The use of multisource spatial data for determining the proliferation of stingless bees in Kenya

David Masereti Makori*, Elfatih M. Abdel-Rahman*, Nelly Ndungu*, John Odindi*, Onisimo Mutanga*, Tobias Landmann*, Henri E. Z. Tonnang* and Nkoba Kiatoko*

*Data Management, Modelling, and Geoinformation (DMMG) Unit, International Centre of Insect Physiology and Ecology (ICIPE), Nairobi, Kenya; †Discipline of Geography and Environmental Science, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

ABSTRACT
Stingless/meliponine bees are eusocial insects whose polyolactic nature enables interaction with a wide variety of wild plants and crops that enhance pollination and, 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% logit), 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 was 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 partitioning 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.

1 Introduction
Stingless bees, also known as meliponine bees, are eusocial insects that comprised hundreds of highly diverse and abundant species (Michener 2000; Iraheta et al. 2015). 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 Kwaapong 2013; Michener 2013), with over 20 species documented and endemic to the African continent and 12 species documented in Kenya (Michener 2000; Eardley 2004). The polyolactic nature of meliponine bees is of particular importance with regard to pollination services they provide to numerous crops and wild plants (Slaa et al. 2006; Njoya 2010). They are known to supplement honeybees (*Apis mellifera*), which have over the years been the major pollinators of wild plants and crops (Winfree et al. 2007; Giannini et al. 2015; Garantonakis et al. 2016). 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 (Muli et al. 2014; Pirk et al. 2015;
Makori et al. 2017). 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 (Slaa et al. 2006; Kiatoko et al. 2014, 2022); hence, the interest in meliponiculture has substantially grown around the African continent.

Moreover, livelihoods of many rural African communities have improved due to the production of high-quality meliponine bee honey and other beekeeping products (Cortopassi-Laurino et al. 2006; Kiatoko et al. 2014) used as food, medicine, and a source of income. Stingless bee honey has anti-biotic 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 antioxidants, immunomodulatory effect, antibacterial, and anti-inflammatory properties that are important for medical purposes (Libério et al. 2009; Araujo et al. 2012).

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 proliferation of stingless bees remain largely undocumented (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, Regetz, and Kremen 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 (Roulston and Goodell 2011; Fierro et al. 2012). 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 factor variations on stingless bees distribution, artificial intelligence (AI), machine learning (ML), and ecological niche (EN) models are commonly used to provide a statistical pathway that influence their proliferation in space and time (Kearney and Porter 2009; Fernández, Hamilton, and Liu 2015). 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 modeling approaches. Therefore, response and predictor variables used in these models should be carefully selected to increase prediction accuracy (Phillips and Dudik 2008; Peterson 2011). In this context, overfitting of EN models should be avoided to achieve reliable prediction of species’ spread (Cord et al. 2014). 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 (Pau, Gillespie, and Wolkovich 2012; Cord et al. 2014).

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. 2014; Naimi and Araújo 2016). In addition, since future bioclimatic conditions are already modeled 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 proliferation and biodiversity loss. Therefore, this study sought to establish the influence of bioclimatic,
topographic, and vegetation seasonality variables on the spatial proliferation of stingless bees in Kenya. This study also modeled the future distribution scenarios and patterns of stingless bees using simulated bioclimatic data.

2 Methods

2.1 Study site

The study area spans four agroecological zones with a representative climatic gradient comprising 40 administrative counties in Kenya. The area covers the coastal, eastern, Mount Kenya, Nyanza, Rift Valley, and western regions of the country, totaling 256,638.49 square kilometers (Figure 1). The stingless bee occurrence data were collected from 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 m above sea level at the coastal region to 2,045 m 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 (Githii et al. 2009; Schürmann et al. 2020).

### 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 | Description | Units |
|-----------|-------------|-------|
| **Temperature variables (n = 10)** | | |
| Bio 1 | mean annual temperature | °C |
| Bio 2 | mean diurnal range in temperature isothermality | °C |
| Bio 3 | isothermality | °C |
| Bio 4 | temperature seasonality | °C |
| Bio 5 | maximum temperature warmest month | °C |
| Bio 6 | minimum temperature coolest month | °C |
| Bio 7 | annual temperature range | °C |
| Bio 8 | mean temperature warmest quarter | °C |
| Bio 9 | mean temperature coolest quarter | °C |
| **Potential evapotranspiration** | potential evapotranspiration | mm |
| **Precipitation variables (n = 11)** | | |
| Bio 10 | mean annual rainfall | mm |
| Bio 11 | rainfall wettest month | mm |
| Bio 12 | rainfall driest month | mm |
| Bio 13 | rainfall seasonality | mm |
| Bio 14 | rainfall wettest quarter | mm |
| Bio 15 | rainfall driest quarter | mm |
| Moisture index | annual moisture index | n/a |
| Moisture index moist quarter | moisture index moist quarter | n/a |
| Moisture index arid quarter | moisture index arid quarter | n/a |
| Dry months | number of dry months | months |
| Length of the longest dry season | length of longest dry season | months |
| **Topographical variables (n = 6)** | | |
| Slope | angle of inclination to the horizontal | % rise |
| Aspect | slope direction | degrees |
| Hillshade | shading effect | n/a |
| Roughness | degree of surface irregularity | n/a |
| Topographical position index | position of surface inclination | n/a |
| Terrain ruggedness index | ground ruggedness | n/a |
| **Vegetation phenological variables (n = 13)** | | |
| Start of season time | time for the start of season | decades |
| End of season time | time for end of season | decades |
| Length of season | length of time from start to end of season | decades |
| Base value | average minimum EVI value | n/a |
| Time of middle of season | time of middle of season | decade |
| Maximum value | maximum value of fitted data | n/a |
| Amplitude | difference between maximum and base level | n/a |
| Left derivative | rate of increase of beginning of season | % |
| Right derivative | rate of decrease of end of season | % |
| Large integral | integral from season start to season end | n/a |
| Small integral | integral difference between season and base level | n/a |
| Start of season value | value at the start of season | n/a |
| End of season value | value at end of season | n/a |
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. (2020a) 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.

