Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting the Climatic Suitability of the Melon Fly (Zeugodacus cucurbitae) in Southeast Africa Using Ecological Niche Models

Luis Bota

luisbota@yahoo.com.br

Eduardo Mondlane University

Dércio Romão

Eduardo Mondlane University

Elemia Chidakwa

Ministry of Agriculture Irrigation and Water Development

Laura Canhanga

Eduardo Mondlane University

Lore Esselens

Royal Museum for Central Africa

Jenipher Tairo

Sokoine University of Agriculture

Joseph Ruboha

Sokoine University of Agriculture

Tomás Chiconela

Eduardo Mondlane University

Maulid Mwatawala

Sokoine University of Agriculture

Domingos Cugala

Eduardo Mondlane University

Massimiliano Virgilio

Royal Museum for Central Africa

Diana Erazo

Royal Museum for Central Africa

Research Article

Keywords: Zeugodacus cucurbitae, Ecological Niche Modelling, Boosted Regression Trees, MaxEnt, Southeast Africa

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11140294/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The melon fly, *Zeugodacus cucurbitae*, is a highly invasive tephritid pest of cucurbit crops and represents an increasing threat to horticultural production in Southeast Africa. Despite its continued spread, large areas of the region remain free of the pest, highlighting the need for spatially explicit risk assessments to support surveillance and preventive management. We assessed the potential climatic suitability of *Z. cucurbitae* across Southeast Africa using ecological niche models calibrated with 344 global occurrence records representing both the native and invaded ranges of the species. Two machine-learning algorithms, Boosted Regression Trees (BRT) and Maximum Entropy (MaxEnt), were developed using eight WorldClim bioclimatic variables selected following variance inflation factor (VIF) analysis. To reduce spatial sampling bias and account for spatial autocorrelation, we implemented a buffered pseudo-absence approach and evaluated model performance using spatial block cross-validation across 30 independent replicates. Model performance was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC) and complementary evaluation metrics. Both algorithms demonstrated high predictive performance, with mean AUC values of 0.924 (95% CI: 0.921–0.926) for BRT and 0.923 (95% CI: 0.920–0.925) for MaxEnt. Both models predicted a pronounced northeast–southwest gradient in climatic suitability, identifying eastern and northern Mozambique, southern Tanzania, Malawi, much of Zambia, and eastern Zimbabwe as the areas with the highest climatic suitability. Although the BRT model captured greater fine-scale spatial heterogeneity and MaxEnt produced smoother suitability gradients, both algorithms identified highly consistent regional patterns. Predictor rankings differed between algorithms, with BRT assigning greater importance to precipitation variables and MaxEnt emphasizing the temperature-related predictors. The strong agreement between modelling approaches indicates robust predictions and provides a scientific basis for strengthening surveillance, pest risk assessment, and preventive management, particularly in climatically suitable areas where the species has not yet been reported.

INTRODUCTION

The melon fly (*Zeugodacus cucurbitae* (Coquillett, 1899)), formerly known as *Dacus cucurbitae* Coquillett, 1899 and later *Bactrocera cucurbitae* (Coquillett, 1899) (De Meyer et al. 2015), is a true fruit fly species (Diptera: Tephritidae) of Central Asian origin. From there, it dispersed to a number of regions and countries in the rest of the Asian continent, Pacific, Oceania, Indian Ocean and Africa (Virgilio et al. 2010; De Meyer et al. 2015; Boontop et al. 2017). The species was first recorded in Africa in Tanzania in 1936, it was reported in 24 countries, including Mozambique and Malawi in the Southeast region (De Meyer et al. 2015; CABI 2024).

Zeugodacus cucurbitae is a polyphagous pest with a vast range of hosts, targeting mostly cucurbits (De Meyer et al. 2015). Minor host families include Anacardiaceae, Annonaceae, Cannellaceae, Oxalidaceae, Passifloraceae, Rutaceae and Solanaceae (Dhillon et al. 2005; De Meyer et al. 2015; CABI 2024). Within the Cucurbitaceae family, the melon fly exhibits varying preferences for each species depending on the region where it occurs (Mwatawala et al. 2006; Vayssières et al. 2007, 2008; Jacquard et al. 2013).

However, *Cucumis melo* L., *Cucurbita maxima* Duchesne, *Cucurbita pepo* L., and *Trichosanthes cucumerina* L. stands as the main host species (CABI 2024). The Cucurbitaceae family plays an important role in food security and in generating income for many households in the Southeast Africa, including Mozambique. In 2022, the cucurbit sector in this region generated approximately USD 154,231,000 from domestic sales and intra-regional exports (FAOSTAT 2024). Beyond food, cucurbit crops have been used medicinally by African families for decades (Olarewaju et al. 2021).

