

Understanding climate change effects on the potential distribution of an important pollinator species, *Ceratina moerenhouti* (Apidae: Ceratinini), in the Eastern Afromontane biodiversity hotspot, Kenya

Mukundi Mukundamago^{a,b}, Timothy Dube^b, Bester Tawona Mudereri^{a,c,*}, Regis Babin^{d,e,f}, H. Michael G. Lattorff^{a,g}, Henri E.Z. Tonnang^a

^a International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772 00100, Nairobi, Kenya

^b Department of Earth Sciences, University of the Western Cape, Bellville, South Africa

^c School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private Bag 3, Wits, 2050, South Africa

^d UMR PHIM Plant Health Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France

^e CIRAD, UMR PHIM, Abidjan, Republic of Côte d'Ivoire

^f Université Houphouët-Boigny - Cocody Batiment, IRD - 01, BP 6483, Abidjan, Republic of Côte d'Ivoire

^g Department of Chemistry, University of Nairobi, P. O. Box 30197-00100, Nairobi, Kenya

ARTICLE INFO

Keywords:

Climate variability
Landscape dynamics
Pollinator
Niche modeling
Ceratina

ABSTRACT

Monitoring key pollinator taxa such as the genus *Ceratina* requires precise near real-time predictions to facilitate better surveillance. The potential habitat suitability of *Ceratina moerenhouti* was predicted in the Eastern Afromontane biodiversity hotspot (EABH) in Kenya using presence-only data, to identify their potential distribution and vulnerability due to climate change. Bioclimatic, edaphic, terrain, land surface temperature, and land use and land cover (LULC) variables were used as predictors. Three machine learning techniques, together with their ensemble model, were evaluated on their suitability to predict current and future (the shared socioeconomic pathways (SSPs), i.e., SSP245 and SSP585) habitat suitability. Predictors were subjected to variable selection using the variance inflation factor resulting in a few ($n = 9$) optimum variables. The area under the curve (AUC) and true skill statistic (TSS) were used for the accuracy assessment of the modeling outputs. The results indicated that 30% and 10% of the EABH in Murang'a and Taita Taveta counties are currently suitable for *C. moerenhouti* occurrence, respectively. However, future projections show a $\pm 5\%$ decrease in *C. moerenhouti* habitats in the two counties. Further, the ensemble model harnessed the algorithm differences while the random forest had the highest individual predictive power (AUC = 0.97; TSS = 0.96). Clay content, LULC, and the slope were the most relevant variables together with temperature and precipitation. Integrating multi-source data in predicting suitable habitats improves model prediction capacity. This study can be used to support the maintenance of flowering plant communities around agricultural areas to improve pollination services.

1. Introduction

Pollinators offer a crucial service to the environment and livelihoods as they are vital contributors to global food security (Landmann et al., 2015; van der Sluijs and Vaage, 2016). Approximately 75% of global food crops depend on natural insect pollination (Gallai et al., 2009). Unfortunately, earlier research showed that over 40% of invertebrate pollinators are threatened with global extinction, and about 9% of wild bee and butterfly species face local extinction (Dicks et al., 2021; Zattara and Aizen, 2021). The decline and population losses in most of these

crucial pollinators have been attributed to climate change, diseases, habitat loss, increased anthropogenic use of pesticides, and monoculture (Janzen and Hallwachs, 2019; Ochungo et al., 2019).

Bees are the largest and most important group of insect pollinators, especially for crop production (Schatz et al., 2021). These pollinators are diverse, particularly in Africa's eight biodiversity hotspots including the Eastern Afromontane biodiversity hotspot (EABH) (Eardley et al., 2009). However, studies on improving pollinator diversity, population size, and pollination services are scarce in Africa, despite the abundance of pollinator-dependent crops (Korpela et al., 2013). Of the few studies

* Corresponding author. International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772 00100, Nairobi, Kenya.

E-mail address: bmudereri@icipe.org (B.T. Mudereri).

<https://doi.org/10.1016/j.pce.2023.103387>

Received 22 August 2022; Received in revised form 17 January 2023; Accepted 20 February 2023

Available online 24 February 2023

1474-7065/© 2023 Elsevier Ltd. All rights reserved.

available, most have mainly focused on the honeybee (*Apis mellifera*) (Ochungo et al., 2019; Sagwe et al., 2021; Tola et al., 2020) while pollination by wild bees may be more important (Garibaldi et al., 2013).

The small *Ceratina moerenhouti* Vachal 1903 is one of the most important bee pollinators for both wild and agricultural plants (Garibaldi et al., 2013), but its ecology is largely understudied (Eardley and Daly, 2007). Earlier studies have reported *Ceratina* genus as an important pollinator of a spectrum of wild and agricultural plant families (Altieri et al., 2015; Eardley and Daly, 2007; Eardley et al., 2009). The genus *Ceratina* is polylectic and has also been reported to visit flowers from different families, including Zygophyllaceae, Vitaceae, Rutaceae, Rosaceae, Resedaceae, Tamaricaceae, Simaroubaceae, Scrophulariaceae, Plumbaginaceae, Oleaceae, Molluginaceae, Malvaceae, Liliaceae, Lamiaceae, Fabaceae, Euphorbiaceae, Dipsacaceae, Cucurbitaceae, Campanulaceae, Convolvulaceae, Chenopodiaceae, Cynareae, Boraginaceae, Brassicaceae, Asteraceae, Araliaceae, Anacardiaceae, and Apiaceae (Terzo and Rasmont 2011).

The genus *Ceratina* is highly diverse and has approximately 380 species distributed throughout the world, with about 29 species occurring in Africa (Discover, 2022; ITIS, 2022). Although they are not honey producers, they are an essential part of the pollination process for crops and wild plants (Eardley and Daly, 2007). Some *Ceratina* species seem to have nests consisting of several adult females, and most live solitary or in small colonies (Rehan, 2020). Most *Ceratina* bee species make nests in deadwood, stems, or piths in various locations mostly in shaded environments (Eardley et al., 2009). However, very few studies have investigated the effect of climate change on the distribution of suitable habitats for most of these species including *C. moerenhouti* (Tabor and Koch, 2021).

In general, the global geographic occupancy of many pollinators such as *C. moerenhouti* has declined in the last 5 decades (Dicks et al., 2021; Gallai et al., 2009). Pollinators and the host plants of pollinators will be significantly impacted by climate change in natural and agricultural ecosystems (Mudereri et al., 2019; Sango and Godwell, 2015). Anthropogenic-induced climate change may increase temperature and alter rainfall patterns. As the atmospheric CO₂ and temperature increase or decrease, they will possibly result in warmer, colder, wetter, or dryer conditions (IPCC, 2014; Niang et al., 2014). Thus, habitat suitability may change for specific pollinators or vegetations, leading to changes in population size, hence altering potential inter- and intraspecific competition (Duan et al., 2020; Okello et al., 2021). A mean temperature increase or decrease will thus likely alter the known life cycles, reproductive success, and mobility of most insects across the globe, including the *C. moerenhouti* species (Ghisbain et al., 2021). Ultimately, promoting evidence-based studies that guide pollinator management and monitoring policies that target the improvement and increase in wild populations is key to reducing pollination deficit in smallholder production systems.

In this study, we used three species distribution models (SDMs) based on machine learning (ML) tools i.e., maximum entropy (MaxEnt), support vector machines (SVM), random forest (RF), and their ensemble combination to predict the best suitable habitat for the *C. moerenhouti* species. These ML algorithms and field-gathered presence-only reference occurrence data were used to predict the current and future habitat suitability within one of the eight biodiversity hotspot areas in Murang'a and Taita Taveta counties in Kenya. SDMs have broadly been used for numerous objectives in conservation biology (Moshobane et al., 2022), biogeography (Bradie and Leung, 2017), and ecology (Mohammadi et al., 2019; Mudereri et al., 2021; Otunga et al., 2017) across different continents, climates, and ecologies (Eshetae et al., 2019). ML SDMs such as the RF, MaxEnt, and SVM have demonstrated their strength and capability to deliver excellent predictive performances as well as their flexibility to cope with autocorrelations (Muthoni et al., 2021; Naimi et al., 2014). In contrast, earlier studies have shown concern over the uncertainties caused by different algorithms. As a result, the ensemble modeling approach is increasingly popular due to its ability to combine

multiple models' predictive strengths and lessen their weaknesses (Chemura et al., 2021).

Of particular concern to modeling biological species is that many species, mainly insects, will likely depend on the prevailing climate conditions (Volis and Blecher, 2021). These conditions usually define most geographical distribution noted in numerous species' distribution modeling outputs (Mutamiswa et al., 2022; Otieno et al., 2019). Among others, vegetation composition, precipitation, temperature, and altitude have been reported by earlier studies as critical environmental factors that affect the ability of insect pests to adapt to an area and eventually their distribution depending on their tolerable thresholds (Azrag et al., 2018; Otieno et al., 2019). Climate factors are the determinants of most arthropod species, including *C. moerenhouti*, and they are presented as the key factors that affect the abundance and distribution of these species. As a result, climate change will pose a fair amount of risk to the species, making an immediate understanding of the consequences necessary.

Recent studies have used the shared socio-economic pathways (SSPs), land use, energy, and emissions implication to determine and predict the potential future climate scenarios (Chemura et al., 2021; Riahi et al., 2017). SSPs are part of a new scenario framework succeeding the international panel on climate change (IPCC), representative concentration pathways (RCPs), to facilitate the integrated analysis of future climate impacts, vulnerabilities, adaptation, and mitigation. The SSPs explain alternative socio-economic developments in five narratives that describe sustainable development, inequality, regional rivalry, fossil-fuelled development, and middle-of-the-road development (Riahi et al., 2017). In addition to the commonly used bioclimatic predictors, an earlier study found that incorporating remotely sensed data and environmental variables such as topography, land cover, land surface temperature, and other environmental variables improved the predictive capacity of models (Jetz et al., 2019; Ropars et al., 2020).

Thus, this study contributes to understanding of the Kenya scenario in *C. moerenhouti* occurrence and contributes to the overall body of knowledge that facilitates the understanding of the connectivity of the global ecosystems and the importance of the EABH for *C. moerenhouti* conservation and management. The study's specific objectives were to: (1) predict and understand the determining factors of the suitability of *C. moerenhouti* habitat at a landscape scale in Kenya and (2) estimate the suitable areas using the future scenarios i.e., SSP245 and SSP585.

1.1. Study area

The study was conducted in two separate counties of Kenya, i.e., Murang'a and Taita Taveta, which lie within the EABH. The EABH stretches from the north in the Middle East to the south in southern Africa covering over 7000 km from Saudi Arabia through widely scattered biogeographically similar mountains in East Africa until the eastern highlands between the border of Mozambique and Zimbabwe. The EABH is one of the 36 global biodiversity hotspots and is one of the highly biologically diverse regions among the world's biodiversity hotspots. Kenya is rich in plant diversity within several unique habitats, with an estimated ± 7000 indigenous plant species (Mahiga et al., 2019).

