopes of Mount Kenya and their implication for crop risks. Part 1: climate variability. Int. J. Climatol. 34, 2136–2152. https://doi.org/10.1002/joc.3427
Cameron, S.A., Sadd, B.M., 2020. Global Trends in Bumble Bee Health. Annu. Rev. Entomol. 65, 209–232. https://doi.org/10.1146/annurev-ento-011118-111847
CGIAR-CSI, 2020. CGIAR-CSI SRTM – SRTM 90m DEM Digital Elevation Database [WWW Document]. URL http://srtm.csi.cgiar.org/ (accessed 11.8.20).
Champetier, A., Sumner, D.A., Wilen, J.E., 2014. The Bioeconomics of Honey Bees and Pollination. Environ. Resour. Econ. 60, 143–164. https://doi.org/10.1007/s10640-014-9761-4
Chaudhary, A., Verones, F., de Baan, L., Hellweg, S., 2015. Quantifying Land Use Impacts on Biodiversity: Combining Species–Area Models and Vulnerability Indicators. Environ. Sci. Technol. 49, 9987–9995. https://doi.org/10.1021/acs.est.5b02507
Chiawo, D.O., Kioko, E.N., Ogol, C.K., Gikungu, M.W., 2011. Bee diversity and floral resources along a disturbance gradient in Kaya Muhaka forest and surrounding farmlands of coastal Kenya.
Cho, M.A., Mathieu, R., Asner, G.P., Naidoo, L., van Aardt, J., Ramoelo, A., Debba, P., Wessels, K., Main, R., Smit, I.P.J., Erasmus, B., 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. Remote Sens. Environ. 125, 214–226. https://doi.org/10.1016/j.rse.2012.07.010

Clark, M.L., Aide, T.M., Grau, H.R., Riner, G., 2010. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. Remote Sens. Environ. 114, 2816–2832. https://doi.org/10.1016/j.rse.2010.07.001

Clasen, A., Somers, B., Pipkins, K., Tits, L., Segl, K., Brell, M., Kleinschmit, B., Spengler, D., Lausch, A., Förster, M., 2015. Spectral Unmixing of Forest Crown Components at Close Range, Airborne and Simulated Sentinel-2 and EnMAP Spectral Imaging Scale. Remote Sens. 7, 15361–15387. https://doi.org/10.3390/rs71115361

Clifford, G.D., Tarassenko, L., 2005. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. IEEE Trans. Biomed. Eng. 52, 630–638. https://doi.org/10.1109/TBME.2005.844028

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychosocial Measurement, 20, 37-46.

Colgan, M.S., Baldeck, C.A., Féret, J.-B., Asner, G.P., 2012. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. Remote Sens. 4, 3462–3480.

Congalton, R.G., Green, K., 2008. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Second Edition. CRC Press.

Cooley, A.M., Carvallo, G., Willis, J.H., 2008. Is Floral Diversification Associated with Pollinator Divergence? Flower Shape, Flower Colour and Pollinator Preference in Chilean Mimulus. Ann. Bot. 101, 641–650. https://doi.org/10.1093/aob/mcn014

Cord, A.F., Klein, D., Gernandt, D.S., de la Rosa, J.A.P., Dech, S., 2014a. Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. J. Biogeogr. 41, 736–748. https://doi.org/10.1111/jbi.12225

Cord, A.F., Klein, D., Gernandt, D.S., de la Rosa, J.A.P., Dech, S., 2014b. Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. J. Biogeogr. 41, 736–748.

Cortopassi-Laurino, M., Imperatriz-Fonseca, V.L., Roubik, D.W., Dollin, A., Heard, T., Aguilar, I., Venturieri, G.C., Eardley, C., Nogueira-Neto, P., 2006. Global meliponiculture: challenges and opportunities. Apidologie 37, 275–292. https://doi.org/10.1051/apido:2006027

Dainat, B., Vanengelsdorp, D., Neumann, P., 2012. Colony collapse disorder in Europe. Environ. Microbiol. Rep. 4, 123–125.

Dallas, T.A., Hastings, A., 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. Glob. Ecol. Biogeogr. 27, 1448–1456. https://doi.org/10.1111/geb.12820

Dalponte, M., Ørka, H.O., Gobakken, T., Gianelle, D., Næsset, E., 2013. Tree Species Classification in Boreal Forests With Hyperspectral Data. IEEE Trans. Geosci. Remote Sens. 51, 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272

Darst, B.F., Malecki, K.C., Engelman, C.D., 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet. 19, 1–6.

de la Fuente, A., Hirsch, B.T., Cernusak, L.A., Williams, S.E., 2021. Predicting species abundance by implementing the ecological niche theory. Ecography 44, 1723–1730. https://doi.org/10.1111/ecog.05776

Delaplane, K.S., 2010. Honey bees and beekeeping.

DeMarche, M.L., Doak, D.F., Morris, W.F., 2019. Incorporating local adaptation into forecasts of species' distribution and abundance under climate change. Glob. Change Biol. 25, 775–793. https://doi.org/10.1111/gcb.14562

Dietemann, V., Nazzi, F., Martin, S.J., Anderson, D.L., Locke, B., Delaplane, K.S., Wauquiez, Q., Tannahill, C., Frey, E., Ziegelmann, B., 2013. Standard methods for varroa research. J. Apic. Res. 52, 1–54.

Dikshit, O., Roy, D.P., 1996. An empirical investigation of image resampling effects upon the spectral and textural supervised classification of a high spatial resolution multispectral image. Photogramm. Eng. Remote Sens. 62, 1085–1092.

Dormann, C.F., 2020. Calibration of probability predictions from machine-learning and statistical models. Glob. Ecol. Biogeogr. 29, 760–765. https://doi.org/10.1111/geb.13070

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Du, Z., Wang, Z., Liu, Yunxia, Wang, H., Xue, F., Liu, Yanxun, 2014. Ecological niche modeling for predicting the potential risk areas of severe fever with thrombocytopenia syndrome. Int. J. Infect. Dis. 26, 1–8.