2.2.1 Response variables
As aforementioned, stingless bees’ occurrence data were collected from 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 modeling (sdm) package (Naimi and Araújo 2016) in R software (R. Core Team 2020) 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 were collected span across four main agroecological zones, which had a representative climatic gradient. Data on six stingless bee species were randomly collected whereby a meliponary with more than one domesticated colony placed in one location or a wild nest with a colony was 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 m (±3).

2.2.2 Predictor variables
2.2.2.1 Vegetation phenology. The 16-day enhanced vegetation index (EVI) observations acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) at 250-m spatial resolution was used to derive vegetation phenological variables (Table 1). EVI over 20-year period (from 2001 to 2020) was processed using TIMESAT software (Eklundh 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 toward 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 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 arcsec resolution (~90 m 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.

2.2.2.3 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 (Platts, Omeny, and Marchant 2015; Fick and Hijmans 2017) 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 CO2 emissions, representative concentration pathway scenario (RCPs) 4.5 watt/m² (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; Muderer et al. 2020b). For each of the timesteps, 21 bioclimatic variables (Table 1) were utilized. Specifically, 10 temperature and 11 precipitation variables were generated for each timestep.

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 modeling 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 (Plant 2012; Dormann et al. 2013). 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, and then iteratively those with high linear regression coefficients are eliminated.

This study first 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 (Plant 2012; Dormann et al. 2013; Araújo et al. 2019). Second, 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 Dormann et al. (2013) that VIF values of 10 or higher between variables indicate collinearity. The correlation matrix presented in Figure 2 indicated that some of the essential and relevant variables to the stingless bees’ distribution such as
isothermality and mean annual rainfall were highly correlated. Therefore, selection of such variables 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 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 (Wisz et al. 2008; Araújo et al. 2019). 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.

2.3 Model fitting

Machine learning EN models in the “sdm” package (Naimi and Araújo 2016) in the R software (R Core Team 2020) 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.

Using a good number of response variables to obtain acceptable prediction accuracies is paramount in EN modeling (Stockwell and Peterson 2002; Wisz et al. 2008). As aforementioned, it has been established that 90% prediction accuracies for machine learning algorithms can be obtained using as low as 10 response variable, but higher number (from 50 sample points) would ultimately yield better or near maximal accuracies (Wisz et al. 2008; Araújo et al. 2019). Given that the response variable consisted of over 10 sample points for each species, it was predicted that the EN models developed in this study would yield “acceptable” accuracies.

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 modeling 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 (Naimi and Araújo 2016; Mudereri et al. 2020a). 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 modeling methods based on the ranking of their predicted accuracy.

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 modeling approach was employed to harmonize the variations among the six EN algorithms. Furthermore, previous studies have demonstrated that ensemble modeling approach usually has a superior predictive power than individual modeling methods (Naimi and Araújo 2016; Araújo et al. 2019; Hao et al. 2019; Makaya et al. 2019). 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 modeling 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.

| S/N | Modeling 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, Leathwick, and Hastie (2008) |
| 6   | Recursive portioning and regression trees | “RPART” | Therneau, Atkinson, and Ripley (2010) |
“ensemble” function in the “sdm” package under the “R” environment was used to predict the spatial distribution of stingless bees in future.

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 (Naimi and Araújo 2016; Araújo et al. 2019), and 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, while 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 (Swets 1988; Mohammadi et al. 2019), which ranks the models’ performance in a way that (i) excellent ≥ 0.90, (ii) good = 0.80–0.89, (iii) fair = 0.70–0.79, (iv) poor = 0.60–0.69, and (v) fail ≤ 0.59.

As previously mentioned, ensemble modeling utilizes strengths from the individual models while minimizing their weaknesses (Araújo and New 2007; Araújo et al. 2019); hence, the 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, while values lower than zero (TSS ≤ 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) (Rapacciuolo et al. 2014; Araújo et al. 2019).

\[
TSS = (sensitivity + specificity) − 1 \quad (1)
\]

where sensitivity represents the presence observation and specificity is absence.

\[
Sensitivity = \frac{a}{a + b} \quad (2)
\]

where \(a\) is true positive presence and \(b\) is false negative presence.

\[
Specificity = \frac{d}{c + d} \quad (3)
\]

where \(c\) is false positive presence and \(d\) is true negative presence.

3 Results
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 collinear ones (Figure 4). In addition, selected bioclimatic variables had the highest relative influence (64.27%) on
Figure 3. Relative variable importance of all 40 variables (refer to Table 1 for more details on the variables) used in the current EN models for predicting the distribution of stingless bees. Mean annual rainfall (bio12), isothermality (bio3), slope, and temperature seasonality (bio4) poised as the most relevant before the highly correlated variables were eliminated (see Table 1 for full variables list).