Murang'a county lies at latitude 0.7839° S and longitude 37.0400° E covering an area of 2326 km². It is a major tea and coffee growing region located in central Kenya bordering Nyeri and Kirinyaga counties in the north, Machakos and Embu counties in the east, and Kiambu county in the south (Fig. 1). The mean annual temperature and rainfall range from 18 °C to 21 °C and 1000 to 1500 mm, respectively. The rainfall pattern is bimodal with 'long rains' occurring from March to May and 'short rains' from October to December (Ndoyisaba et al., 2021). Furthermore, Murang'a is characterized by forest fragments, coffee, and tea with other fruit trees such as avocado (*Persea americana*), macadamia (*Macadamia integrifolia*), mango (*Mangifera indica*), and hedgerow trees like grevillea (*Grevillea robusta*) including intercrops such as bean, maize, sweet potato, banana, and tree tomato on an average farm size of 0.5 ha

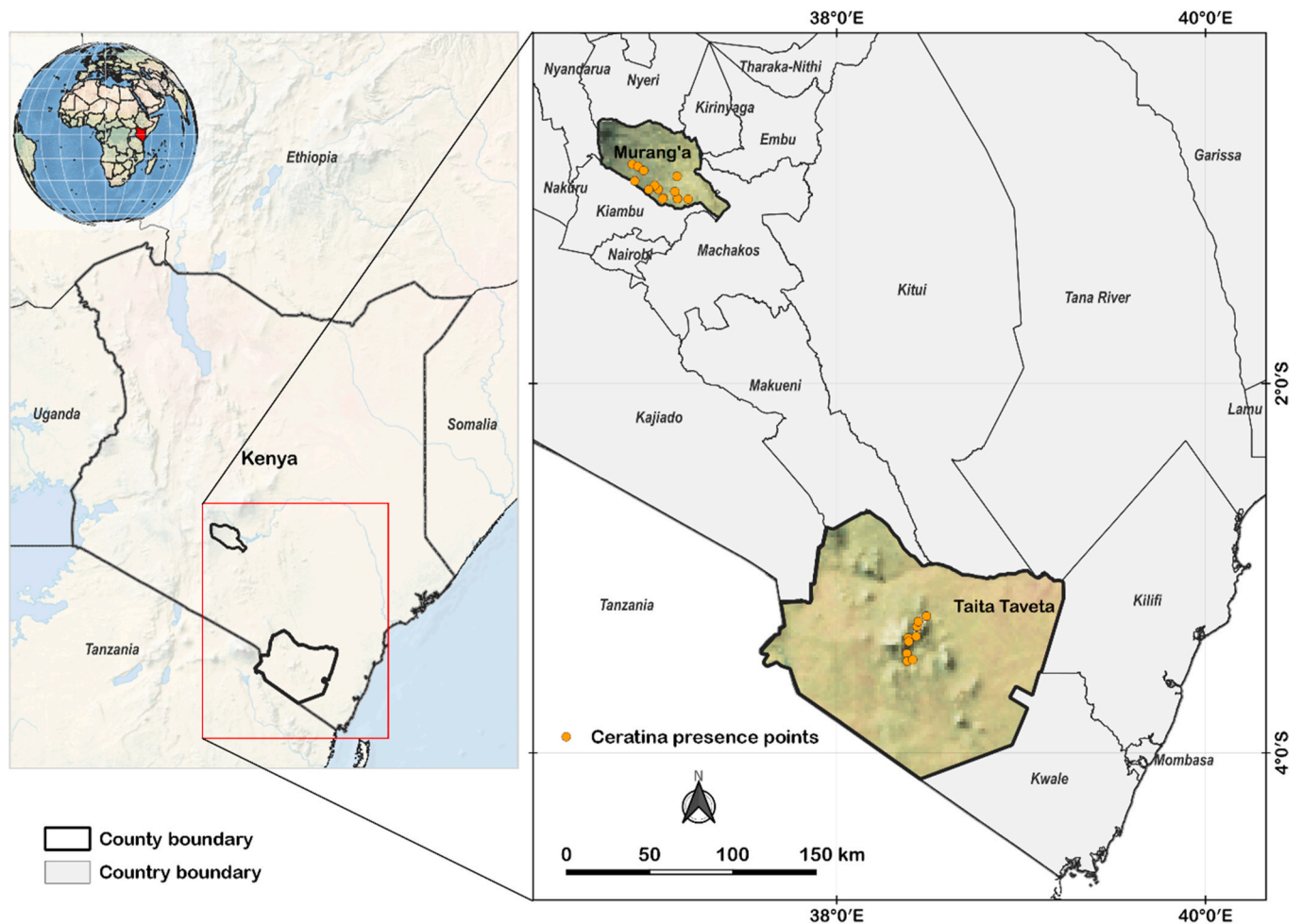


Fig. 1. Location of the study area in Africa and Kenya. The two counties of Murang'a and Taita Taveta with the *Ceratina moerenhouti* ($n = 55$) occurrence points superimposed on the shaded relief Natural earth data (<https://www.naturalearthdata.com/features/>).

(Aduvukha et al., 2021; Murang'a County Government, 2021).

On the other hand, Taita Taveta covers a spatial area of 17,084 km² and lies between latitude 3.3831° S and longitude 38.3625° E located in southern Kenya which borders Kajiado and Makuani on the northwest, Kitui, and Tana River on the North while Kilifi and Kwale are in the East (Fig. 1). Taita Taveta is characterized by a very high diversity of flora and fauna with a high level of endemism. The dominant land use characteristic in the study area is subsistence small-scale farming. Within these small-scale farms, the common crops in Taita Taveta include beans, peas, potatoes, maize, bananas, tomatoes, cabbages, sorghum, cassava, millet, and avocados. Both high population pressure and dominant subsistence farming have caused dynamic changes in land-use patterns resulting in significant levels of land degradation. Since the 1950s indigenous forests have suffered significant levels of deforestation and degradation and it is assumed that approximately 1% of the original forest remains. This forest loss is mainly due to the clearing of indigenous forests for exotic timber plantations and agricultural expansion, hence affecting the diversity and abundance of insects in the area. In the greater part of this study area, indigenous forests have been replaced by exotic plantations.

2. Methodology

2.1. Reference occurrence data collection

A total of sixty farms for both Murang'a and Taita Taveta that had a standard separation distance of at least 3 km were used. The survey units

were farms, hedgerows, forest fragments, and roadsides (Baard and Kraaij, 2019). Reference occurrence data collection was conducted twice i.e., for five consecutive days between the 11th to the 15th of November 2019 and again from the 26th to the 30th of October 2020. The data collection phase was purposefully matched with the peak flowering of crops and weeds within the farms in the study area. The *C. moerenhouti* were captured, using sweep nets between 07:00 h to 17:00 h to match their most active foraging time. A sampling unit of 5 × 100 m transect was set up on each farm following the method of Roulston et al. (2007). Direct observations and recording of bee-flower interactions were done for 30 min repeatedly at ten different random locations along the transect before moving to the next farm (Roulston et al., 2007). Plants were sampled within the same transects, using 4 replicates of 2 × 2 m quadrats where plant species and cover were determined. A handheld global positioning system (GPS) device with an error margin of ±3 m was used to capture the coordinates of the 10 sampled locations as well as the central coordinate of the transect. A total of 55 "presence-only" *C. moerenhouti* reference data were collected covering the selected farms in the two study sites of Murang'a and Taita Taveta. The sampling design to scout for the occurrence of *C. moerenhouti* followed an elevation gradient from the high to the low elevation areas. However, most of the occurrences were observed within high-altitude areas.

The collected bees and plant samples were stored in the freezer and plant presser, respectively. These bee specimens were identified using a camera microscope in tandem with bee identification manuals and virtual insect collections using wing pattern morphology and other

distinguishing morphological features (terga, sterna, clypeus, mandibles, metanotum, propodeum, etc). The collected insects were sorted at the International centre of insect physiology and ecology (*icipe*) African reference laboratory for bee health (ARLBH) according to their order, family, genus, and species level. The sorted specimens were stored and labeled separately from each farm location and were matched to the central coordinate of the transect where they were collected from and were used in this analysis as the presence-only data.

2.2. Predictor variables

The predictor variables that were used to predict the suitability of the *C. moerenhouti* habitat were derived from the bioclimatic, topographic, edaphic, land surface temperature (LST), and land use and land cover (LULC). To match the size and extent of edaphic variables, all variables were clipped to the study area boundary and resampled to 250×250 m pixels. Thus, the spatial and temporal resolution of our models should not impact their accuracy since the farms were at least 3 km apart.

2.2.1. Bioclimatic variables

One of the benefits of SDMs is that predictive models fitted in current conditions are transferable to novel periods under emission scenario projections (Mesgaran et al., 2014; Zurell et al., 2012). This provides opportunities to understand and compute the potential variations in species ranges due to climate change. In our first analysis, we used 19 bioclimatic variables that were downloaded from the WorldClim platform (www.worldclim.org) at approximately 1×1 km spatial resolution (Fick and Hijmans, 2017). The data has been interpolated from data obtained from weather stations around the world, using the thin plate smoothing spline algorithm implemented in ANUSPLIN (Fick and Hijmans, 2017).

2.2.2. Edaphic variables

Five soil properties, from the 'AfSoilGrids250m', were downloaded from <https://www.isric.org/explore> and used in the predictive models (Hengl et al., 2015). These soil grids are provided at different soil depths, and we opted to use the 0–30 cm depth with a spatial resolution of 250×250 m. It was assumed that this depth can influence the species of plants and their growth form and rate. Thus, the types of available plants either provide or limit the areas for foraging and nesting for the *C. moerenhouti*. Specifically, we used total soil nitrogen (N) (mg/kg: ppm), soil pH, soil organic carbon (SOC) in g/kg, clay content (g/100 g), and sand content (50–2000 μ m) in g/100 g (Hengl et al., 2015). These variables broadly influence soil fertility for vegetation growth, and thus the potential production of flowers for the foraging of the *C. moerenhouti* as well as the production of soil organic matter to enhance interaction with other insects (Eardley and Daly, 2007; Okello et al., 2021).

2.2.3. Land surface temperature

The LST from the climate modeling Grid product (LST_Day_CMG), available in K and simulated from moderate resolution imaging spectroradiometer (MODIS) data, was downloaded at <https://lpdaac.usgs.gov/products/mod11c2v006/> (Wan et al., 2015). Specifically, we used the 'multi-day' MOD11C2 LST product of 5.6×5.6 km spatial resolution available from the year 2000 to the present. The LST variable was chosen for inclusion in the analysis as the temperatures are at the ground surface. Thus, LST together with atmospheric temperature must be within the threshold tolerable temperatures of the insects and hence would influence the suitability of the habitat (Azrag et al., 2017). Therefore, we hypothesized that the surface fluxes measured by LST would be one of the proxy key variables that greatly predict the potential occurrence of the *C. moerenhouti* species.

2.2.4. Terrain variables

The terrain variables were derived from the shuttle radar topographic mission (SRTM) data that is provided at 1 arc sec (~ 30 m

resolution) digital elevation model (DEM) with a 16 m vertical error (Farr et al., 2007). We assumed that this error margin would not significantly influence the performance of our models. In addition to the elevation, we derived six other terrain variables i.e., aspect, slope, flow direction, roughness, terrain ruggedness index (TRI), and terrain position index (TPI) using the 'terrain' function available in the raster package (Hijmans, 2020) in R version 4.0.5 (R Core Team, 2021). The influence of terrain variables on bees' occurrence was reported by earlier studies and was anticipated to influence the occurrence and propagation of *C. moerenhouti* by altering precipitation, temperature, vegetation including crops, and the angle, direction, and intensity of the sun on the earth's surface (Azrag et al., 2017; Makori et al., 2017).

2.2.5. Land use and land cover (LULC)

The 20×20 m pixel size Sentinel-2 prototype land cover data for Africa is provided by the European Space Agency (ESA) and freely available from <http://2016africalandcover20m.esrin.esa.int/download.php> was used as the LULC variable in this study (ESA, 2020). The data include a legend comprising ten generic classes i.e., 'trees cover areas', 'shrubs cover areas', 'grassland', 'cropland', 'vegetation aquatic or regularly flooded', 'lichen and mosses or sparse vegetation', 'bare areas', 'built-up areas', 'snow or ice' and 'open water'. Of these ten classes, only nine are available in the study area except the 'snow or ice' class. This ESA-derived LULC dataset was derived from a one-year of Sentinel-2A observations from December 2015 to December 2016 classified using the RF algorithm (ESA, 2020). Other studies on bees or other insect species have demonstrated the interdependence of land cover patterns and insect species, regarding their relative potential spatial distribution (Adan et al., 2021; Nyabako et al., 2020; Ropars et al., 2020). However, these regionally developed datasets such as the Sentinel-2 prototype land cover data for Africa used in this study often have accuracy limitations. However, they provide a critical gridded view of the spatial location of various LULC classes at scale but must be treated with caution.