Eardley, C., Kwapong, P., 2013. Taxonomy as a Tool for Conservation of African Stingless Bees and Their Honey, in: Vit, P., Pedro, S.R.M., Roubik, D. (Eds.), Pot-Honey: A Legacy of Stingless Bees. Springer, New York, NY, pp. 261–268. https://doi.org/10.1007/978-1-4614-4960-7_18

Eardley, C.D., 2004. Taxonomic revision of the African stingless bees (Apoidea : Apidae : Apinae : Meliponini). Afr. Plant Prot. 10, 63–96.

Eklundha, L., Jönsson, P., 2017. TIMESAT 3.3 with Seasonal Trend Decomposition and Parallel Processing Software Manual. Lund University.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Endo, T., Watanabe, T., Yamamoto, A., 2015. Confidence interval estimation by bootstrap method for uncertainty quantification using random sampling method. J. Nucl. Sci. Technol. 52, 993–999. https://doi.org/10.1080/00223131.2015.1034216

Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. Appl. Ecol. 41, 263–274. https://doi.org/10.1111/j.0021-8901.2004.00881.x

Entine, J., 2020. The Varroa destructor mite is the greatest threat to US honeybees, and why things could get worse. Genet. Lit. Proj. URL https://geneticliteracyproject.org/2020/01/22/the-varroa-destructor-mite-is-the-greatest-threat-to-us-honeybees-and-why-things-could-get-worse/ (accessed 7.8.22).

Evans, J.D., Schwarz, R.S., 2011. Bees brought to their knees: microbes affecting honey bee health. Trends Microbiol. 19, 614–620.

Exelis, V.I.S., 2015. ENVI 5.3. Exelis VIS Boulder CO USA.

Fagerland, M.W., Sandvik, L., 2009. The Wilcoxon–Mann–Whitney test under scrutiny. Stat. Med. 28, 1487–1497. https://doi.org/10.1002/sim.3561

Farley, S.S., 2017. A General Framework for Predicting the Optimal Computing Configurations for Climate-Driven Ecological Forecasting Models (Thesis).

Fauvel, M., Chanussot, J., Benediktsson, J.A., 2008. Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel, in: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference On. IEEE, pp. 1884–1887.

Felton, A.J., Shriver, R.K., Bradford, J.B., Suding, K.N., Allred, B.W., Adler, P.B., 2021. Biotic versus abiotic controls on temporal sensitivity of primary production to precipitation across North American drylands. New Phytol. n/a. https://doi.org/10.1111/nph.17543

Fernández, M., Hamilton, H., 2015. Ecological Niche Transferability Using Invasive Species as a Case Study. PLoS ONE 10. https://doi.org/10.1371/journal.pone.0119891

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. null, 38–49. https://doi.org/null

Fierro, MM, Cruz-López, L., Sánchez, D., Villanueva-Gutiérrez, R., Vandame, R., 2012. Effect of Biotic Factors on the Spatial Distribution of Stingless Bees (Hymenoptera: Apidae, Meliponini) in Fragmented Neotropical Habitats. Neotrop. Entomol. 41, 95–104. https://doi.org/10.1007/s13744-011-0009-5

Fierro, Mm, Cruz-López, L., Sánchez, D., Villanueva-Gutiérrez, R., Vandame, R., 2012. Effect of Biotic Factors on the Spatial Distribution of Stingless Bees (Hymenoptera: Apidae, Meliponini) in Fragmented Neotropical Habitats. Neotrop. Entomol. 41, 95–104. https://doi.org/10.1007/s13744-011-0009-5

Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K., 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. Image Process. IEEE Trans. On 17, 1737–1754.

Fombong, A.T., Mumoki, F.N., Muli, E., Masiga, D.K., Arbogast, R.T., Teal, P.E.A., Torto, B., 2012. Occurrence, diversity and pattern of damage of Oplostomus species (Coleoptera: Scarabaeidae), honey bee pests in Kenya. Apidologie 44, 11–20. https://doi.org/10.1007/s13592-012-0149-6

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. Ann. Stat. 19, 1–67.

Fu, X., Wang, L., 2003. Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. Syst. Man Cybern. Part B Cybern. IEEE Trans. On 33, 399–409.

Gao, Y., Tian, L., 2021. Confidence interval estimation for sensitivity and difference between two sensitivities at a given specificity under tree ordering. Stat. Med. 40, 3695–3723. https://doi.org/10.1002/sim.8993

Garantonakis, N., Varikou, K., Birouraki, A., Edwards, M., Kalliakaki, V., Andrinopoulos, F., 2016. Comparing the pollination services of honey bees and wild bees in a watermelon field. Sci. Hortic. 204, 138–144. https://doi.org/10.1016/j.scienta.2016.04.006

Giannini, T.C., Boff, S., Cordeiro, G.D., Cartolano, E.A., Veiga, A.K., Imperatriz-Fonseca, V.L., Saraiva, A.M., 2015. Crop pollinators in Brazil: a review of reported interactions. Apidologie 46, 209–223. https://doi.org/10.1007/s13592-014-0316-z

Githui, F., Gitau, W., Mutua, F., Bauwens, W., 2009. Climate change impact on SWAT simulated streamflow in western Kenya. Int. J. Climatol. 29, 1823–1834. https://doi.org/10.1002/joc.1828

Giusti, M., 2017. Aethina tumida, new outbreak: the red zone expanded [WWW Document]. URL https://agronotizie.imagelinenetwork.com/zootecnia/2017/06/09/aethina-tumida-nuovo-focolaio-ampliata-la-zona-rossa/54432 (accessed 7.8.22).

Gomollón-Bel, F., 2020. Rare trehalulose sugar found in stingless bee's honey [WWW Document]. Chem. World. URL https://www.chemistryworld.com/news/rare-trehalulose-sugar-found-in-stingless-bees-honey/4012235.article (accessed 7.8.22).

Govender, M., Chetty, K., Naiken, V., Bulcock, H., 2008. A comparison of satellite hyperspectral and multispectral remote sensing imagery for improved classification and mapping of vegetation. Water SA 34, 147-154–154. https://doi.org/10.4314/wsa.v34i2.183634

Granitto, P.M., Furlanello, C., Biasioli, F., Gasperi, F., 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemom. Intell. Lab. Syst. 83, 83–90.