Figure 4. Recursive feature elimination (RFE) model indicating that a few selected variables (14 or less) could yield comparable prediction model accuracy, compared to pooling predictor variables as revealed by the root mean square error (RMSE) of cross-validated predictor models.
the EN models developed to predict the stingless bee occurrence, while vegetation phenological and topographical predictor variables contributed to 21.36\% and 14.36\% logit, respectively (Figure 5).

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 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.

3.2 Stingless bees’ prediction models 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 and 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) for 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, indicating 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. Moreover, some individual replicate runs were demonstrated to be closer to the one-to-one

Table 3. AUC and TSS of the six EN modeling methods used to predict the distribution of stingless bees in Kenya. Both AUC and TSS indicate that random forest (RF) and RANGER have high accuracy, while RPART attained the least TSS accuracy.

| 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 the six EN modeling methods used to predict distribution of stingless bees. The red and blue curves indicate smoothed mean area under curve (AUC) of the training data and test data, respectively, with 10-fold cross validation sampling. Light blue/cyan curves indicate 10-fold replicate runs for individual models with training data. The black dotted line shows the one-to-one mid-point interaction between sensitivity and specificity for each model.
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. Moreover, presence–absence data of response variables presented limited dispersion from the mean, indicating good separability between classes and high integrity within classes.

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 (Bryan et al. 2010; Camberlin et al. 2014; Ochieng, Kirimi, and Mathenge 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 will be 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.

The change in map based on probability of stingless bee occurrence (Figure 9) revealed that many areas especially Kisii (in Nyanza region), Mt. Elgon

![Figure 9](image-url)  
*Figure 9.* Stingless bees distribution change map (c) indicating difference between the current and future probability of occurrence. Dark green and green colors indicate positive change, white denotes no change, while red and deep red denotes negative change.
4 Discussion

Stingless bees are important in fragmented landscapes due to their polylectic 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 proliferation of stingless bees is an important element in the determination of their future spatial distribution.

4.1 Variable interaction and EN model development

Despite stingless bees’ diverse and generalist foraging ability, climate and vegetation phenology influence their proliferation (Slaa et al. 2006; Njoya 2010). 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 (Fierro et al. 2012; Lovett 2015; Mwalusepo et al. 2015; Platts, Omeny, and Marchant 2015; Makori et al. 2017) that note a positive correlation between precipitation and the proliferation of stingless bees.

Table 4. Probability of change for stingless bees’ distribution maps between the current and future predicted areas. The classes indicate areas in square kilometers (km²) predicted to have diminishing and proliferation probability of change.

| 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 |

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’ proliferation increased in regions totaling 35,958.68 km², while a total of 99,363.70 km² was predicted to have diminishing probability of occurrence (Table 4) with unaltered predicted future climate change pathways. Most interclass variations were centered 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 km² in the study area.

The recursive feature elimination (RFE) model (Granitto et al. 2006; Yan and Zhang 2015; Darst, Malecki, and Engelman 2018) 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 determines niche colonization of stingless bees and facilitate their proliferation 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 (Torto et al. 2010; Pau, Gillespie, and Wolkovich 2012). 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 proliferation. 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 and Appendix S1). Specifically, suitable ENs spanned through regions with high altitudinal gradients across the study site (Appendix S2). Previous studies (Mani 2013; Jackson et al. 2018; Birrell et al. 2020; Kimathi et al. 2020) 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 proliferation 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.

### 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 vis-a-vis response variables (Aranda and Lobo, 2011; Naimi and Araújo 2016; Araújo et al. 2019). Also, they are largely dependent on the selection of important non-conflicting 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 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 (Kajita, Aihara, and Kobayashi 2017; Gao and Tian 2021). 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.

### 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 km², 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 (Platts, Omeny, and Marchant 2015; Fick and Hijmans 2017). 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 km², 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 (Williams et al. 2010; Dainat, Vanengelsdorp, and Neumann 2012).

The generalist behavior of stingless bees suggests that over time, they have adapted to shifting vegetation density and diversity (Slaa et al. 2006; Njoya 2010). 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 neighboring Arabuko Sokoke Forest in Kilifi county were predicted to be more suitable with high accuracy. Besides, their colonization rates were demonstrated to be higher than the neighboring Kwale County.

Moreover, 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 (Jackson et al. 2018; Birrell et al. 2020). 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’ proliferation as phenology and climate.

5 Conclusions

This study developed the first ever predictive EN models for the spatial distribution of stingless bees in Kenya using an ensemble modeling
approach of six machine learning EN algorithms. The polylactic nature of stingless bees is particularly important due to nonselective 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 modeling 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 toward 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 modeled climatic pathways by the IPCC indicated that there was more spatial diminishing of stingless bee suitable habitats than proliferation. 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 toward 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 Mudneri 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.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the European Commission [DCI-FOOD-2011/023-520]; Norwegian Agency for Development Cooperation-NORAD [RAF-3058 KEN-18/0005]; and Bayer CropScience AG [NC20662450].

ORCID

David Masereti Makori http://orcid.org/0000-0003-0849-778X
Data Availability Statement

The authors confirm that the data supporting the findings of this study are available within the article and any additional data are available from the corresponding author, upon reasonable request on davidmakori@gmail.com.