The invasion of *Z. cucurbitae* in Africa follows a spatially discontinuous and temporally prolonged pattern. After its first record in Tanzania in 1936, it was recorded in West Africa in 1999, but only dispersed into Southeast Africa decades later (Malawi: 2010 and Mozambique: 2013) (De Meyer et al. 2015; CABI 2024). In Mozambique, it remained confined to the north until expansion into central regions in 2021, while the south remains pest-free (De Meyer et al. 2015; Cugala & Canhanga 2023). Notably, it has not yet been reported from several countries in Southeast African region despite climatic and agricultural connectivity to the infested areas.

Temperature and Precipitation are recognized as the principal climatic factors regulating the development and population dynamics of *Z. cucurbitae*. Laboratory studies have shown that increasing temperatures accelerate immature development and increase survival, whereas lower temperatures prolong development and reduce survival (Vayssières et al. 2008; Mkiga & Mwatawala 2015; Ahn et al. 2022). Field studies further indicate that population abundance varies seasonally and geographically with climatic conditions, while rainfall influences host availability and seasonal population dynamics (Mwatawala et al. 2006; Vayssières et al. 2008; De Meyer et al. 2015). This delayed and nonlinear expansion for a highly invasive pest suggests that ecological and anthropogenic barriers may be limiting its establishment and spread. Understanding these factors is essential for predicting future risk scenarios, especially under climate change.

Ecological Niche Models (ENMs) are widely used to predict the potential distribution and invasion risk of animal and plant species (Elith & Leathwick 2009; Ghisbain et al. 2023; He et al. 2025). ENMs are generally classified into correlative and mechanistic approaches. Correlative models infer species-environment relationships from occurrence records and environmental predictors, whereas mechanistic models rely on physiological tolerances and functional traits. Among correlative approaches, MaxEnt (a machine learning algorithm based on the principle of maximum entropy) is one of the most widely applied methods because of its high predictive performance, robustness, and ability to model species distributions using presence-only data (Phillips et al. 2006; Duan et al. 2014; Kumar et al. 2014). Another powerful machine learning approach is Boosted Regression Trees (BRT), which combines decision trees with boosting techniques to sequentially improve predictive performance (Elith et al. 2008; Erazo et al. 2024). BRT can automatically capture complex nonlinear relationships and interactions among environmental variables, accommodate different data types, and is relatively robust to overfitting (Elith et al. 2008). Consequently, BRT is recognized as one of the best-performing algorithms for species distribution modelling and often performs as well as, or better than, traditional statistical methods, particularly when presence-absence data are available (Elith et al. 2008; Beery et al. 2021). MaxEnt has

been widely applied to predict the potential distribution of invasive tephritid fruit flies (Qin et al. 2015). In contrast, the application of BRT to this group appears to be less common despite its strong predictive performance in species distribution modelling. The complementary strengths of MaxEnt and BRT make their combined application valuable for identifying environmentally suitable areas and evaluating invasion risk.

Despite the successful application of ecological niche models to predict the potential distribution of *Z. cucurbitae* and other tephritid species (Kumar et al. 2014; Qin et al. 2015), studies specifically focused on Southeast Africa remain limited. This region is particularly relevant because the invasion has been delayed and spatially heterogeneous, with large pest-free areas persisting despite extensive agricultural production and active trade networks that could facilitate further spread. Identifying environmentally suitable areas and the environmental factors limiting establishment is therefore essential for early detection, surveillance, and management of future invasions. In this study, we applied MaxEnt and BRT models to (i) map the potential distribution of *Z. cucurbitae* in Southeast Africa and (ii) identify the key environmental factors constraining its distribution.

MATERIALS AND METHODS

Geographic Delimitation

This study focused on the Southeast Africa. We defined Southeast Africa as comprising the following countries: Mozambique, Malawi, Tanzania, Zambia, Zimbabwe, Botswana, Namibia, South Africa, eSwatini and Lesotho (Fig. 1).

Occurrence Data

A total of 557 georeferenced occurrence records of *Z. cucurbitae* were compiled from multiple sources: authors-led field survey conducted in Mozambique and Malawi from 2020 to 2025 (70 records), CABI invasive species compendium (107 records) (CABI 2024), and The Royal Museum for Central Africa database (380 records) (https://fruitflies.africamuseum.be/outputs/database). These records show the confirmed presences of the pest in Africa, Asia, Oceania and the Pacific. The data were cleaned by removing duplicate records and records with invalid or missing coordinates. The remaining 344 georeferenced occurrence records were used for model calibration (Table 1).