Grundel, R., Jean, R.P., Frohnapple, K.J., Glowacki, G.A., Scott, P.E., Pavlovic, N.B., 2010. Floral and nesting resources, habitat structure, and fire influence bee

distribution across an open-forest gradient. Ecol. Appl. 20, 1678–1692. https://doi.org/10.1890/08-1792.1
Guan, B., Guo, H., Chen, S., Li, D., Liu, X., Gong, X., Ge, G., 2020. Shifting ranges of eleven invasive alien plants in China in the face of climate change. Ecol. Inform. 55, 101024. https://doi.org/10.1016/j.ecoinf.2019.101024
Habel, J.C., Casanova, I.C.C., Zamora, C., Teucher, M., Hornetz, B., Shauri, H., Mulwa, R.K., Lens, L., 2017. East African coastal forest under pressure. Biodivers. Conserv. 26, 2751–2758.
Hallman, T.A., Robinson, W.D., 2020. Comparing multi- and single-scale species distribution and abundance models built with the boosted regression tree algorithm. Landsc. Ecol. 35, 1161–1174. https://doi.org/10.1007/s10980-020-01007-7
Hao, T., Elith, J., Guillera-Arroita, G., Lahoz-Monfort, J.J., 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Divers. Distrib. 25, 839–852. https://doi.org/10.1111/ddi.12892
Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. CRC Press.
Hedley, J., Roelfsema, C., Koetz, B., Phinn, S., 2012. Capability of the Sentinel 2 mission for tropical coral reef mapping and coral bleaching detection. Remote Sens. Environ. 120, 145–155.
Hempel de Ibarra, N., Vorobyev, M., Menzel, R., 2014. Mechanisms, functions and ecology of colour vision in the honeybee. J. Comp. Physiol. A 200, 411–433. https://doi.org/10.1007/s00359-014-0915-1
Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29, 773–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.
Hines, H.M., Hendrix, S.D., 2005. Bumble Bee (Hymenoptera: Apidae) Diversity and Abundance in Tallgrass Prairie Patches: Effects of Local and Landscape Floral Resources. Environ. Entomol. 34, 1477–1484. https://doi.org/10.1603/0046-225X-34.6.1477
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the Dimensionality of Data with Neural Networks. Science 313, 504–507. https://doi.org/10.1126/science.1127647
Holzschuh, A., Steffan-Dewenter, I., Kleijn, D., Tscharntke, T., 2007. Diversity of flower-visiting bees in cereal fields: effects of farming system, landscape composition and regional context. J. Appl. Ecol. 44, 41–49. https://doi.org/10.1111/j.1365-2664.2006.01259.x
Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. Remote Sens. 4, 2661–2693. https://doi.org/10.3390/rs4092661

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. Remote Sens. 8, 166. https://doi.org/10.3390/rs8030166

Iraheta, C.E.R., Martínez, M.Á.H., Romero, L.A.A., Álvarez, M.E.C., Arévalo, D.R., González, V.A.R., 2015. Stingless bee distribution and richness in El Salvador (Apidae, Meliponinae). J. Apic. Res. 54, 1–10. https://doi.org/10.1080/00218839.2015.1029783

Jackson, J.M., Pimsler, M.L., Oyen, K.J., Koch-Uhuad, J.B., Herndon, J.D., Strange, J.P., Dillon, M.E., Lozier, J.D., 2018. Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. Mol. Ecol. 27, 2926–2942. https://doi.org/10.1111/mec.14735

Jamali, S., Jönsson, P., Eklundh, L., Ardö, J., Seaquist, J., 2015. Detecting changes in vegetation trends using time series segmentation. Remote Sens. Environ. 156, 182–195. https://doi.org/10.1016/j.rse.2014.09.010

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe Version 4. Available CGIAR-CSI SRTM 90m Database Httpsrtm Csi Cgiar Org.

Jia, T., Barabási, A.-L., 2013. Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks. Sci. Rep. 3, 2354. https://doi.org/10.1038/srep02354

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., Lobo, J.M., 2011. Use of niche models in invasive species risk assessments. Biol. Invasions 13, 2785–2797.

Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845. https://doi.org/10.1016/j.cageo.2004.05.006

Jung-Rothenhäusler, F., Weichelt, H., Pach, M., 2007. RAPIDEYE – A NOVEL APPROACH TO SPACE BORNE GEO-INFORMATION SOLUTIONS 4.

Kajita, M.K., Aihara, K., Kobayashi, T.J., 2017. Balancing specificity, sensitivity, and speed of ligand discrimination by zero-order ultraspecificity. Phys. Rev. E 96, 012405. https://doi.org/10.1103/PhysRevE.96.012405

Kasina, J.M., Mburu, J., Kraemer, M., Holm-Mueller, K., 2009. Economic benefit of crop pollination by bees: a case of Kakamega small-holder farming in western Kenya. J. Econ. Entomol. 102, 467–473.

Kearney, M., Porter, W., 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecol. Lett. 12, 334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x

Kenya GIS Data [WWW Document], 2007. . World Resour. Inst. URL https://www.wri.org/resources/data-sets/kenya-gis-data (accessed 6.25.20).

Kiatoko, N., Pozo, M.I., Van Oystaeyen, A., Musonye, M., Kika, J., Wäckers, F., van Langevelde, F., Hundt, B., Jaramillo, J., 2022. African endemic stingless bees as an efficient alternative pollinator to honey bees in greenhouse cucumber