2.3. Future projection variables

In this study, the topographic and edaphic, LST, and LULC were assumed and considered to remain static to project the future potentially suitable habitat for the *C. moerenhouti*, while the independent climatic predictors were those extracted from the global climatic model (GCM) provided by the center for climate system research (Japan, MIROC6). This GCM was selected because it provides completely bias-corrected data for all the periods including 2050 (2041–2060) data that was used in this analysis (Chemura et al., 2021).

The impact of climate change was evaluated from the projected climate and socio-economic conditions defined by the combination of the RCPs and SSPs (Riahi et al., 2017). The RCPs consider the impact of potential future greenhouse gas emission trajectories on the climate system based on the forcing until the end of the twenty-first century (IPCC, 2014). Four levels of RCP have been proposed by the IPCC i.e., RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5. In contrast, the SSP concept narrates the future of the world using trajectories of population growth, economic growth, trade development, technological development, and implementation of environmental policies (Riahi et al., 2017). The five SSPs are commonly referred to as (i) the most sustainable development (SSP1), (ii) middle-of-the-road development (SSP2), (iii) regional rivalry (SSP3), (iv) inequality (SSP4), and (v) full fossil-fuelled development (SSP5) pathways (Chemura et al., 2021). Therefore, the combination of RCPs and SSPs simulates a more probable integration scenario matrix of radiative forcing and socioeconomic development influences because each SSP is broadly aligned with one or two RCPs. For analysis of the future impact of climate change on the *C. moerenhouti*, we used two of the four possible combinations i.e., SSP2-RCP4.5 (SSP245) and SSP5-RCP8.5 (SSP585) to enable comparison in similar studies that apply the same combinations.

2.4. Collinearity test of variables used in the species distribution models (SDM)

Collinearity amongst the predictor variables in most SDMs causes instability and overfitting of models thus reducing the reliability of the outputs (Dormann et al., 2013). The variance inflation factor (VIF), which detects multicollinearity by taking each predictor and regressing it against the other variables in multiple linear regression analysis was used for selecting optimum variables for the modeling (Plant, 2012). The resulting coefficients of determination (R^2) value attained from these multiple regressions are then replaced in the VIF calculation formula for each pair as shown in equation (1).

$$VIF_i = \frac{1}{1 - R_i^2} \quad (1)$$

Where i is the predictor.

In this study, the 'vifcor' function in the 'usdm' package available in R version 4.0.5 (Naimi et al., 2014; R Core Team, 2021) was used to calculate the VIF values for each of the variables. The 'vifcor' function iteratively selects pairs of variables with high linear correlation, then eliminates the one with the highest VIF. The threshold for elimination was set as $th = 0.7$, which represents a Pearson correlation coefficient ($r \geq 0.7$) following the recommendation of Kyalo et al. (2018). Predictor variables that have VIF values that are greater than 10 demonstrate evidence of high collinearity with other variables within a model and hence are often eliminated (Dormann et al., 2013). A correlation matrix was then used to test the level of correlation on the retained variables to enhance the empirical selection and retention of key variables that have been reported in the literature to be of ecological significance to bees (Makori et al., 2017). Our variable elimination procedure resulted in a selection of nine optimum variables from a total of 32 bioclimatic and remotely sensed variables. The nine variables that were used in the final modeling procedure are highlighted in Table 1.

2.5. Species distribution models implementation

The 'sdm' package (Naimi and Araújo, 2016) in R version 4.0.5 (R Core Team, 2021) was used to predict the current distribution of the *C. moerenhouti* as well as to project the suitability of the habitat into the future. The 55 presence-only points data that were collected in the field were used together with 10,000 pseudo-absence points generated using the 'sdmdata' function inherent in the 'sdm' package to build the

Table 1

The nine bioclimatic and remotely sensed variables that were retained after the correlation test elimination and were used in the species distribution models (SDMs) for *C. moerenhouti* occurrence prediction and their variance inflation factor (VIF) values.

Variable	Description	Units	V.I.F Value
<i>Bioclimatic variable</i>			
Bio1	Annual mean temperature	°C	9.58
Bio4	Temperature seasonality (standard deviation × 100)	–	9.73
Bio12	Annual precipitation	mm	9.81
Bio15	Precipitation seasonality (coefficient of variation)	–	3.00
<i>Edaphic variables</i>			
Clay content	Quantity of clay in the soil	g/100g	4.21
<i>Land surface temperature (LST)</i>			
LST	Land surface temperature	K	5.50
<i>Terrain variables</i>			
Aspect	Slope direction	Degrees	1.55
Slope	Ground steepness	%	2.27
<i>Land use and land cover</i>			
LULC	The land use and land cover	categorical	2.00

models' reference base for the presence and absence of the *C. moerenhouti* species. The number of points used as presence data was within the sufficient sample size required for accurate predictions when robust ML algorithms are used such as those employed in this study (Stockwell and Peterson, 2002). The 'sdm' package provides a single platform with the possibility to run 15 ML algorithms using the same presence-only and pseudo-absence data by applying an object-oriented reproducible and extensible framework for SDM in R (Naimi and Araújo, 2016). In the present study, we selected and inter-compared only three of the 15 modeling techniques in 'sdm' as follows: RF, MaxEnt, and SVM.

In the RF model, the prediction is achieved by selecting the highest probability occurrence value from multiple decision trees (Muthoni et al., 2021). On the other hand, MaxEnt predicts the species occurrence by finding the largest spread (maximum entropy) (Mutamiswa et al., 2022; Phillips et al., 2017) while SVM uses a hyperplane to estimate the divergence of class groupings for the prediction (Hastie et al., 1994; Vapnik, 1979). These three algorithms were selected in this study because they are widely used in conducting complex output predictions with relatively high modeling accuracies for regression and classification (Abdel-Rahman et al., 2013). A summary of these models' execution syntax and their corresponding packages used by 'sdm' in the parallel model simulations is provided in Table 2.

2.6. Rank sum test for variable importance

The ranking of variables to predict their relative importance was done using a rank sum test. Within each of the ML methods, variables were ranked according to their degree of variable importance. For each variable, ranks were summed across the three ML methods and those rank sums were sorted to predict the importance of variables across different methods. A quantitative method was also used to create a contingency table of the observed values. The row and column products were then standardized for the whole sample dataset as a table of expected values. For each variable in each ML method, we calculated the deviation observed from the expected value $(obs - exp)^2 / exp$. The least deviating values were predicted as the best-performing variables and ML methods. Thus, the variance in variable importance showed that the MaxEnt is more strongly deviating from the SVM and RF.

2.7. Ensemble modeling

As ensemble modeling binds together different models that have the highest precision and accuracy, it maximizes the prediction accuracy with a higher degree of reliability. As a result of the present study, the function 'ensemble' within the 'sdm' package was used to harmonize the results of the three occurrence prediction algorithms by using the area under the curve (AUC) weighted average approach. (Naimi and Araújo, 2016).

2.8. Models' accuracy validation

A 10-fold cross-validation approach was used to test the validity of our models. These folds were randomly split using the semi-automated

Table 2

R software packages used by 'sdm' in the parallel execution of the three models: namely random forest (RF), maximum entropy (MaxEnt), and support vector machines (SVM).

Algorithm	Syntax code in 'sdm'	Package used	Reference
Random forest	'rf'	randomForest	Liaw et al. (2002)
Maximum entropy	'maxent'	dismo	Hijmans et al. (2021)
Support vector machines	'svm'	Kernlab	Karatzoglou et al. (2004)

sampling approach in the ‘sdm’ package. The performances of the three models were evaluated using the receiver operating curve (ROC) by analyzing the AUC and true skills statistic values (TSS: Allouche et al., 2006). The values for the AUC range between 0 and 1. Imprecise models have an AUC of 0, while accurately predicting models have an AUC of 1. In principle, high model prediction performances are demonstrated by models with AUC > 0.7 (Mohammadi et al., 2019). In contrast, TSS (equations (2)–(4)) merges the sensitivity and specificity of the models to account for the model commission and omission errors (Kyalo et al., 2018). The values of TSS range between -1 to +1, where +1 demonstrates a perfect agreement between the observed and the predicted *C. moerenhouti* occurrence, while values ≤ 0 indicate no agreements or that most of the predictions for the *C. moerenhouti* occurrence were produced by chance (Allouche et al., 2006).

$$TSS = Sensitivity + Specificity - 1 \tag{2}$$

$$Sensitivity = \frac{a}{a + b} \tag{3}$$

$$Specificity = \frac{d}{c + d} \tag{4}$$

where *a* is true positive, *b* is a false negative, *c* is false positive, and *d* is true negative.

2.9. Model outputs visualization

The QGIS software version 3.20 was used to visualize the outputs of the three models and their respective ensembles (QGIS Development Team, 2021). Based on a suggestion by (Abdelal et al. (2019)), we reclassified our probability maps into five classes of *C. moerenhouti* habitat suitability and probability of occurrence. These classes were: (i)

very low probability (≤ 0.05), (ii) low probability (0.051–0.10), (iii) moderate probability (0.11–0.30), (iv) high probability (0.31–0.50), and (v) very high probability (≥ 0.50).

Using these five classes of suitability as highlighted in (i – v) above, we reclassified the pixel values in each of the ensemble images of the current, SSP245, and SSP585 using numerical values of 1–5 to allow for standardized numerical evaluation of the suitability losses caused by the potential climatic changes using the QGIS software version 3.20 (QGIS Development Team, 2021). Image differencing was used by subtracting the current image from the future (SSP245 and SSP585) probability of habitat suitability. Classes of gains (+) and losses (–) in suitability were obtained from the image differencing per each pixel and were mapped within the two counties i.e., Murang’a and Taita Taveta.

3. Results

3.1. Models’ accuracy, comparison, and validation

Details of the VIF values for the variables are provided in Table 1. The lowest values of VIF were obtained from aspect (1.55), LULC (2.00), and slope (2.27) while higher values of VIF were observed in bioclimatic variables (VIF > 9). The R² values for the selected variables were relatively low while high correlation values were observed against Bio4, Bio12, or LST (Fig. 2). Despite these values, these variables were not large enough to be eliminated from the model. The variables that had VIF values greater than 10, were excluded from our modeling experiments.