References

Ackerley, D. D., S. R. Loarie, W. K. Cornwell, S. B. Weiss, H. Hamilton, R. Branciforte, N. J. B. Kraft, et al. 2010. “The Geography of Climate Change: Implications for Conservation Biogeography.” Diversity & Distributions 16 (3): 476–487. doi:10.1111/j.1472-4642.2010.00654.x.

Aranda, A. C., and J. M. Lobo. 2011. “How Well Does Presence-only-based Species Distribution Modelling Predict Assemblage Diversity? A Case Study of the Tenerife Flora.” Ecography 34 (1): 31–38. doi:10.1111/j.1600-0587.2010.01634.x.

Araújo, M. B., R. P. Anderson, A. Márcia Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, et al. 2019. “Standards for Distribution Models in Biodiversity Assessments.” Science Advances 5 (1): eaat4858. doi:10.1126/sciadv.aat4858.

Araújo, M. A. R., S. A. Libério, R. N. M. Guerra, M. N. S. Ribeiro, F. R. F. Nascimento. 2012. “Mechanisms of Action Underlying the Anti-inflamatory and Immunomodulatory Effects of Propolis: A Brief Review.” Revista Brasileira de Farmacognosia 22 (1): 208–219. doi:10.1590/S0102-695X2011005000167.

Araújo, M. B., and M. New. 2007. “Ensemble Forecasting of Species Distributions.” Trends in Ecology & Evolution 22 (1): 42–47. doi:10.1016/j.tree.2006.09.010.

Bijlsma, L., L. L. M. de Bruijn, E. P. Martens, M. J. Sommeijer. 2006. “Water Content of Stingless Bee Honey (Apidae, Meliponini): Interspecific Variation and Comparison with Honey of Apis Mellifera.” Apidologie 37 (4): 480–486. doi:10.1051/apido:2006034.

Birrell, J. H., A. A. Shah, S. Hotaling, J. J. Giersch, C. E. Williamson, D. Jacobsen, H. A. Woods, et al. 2020. “Insects in High-elevation Streams: Life in Extreme Environments Imperiled by Climate Change.” Global Change Biology 26 (12): 6667–6684. doi:10.1111/gcb.15356.

Breiman, L. 2001. “Random Forests.” Machine Learning 45 (1): 5–32. doi:10.1023/A:1010933404324.

Bryan, E., Ringer, C., Okoba, B., Roncoli, C., Silvestri, S. and Herrera, M.T. 2010. Coping with Climate Variability and Adapting to Climate Change in Kenya: Household and Community Strategies and Determinants. Washington, DC: International Food Policy Research Institute (IFPRI).

Camberlin, P., J. Boyard-Micheau, N. Philippin, C. Baron, C. Leclerc, C. Mwongera. 2014. “Climatic Gradients along the Windward Slopes of Mount Kenya and Their Implication for Crop Risks. Part 1: Climate Variability.” International Journal of Climatology 34 (7): 2136–2152. doi:10.1002/joc.3427.

CGIAR-CSI (2020). “CGIAR-CSI SRTM – SRTM 90m DEM Digital Elevation Database “ (Accessed 8 November 2020).http://srtm.csi.cgiar.org/

Cord, A. F., D. Klein, D. S. Gernandt, J. A. P. de la Rosa, S. Dech. 2014. “Remote Sensing Data Can Improve Predictions of Species Richness by Stacked Species Distribution Models: A Case Study for Mexican Pines.” Journal of Biogeography 41 (4): 736–748. doi:10.1111/jbi.12225.

Core Team, R. 2020. R: A Language and Environment for Statistical Computing (Version 3.6.2).Computer software ed. Vienna, Austria: R Foundation for Statistical Computing.

Cortopassi-Laurino, M., V. L. Imperatriz-Fonseca, D. W. Roubik, A. Dollin, T. Heard, I. Aguilar, G. C. Venturieri, et al. 2006. “Global Meliponiculture: Challenges and Opportunities.” Apidologie 37 (2): 275–292. doi:10.1051/apido:2006027.

Dainat, B., D. Vanengelsdorp, and P. Neumann. 2012. “Colony Collapse Disorder in Europe.” Environmental Microbiology Reports 4 (1): 123–125. doi:10.1111/j.1758-2229.2011.00312.x.

Darst, B. F., K. C. Malecki, and C. D. Engelman. 2018. “Using Recursive Feature Elimination in Random Forest to Account for Correlated Variables in High Dimensional Data.” BMC Genetics 19 (1): 1–6. doi:10.1186/s12863-018-0633-8.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Car!, G. Carré, J. R. G. Marquéz, et al. 2013. “Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance.” Ecography 36 (1): 27–46. doi:10.1111/j.1600-0587.2012.07348.x.

Du, Z., Z. Wang, Y. Liu, H. Wang, F. Xue, Y. Liu. 2014. “Ecological Niche Modeling for Predicting the Potential Risk Areas of Severe Fever with Thrombocytopenia Syndrome”. International Journal of Infectious Diseases 26: 1–8. doi:10.1016/j.ijid.2014.04.006.

Eardley, C. D. 2004. “Taxonomic Revision of the African Stingless Bees (Apoidae: Apidae: Apinae: Meliponini).” African Plant Protection 10 (2): 63–96.