(Cucumis sativus L). J. Apic. Res. 0, 1–13. https://doi.org/10.1080/00218839.2021.2013421
Kiatoko, N., Raina, S.K., Muli, E., Mueke, J., 2014. Enhancement of fruit quality in Capsicum annum through pollination by Hypotrigona gribodoi in Kakamega, Western Kenya. Entomol. Sci. 17, 106–110. https://doi.org/10.1111/ens.12030
Kimathi, E., Tonnang, H.E.Z., Subramanian, S., Cressman, K., Abdel-Rahman, E.M., Tesfayohannes, M., Niassy, S., Torto, B., Dubois, T., Tanga, C.M., Kassie, M., Ekesi, S., Mwangi, D., Kelemu, S., 2020. Prediction of breeding regions for the desert locust Schistocerca gregaria in East Africa. Sci. Rep. 10, 11937. https://doi.org/10.1038/s41598-020-68895-2
Kinyanjui, M.J., 2011. NDVI-based vegetation monitoring in Mau forest complex, Kenya. Afr. J. Ecol. 49, 165–174. https://doi.org/10.1111/j.1365-2028.2010.01251.x
Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. Lond. B Biol. Sci. 274, 303–313. https://doi.org/10.1098/rspb.2006.3721
Kremen, C., Williams, N.M., Aizen, M.A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., Packer, L., Potts, S.G., Roulston, T., Steffan-Dewenter, I., Vázquez, D.P., Winfree, R., Adams, L., Crone, E.E., Greenleaf, S.S., Keitt, T.H., Klein, A.-M., Regetz, J., Ricketts, T.H., 2007. Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. Ecol. Lett. 10, 299–314. https://doi.org/10.1111/j.1461-0248.2007.01018.x
Kumar, M.S., Singh, A.J.A.R., Alagumuthu, G., Nguku E., 2012. Traditional beekeeping of stingless bee (Trigona sp) by Kani tribes of Western Ghats, Tamil Nadu, India. IJTK Vol112 April 2012.
Landmann, T., Piiroinen, R., Makori, D.M., Abdel-Rahman, E.M., Makau, S., Pellikka, P., Raina, S.K., 2015. Application of hyperspectral remote sensing for flower mapping in African savannas. Remote Sens. Environ. 166, 50–60. https://doi.org/10.1016/j.rse.2015.06.006
Latif, Z.A., Zamri, I., Omar, H., 2012. Determination of tree species using Worldview-2 data, in: Signal Processing and Its Applications (CSPA), 2012 IEEE 8th International Colloquium On. IEEE, pp. 383–387.
Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sens. Environ. 100, 356–362. https://doi.org/10.1016/j.rse.2005.10.014
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.
Libério, S.A., Pereira, A.L.A., Araújo, M.J.A.M., Dutra, R.P., Nascimento, F.R.F., Monteiro-Neto, V., Ribeiro, M.N.S., Gonçalves, A.G., Guerra, R.N.M., 2009. The potential use of propolis as a cariostatic agent and its actions on mutans

group streptococci. J. Ethnopharmacol. 125, 1–9. https://doi.org/10.1016/j.jep.2009.04.047
Lovett, J.C., 2015. Modelling the effects of climate change in Africa. Afr. J. Ecol. 53, 1–2. https://doi.org/10.1111/aje.12218
Lunau, K., Maier, E.J., 1995. Innate colour preferences of flower visitors. J. Comp. Physiol. A 177, 1–19. https://doi.org/10.1007/BF00243394
Madrigal, J., 2022. How to Photograph Bees - Nature TTL [WWW Document]. URL https://www.naturettl.com/how-to-photograph-bees/ (accessed 7.8.22).
Makaya, N.P., Mutanga, O., Kiala, Z., Dube, T., Seutloali, K.E., 2019. Assessing the potential of Sentinel-2 MSI sensor in detecting and mapping the spatial distribution of gullies in a communal grazing landscape. Phys. Chem. Earth Parts ABC, 18th WaterNet/WARFSA/GWPSA Symposium on Integrated Water Resources Development and Management: Innovative Technological Advances for Water Security in Eastern and Southern Africa - Part B 112, 66–74. https://doi.org/10.1016/j.pce.2019.02.001
Makori, D.M., Abdel-Rahman, E.M., Ndungu, N., Odindi, J., Mutanga, O., Landmann, T., Tonnang, H.E.Z., Kiatoko, N., 2022. The use of multisource spatial data for determining the proliferation of stingless bees in Kenya. GIScience Remote Sens. 59, 648–669. https://doi.org/10.1080/15481603.2022.2049536
Makori, D.M., Fombong, A.T., Abdel-Rahman, E.M., Nkoba, K., Ongus, J., Irungu, J., Mosomtai, G., Makau, S., Mutanga, O., Odindi, J., Raina, S., Landmann, T., 2017. Predicting Spatial Distribution of Key Honeybee Pests in Kenya Using Remotely Sensed and Bioclimatic Variables: Key Honeybee Pests Distribution Models. ISPRS Int. J. Geo-Inf. 6, 66. https://doi.org/10.3390/ijgi6030066
Maney, C., Sassen, M., Hill, S.L.L., 2022. Modelling biodiversity responses to land use in areas of cocoa cultivation. Agric. Ecosyst. Environ. 324, 107712. https://doi.org/10.1016/j.agee.2021.107712
Mani, M.S., 2013. Ecology and Biogeography of High Altitude Insects. Springer Science & Business Media.
Marshall, V., Lewis, M., Ostendorf, B., 2012. Do additional bands (coastal, nir-2, red-edge and yellow) in worldview-2 multispectral imagery improve discrimination of an invasive tussock, buffel grass (*cenchrus ciliaris*)? ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXIX-B8, 277–281. https://doi.org/10.5194/isprsarchives-XXXIX-B8-277-2012
Martin, S.J., Highfield, A.C., Brettell, L., Villalobos, E.M., Budge, G.E., Powell, M., Nikaido, S., Schroeder, D.C., 2012. Global honey bee viral landscape altered by a parasitic mite. Science 336, 1304–1306.
Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Michener, C.D., 2013. The Meliponini, in: Vit, P., Pedro, S.R.M., Roubik, D. (Eds.), Pot-Honey: A Legacy of Stingless Bees. Springer, New York, NY, pp. 3–17. https://doi.org/10.1007/978-1-4614-4960-7_1
Michener, C.D., 2000. The bees of the world. JHU Press.
Miller, T.E.X., Louda, S.M., Rose, K.A., Eckberg, J.O., 2009. Impacts of insect herbivory on cactus population dynamics: experimental demography across an environmental gradient. Ecol. Monogr. 79, 155–172. https://doi.org/10.1890/07-1550.1
Mitchell, J.J., Glenn, N.F., Sankey, T.T., Derryberry, D.R., Germino, M.J., 2012. Remote sensing of sagebrush canopy nitrogen. Remote Sens. Environ. 124, 217–223.
Mohajan, H.K., 2014. Food and Nutrition Scenario of Kenya. Am. J. Food Nutr. 12.
Mohammadi, S., Ebrahimi, E., Shahriari Moghadam, M., Bosso, L., 2019. Modelling current and future potential distributions of two desert jerboas under climate change in Iran. Ecol. Inform. 52, 7–13. https://doi.org/10.1016/j.ecoinf.2019.04.003
Mokaya, H.O., Nkoba, K., Ndunda, R.M., Vereecken, N.J., 2021. Characterization of honeys produced by sympatric species of Afrotropical stingless bees (Hymenoptera, Meliponini). Food Chem. 366, 130597. https://doi.org/10.1016/j.foodchem.2021.130597
Morawetz, L., Svoboda, A., Spaethe, J., Dyer, A.G., 2013. Blue colour preference in honeybees distracts visual attention for learning closed shapes. J. Comp. Physiol. A 199, 817–827. https://doi.org/10.1007/s00359-013-0843-5
Mudereri, Abdel-Rahman, E.M., Dube, T., Landmann, T., Khan, Z., Kimathi, E., Owino, R., Niassy, S., 2020a. Multi-source spatial data-based invasion risk modeling of Striga (Striga asiatica) in Zimbabwe. GIScience Remote Sens. 57, 553–571. https://doi.org/10.1080/15481603.2020.1744250
Mudereri, Mukanga, C., Mupfiga, E.T., Gwatirisa, C., Kimathi, E., Chitata, T., 2020b. Analysis of potentially suitable habitat within migration connections of an intra-African migrant-the Blue Swallow (Hirundo atrocaerulea). Ecol. Inform. 57, 101082. https://doi.org/10.1016/j.ecoinf.2020.101082
Muli, E., Patch, H., Frazier, M., Frazier, J., Torto, B., Baumgarten, T., Kilonzo, J., Kimani, J.N., Mumoki, F., Masiga, D., Tumlinson, J., Grozinger, C., 2014a. Evaluation of the Distribution and Impacts of Parasites, Pathogens, and Pesticides on Honey Bee (Apis mellifera) Populations in East Africa. PLoS ONE 9, e94459. https://doi.org/10.1371/journal.pone.0094459
Muli, E., Patch, H., Frazier, M., Frazier, J., Torto, B., Baumgarten, T., Kilonzo, J., Kimani, J.N., Mumoki, F., Masiga, D., Tumlinson, J., Grozinger, C., 2014b. Evaluation of the Distribution and Impacts of Parasites, Pathogens, and Pesticides on Honey Bee (Apis mellifera) Populations in East Africa. PLOS ONE 9, e94459. https://doi.org/10.1371/journal.pone.0094459
Muli, E., Patch, H., Frazier, M., Frazier, J., Torto, B., Baumgarten, T., Kilonzo, J., Kimani, J.N., Mumoki, F., Masiga, D., Tumlinson, J., Grozinger, C., 2014c.