Using the ROC, the patterns of the smoothed graphs of the ten replicated ROCs showed that all three models were relatively consistent in their prediction amongst the model replicates (Fig. 3). Also, all the models generally showed very high accuracy in predicting *C. moerenhouti* habitat suitability in EABH, with all the models

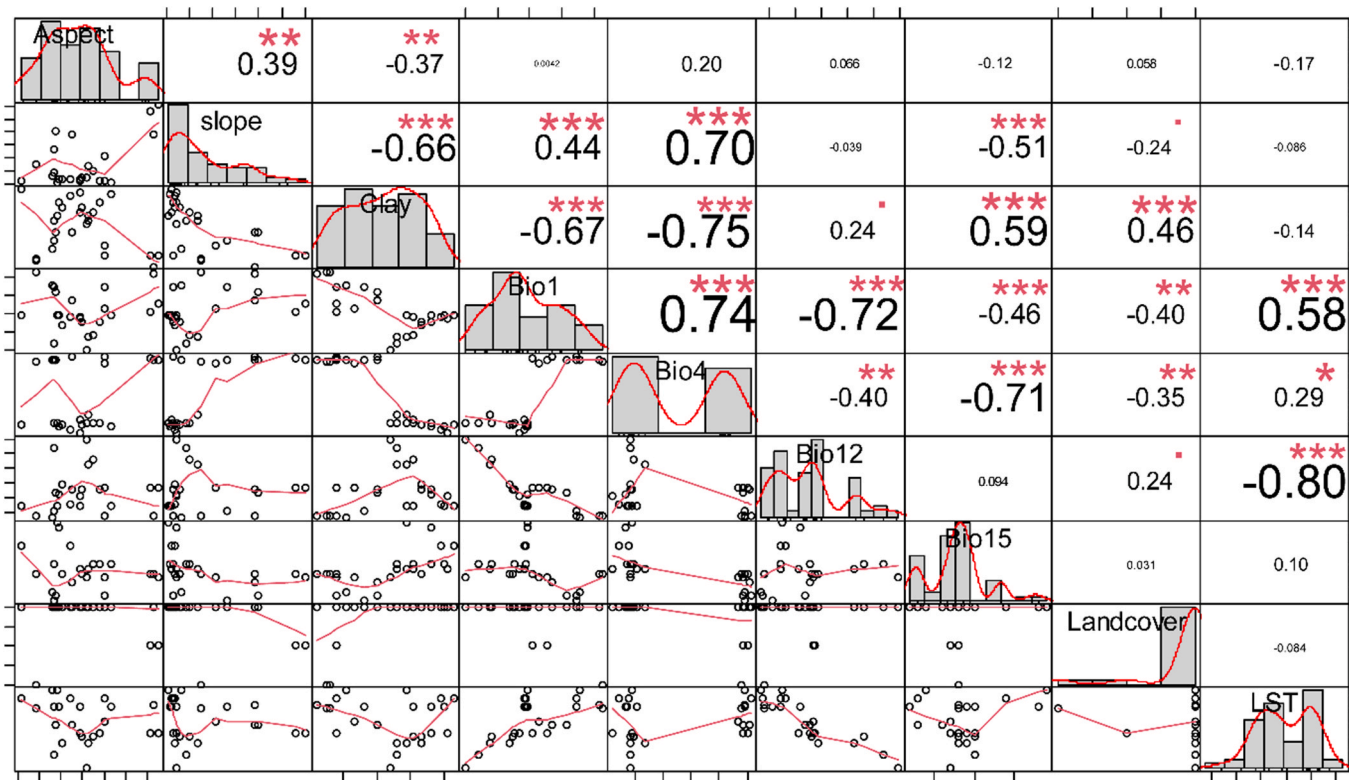


Fig. 2. Pairwise correlation scatter plots, variable histograms, and the correlation coefficient of the selected variables used in the modeling of *C. moerenhouti* habitat suitability in the Eastern Afrotropical biodiversity hotspot (EABH). The red stars represent the level of correlation i.e., 1 red star representing $R^2 < 0.4$, 2 red stars representing $0.41 < R^2 < 0.45$, and 3 red stars representing $R^2 > 0.45$. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

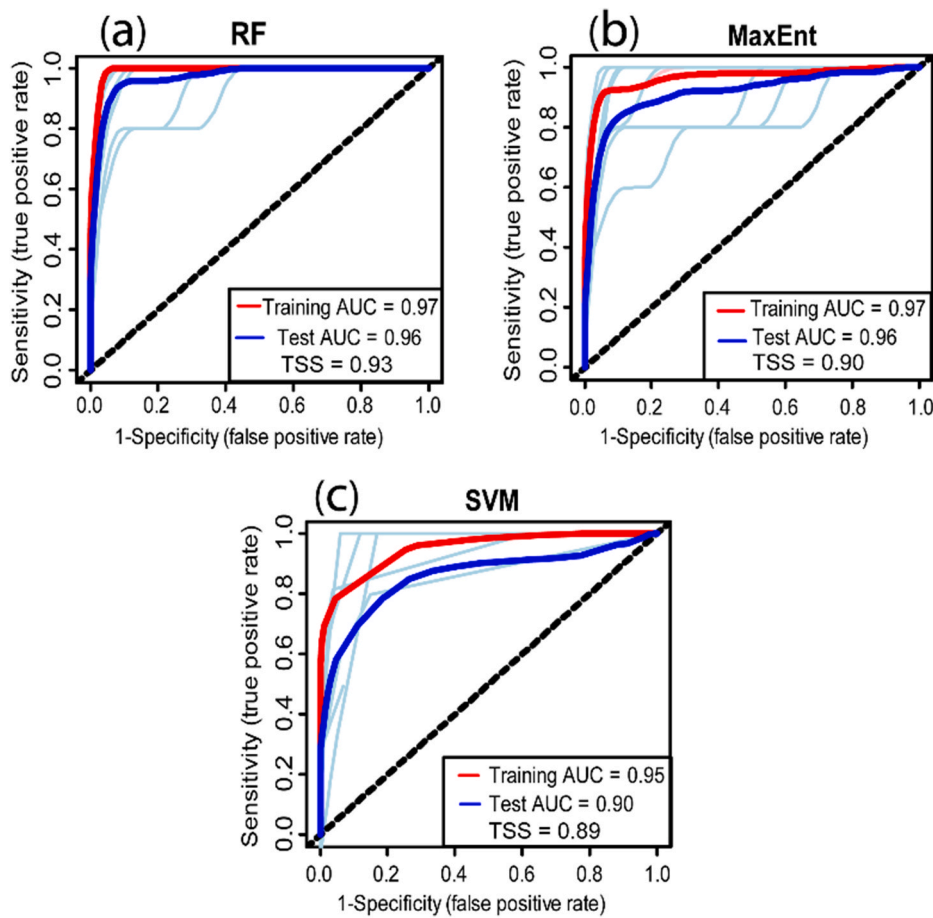


Fig. 3. Results of the receiver operating curve (ROC) for the three-machine learning (ML) and species distribution models (SDMs) used to predict *C. moerenhouti* habitat suitability in the Eastern Afromontane biodiversity hotspot (EABH) namely: (a) random forest (RF), (b) maximum entropy (MaxEnt), and (c) support vector machines (SVM). The red curves symbolize the smoothed mean area under the curve (AUC) using the training data, while the blue curve depicts the smoothed mean AUC using the test data from the 10-fold cross-validation sampling. The cyan curves show the 10-fold replicated model runs using the training data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

producing acceptable accuracies with $AUC \geq 0.90$ and TSS values ≥ 0.89 (Fig. 3).

3.2. Variable importance analysis

Nine out of the 32 predictor variables were selected and used in the final analysis. None of the models had a similar hierarchy of variable importance. However, LULC and Bio 15 appeared twice in the top five of two of the models, i.e., RF and MaxEnt. The three most important variables highlighted by the three models are clay content, LULC, and slope by RF, MaxEnt, and SVM, respectively. Although the three models showed no regular pattern regarding the variables' hierarchy, further analysis showed that bioclimatic variables were often appearing toward the top of the list (Fig. 4). The results of MaxEnt deviated strongly from RF over the aspect, Bio1, and slope, but agreed over LST, Bio4, 15, and Bio12. However, the rank sum approach suggested that the most important variables across the three models were clay, LULC, and Bio1 while slope was the least important.

3.3. Ensemble modeling and spatial predictions of the suitable habitat of the *Ceratina moerenhouti* in Murang'a and Taita Taveta

The three ML models using the nine predictor variables demonstrated diverse results for predicting *C. moerenhouti* probability of occurrence in EABH (Fig. 5). However, all three models predicted the *C. moerenhouti* ecological niche and occurrence to be mostly in the southern region in Murang'a and the central region in Taita Taveta particularly Taita Hills as shown by the warmer colors (yellow, orange, and red) in Fig. 5. Areas represented by the cooler colors (blue) were predicted to be very low or have low habitat suitability for the

C. moerenhouti. All the models show a general trend of decreasing suitability with climate change for both SSP245 and SSP585. In general, the ensemble models show that the *C. moerenhouti* suitable habitat and probability of occurrence is skewed towards the southern Murang'a and around the Taita Hills in central Taita Taveta. The current area suitable (moderate, high, and very high) for *C. moerenhouti* occurrence in Murang'a is 30% of the total area, while it is 10% in Taita Taveta. A decrease from the current suitable habitat of 2% and 4% is likely by the end of 2050, using the SSP245 climate scenario in Murang'a and Taita Taveta, respectively. Similarly, using the SSP585 climate scenario, a decline in the current suitable habitat of 4% and 5% is likely by the end of 2050 using the SSP585 climate scenario in Murang'a and Taita Taveta, respectively. Therefore, the approximate area currently suitable for the occurrence of *C. moerenhouti* in Murang'a is 697.8 km² from a total of 2326 km² while in Taita Taveta the currently suitable area is 1708.4 km² from a total country area of 17,084 km².

The image differencing showed that some potential gains and losses were likely to occur because of climate change. Fig. 6 shows that areas in the south of Murang'a and the peak of Taita Hills will likely experience the greatest losses in habitat suitability while some portions in the east of Murang'a county will gain in suitability. Most of the regions in the study area will remain unchanged within the very low to low suitability for the occurrence of the *C. moerenhouti* species (Fig. 6).

4. Discussion

The conservation of suitable habitats for pollinators enhances overall biodiversity. This also maintains the ecosystem services they provide, protects the soil integrity, improves the water quality by mitigating runoff, safeguards against soil erosion, and enriches rural aesthetics.

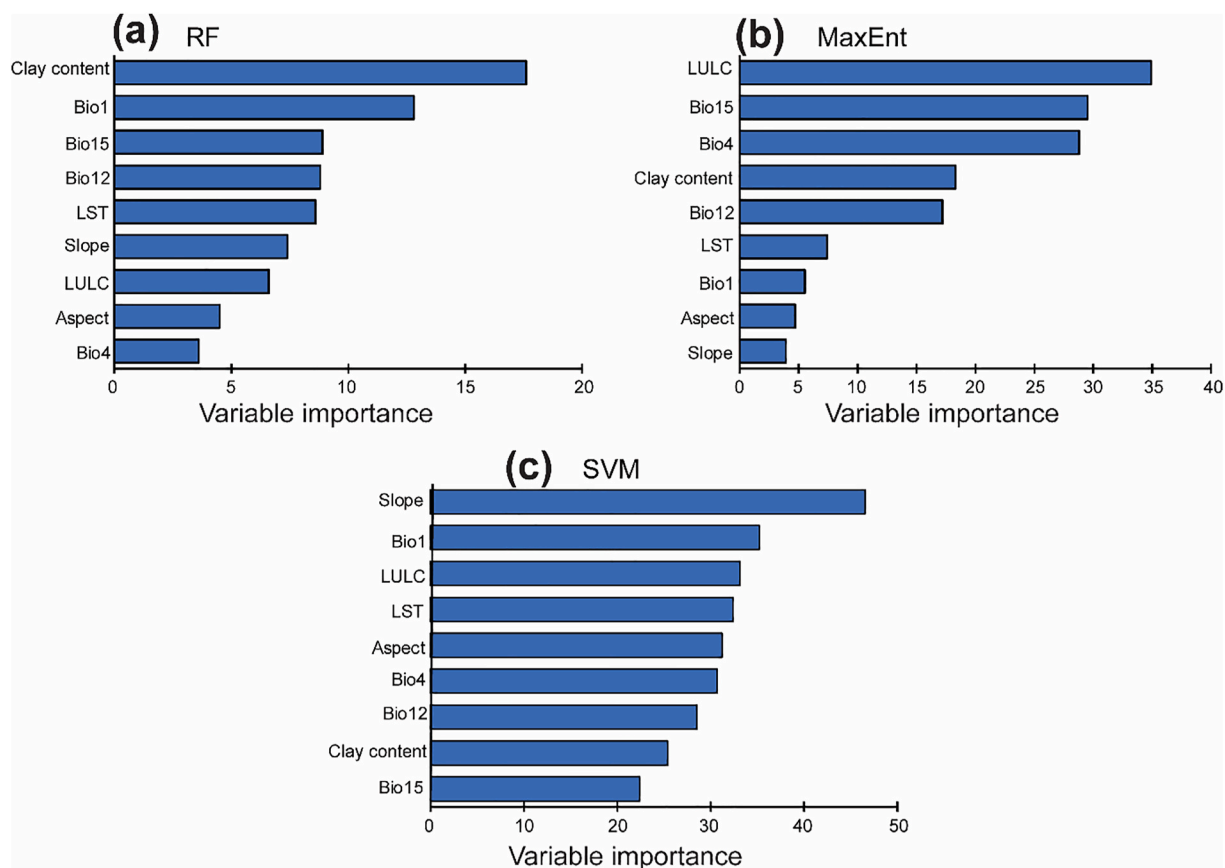


Fig. 4. The nine most important variables that were selected using the variance inflation factor (VIF) variable selection method for the three species distribution models used to predict *Ceratina moerenhouti* occurrence in Eastern Afromontane biodiversity hotspot (EABH) i.e., (a) random forest (RF), (b) maximum entropy (MaxEnt), and (c) support vector machines (SVM).