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

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

Elith, J., J. R. Leathwick, and T. Hastie. 2008. “A Working Guide to Boosted Regression Trees.” Journal of Animal Ecology 77 (4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Farley, S. S. (2017). “A General Framework for Predicting the Optimal Computing Configurations for Climate-Driven Ecological Forecasting Models.” Thesis. (Accessed 22 June 2021).https://minds.wisconsin.edu/handle/1793/78904

Felton, A. J., R. K. Shriver, J. B. Bradford, K. N. Suding, B. W. Allred, P. B. Adler. 2021. “Biotic versus Abiotic Controls on Temporal Sensitivity of Primary Production to Precipitation across North American Drylands.” New Phytologist 231 (6): 2150–2161. doi:10.1111/nph.17543n/a(n/a).
Fernández, M., H. Hamilton, and J. Liu. 2015. “Ecological Niche Transferability Using Invasive Species as a Case Study.” PloS ONE 10 (3): e0119891. doi:10.1371/journal.pone.0119891.

Fick, S. E., and R. J. Hijmans. 2017. “WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas.” International Journal of Climatology 37 (12): 4302–4315. doi:10.1002/joc.5086.

Fierro, M., L. Cruz-López, D. Sánchez, R. Villanueva-Gutiérrez, R. Vandame. 2012. “Effect of Biotic Factors on the Spatial Distribution of Stingless Bees (Hymenoptera: Apidae, Meliponini) in Fragmented Neotropical Habitats.” Neotropical Entomology 41 (2): 95–104. doi:10.1590/s1374-011-0009-5.

Friedman, J. H. 1991. “Multivariate Adaptive Regression Splines.” The Annals of Statistics 19 (1): 1–67.

Gao, Y., and L. Tian. 2021. “Confidence Interval Estimation for Sensitivity and Difference between Two Sensitivities at a Given Specificity under Tree Ordering.” Statistics in Medicine 40 (16): 3695–3723. doi:10.1002/sim.8993.

Garantonaki, N., K. Varikou, A. Biorouraki, M. Edwards, V. Kalliakatos, F. Andrinopoulos. 2016. “Comparing the Pollination Services of Honey Bees and Wild Bees in a Watermelon Field.” Scientia Horticulturae 204:138–144. doi:10.1016/j.scienta.2016.04.006.

Giannini, T.C., S. Boff, G. D. Cordeiro, E. A. Cartolano, A. K. Veiga, V. L. Imperatriz-Fonseca, A. M. Saraiva, et al. 2015. “Crop Pollinators in Brazil: A Review of Reported Interactions.” Apidologie 46 (2): 209–223. doi:10.1051/s13592-014-0316-z.

Githui, F., W. Gitau, F. Mutua, W. Bauwens. 2009. “Climate Change Impact on SWAT Simulated Streamflow in Western Kenya.” International Journal of Climatology 29 (12): 1823–1834. doi:10.1002/joc.1828.

Granitto, P. M., C. Furlanello, F. Biasioli, F. Gasperi. 2006. “Recursive Feature Elimination with Random Forest for PTR-MS Analysis of Agroindustrial Products.” Chemometrics and Intelligent Laboratory Systems 83 (2): 83–90. doi:10.1016/j.chemolab.2006.01.007.

Grundel, R., R. P. Jean, K. J. Frohnapple, G. A. Glowacki, P. E. Scott, N. B. Pavlovic. 2010. “Floral and Nesting Resources, Habitat Structure, and Fire Influence Bee Distribution across an Open-forest Gradient.” Ecological Applications 20 (6): 1678–1692. doi:10.1890/08-1792.1.

Guan, B., H.-J. Guo, -S.-S. Chen, D.-M. Li, X. Liu, X. Gong, G. Ge, et al. 2020. “Shifting Ranges of Eleven Invasive Alien Plants in China in the Face of Climate Change.” Ecological Informatics 55:101024. doi:10.1016/j.ecoinf.2019.101024.

Habel, J. C., I. C. C. Casanova, C. Zamora, M. Teucher, B. Hornetz, H. Shauri, R. K. Mulwa, et al. 2017. “East African Coastal Forest under Pressure.” Biodiversity and Conservation 26 (11): 2751–2758. doi:10.1007/s10531-017-1375-z.

Hao, T., J. Elith, G. Guillera-Arroita, J. J. Lahoz-Monfort. 2019. “A Review of Evidence about Use and Performance of Species Distribution Modelling Ensembles like BIOMOD.” Diversity & Distributions 25 (5): 839–852. doi:10.1111/dad.12892.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized Additive Models. Florida, USA: CRC Press.

Hinton, G. E., and R. R. Salak hutdinov. 2006. “Reducing the Dimensionality of Data with Neural Networks.” Science 313 (5786): 504–507. doi:10.1126/science.1127647.

Iraheta, C. E. R., M. Á. Hernández Martínez, L. A. Alas Romero, M. E. Claros Álvarez, D. R. Arévalo, V. A. Rodríguez González. 2015. “Stingless Bee Distribution and Richness in El Salvador (Apidae, Meliponinae).” Journal of Apicultural Research 54 (1): 1–10. doi:10.1080/02652208.2015.1029783.

Jackson, J. M., M. L. Pimsler, K. J. Oyen, J. B. Koch-Uhuad, J. D. Herndon, J. P. Strange, M. E. Dillon, et al. 2018. “Distance, Elevation and Environment as Drivers of Diversity and Divergence in Bumble Bees across Latitude and Altitude.” Molecular Ecology 27 (14): 2926–2942. doi:10.1111/mec.14735.