Evaluation of the Distribution and Impacts of Parasites, Pathogens, and Pesticides on Honey Bee (Apis mellifera) Populations in East Africa. PLOS ONE 9, e94459. https://doi.org/10.1371/journal.pone.0094459

Mumoki, F. n., Fombong, A., Muli, E., Muigai, A. w. t., Masiga, D., 2014. An Inventory of Documented Diseases of African Honeybees. Afr. Entomol. 22, 473–487. https://doi.org/10.4001/003.022.0313

Muok, B.O., Owuor, B., Dawson, I., Were, J., 2000. The potential of indigenous fruit trees: results of a survey in Kitui District, Kenya. Agrofor. Today 12, 13–16.

Mwalusepo, S., Tonnang, H.E.Z., Massawe, E.S., Okuku, G.O., Khadioli, N., Johansson, T., Calatayud, P.-A., Le Ru, B.P., 2015. Predicting the Impact of Temperature Change on the Future Distribution of Maize Stem Borers and Their Natural Enemies along East African Mountain Gradients Using Phenology Models. PLoS ONE 10, e0130427. https://doi.org/10.1371/journal.pone.0130427

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39, 368–375. https://doi.org/10.1111/ecog.01881

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography 37, 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Naughton, D., Brunn, A., Czapla-Myers, J.S., Douglass, S., Thiele, M., Weichelt, H., Oxfort, M., 2011. Absolute radiometric calibration of the RapidEye multispectral imager using the reflectance-based vicarious calibration method. J. Appl. Remote Sens. 5, 053544. https://doi.org/10.1117/1.3613950

Ndungu, N.N., Yusuf, A.A., Raina, S.K., Masiga, D.K., Pirk, C.W.W., Nkoba, K., 2019. Nest architecture as a tool for species discrimination of Hypotrigona species (Hymenoptera: Apidae: Meliponini). Afr. Entomol. 27, 25–35.

Neuhäuser, M., Ruxton, G.D., 2009. Distribution-free two-sample comparisons in the case of heterogeneous variances. Behav. Ecol. Sociobiol. 63, 617–623. https://doi.org/10.1007/s00265-008-0683-4

Neumann, P., Ellis, J.D., 2008. The small hive beetle (Aethina tumida Murray, Coleoptera: Nitidulidae): distribution, biology and control of an invasive species. J. Apic. Res. 47, 181–183.

Neumann, P., Pettis, J.S., Schäfer, M.O., 2016. Quo vadis Aethina tumida? Biology and control of small hive beetles. Apidologie 47, 427–466. https://doi.org/10.1007/s13592-016-0426-x

Neumann, P., Pirk, C., Hepburn, H., Solbrig, A., Ratnieks, F., Elzen, P., Baxter, J., 2001. Social encapsulation of beetle parasites by Cape honeybee colonies (Apis mellifera capensis Esch.). Naturwissenschaften 88, 214–216. https://doi.org/10.1007/s001140100224

Newbold, T., 2018. Future effects of climate and land-use change on terrestrial vertebrate community diversity under different scenarios. Proc. R. Soc. B Biol. Sci. 285, 20180792. https://doi.org/10.1098/rspb.2018.0792

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverria-Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L.P., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520, 45–50. https://doi.org/10.1038/nature14324

Njoya, M.T.M., 2010. Diversity of Stingless Bees in Bamenda Afromontane Forests – Cameroon. Nest architecture, Behaviour and Labour calendar.