Integrating these secondary benefits into decision-making processes aids stakeholders to assess the trade-offs implicit in supplying ecosystem services (Wratten et al., 2012). Thus, in this current study, three of the most robust ML, MaxEnt, and SDMs were used to predict the potential current and future distribution as well as habitat suitability of one of the most important pollinator species i.e., *C. moerenhouti* in EABH. We followed, the best practice standards when conducting SDMs that include assessing the quality of the reference data, eliminating correlated predictor variables, performing model evaluation ideals, and building multiple models using the same data. Therefore, the results from this study can be relied on for use in decision-making within the study area.

4.1. Habitat suitability for the *C. moerenhouti* in the current climate conditions and under climate change scenario

Essentially, model input data pre-processing improves the robustness, precision, and dependability of the derived outputs (Araújo et al., 2019). All SDMs are designed to reflect the intrinsic interrelations and connections between the species, their niche, and the environment where they exist. The ‘usdm’ package and the ‘vifcor’ function enabled easy and pragmatic ways of eliminating the correlated variables systematically as well as performing the ML SDMs using the same data (Jafarian et al., 2019). In particular, the use of the VIF enabled the elimination of redundant and conflating variables from 32 to nine which improved the computation and accuracy of our modeling process. Other earlier studies have also successfully used the VIF to select a few non-correlated predicted variables (Abdelaal et al., 2019; Muposhi et al., 2016). The non-conflated variables ($n = 9$) that were finally used in the modeling experiments were crucial in explaining the occurrence of *C. moerenhouti*. Nevertheless, the variables that were selected by our

approach are relevant for modeling *C. moerenhouti* in EABH and can change if used elsewhere on the globe.

The results obtained in this study showed that the climatic variables i.e., temperature (Bio1 and Bio4) and precipitation (Bio12 and Bio15) together with clay content and LULC were central in predicting the habitat suitability of the *C. moerenhouti*. This concurred with the results reported by Makori et al. (2017), who attributed the distribution of most bee species to climatic and LULC variables. This can mainly be attributed to the influence of pedoclimatic variables to enable the establishment, growth, propagation, and success of flowering plants (Fox, 2019; Landmann et al., 2018; Petanidou et al., 2014). Using the SSP scenarios envisions the future of the world using trajectories of population growth, economic growth, trade development, technological development, and implementation of environmental policies (Riahi et al., 2017). In Africa, there is little anticipation of massive growth enough to reach the SSP585 climate scenario thus the most practical and expected scenario to affect the *C. moerenhouti* is the SSP245. However, the changes that may be observed in the EABH could be an indirect result of the massive developments and emissions from the developed world enough to warrant the observed potential loss in suitability that is likely to occur because of the SSP585 climate scenario. Climate change might reduce agricultural production by reducing precipitation and increasing temperature and CO₂ in most parts of the world (Niang et al., 2014). This reduction in cropping will result in the loss of flowering hosts for most pollinators (Janzen and Hallwachs, 2019). Additionally, agriculture and urbanization are the main drivers of plant diversity loss, with intensified systems using lethal doses of insecticides, fungicides, and herbicides that could contribute to negative impacts on wild and managed bees, including *C. moerenhouti*. Our results using the SSPs 245 and 585, point to a reduction in the most suitable habitat crucial for the survival of

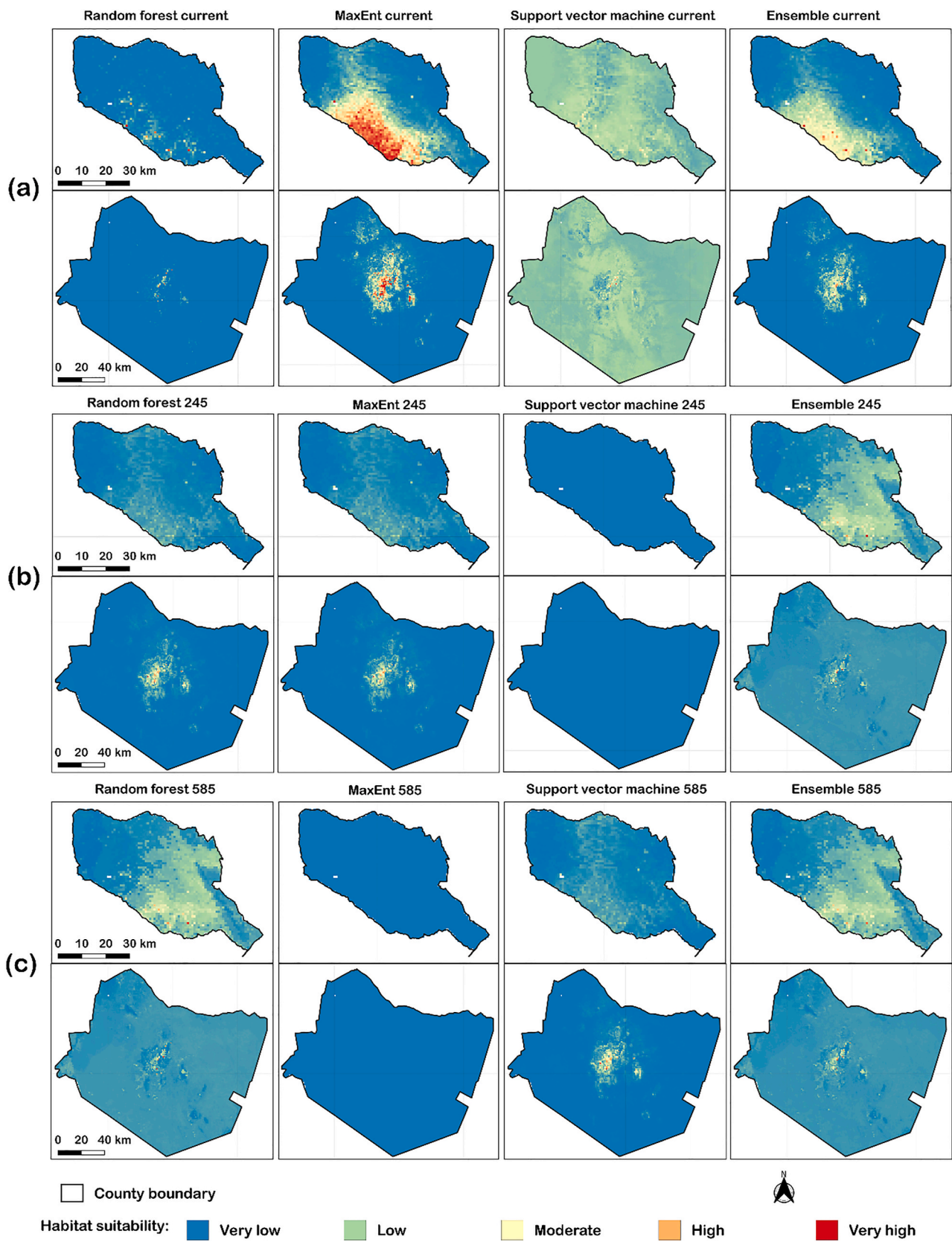


Fig. 5. Spatial probability of *C. moerenhouti* occurrence using nine predictor variables and three machine learning algorithms: (a) random forest (RF), (b) maximum entropy (MaxEnt), and (c) support vector machines (SVM). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

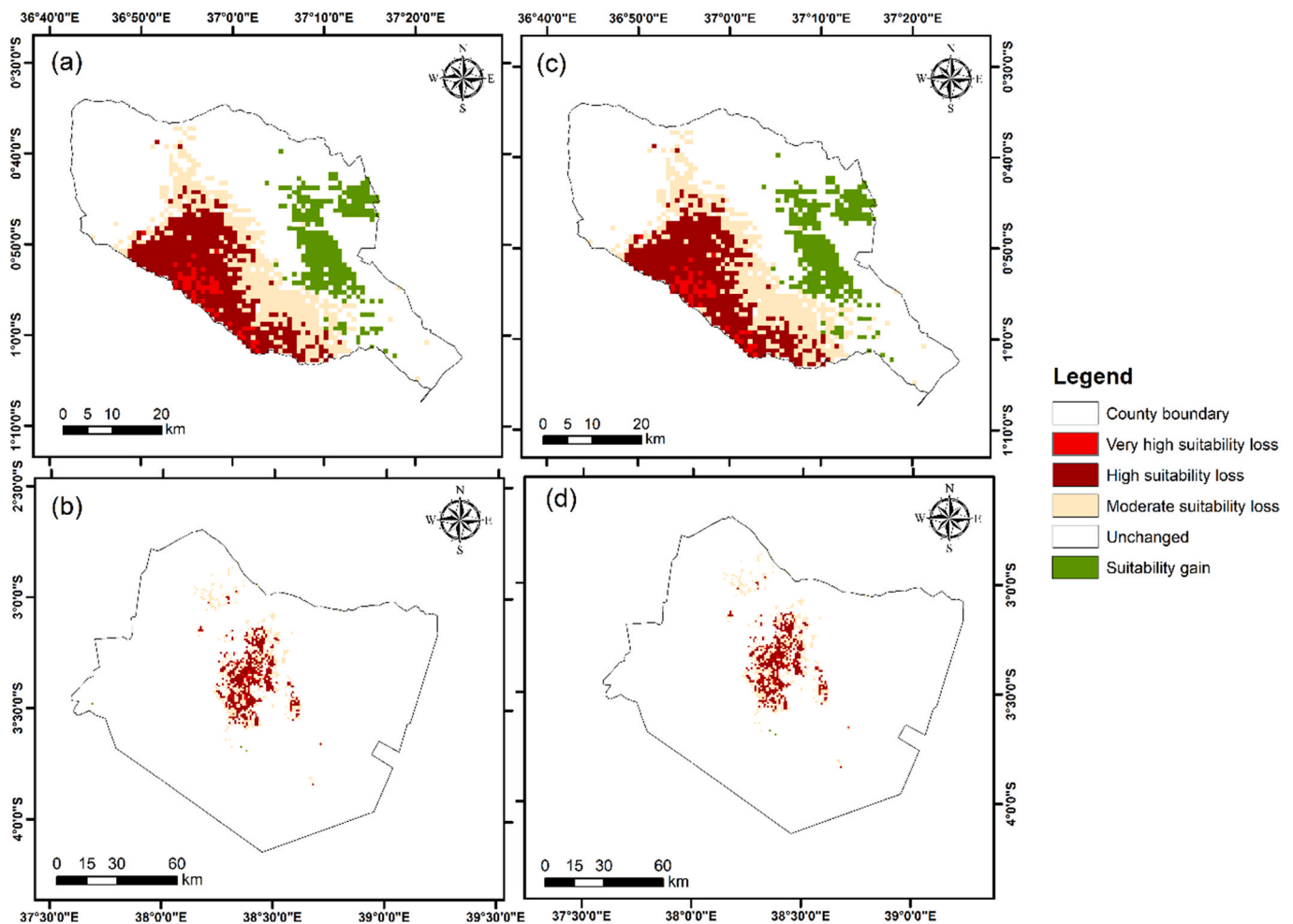


Fig. 6. Suitability losses and gains derived from the ensemble models in Murang'a and Taita Taveta using the two climate scenarios (a) and (b) SSP245 while (c) and (d) were for the SSP585 climate scenario. Red shows the areas that have the potential to experience the greatest loss in suitability while brown and yellow show high and moderate losses, respectively. The white color represents the areas that were unchanged while the green shows areas that positively gained suitability because of climate change. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

C. moerenhouti in the future. Because of the limited amount of forage and feed, *C. moerenhouti* populations and viability may be locally reduced or driven into extinction because of the loss of various floral resources from their diet. Therefore, any efforts targeted at curbing the loss of *C. moerenhouti* habitat in the future should focus on areas with the likelihood of temperature increase and a reduced amount of rainfall.