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P. and Lobo, J.M. 2011. “Use of Niche Models in Invasive Species Risk Assessments.” Biological Invasions 13 (12): 2785–2797. doi:10.1007/s10530-011-9963-4.

Kajita, M. K., K. Aiha, and T. J. Kobayashi. 2017. “Balancing Specificity, Sensitivity, and Speed of Ligand Discrimination by Zero-order Ultraspecificity.” Physical Review E 96 (1): 012405. doi:10.1103/PhysRevE.96.012405.

Kearney, M., and W. Porter. 2009. “Mechanistic Niche Modelling: Combining Physiological and Spatial Data to Predict Species’ Ranges.” Ecology Letters 12 (4): 334–350. doi:10.1111/j.1461-0248.2008.01277.x.

Kitakoto, N., Pozo, M.I., Van Oystaeyen, A., Musonye, M., Kika, J., Wäckers, F., van Langevelde, F., Hundt, B. and Jaramillo, J. 2022. “African Endemic Stingless Bees as an Efficient Alternative Pollinator to Honey Bees in Greenhouse Cucumber (Cucumis Sativus L.).” Journal of Apicultural Research 61 (2): 1–13. doi:10.1080/00218839.2021.193421.

Kitakoto, N., S. K. Raina, E. Muli, J. Mueke. 2014. “Enhancement of Fruit Quality in Capsicum Annum through Pollination by Hypotrigona Gribodoi in Kakamaga, Western Kenya.” Entomological Science 17 (1): 106–110. doi:10.1111/ens.12030.

Kimathi, E. H. E. Z. Tonnanq, S. Subramanian, K. Cressman, E. M. Abdel-Rahman, M. Tesfayohannes, S. Niassy, et al. 2020. “Prediction of Breeding Regions for the Desert Locust Schistocerca Gregaria in East Africa.” Scientific Reports 10 (1): 11937. doi:10.1038/s41598-020-68895-2.

Kinyanjui, M. J. 2011. “NDVI-based Vegetation Monitoring in Mau Forest Complex, Kenya.” African Journal of Ecology 49 (2): 165–174. doi:10.1111/j.1365-2028.2010.01251.x.

Kumar, M. S., Singh, A.J.A. and Alagumuthu, G. 2012. “Traditional Beekeeping of Stingless Bee (Trigona Sp) by Kani Tribes of Western Ghts, Tamil Nadu, India.” Jtk 11 (2): [April 2012]. [Preprint]. Available at:Accessed . http://nopr.niscar.res.in/handle/123456789/13867.

Libério, S. A., A. L. A. Pereira, M. J. A. M. Araújo, R. P. Dutra, F. R. F. Nascimento, V. Monteiro-Neto, M. N. S. Ribeiro, et al. 2009. “The Potential Use of Propolis as a Cariostatic Agent and Its Actions on Mutans Group Streptococci.” Journal of Ethnopharmacology 125 (1): 1–9. doi:10.1016/j. jep.2009.04.047.
Lovett, J. C. 2015. “Modelling the Effects of Climate Change in Africa.” *African Journal of Ecology* 53 (1): 1–2. doi:10.1111/aje.12218.

Makaya, N. P., O. Mutanga, Z. Kiala, T. Dube, K. E. Seutloali. 2019. “Assessing the Potential of Sentinel-2 MSI Sensor in Detecting and Mapping the Spatial Distribution of Gullies in a Communal Grazing Landscape.” *Physics and Chemistry of the Earth, Parts A/ B/C* 112:66–74. doi:10.1016/j.pce.2019.02.001.

Makori, D. M., A. Fombong, E. Abdel-Rahman, K. Nkoba, J. Ongus, J. Irungu, G. Mosomtai, et al. 2017. “Predicting Spatial Distribution of Key Honeybee Pests in Kenya Using Remotely Sensed and Bioclimatic Variables: Key Honeybee Pests Distribution Models.” *ISPRS International Journal of Geo-Information* 6 (3): 66. doi:10.3390/ijgi6030066.

Mani, M. S. 2013. *Ecology and Biogeography of High Altitude Insects.* Germany: Springer Science & Business Media.

Michener, C. D. 2000. *The Bees of the World.* Maryland, USA: JHU Press.

Michener, C. D. 2013. “The Meliponini.” In *Pot-Honey: A Legacy of Stingless Bees*, edited by P. Vit, S. R. M. Pedro, and D. Roubik, 3–17. New York, NY: Springer. doi:10.1007/978-1-4614-4960-7_1.

Miller, T. E. X., S. M. Louda, K. A. Rose, J. O. Eckberg. 2009. “Impacts of Insect Herbivory on Cactus Population Dynamics: Experimental Demography across an Environmental Gradient.” *Ecological Monographs* 79 (1): 155–172. doi:10.1890/07-1550.1.

Mohammadi, S., E. Ebrahimi, M. Shahriari Moghadam, L. Bosso. 2019. “Modelling Current and Future Potential Distributions of Two Desert Jerboas under Climate Change in Iran.” *Ecological Informatics* 52:7–13. doi:10.1016/j.ecoinf.2019.04.003.

Mokaya, H. O., K. Nkoba, R. M. Ndunda, N. J. Vereecken. 2021. “Characterization of Honey Produced by Sympatric Species of Afrotropical Stingless Bees (Hymenoptera, Meliponini).” *Food Chemistry* 366:130597. doi:10.1016/j.foodchem.2021.130597.