Nkoba, K., Raina, S.K., Muli, E., Mithöfer, K., Mueke, J., 2012. Species richness and nest dispersion of some tropical meliponine bees (Apidae: Meliponinae) in six habitat types in the Kakamega forest, western Kenya. Int. J. Trop. Insect Sci. 32, 194–202. https://doi.org/10.1017/S1742758412000355

Ochieng, J., Kirimi, L., Mathenge, M., Amolo, 2016. Effects of climate variability and change on agricultural production: The case of small scale farmers in Kenya. NJAS - Wagening. J. Life Sci., Social science perspectives on the bio-economy 77, 71–78. https://doi.org/10.1016/j.njas.2016.03.005

Odawa, S., Seo, Y., 2019. Water Tower Ecosystems under the Influence of Land Cover Change and Population Growth: Focus on Mau Water Tower in Kenya. Sustainability 11, 3524. https://doi.org/10.3390/su11133524

Oladimeji, Y., 2018. ANALYSIS OF HONEYBEE PRODUCTION AMONG WOMEN BEEKEEPERS IN KADUNA STATE, NIGERIA.

Ongus, J.R., Irungu, J., Raina, S., 2017. Correlation between Pest Abundance and Prevalence of Honeybee Pathogens at Selected Apiaries in Kenya, 2013/2014 10.

Pachauri, R.K., Mayer, L., 2015. Climate change 2014: synthesis report. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Pau, S., Gillespie, T.W., Wolkovich, E.M., 2012. Dissecting NDVI–species richness relationships in Hawaiian dry forests. J. Biogeogr. 39, 1678–1686. https://doi.org/10.1111/j.1365-2699.2012.02731.x

Pernal, S.F., Currie, R.W., 2001. The influence of pollen quality on foraging behavior in honeybees (Apis mellifera L.). Behav. Ecol. Sociobiol. 51, 53–68. https://doi.org/10.1007/s002650100412

Peters, J., Baets, B.D., Verhoest, N.E.C., Samson, R., Degroeve, S., Becker, P.D., Huybrechts, W., 2007. Random forests as a tool for ecohydrological distribution modelling. Ecol. Model. 207, 304–318. https://doi.org/10.1016/j.ecolmodel.2007.05.011

Peterson, A.T., 2011. Ecological niches and geographic distributions (MPB-49). Princeton University Press.

Peterson, A.T., Ball, L.G., Cohoon, K.P., 2002. Predicting distributions of Mexican birds using ecological niche modelling methods. Ibis 144, E27–E32. https://doi.org/10.1046/j.0019-1019.2001.00031.x

Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecol. Model. 117, 159–164. https://doi.org/10.1016/S0304-3800(99)00023-X

Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. Glob. Ecol. Biogeogr. 17, 135–144. https://doi.org/10.1111/j.1466-8238.2007.00347.x

Peterson, A.T., Papeş, M., Eaton, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30, 550–560.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197. https://doi.org/10.1890/07-2153.1

Pickering, C.M., Stock, M., 2003. Insect colour preference compared to flower colours in the Australian Alps. Nord. J. Bot. 23, 217–223. https://doi.org/10.1111/j.1756-1051.2003.tb00384.x

Pirk, C.W.W., Strauss, U., Yusuf, A.A., Démares, F., Human, H., 2016. Honeybee health in Africa—a review. Apidologie 47, 276–300. https://doi.org/10.1007/s13592-015-0406-6

Pirk, C.W.W., Strauss, U., Yusuf, A.A., Démares, F., Human, H., 2015. Honeybee health in Africa—a review. Apidologie 1–25. https://doi.org/10.1007/s13592-015-0406-6

Plant, R.E., 2012. Spatial Data Analysis in Ecology and Agriculture Using R. CRC Press.

Platts, P.J., Omeny, P.A., Marchant, R., 2015. AFRICLIM: high-resolution climate projections for ecological applications in Africa. Afr. J. Ecol. 53, 103–108. https://doi.org/10.1111/aje.12180

Pohl, N.B., Van Wyk, J., Campbell, D.R., 2011. Butterflies show flower colour preferences but not constancy in foraging at four plant species. Ecol. Entomol. 36, 290–300. https://doi.org/10.1111/j.1365-2311.2011.01271.x

Pontius, R.G.Jr., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J. Remote Sens. 32, 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol. Evol. 25, 345–353. https://doi.org/10.1016/j.tree.2010.01.007

Potts, S.G., Petanidou, T., Roberts, S., O'Toole, C., Hulbert, A., Willmer, P., 2006. Plant-pollinator biodiversity and pollination services in a complex Mediterranean landscape. Biol. Conserv. 129, 519–529. https://doi.org/10.1016/j.biocon.2005.11.019

QGIS Development Team, 2021. Quantum Geographical Information System [WWW Document]. URL https://qgis.org/en/site/ (accessed 5.10.21).

QGIS Development Team, 2020. Quantum Geographical Information System [WWW Document]. URL https://qgis.org/en/site/ (accessed 4.2.20).

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vigorous Calisthenics.

Rabus, B., Eineder, M., Roth, A., Bamler, R., 2003. The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. ISPRS J. Photogramm. Remote Sens. 57, 241–262. https://doi.org/10.1016/S0924-2716(02)00124-7

Raes, N., ter Steege, H., 2007. A null-model for significance testing of presence-only species distribution models. Ecography 30, 727–736. https://doi.org/10.1111/j.2007.0906-7590.05041.x

Rai, V.L., Singh, R.P., Singh, D.K., 2021. BEEKEEPING: ADDITIONAL SOURCE OF INCOME FOR FARMERS OF UTTAR PRADESH 3, 6.