The genus *Ceratina* has a diverse distribution and can be found on every continent; with rare and limited distribution in Australia (Michener 2007). In Africa, *Ceratina* is widely distributed in South Africa, Liberia, Angola, Ethiopia, Sudan, Malawi, Botswana, Zimbabwe, Zambia, the coast of Mozambique, Burundi, Ghana, Uganda, Kenya, and Tanzania (Daly 1988; Eardley and Daly 2007; Eardley et al., 2009). The *C. moerenhouti* bees can be found in numerous habitats in East Africa such as rangelands, grasslands, woodlands, natural forests, open habitats, protected areas, farmlands, wetlands, marshlands, and riparian areas. The genus *Ceratina* comprises xylophilous mass-provisioners that nest in deadwood, stems, or piths by creating linear burrows in various locations mostly in shaded environments (Raju and Rao, 2006). The nesting biology of several species of *Ceratina* such as *C. smaragdula* (Kapil and Kumar, 1969), *C. calcarata* (Rehan and Richards, 2010) and *C. binghami* (Udayakumar and Shivalingaswamy, 2019) in *Rubus idaeus* L., *Rhus typhina* L., *Syzygium cumini* and *Cassia siamea* is well documented and conservation in different habitats can be achieved. On the

other hand, *C. moerenhouti* is known to nest on *Aloe littoralis*, *Hibiscus cannabinus*, *Hyparrhenia dissoluta*, *Jacaranda mimosifolia*, Poaceae (roof thatch), *Lantana camara*, *Moringa oleifera*. Thus, the damage to nesting habitats for these native bees due to deforestation, clearing of deadwood, and intense agricultural activities will have an immediate effect on the valuable pollination services they provide (Eardley et al., 2009). Declines in bee abundance and richness in agricultural landscapes, driven by habitat loss, have been observed several times (Kevan, 1999; Ricketts, 2004). Moreover, studies have also reported the loss of natural vegetation from exotic species, particularly in Taita Taveta where the exotic species are introduced for planned and purposeful agroforestry purposes. This reduces the potential plant biodiversity that can be realized within an area as the exotic species often become invasive.

Consequently, the indicative impact of climate change (weather variability) on the plant-pollinator interaction is the influence on the plant flowering period, which in turn can alter the pollinators' primary food sources (Petanidou et al., 2014). Thus, the co-dependency of plants and pollinators can be interrupted by climate changes when their life cycles are altered in diverse ways in response to the environment and human socioeconomic shifts (Schweiger et al., 2010). The relationship between plants and their pollinators is at stake for desynchronizing when environmental signals are distinct and when there is not enough overlap of generalist pollinator populations. For example, some insects'

larval stage is programmed to mature when local flowers first bloom and nectar begins to flow (Garibaldi et al., 2013; Forrest and Thomson, 2011). Even a polylectic bee like *C. moerenhouti* may be affected as both larval development and survival can differ by available pollen diet as influenced by the prevailing climate and climate variability. A 2007 study by Lindsey on the impacts of rising temperatures on a plant-pollinator network showed that seasonal cycle shifts reduced floral resources for 17%–50% of pollinator species. These studies reveal that temperature-based temporal mismatches between plants and pollinators do occur but vary by species and region (Lindsey 2007). Additionally, these bees are also efficient pollinators of crops such as apples, coffee, sesame, cowpeas, and beans which may also be affected by changes in LULC because of anthropogenic activities (Janzen and Hallwachs, 2019). The loss of a diversity of host plants for nesting and foraging may be detrimental to the *C. moerenhouti* as research has reported that disruption of visitations for pollination purposes and nesting resources by both biotic and abiotic factors can threaten their existence (Garibaldi et al., 2013). This is also supported by the importance of clay content to *C. moerenhouti* in this study as most high clay content areas have better fertility to host most crops hence the relevance of clay content identified in this study.

4.2. Model performances

Generally, SDMs with AUC and TSS beyond 0.7 are indicative of credible predictive and simulation performance (Elith et al., 2010). The AUC and TSS values obtained from the three models used in this study were above the 0.9 threshold, suggesting that the models performed well in simulating the distribution of *C. moerenhouti* in EABH. As would have been expected from models that run different mathematical functions, the model accuracies and the predicted areas differed across the three models (Araújo et al., 2019), hence must be treated with caution. Results from this study pointed to RF as the best predictive model for the *C. moerenhouti* distribution. However, the values of AUC and TSS for the MaxEnt and SVM did not differ much (± 1) hence these models can all be used to deliver similar accuracy within the EABH. These methods (RF, MaxEnt, and SVM) have also been used and suggested by many researchers as the best for simulating predictions for occurrence and mapping their geographical niches (Mudereri et al., 2020). However, the spatial outputs were visually different hence the use of the ensemble approach.

The varied outputs obtained in this study have been reported by other studies that have used multiple ML models in SDMs (Jafarian et al., 2019; Mohammadi et al., 2019). For instance, Jafarian et al. (2019) concluded that the ensemble method yielded high analytical and robust strength compared to the four predictive models that they used to simulate the occurrence of five dominant plant species in Iran. Again, Mohammadi et al. (2019) predicted two rodent species using MaxEnt and 'sdm' and they determined that all models were similar, and all showed high predictive power. Comparably, in this study, there is no convincing evidence to substantiate that one of these models is considerably better than the other. Therefore, the ensemble modeling approach is recommended to reduce the modeling uncertainties (Mtengwana et al., 2021; Schulz et al., 2021).

4.3. Variation of variable importance hierarchy from the three models

The mystery in using ML algorithms is their ability to provide accurate predictions without a realistic interpretation of the context of the application. Herein we dealt with a species in its ecological niche and were expecting all algorithms to provide similar variable importance and hierarchy to the contribution of the realized niche. Unfortunately, this was not the case as each algorithm presented a different level and hierarchy of variable importance. Moreover, this is brought about by the optimization problem in which we used supervised learning by inputting data with a set target. Therefore, a loss function ($Loss\ function = set\ target$

- *model outputs*) is established during computation, when the loss function estimates are largely diverted from the target, the errors are big, and hence algorithms provide poor predictions (Wang et al., 2020). This happens because, in all optimization problems where we need to minimize the errors, obtaining the global minimum is a challenge; most algorithms converge when a local minimum is found (Cassoli et al., 2012). To overcome such difficulty, an automated search of the best parameters can be used or a rank sum approach such as that used in this study. However, this process is time-consuming and requires high-performance computers (HPC) which are not always readily available.

4.4. Implications of our study

Modeling the potential habitat suitability of insects such as *C. moerenhouti* is useful in agricultural management systems. Our study supports national scale management strategies for such key pollinators that enhance productivity. Pollination by bees promotes sustainable development goals (SDGs) through food security (SDG2) and biodiversity (SDG15) and other less explored SDGs (Patel et al., 2021). This is because they offer an array of ecosystem services that support the welfare of people whilst sustaining the planet's life support systems. There is therefore a need for the establishment of flower-rich habitats within or around intensively farmed landscapes to increase the availability of pollen and nectar resources. Furthermore, results from the present study show that using SDMs is one of the most reliable and central tools for determining the fundamental and realized niche of species such as *C. moerenhouti* within a geographical space.

5. Conclusions

Our results showed that the type of LULC, clay content, temperature, and precipitation are the key drivers of the occurrence of *C. moerenhouti* that will result in the loss of suitable habitat in the future. Therefore, immediate action is critical to promote the growth of preferred flowering plants and suitable nesting sites for *C. moerenhouti*. This can be achieved by promoting flower-rich habitats within or around intensively farmed landscapes. Also, because of the loss of suitable habitat because of climate change, our results highlight the need to quickly employ and implement effective and early *C. moerenhouti* management options to reduce the chances of this loss of habitat before they go into extinction. Additionally, since the LULC emerged as one of the most relevant variables, studies that predict future scenarios of the *C. moerenhouti* occurrence or habitat suitability must consider the future LULC changes that are likely to occur. Although we used only part of Kenya as a case study, our modeling approach can be upscaled to the rest of the other biodiversity hotspots or countries that possess similar agroecological characteristics. Further, we assessed the utility of the ML SDMs i.e., MaxEnt, RF, and SVM, and their respective ensemble for predicting the probability of *C. moerenhouti* occurrence in EABH using multi-source bioclimatic and remotely sensed data. We established that RF, MaxEnt, SVM, and the ensemble approach, yield the most accurate *C. moerenhouti* occurrence prediction results in EABH.

Funding

"The authors gratefully acknowledge the financial support for this research by the following organizations and agencies: JRS Biodiversity Foundation (grant number: 60610); the Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Australian Centre for International Agricultural Research (ACIAR); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. Mukundi Mukundamago was supported by the German Academic Exchange Service (DAAD) In-Region Postgraduate Scholarship. The views expressed herein do not necessarily reflect the official opinion of the donors".

Ethical approval

Permission to work on farms within Kenya (Murang'a and Taita Taveta) was provided by National Commission for Science, Technology, and Innovation (NACOSTI) and Kenya Forest Service (KFS).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Michael G. Lattorff reports financial support was provided by JRS Biodiversity Foundation.

Data availability

<https://dmmg.icipe.org/dataportal/dataset/understanding-climate-change-effects-on-the-potential-distribution-of-an-important-pollinator>

Acknowledgments

We are grateful to the local community of villages in Murang'a and Taita Taveta counties for granting access to their farms and sharing knowledge. The following helped with fieldwork: Ephantus Guandaru Kimani, James Ng'ang'a Kimani, Joseph W. Kilongo, James Kimani Kinuthia, Darius K. Mwambala, and Benson M. Lombo. Our gratitude is extended to the Taita Taveta Research Station (TERRA) and Wundanyi, Kenya for research logistics and to Mr. Benard Malenge for structuring the maps.