Muderer, B. T., E. M. Abdel-Rahman, T. Dube, T. Landmann, Z. Khan, E. Kimathi, R. Owino, et al. 2020a. “Multi-source Spatial Data-based Invasion Risk Modeling of Striga (Striga Asiatica) in Zimbabwe.” *GIScience & Remote Sensing* 57 (4): 553–571. doi:10.1080/15481603.2020.1744250.

Muderer, B. T., C. Mukanga, E. T. Mupfiga, C. Gwattirisa, E. Kimathi, T. Chitata. 2020b. “Analysis of Potentially Suitable Habitat within Migration Connections of an intra-African Migrant-the Blue Swallow (Hirundo Atrocaerulea).” *Ecological Informatics* 57:101082. doi:10.1016/j.ecoinf.2020.101082.

Muli, E., H. Patch, M. Frazier, J. Frazier, B. Torto, T. Baumgarten, J. Kilonzo, et al. 2014. “Evaluation of the Distribution and Impacts of Parasites, Pathogens, and Pesticides on Honey Bee (Apis Mellifera) Populations in East Africa.” *PLOS ONE* 9 (4): e94459. doi:10.1371/journal.pone.0094459.

Mwaleshi, S., H. E. Z. Tonnang, E. S. Massawe, G. O. Okuku, N. Khadioli, T. Johansson, P.-A. Calatayud, et al. 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 (6): e0130427. doi:10.1371/journal.pone.0130427.

Naimi, B., and M. B. Araújo. 2016. “Sdm: A Reproducible and Extensible R Platform for Species Distribution Modelling.” *Ecography* 39 (4): 368–375. doi:10.1111/ecog.01881.

Naimi, B., N. A. S. Hamm, T. A. Groen, A. K. Skidmore, A. G. Toxoepus. 2014. “Where Is Positional Uncertainty a Problem for Species Distribution Modelling?” *Ecography* 37 (2): 191–203. doi:10.1111/1600-0587.1200205.x.

Ndungu, N. N., A. A. Yusuf, S. K. Raina, D. K. Masiga, C. W. W. Pirk, K. Nkoba. 2019. “Nest Architecture as a Tool for Species Discrimination of Hypotrigona Species (Hymenoptera: Apidae: Meliponini).” *African Entomology* 27 (1): 25–35. doi:10.4001/003.027.0025.

Njaya, M. T. M. (2010). “Diversity of Stingless Bees in Bamenda Afromontane Forests – Cameroon.” (Accessed 9 July 2020). https://bonndoc.unibonn.de/xmlui/handle/20.500.11811/4187

Nkoba, K., S. K. Raina, E. Muli, K. Mithöfer, J. Mueke. 2012. “Species Richness and Nest Dispersion of Some Tropical Meliponine Bees (Apidae: Meliponinae) in Six Habitat Types in the Kakamega Forest, Western Kenya.” *International Journal of Tropical Insect Science* 32 (4): 194–202. doi:10.1017/S1742758412000355.

Ochieng, J., L. Kimi, M. Mathenge. 2016. “Effects of Climate Variability and Change on Agricultural Production: The Case of Small Scale Farmers in Kenya.” *NJAS - Wageningen Journal of Life Sciences* 77:71–78. doi:10.1016/j.njas.2016.03.005.

Odawa, S., and Y. Seo. 2019. “Water Tower Ecosystems under the Influence of Land Cover Change and Population Growth: Focus on Mau Water Tower in Kenya.” *Sustainability* 11 (13): 3524. doi:10.3390/su11133524.

Pachauri, R. K., and L. Mayer (2015). “Climate Change 2014: Synthesis Report.” Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Pau, S., T. W. Gillespie, and E. M. Wolkovich. 2012. “Dissecting NDVI–species Richness Relationships in Hawaiian Dry Forests.” *Journal of Biogeography* 39 (9): 1678–1686. doi:10.1111/j.1365-2699.2012.02731.x.

Peterson, A. T. 2011. *Ecological Niches and Geographic Distributions (MPB-49).* New Jersey, USA: Princeton University Press.

Phillips, S. J., and M. Dudik. 2008. “Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation.” *Ecography* 31 (2): 161–175. doi:10.1111/j.0906-7590.2008.05203.x.

Pirk, C. W., Strauss, U., Yusuf, A.A., Démares, F. and Human, H. 2015. “Honeybee Health in Africa—a Review.” *Apidologie* 1–25. doi:10.1007/s13592-015-0406-6.

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

Platts, P. J., P. A. Omeny, and R. Marchant. 2015. “AFRICLIM: High-resolution Climate Projections for Ecological Applications in Africa.” *African Journal of Ecology* 53 (1): 103–108. doi:10.1111/aje.12180.
QGIS Development Team (2021). “Quantum Geographical Information System.” Accessed 10 May 2021. https://qgis.org/en/site/

Rapacciuolo, G., D. B. Roy, S. Gillings, A. Purvis. 2014. “Temporal Validation Plots: Quantifying How Well Correlative Species Distribution Models Predict Species’ Range Changes over Time.” Methods in Ecology and Evolution 5 (5): 407–420. doi:10.1111/2041-210X.12181.