Raina, S. k., Kimbu, D. m., 2005. Variations in races of the honeybee Apis mellifera (Hymenoptera: Apidae) in Kenya. Int. J. Trop. Insect Sci. 25, 281–291. https://doi.org/10.1079/IJT200588

Raina, S.K., Kioko, E., Zethner, O., Wren, S., 2011. Forest Habitat Conservation in Africa Using Commercially Important Insects. Annu. Rev. Entomol. 56, 465–485. https://doi.org/10.1146/annurev-ento-120709-144805

Rapacciuolo, G., Roy, D.B., Gillings, S., Purvis, A., 2014. Temporal validation plots: quantifying how well correlative species distribution models predict species' range changes over time. Methods Ecol. Evol. 5, 407–420. https://doi.org/10.1111/2041-210X.12181

Richard, K., Abdel-Rahman, E.M., Mohamed, S.A., Ekesi, S., Borgemeister, C., Landmann, T., 2018. Importance of Remotely-Sensed Vegetation Variables for Predicting the Spatial Distribution of African Citrus Triozid (Trioza erytreae) in Kenya. ISPRS Int. J. Geo-Inf. 7, 429. https://doi.org/10.3390/ijgi7110429

Ricketts, T.H., Regetz, J., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Bogdanski, A., Gemmill-Herren, B., Greenleaf, S.S., Klein, A.M., Mayfield, M.M., Morandin, L.A., Ochieng', A., Viana, B.F., 2008. Landscape effects on crop pollination services: are there general patterns? Ecol. Lett. 11, 499–515. https://doi.org/10.1111/j.1461-0248.2008.01157.x

Rodríguez-Castañeda, G., Hof, A.R., Jansson, R., Harding, L.E., 2012. Predicting the Fate of Biodiversity Using Species' Distribution Models: Enhancing Model

Comparability and Repeatability. PLOS ONE 7, e44402. https://doi.org/10.1371/journal.pone.0044402
Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C., 2012. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens. Environ. 121, 93–107.
Roulston, T.H., Goodell, K., 2011. The Role of Resources and Risks in Regulating Wild Bee Populations. Annu. Rev. Entomol. 56, 293–312. https://doi.org/10.1146/annurev-ento-120709-144802
Saatchi, S., Buermann, W., Ter Steege, H., Mori, S., Smith, T.B., 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. Remote Sens. Environ. 112, 2000–2017.
Sande, S.O., Crewe, R.M., Raina, S.K., Nicolson, S.W., Gordon, I., 2009. Proximity to a forest leads to higher honey yield: Another reason to conserve. Biol. Conserv. 142, 2703–2709. https://doi.org/10.1016/j.biocon.2009.06.023
Sarhrouni, El., Hammouch, A., Aboutajdine, D., 2012. Dimensionality reduction and classification feature using mutual information applied to hyperspectral images: a filter strategy based algorithm. ArXiv Prepr. ArXiv12100052.
Schmidt, K.S., Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. Remote Sens. Environ. 85, 92–108. https://doi.org/10.1016/S0034-4257(02)00196-7
Schroeder, C., Onyango, T.K., Nar, R.B., Jick, N.A., Parzies, H.K., Gemenet, D.C., 2013. Potentials of hybrid maize varieties for small-holder farmers in Kenya: a review based on Swot analysis. Afr. J. Food Agric. Nutr. Dev. 13. https://doi.org/10.4314/ajfand.v13i2
Schürmann, A., Kleemann, J., Fürst, C., Teucher, M., 2020. Assessing the relationship between land tenure issues and land cover changes around the Arabuko Sokoke Forest in Kenya. Land Use Policy 95, 104625.
Shen, M., Chen, J., Zhu, X., Tang, Y., Chen, X., 2010. Do flowers affect biomass estimate accuracy from NDVI and EVI? Int. J. Remote Sens. 31, 2139–2149. https://doi.org/10.1080/01431160903578812
Shen, Y., Wu, L., Di, L., Yu, Genong, Tang, H., Yu, Guoxian, Shao, Y., 2013. Hidden Markov models for real-time estimation of corn progress stages using MODIS and meteorological data. Remote Sens. 5, 1734–1753.
Simone-Finstrom, M.D., Spivak, M., 2012. Increased Resin Collection after Parasite Challenge: A Case of Self-Medication in Honey Bees? PLOS ONE 7, e34601. https://doi.org/10.1371/journal.pone.0034601
Slaa, E.J., Chaves, L.A.S., Malagodi-Braga, K.S., Hofstede, F.E., 2006. Stingless bees in applied pollination: practice and perspectives. Apidologie 37, 293–315. https://doi.org/10.1051/apido:2006022
Speranza, C.I., Kiteme, B., Ambenje, P., Wiesmann, U., Makali, S., 2009. Indigenous knowledge related to climate variability and change: insights from droughts

in semi-arid areas of former Makueni District, Kenya. Clim. Change 100, 295–315. https://doi.org/10.1007/s10584-009-9713-0
Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X
Stratoulias, D., Balzter, H., Sykioti, O., Zlinszky, A., Tóth, V.R., 2015. Evaluating Sentinel-2 for Lakeshore Habitat Mapping Based on Airborne Hyperspectral Data. Sensors 15, 22956–22969. https://doi.org/10.3390/s150922956
Strauss, U., Human, H., Gauthier, L., Crewe, R.M., Dietemann, V., Pirk, C.W.W., 2013. Seasonal prevalence of pathogens and parasites in the savannah honeybee (Apis mellifera scutellata). J. Invertebr. Pathol. 114, 45–52. https://doi.org/10.1016/j.jip.2013.05.003
Strebel, N., Kéry, M., Guélat, J., Sattler, T., 2022. Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model. J. Biogeogr. 49, 563–575. https://doi.org/10.1111/jbi.14335
Suso, M.J., Bebeli, P.J., Christmann, S., Mateus, C., Negri, V., Pinheiro de Carvalho, M.A.A., Torricelli, R., Veloso, M.M., 2016. Enhancing Legume Ecosystem Services through an Understanding of Plant–Pollinator Interplay. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.00333
Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293. https://doi.org/10.1126/science.3287615
Therneau, T.M., Atkinson, B., Ripley, M.B., 2010. The rpart package. Oxford, UK.
Thulin, S., Hill, M.J., Held, A., Jones, S., Woodgate, P., 2012. Hyperspectral determination of feed quality constituents in temperate pastures: Effect of processing methods on predictive relationships from partial least squares regression. Int. J. Appl. Earth Obs. Geoinformation 19, 322–334.
Torné-Noguera, A., Rodrigo, A., Arnan, X., Osorio, S., Barril-Graells, H., da Rocha-Filho, L.C., Bosch, J., 2014. Determinants of Spatial Distribution in a Bee Community: Nesting Resources, Flower Resources, and Body Size. PLoS ONE 9. https://doi.org/10.1371/journal.pone.0097255
Torto, B., Fombong, A.T., Mutyambai, D.M., Muli, E., Arbogast, R.T., Teal, P.E.A., 2010. Aethina tumida (Coleoptera: Nitidulidae) and Oplostomus haroldi (Coleoptera: Scarabaeidae): Occurrence in Kenya, Distribution within Honey Bee Colonies, and Responses to Host Odors. Ann. Entomol. Soc. Am. 103, 389–396. https://doi.org/10.1603/AN09136
Ushimaru, A., Watanabe, T., Nakata, K., 2007. Colored floral organs influence pollinator behavior and pollen transfer in Commelina communis (Commelinaceae). Am. J. Bot. 94, 249–258. https://doi.org/10.3732/ajb.94.2.249
Vanbrabant, Y., Delalieux, S., Tits, L., Pauly, K., Vandermaesen, J., Somers, B., 2020. Pear Flower Cluster Quantification Using RGB Drone Imagery. Agronomy 10, 407. https://doi.org/10.3390/agronomy10030407
Waldock, C., Stuart-Smith, R.D., Albouy, C., Cheung, W.W.L., Edgar, G.J., Mouillot, D., Tjiputra, J., Pellissier, L., 2022. A quantitative review of abundance-based