References

- Abdelaal, M., Fois, M., Fenu, G., Bacchetta, G., 2019. Using MaxEnt modeling to predict the potential distribution of the endemic plant *Rosa arabica* Crep. in Egypt. *Ecol. Inf.* 50, 68–75. <https://doi.org/10.1016/j.ecoinf.2019.01.003>.
- 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. Rem. Sens.* 34, 712–728. <https://doi.org/10.1080/01431161.2012.713142>.
- Adan, M., Abdel-Rahman, E.M., Gachoki, S., Muriithi, B.W., Lattorff, H.M.G., Kerubo, V., Landmann, T., Mohamed, S.A., Tonnang, H.E.Z., Dubois, T., 2021. Use of earth observation satellite data to guide the implementation of integrated pest and pollinator management (IPPM) technologies in an avocado production system. *Remote Sens. Appl.: Soc. Environ.* 23, 100566 <https://doi.org/10.1016/j.rsase.2021.100566>.
- Aduvukha, G.R., Abdel-rahman, E.M., Sichangi, A.W., Makokha, G.O., Landmann, T., Mudereri, B.T., Tonnang, H.E.Z., Dubois, T., 2021. Cropping pattern mapping in an agro-natural heterogeneous landscape using Sentinel-2 and Sentinel-1 satellite datasets. *Agriculture* 11, 1–22. <https://doi.org/10.3390/agriculture11060530>.
- Allouche, O., Tsoar, A., 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>.
- Altieri, M.A., Nicholls, C.L., Gillespie, M., Waterhouse, B., Wratten, S., Gbehounou, G., Gemmill-Herren, B., 2015. *Crops, Weeds and Pollinators: Understanding Ecological Interactions for Better Management*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Araujo, M.B., Anderson, R.P., Barbosa, M.A., Beale, C.M., Dormann, C.F., Regan, E., 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, 1–12. <https://doi.org/10.1126/sciadv.aat4858>.
- Azrag, A.G.A., Murungi, L.K., Tonnang, H.E.Z., Mwenda, D., Babin, R., 2017. Temperature-dependent models of development and survival of an insect pest of African tropical highlands, the coffee Antestia bug *Antestiopsis thunbergii* (Hemiptera: pentatomidae). *J. Therm. Biol.* 70, 27–36. <https://doi.org/10.1016/j.jtherbio.2017.10.009>.
- Azrag, A.G.A., Pirk, C.W.W., Yusuf, A.A., Pinard, F., Niassy, S., Mosomtai, G., Babin, R., 2018. Prediction of insect pest distribution as influenced by elevation: combining field observations and temperature-dependent development models for the coffee stink bug, *Antestiopsis thunbergii* (Gmelin). *PLoS One* 13, 1–18. <https://doi.org/10.1371/journal.pone.0199569>.
- Baard, J.A., Kraaij, T., 2019. Use of a rapid roadside survey to detect potentially invasive plant species along the garden route, South Africa. *Koedoe* 61. <https://doi.org/10.4102/koedoe.v61i1.1515>.
- Bradie, J., Leung, B., 2017. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *J. Biogeogr.* 44, 1344–1361. <https://doi.org/10.1111/jbi.12894>.
- Cassoli, A., Di Lorenzo, D., Locatelli, M., Schoen, F., Sciandrone, M., 2012. Machine learning for global optimization. *Comput. Optim. Appl.* 51, 279–303. <https://doi.org/10.1007/s10589-010-9330-x>.
- Chemura, A., Mudereri, B.T., Yalaw, A.W., Gornott, C., 2021. Climate change and specialty coffee potential in Ethiopia. *Sci. Rep.* 11, 1–13. <https://doi.org/10.1038/s41598-021-87647-4>.
- Daly, H.V., 1988. *Bees of the New Genus Ctenoceratina in Africa, South of the Sahara (Hymenoptera: Apoidea)*, vol. 108. Univ of California Press.
- Dicks, L.V., Breeze, T.D., Ngo, H.T., Senapathi, D., An, J., Aizen, M.A., Basu, P., Buchori, D., Galetto, L., Garibaldi, L.A., Gemmill-Herren, B., Howlett, B.G., Imperatriz-Fonseca, V.L., Johnson, S.D., Kovacs-Hostyanszki, A., Kwon, Y.J., Lattorff, H.M.G., Lungharwo, T., Seymour, C.L., Vanbergen, A.J., Potts, S.G., 2021. A global-scale expert assessment of drivers and risks associated with pollinator decline. *Nat. Ecol. Evol.* 5, 1453–1461. <https://doi.org/10.1038/s41559-021-01534-9>.
- Discover, Life, 2022. Discover Life. <https://www.discoverlife.org>.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquez, J.R.G., Gruber, B., Lafourcade, B., Leitaó, P.J., Munkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schroder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Duan, J.J., Bauer, L.S., Van Driesche, R., Schmude, J.M., Petrice, T., Chandler, J.L., Elkinton, J., 2020. Effects of extreme low winter temperatures on the overwintering survival of the introduced larval parasitoids *Spathius galinae* and *Tetrastichus planipennis*: implications for biological control of Emerald ash borer in North America. *J. Econ. Entomol.* 113, 1145–1151. <https://doi.org/10.1093/jee/toaa048>.
- Eardley, C., Daly, H.V., 2007. *Bees of the Genus Ceratina Latreille in Southern Africa (Hymenoptera, Apoidea)*, vol. 13 (M. Schwarz).
- Eardley, C.D., Gikungu, M., Schwarz, M.P., 2009. Bee conservation in sub-saharan Africa and Madagascar: diversity, status, and threats. *Apidologie* 40, 355–366. <https://doi.org/10.1051/apido/2009016>.
- Elith, J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., Phillips, S.J., 2010. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.
- ESA, 2020. CCI Land Cover - S2 Prototype Land Cover 20m Map of Africa 2016 [WWW Document]. URL <http://2016africallandcover20m.esrin.esa.int/>. accessed 1.13.20.
- Eshetae, M.A., Hailu, B.T., Demissew, S., 2019. Spatial characterization and distribution modelling of *Ensete ventricosum* (wild and cultivated) in Ethiopia 36, 60–75. <https://doi.org/10.1080/10106049.2019.1588392>.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.E., 2007. The shuttle radar topography mission. *Rev. Geophys.* 45 <https://doi.org/10.1029/2005RG000183/FORMAT/PDF>.
- 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>.
- Forrest, J.R.K., Thomson, J.D., 2011. An examination of synchrony between insect emergence and flowering in Rocky Mountain meadows. *Ecol. Monogr.* 81 (3), 469–491.
- Fox, N., 2019. *Plant phenology and climate change: possible effect on the onset of various wild plant species first flowering day in the UK*. Master Thesis in Geographical Information Science. Department of Physical Geography and Ecosystem Science, Lund University.
- Gallai, N., Salles, J.M., Settele, J., Vaissiere, B.E., 2009. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* 68, 810–821. <https://doi.org/10.1016/j.ecolecon.2008.06.014>.
- Garibaldi, L.A., Steffan-Dewenter, I., Winfree, R., Aizen, M.A., Bommarco, R., Cunningham, S.A., Kremen, C., Carvalheiro, L.G., Harder, L.D., Afik, O., Bartomeus, I., Benjamin, F., Boreux, V., Cariveau, D., Chacoff, N.P., Dudenhofer, J. H., Freitas, B.M., Ghazoul, J., Greenleaf, S., Hipolito, J., Holzschuh, A., Howlett, B., Isaacs, R., Javorek, S.K., Kennedy, C.M., Krewenka, K.M., Krishnan, S., Mandelik, Y., Mayfield, M.M., Motzke, I., Munyuli, T., Nault, B.A., Otieno, M., Petersen, J., Pisanty, G., Potts, S.G., Rader, R., Ricketts, T.H., Rundlof, M., Seymour, C.L., Schuepp, C., Szentgyorgyi, H., Taki, H., Tscharntke, T., Vergara, C.H., Viana, B.F., Wanger, T.C., Westphal, C., Williams, N., Klein, A.M., 2013. Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* 340, 1608–1611. <https://doi.org/10.1126/science.1230200>, 1979.
- Ghisbain, G., Gerard, M., Wood, T.J., Hines, H.M., Michez, D., 2021. Expanding insect pollinators in the Anthropocene. *Biol. Rev.* 96, 2755–2770. <https://doi.org/10.1111/BRV.12777>.
- Hastie, T., Tibshirani, R., Buja, A., 1994. Flexible discriminant analysis by optimal scoring. *J. Am. Stat. Assoc.* 89, 1255–1270. <https://doi.org/10.1080/01621459.1994.10476866>.
- Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* 10, 1–26. <https://doi.org/10.1371/journal.pone.0125814>.
- Hijmans, R.J., 2020. raster: geographic data analysis and modeling. R package version 3.3-7. <https://CRAN.R-project.org/package=raster>.
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2021. Dismo: species distribution modeling. R package version 1.3-3. <https://CRAN.R-project.org/package=dismo>.
- IPCC, 2014. *Climate Change 2014. Synthesis Report*. Versión Inglés, Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth

- Assessment Report of the Intergovernmental Panel on Climate Change. <https://doi.org/10.1017/CBO9781107415324>.
- ITIS [Interagency Taxonomic Information System], 2022. Ceratina Latreille, 1802. On-Line Database. <http://www.itis.gov>. (Accessed 17 March 2022). accessed.
- Jafarian, Z., Kargar, M., Bahreini, Z., 2019. Which spatial distribution model best predicts the occurrence of dominant species in semi-arid rangeland of northern Iran? *Ecol. Inf.* 50, 33–42. <https://doi.org/10.1016/j.ecoinf.2018.12.011>.
- Janzen, D.H., Hallwachs, W., 2019. Perspective: where might be many tropical insects? *Biol. Conserv.* 233, 102–108. <https://doi.org/10.1016/j.biocon.2019.02.030>.
- Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan, E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-019-0826-1>.
- Kapil, R.P., Kumar, S., 1969. *Biology of Ceratina (Ceratinini: Hymenoptera)*. J. Res. Punjab Agric. Univ.
- Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. Kenlab - an S4 package for kernel methods in R. *J. Stat. Software* 11, 1–20.
- Kevan, P.G., 1999. Pollinators as bioindicators of the state of the environment: species, activity, and diversity. *Agric. Ecosyst. Environ.* 74, 373–393. [https://doi.org/10.1016/S0167-8809\(99\)00044-4](https://doi.org/10.1016/S0167-8809(99)00044-4).
- Korpela, E.L., Hyvonen, T., Lindgren, S., Kuussaari, M., 2013. Can pollination services, species diversity and conservation be simultaneously promoted by sown wildflower strips on farmland? *Agric. Ecosyst. Environ.* 179, 18–24. <https://doi.org/10.1016/j.agee.2013.07.001>.
- Kyalo, R., Abdel-Rahman, E., Mohamed, S., Ekesi, S., Christian, B., Landmann, T., 2018. Importance of remotely-sensed vegetation variables for predicting the spatial distribution of African citrus triozid (*Triosa erytraea*) in Kenya. *ISPRS Int. J. Geo-Inf.* 7, 429. <https://doi.org/10.3390/ijgi7110429>.
- Landmann, T., Feilhauer, H., Shen, M., Chen, J., Raina, S., 2018. Mapping the distribution and abundance of flowering plants using hyperspectral sensing. *Advanced Applications in Remote Sensing of Agricultural Crops and Natural Vegetation* 69–78. <https://doi.org/10.1201/9780429431166-4>.
- Landmann, T., Piironen, 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. *Rem. Sens. Environ.* 166, 50–60. <https://doi.org/10.1016/j.rse.2015.06.006>.
- Liaw, A., Wiener, M., Weiner, M., 2002. Classification and regression by random forest. *R. News* 2, 18–22. <https://doi.org/10.1177/154405910408300516>.
- Lindsey, R., 2007. Buzzing about Climate Change. National Aeronautics and Space Administration, Earth Observatory. <http://earthobservatory.nasa.gov/Features/Bees/bees.php>.
- Mahiga, S.N., Webala, P., Mware, M.J., Ndag'ang'a, P.K., 2019. Influence of land-use type on forest bird community composition in Mount Kenya forest. *Int. J. Ecol.* 1–8. <https://doi.org/10.1155/2019/8248270>, 2019.
- Makori, D., Mutanga, O., Irungu, J., Mosomtai, G., Odindi, J., Makau, S., Nkoba, K., Abdel-Rahman, E., Fombong, A., Ongus, 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>.
- Mesgaran, M.B., Cousens, R.D., Webber, B.L., 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Divers. Distrib.* 20, 1147–1159. <https://doi.org/10.1111/ddi.12209>.
- Michener, C.D., 2007. *The Bees of the World*, second ed. Johns Hopkins University Press, Baltimore, p. 953 <https://jhupbooks.press.jhu.edu/title/bees-world>.
- 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. Inf.* 52, 7–13. <https://doi.org/10.1016/j.ecoinf.2019.04.003>.
- Moshobane, M.C., Mudereri, B.T., Mukundamago, M., Chitata, T., 2022. Predicting future distribution patterns of *Jatropha gossypifolia* L. in South Africa in response to climate change. *South Afr. J. Bot.* 146, 417–425. <https://doi.org/10.1016/J.SAJB.2021.11.031>.
- Mtengwana, B., Dube, T., Mudereri, B.T., Dube, T., 2021. Modeling the geographic spread and proliferation of invasive alien plants (IAPs) into new ecosystems using multi-source data and multiple predictive models in the Heuningnes catchment, South Africa. *GIScience Remote Sens.* 1–18. <https://doi.org/10.1080/15481603.2021.1903281>, 00.
- Mudereri, B.T., Abdel-Rahman, E.M., Dube, T., Landmann, T., Khan, Z.R., Kimathi, E., Owino, R., Niassy, S., 2020. 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, B.T., Chitata, T., Chemura, A., Makaure, J., Mukanga, C., Abdel-Rahman, E.M., 2021. Is the protected area coverage still relevant in protecting the Southern Ground-hornbill (*Bucorvus leadbeateri*) biological niche in Zimbabwe? Perspectives from ecological predictions. *GIScience Remote Sens.* 58, 405–424. <https://doi.org/10.1080/15481603.2021.1883947>.
- Mudereri, B.T., Dube, T., Adel-Rahman, E.M., Niassy, S., Kimathi, E., Khan, Z., Landmann, T., 2019. A comparative analysis of PlanetScope and Sentinel-2 spaceborne sensors in mapping *Striga* weed using Guided Regularised Random Forest classification ensemble. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W13* 701–708. <https://doi.org/10.5194/isprs-archives-XLII-2-W13-701-2019>.
- Muposhi, V.K., Gandiwa, E., Chemura, A., Bartels, P., Makuza, S.M., Madiri, T.H., 2016. Habitat heterogeneity variably influences habitat selection by wild herbivores in a semi-arid tropical savanna ecosystem. *PLoS One* 11. <https://doi.org/10.1371/journal.pone.0163084>.
- Murang' a County Government, 2021. Murang' a county government: the official website for Murang' a county government [WWW Document]. URL <https://muranga.go.ke/>. accessed 11.1.21.
- Mutamiswa, R., Chikowore, G., Nyamukondiwa, C., Mudereri, B.T., Khan, Z.R., Chidawanyika, F., 2022. Biogeography of cereal stem borers and their natural enemies: forecasting pest management efficacy under changing climate. *Pest Manag. Sci.* <https://doi.org/10.1002/ps.7062>.
- Muthoni, F., Thierfelder, C., Mudereri, B.T., Manda, J., Bekunda, M., Hoeschle-Zeledon, I., 2021. Machine Learning Model Accurately Predicts Maize Grain Yields in Conservation Agriculture Systems in Southern Africa. 2021 9th International Conference on Agro-Geoinformatics. <https://doi.org/10.1109/Agro-Geoinformatics50104.2021.9530335>. *Agro-Geoinformatics 2021* 1–5.
- Naimi, B., Araujo, 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>.
- Ndayisaba, P.C., Kuyah, S., Midega, C.A.O., Mwangi, P.N., Khan, Z.R., 2021. Intercropping desmodium and maize improves nitrogen and phosphorus availability and performance of maize in Kenya. *Field Crop. Res.* 263, 108067 <https://doi.org/10.1016/j.fcr.2021.108067>.
- Niang, I., Ruppel, O.C., Abdrabo, M.A., Essel, A., Lennard, C., Padgham, J., Urquhart, P., 2014. In: Barros, V.R., Field, C.B., Dokken, D.J., Mastrandrea, M.D., Mach, K.J., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., W., L.L. (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Barros. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199–1265. <https://doi.org/10.1017/CBO9781107415386.002>. Africa.
- Nyabako, T., Mvumi, B.M., Stathers, T., Mlambo, S., Mubayiwa, M., 2020. Predicting *Prostephanus truncatus* (Horn) (Coleoptera: bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: a machine learning approach. *J. Stored Prod. Res.* 87, 101592 <https://doi.org/10.1016/j.jspr.2020.101592>.
- Ochungo, P., Veldtman, R., Abdel-Rahman, E.M., Raina, S., Muli, E., Landmann, T., 2019. Multi-sensor mapping of honey bee habitats and fragmentation in agro-ecological landscapes in Eastern Kenya. *Geocarto Int.* 1–22. <https://doi.org/10.1080/10106049.2019.1629645>, 0.
- Okello, E.N., Amugene, N.O., Mukiama, T.K., Lattorff, H.M.G., 2021. Abundance and community composition of flower visiting insects of avocado (*Persea americana* Mill) in the East African region. *Int. J. Trop. Insect Sci.* 41, 2821–2827. <https://doi.org/10.1007/s42690-021-00463-1>.
- Otieno, M.H.J., Aiyeko, M.A., Niassy, S., Salifu, D., Abdelmutalab, A.G.A., Fathiya, K.M., Subramanian, S., Fiaboe, K.K.M., Roos, N., Ekesi, S., Tanga, C.M., 2019. Integrating temperature-dependent life table data into Insect Life Cycle Model for predicting the potential distribution of *Scapsipedeus icipe* Hugel & Tanga. *PLoS One* 14, 1–27. <https://doi.org/10.1371/journal.pone.0222941>.
- Otunga, C., Odindi, J., Mutanga, O., Adjorlolu, C., Botha, J., 2017. Predicting the distribution of C3 (*Festuca* spp.) grass species using topographic variables and binary logistic regression model 489–504. <https://doi.org/10.1080/10106049.2016.1265598>.
- Petanidou, T., Kallimanis, A.S., Sgardelis, S.P., Mazaris, A.D., Pantis, J.D., Waser, N.M., 2014. Variable flowering phenology and pollinator use in a community suggest future phenological mismatch. *Acta Oecol.* 59, 104–111. <https://doi.org/10.1016/j.actao.2014.06.001>.
- Phillips, S.J., Anderson, R.P., Dudik, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40, 887–893. <https://doi.org/10.1111/ecog.03049>.
- Plant, R.E., 2012. *Spatial Data Analysis in Ecology and Agriculture Using R*. CRC Press: Taylor and Francis Group, California.
- QGIS Development Team, 2021. QGIS geographic information system. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rehan, S.M., 2020. Small carpenter bees (*Ceratina*). *Encycl. Soc. Insects* 1–4. https://doi.org/10.1007/978-3-319-90306-4_106-1.
- Rehan, S.M., Richards, M.H., 2010. Nesting biology and subsociality in *Ceratina calcarata* (Hymenoptera: apidae). *Can. Entomol.* 142, 65–74. <https://doi.org/10.4039/n09-056>.
- Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuarema, J.C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humenberger, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Streler, J., Drouot, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environ. Change* 42, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>.
- Ricketts, T.H., 2004. Tropical forest fragments enhance pollinator activity in nearby coffee crops. *Conserv. Biol.* 18, 1262–1271. <https://doi.org/10.1111/J.1523-1739.2004.00227.X>.

- Ropars, L., Affre, L., Schurr, L., Flacher, F., Genoud, D., Mutillod, C., Geslin, B., 2020. Land cover composition, local plant community composition and honeybee colony density affect wild bee species assemblages in a Mediterranean biodiversity hot-spot. *Acta Oecol.* 104 <https://doi.org/10.1016/j.actao.2020.103546>.
- Roulston, T.H., Smith, S.A., Brewster, A.L., 2007. A Comparison of Pan trap and Intensive net sampling techniques for documenting a bee (Hymenoptera: apiformes) fauna. *J. Kans. Entomol. Soc.* 80 (2), 179–181. [https://doi.org/10.2317/0022-8567\(2007\)80\[179:ACOPTAJ\]2.0.CO;2](https://doi.org/10.2317/0022-8567(2007)80[179:ACOPTAJ]2.0.CO;2).
- Sagwe, R.N., Peters, M.K., Dubois, T., Steffan-Dewenter, I., Lattorff, H.M.G., 2021. Pollinator supplementation mitigates pollination deficits in smallholder avocado (*Persea americana* Mill.) production systems in Kenya. *Basic Appl. Ecol.* 56, 392–400. <https://doi.org/10.1016/j.baae.2021.08.013>.
- Sango, I., Godwell, N., 2015. Climate change trends and environmental impacts in the Makonde Communal lands, Zimbabwe. *South Afr. J. Sci.* 111, 1–6. <https://doi.org/10.17159/sajs.2015/20140266>.
- Schatz, B., Drossart, M., Henry, M., Geslin, B., Allier, F., Savajol, C., Gerard, M., Michez, D., 2021. Pollinator conservation in the context of global changes with a focus on France and Belgium. *Acta Oecol.* 112 <https://doi.org/10.1016/j.actao.2021.103765>.
- Schulz, D., Yin, H., Tischbein, B., Verleysdonk, S., Adamou, R., Kumar, N., 2021. Land use mapping using Sentinel-1 and Sentinel-2 time series in a heterogeneous landscape in Niger, Sahel. *ISPRS J. Photogrammetry Remote Sens.* 178, 97–111. <https://doi.org/10.1016/j.isprsjprs.2021.06.005>.
- Schweiger, O., Biesmeijer, J.C., Bommarco, R., Hickler, T., Hulme, P.E., Klotz, S., Kuhn, I., Moora, M., Nielsen, A., Ohlemuller, R., Petanidou, T., Potts, S.G., Pysek, P., Stout, J.C., Sykes, M.T., Tscheulin, T., Vila, M., Walther, G.R., Westphal, C., Winter, M., Zobel, M., Settele, J., 2010. Multiple stressors on biotic interactions: how climate change and alien species interact to affect pollination. *Biol. Rev.* 85, 777–795. <https://doi.org/10.1111/J.1469-185X.2010.00125.X>.
- 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](https://doi.org/10.1016/S0304-3800(01)00388-X).
- Tabor, J.A., Koch, J.B., 2021. Ensemble models predict invasive bee habitat suitability will expand under future climate scenarios in Hawai'i, 2021 *Insects* 12. <https://doi.org/10.3390/INSECTS12050443>, 443 12, 443.
- Terzo, M., Rasmont, P., 2011. Atlas of the European bees: genus *Ceratina*. Atlas hymenoptera-atlas of the European bees-STEP project. Mons, Gembloux. <http://www.zoologie.umh.ac.be/hymenoptera/page.aspx?ID=192>.
- Tola, Y.H., Waweru, J.W., Hurst, G.D.D., Slippers, B., Paredes, J.C., 2020. Characterization of the Kenyan honey bee (*Apis mellifera*) gut microbiota: a first look at tropical and sub-Saharan African bee associated microbiomes. *Microorganisms* 8, 1–14. <https://doi.org/10.3390/microorganisms8111721>.
- Udayakumar, A., Shivalingaswamy, T.M., 2019. Nest architecture and life cycle of small carpenter bee, *Ceratina binghami* cockerell (xylocopinae: apidae: hymenoptera). *Sociobiology* 66, 61–65. <https://doi.org/10.13102/sociobiology.v66i1.3558>.
- Vapnik, V., 1979. *Estimation of Dependences Based on Empirical Data*. Nauka, Moscow, Translation, vol. 27. Springer Verlag, New York, pp. 5165–5184, 1982.
- Volis, S., Blecher, M., 2021. Present and future of *Iris atrofusca* populations in the Negev Desert analyzed by population viability analysis and species distribution modeling. *Acta Oecol.* 111, 103718 <https://doi.org/10.1016/j.actao.2021.103718>.
- Wan, Z., Hook, S., Hulley, G., 2015. MOD11C2 MODIS/Terra land surface temperature/emissivity 8-day L3 global 0.05Deg CMG V006 [Data set]. NASA EOSDIS land proc. DAAC. <https://doi.org/10.5067/MODIS/MOD11C2.006>.
- Wang, Q., Ma, Y., Zhao, K., Tian, Y., 2020. A Comprehensive survey of loss functions in machine learning. *Ann. Data Sci.* <https://doi.org/10.1007/s40745-020-00253-5>.
- Zattara, E.E., Aizen, M.A., 2021. Worldwide occurrence records suggest a global decline in bee species richness. *One Earth* 4, 114–123. <https://doi.org/10.1016/J.ONEEAR.2020.12.005>.
- Zurell, D., Elith, J., Schroder, B., 2012. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Divers. Distrib.* 18, 628–634. <https://doi.org/10.1111/j.1472-4642.2012.00887.x>.