Richard, K., E. Abdel-Rahman, S. Mohamed, S. Ekesi, C. Borgemeister, T. Landmann. 2018. “Importance of Remotely-Sensed Vegetation Variables for Predicting the Spatial Distribution of African Citrus Triozid (Trioza Erytreae) in Kenya.” ISPRS International Journal of Geo-Information 7 (11): 429. doi:10.3390/ijgi7110429.

Roulston, T. H., and K. Goodell. 2011. “The Role of Resources and Risks in Regulating Wild Bee Populations.” Annual Review of Entomology 56 (1): 293–312. doi:10.1146/annurev-ento-120709-144802.

Saatchi, S., W. Buermann, H. ter Steege, S. Mori, T. B. Smith. 2008. “Modeling Distribution of Amazonian Tree Species and Diversity Using Remote Sensing Measurements.” Remote Sensing of Environment 112 (5): 2000–2017. doi:10.1016/j.rse.2008.01.008.

Schürmann, A., J. Kleemann, C. Fürst, M. Teucher. 2020. “Assessing the Relationship between Land Tenure Issues and Land Cover Changes around the Arabuko Sokoke Forest in Kenya”. Land Use Policy 95: 104625. doi:10.1016/j.landusepol.2020.104625.

Shen, Y., L. Wu, L. Di, G. Yu, H. Tang, G. Yu, Y. Shao, et al. 2013. “Hidden Markov Models for Real-time Estimation of Corn Progress Stages Using MODIS and Meteorological Data.” Remote Sensing 5 (4): 1734–1753. doi:10.3390/rs5041734.

Slaa, E. J., L. A. Sánchez Chaves, K. S. Malagodi-Braga, F. E. Hofstede. 2006. “Stingless Bees in Applied Pollination: Practice and Perspectives.” Apidologie 37 (2): 293–315. doi:10.1051/apido:2006022.

Speranza, C. I., B. Kiteme, P. Ambenje, U. Wiesmann, S. Makali. 2009. “Indigenous Knowledge Related to Climate Variability and Change: Insights from Droughts in Semi-arid Areas of Former Makueni District, Kenya.” Climatic Change 100 (2): 295–315. doi:10.1007/s10584-009-9713-0.

Stockwell, D. R. B., and A. T. Peterson. 2002. “Effects of Sample Size on Accuracy of Species Distribution Models.” Ecological Modelling 148 (1): 1–13. doi:10.1016/S0304-3800(01)00388-X.

Swets, J. A. 1988. “Measuring the Accuracy of Diagnostic Systems.” Science 240 (4857): 1285–1293. doi:10.1126/science.3287615.

Therneau, T. M., B. Atkinson, and M. B. Ripley. 2010. The Rpart Package. Oxford, UK: Oxford University Press.

Torné-Noguera, A., A. Rodrigo, X. Arnau, S. Osorio, H. Barril-Graells, L. C. da Rocha-filho, J. Bosch, et al. 2014. “Determinants of Spatial Distribution in a Bee Community: Nesting Resources, Flower Resources, and Body Size.” PLoS ONE 9 (5): e97255. doi:10.1371/journal.pone.0097255.

Torto, B., A. T. Fombong, D. M. Mutyambai, E. Muli, R. T. Arboagost, P. E. A. Teal. 2010. “Aethina Tumida (Coleoptera: Nitidulidae) and Opiostomus Haroldi (Coleoptera: Scarabaeidae): Occurrence in Kenya, Distribution within Honey Bee Colonies, and Responses to Host Odors.” Annals of the Entomological Society of America 103 (3): 389–396. doi:10.1603/AN09136.

Williams, N. M., J. Regetz, and C. Kremen. 2012. “Landscape-scale Resources Promote Colony Growth but Not Reproductive Performance of Bumble Bees.” Ecology 93 (5): 1049–1058. doi:10.1890/11-1006.1.

Williams, G. R., D. R. Tarpy, D. vanEngelsdorp, M.-P. Chauzat, D. L. Cox-Foster, K. S. Delaplane, P. Neumann, J. S. Pettis, R. E. L. Rogers, and D. Shutler. 2010. “Colony Collapse Disorder in Context.” Bioessays 32 (10): 845. doi:10.1002/bies.201000075.

Winfree, R., N. M. Williams, J. Dushoff, C. Kremen. 2007. “Native Bees Provide Insurance against Ongoing Honey Bee Losses.” Ecology Letters 10 (11): 1105–1113. doi:10.1111/j.1461-0248.2007.01110.x.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan. 2008. “Effects of Sample Size on the Performance of Species Distribution Models.” Diversity & Distributions 14 (5): 763–773. doi:10.1111/j.1472-4642.2008.00482.x.

Wright, M. N., Wager, S., Probst, P. and Wright, M.M.N. 2018. “Package Ranger.” Version 0.11 2

Yackulic, C. B., R. Chandler, E. F. Zipkin, J. A. Royle, J. D. Nichols, E. H. Campbell Grant, S. Veran, et al. 2013. “Presence-only Modelling Using MAXENT: When Can We Trust the Inferences?” Methods in Ecology and Evolution 4 (3): 236–243. doi:10.1111/2041-210x.12004.

Yan, K., and D. Zhang. 2015. “Feature Selection and Analysis on Correlated Gas Sensor Data with Recursive Feature Elimination.” Sensors and Actuators. B, Chemical 212: 353–363. doi:10.1016/j.snb.2015.02.025.