species distribution models. Ecography 2022. https://doi.org/10.1111/ecog.05694
Wang, A., Chen, J., Jing, C., Ye, G., Wu, J., Huang, Z., Zhou, C., 2015. Monitoring the Invasion of Spartina alterniflora from 1993 to 2014 with Landsat TM and SPOT 6 Satellite Data in Yueqing Bay, China. PLOS ONE 10, e0135538. https://doi.org/10.1371/journal.pone.0135538
Ward, G., Hastie, T., Barry, S., Elith, J., Leathwick, J.R., 2009. Presence-Only Data and the EM Algorithm. Biometrics 65, 554–563. https://doi.org/10.1111/j.1541-0420.2008.01116.x
Warrens, M.J., 2015a. Properties of the quantity disagreement and the allocation disagreement. Int. J. Remote Sens. 36, 1439–1446. https://doi.org/10.1080/01431161.2015.1011794
Warrens, M.J., 2015b. Relative quantity and allocation disagreement measures for category-level accuracy assessment. Int. J. Remote Sens. 36, 5959–5969. https://doi.org/10.1080/01431161.2015.1110265
Warui, M.W., Gikungu, M., Bosselmann, A.S., Hansted, L., 2018a. Pollination of Acacia woodlands and honey production by honey bees in Kitui, Kenya. Future Food J. Food Agric. Soc. 6, 40–50.
Warui, M.W., Gikungu, M., Bosselmann, A.S., Hansted, L., 2018b. Pollination of Acacia woodlands and honey production by honey bees in Kitui, Kenya. Future Food J. Food Agric. Soc. 6, 40–50.
Waske, B., Benediktsson, J.A., Árnason, K., Sveinsson, J.R., 2009. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. Can. J. Remote Sens. 35, S106–S116. https://doi.org/10.5589/m09-018
Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R., Stewart, A.M., 2003. Niche modeling perspective on geographic range predictions in the marine environment using a machine-learning algorithm.
Wilkins, S., Brown, M.A., Cuthbertson, A.G., 2007. The incidence of honey bee pests and diseases in England and Wales. Pest Manag. Sci. 63, 1062–1068. https://doi.org/10.1002/ps.1461
Williams, G.R., Tarpy, D.R., 2010. Colony collapse disorder in context. Bioessays 32, 845.
Williams, G.R., Tarpy, D.R., Vanengelsdorp, D., Chauzat, M.-P., Cox-Foster, D.L., Delaplane, K.S., Neumann, P., Pettis, J.S., Rogers, R.E., Shutler, D., 2010. Colony collapse disorder in context. Bioessays 32, 845–846.
Williams, N.M., Regetz, J., Kremen, C., 2012. Landscape-scale resources promote colony growth but not reproductive performance of bumble bees. Ecology 93, 1049–1058. https://doi.org/10.1890/11-1006.1
Winfree, R., Aguilar, R., Vázquez, D.P., LeBuhn, G., Aizen, M.A., 2009. A meta-analysis of bees' responses to anthropogenic disturbance. Ecology 90, 2068–2076. https://doi.org/10.1890/08-1245.1

Winfree, R., Williams, N.M., Dushoff, J., Kremen, C., 2007. Native bees provide insurance against ongoing honey bee losses. Ecol. Lett. 10, 1105–1113. https://doi.org/10.1111/j.1461-0248.2007.01110.x

Winfree, R., Williams, N.M., Gaines, H., Ascher, J.S., Kremen, C., 2008. Wild bee pollinators provide the majority of crop visitation across land-use gradients in New Jersey and Pennsylvania, USA. J. Appl. Ecol. 45, 793–802. https://doi.org/10.1111/j.1365-2664.2007.01418.x

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Wright, M.N., Wager, S., Probst, P., Wright, M.M.N., 2018. Package 'ranger.'

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol. Evol. 4, 236–243. https://doi.org/10.1111/2041-210x.12004

Yan, K., Zhang, D., 2015. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens. Actuators B Chem. 212, 353–363.

Yin, J., Wang, Y., Hu, J., 2012. A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy. IEEE Trans. Ind. Inform. 8, 935–943.

Zayed, A., 2009a. Bee genetics and conservation. Apidologie 40, 237–262. https://doi.org/10.1051/apido/2009026

Zayed, A., 2009b. Bee genetics and conservation. Apidologie 40, 237–262. https://doi.org/10.1051/apido/2009026

Zhang, J., Han, C., Liu, Z., 2009. Absorption spectrum estimating rice chlorophyll concentration: Preliminary investigations.

Zheng, Q., Huang, W., Cui, X., Shi, Y., Liu, L., 2018. New Spectral Index for Detecting Wheat Yellow Rust Using Sentinel-2 Multispectral Imagery. Sensors 18, 868. https://doi.org/10.3390/s18030868

Zulian, G., Maes, J., Paracchini, M.L., 2013. Linking Land Cover Data and Crop Yields for Mapping and Assessment of Pollination Services in Europe. Land 2, 472–492. https://doi.org/10.3390/land2030472