Table 1
Number of *Z. cucurbitae* occurrence records compiled and retained after data cleaning for ecological niche modelling by continent.

Continent	Compiled occurence records	Final occurrence records
Africa	474	263
Asia	73	71
Oceania	8	8
Pacific	2	2
Total	**557**	**344**

The data span included the native and global occurrence data recorded from 1895 to 2025. We used this dataset to capture the species realized ecological niche and support robust prediction in Southeast Africa. Similar approach was used by De Meyer et al. (2009), who used occurrence records and climatic predictors to model the potential geographic distribution of an invasive tephritid species. The models were trained using occurrence records from both native and invaded ranges, under the assumption that this combined dataset more completely defines the niche envelope of the species.

Environmental Data

We obtained nineteen bioclimatic variables, from WorldClim version 2.1 (Fick & Hijmans 2017), at a 10 arc-minute resolution using the geodata package on R studio (Hijmans et al. 2024). These variables represent long term climatic averages for the 1970–2000 period and include temperature and precipitation derived variables known to influence insect physiology and distribution (Table 2). Because the objective was to characterize the climatic niche of the species, the same bioclimatic layers were used to extract environmental values for all occurrence records irrespective of their collection year. The bioclimatic variables were retained at their global extent for model calibration, and the final ensemble predictions were subsequently cropped to Southeast Africa.

To reduce multicollinearity among the predictors, we performed a Variance Inflation Factor (VIF) analysis using the usdm package (Naimi et al. 2014), allowing the selection of the most informative variables. Further, we applied a conservative threshold of VIF < 5, and variables exceeding this threshold were removed. Eight environmental predictors were retained for modelling, namely Bio2, Bio3, Bio 8, Bio9, Bio13, Bio15, Bio18 and Bio19 (Table 3). In addition, we conducted a pairwise Pearson correlation analyses among the remaining predictors to ensure that no pairwise relationship exceed $|r| = 0.7$, indicating low redundancy among variables (Fig. 2).

Table 2
List of the 19 environmental variables

Variable	Symbol
Annual mean temperature (^{o}C)	Bio 1
Mean diurnal range (Mean of monthly (max temp - min temp)) (^{o}C)	Bio 2
Isothermality (%)	Bio 3
Temperature seasonality (standard deviation6100) (^{o}C)	Bio 4
Max temperature of warmest month (^{o}C)	Bio 5
Min temperature of coldest month (^{o}C)	Bio 6
Temperature annual range (^{o}C)	Bio 7
Mean temperature of wettest quarter (^{o}C)	Bio 8
Mean temperature of driest quarter (^{o}C)	Bio 9
Mean temperature of warmest quarter (^{o}C)	Bio 10
Mean temperature of coldest of quarter (^{o}C)	Bio 11
Annual precipitation (mm)	Bio 12
Precipitation of wettest month (mm)	Bio 13
Precipitation of driest month (mm)	Bio 14
Precipitation seasonality (coefficient of variation) (mm)	Bio 15
Precipitation of wettest quarter (mm)	Bio 16
Precipitation of driest quarter (mm)	Bio 17
Precipitation of warmest quarter (mm)	Bio 18
Precipitation of coldest quarter (mm)	Bio 19

Table 3
List of the 8 remaining environmental variables

Variable	Symbol	VIF
Mean Diurnal Range	Bio 2	3.95
Isothermality	Bio 3	1.98
Temperature of Wettest Quarter	Bio 8	1.82
Temperature of Driest Quarter	Bio 9	3.8
Precipitation of Wettest Month	Bio 13	3.77
Precipitation Seasonality	Bio 15	4.08
Precipitation of Warmest Quarter	Bio 18	2.07
Precipitation of Coldest Quarter	Bio 19	2.97

Software

All analysis and mapping were conducted in R version 4.3.3 (R Core Team 2024) using R studio environment.

Ecological Niche Modelling

Boosted Regression Trees (BRT)

We implemented the BRT model using the gbm package (Ridgeway & Developers 2024). BRT combines regression trees and boosting techniques to model complex nonlinear relationships and interactions among environmental predictors (Elith et al. 2008). The response variable follows Bernoulli distribution, representing species presence (1) and pseudo-absence (0).

We generated pseudo absence points across the global calibration extent using a buffered pseudo absence approach to reduce sampling bias. To account for uncertainty associated with pseudo-absence selection and model fitting, we repeated the analysis across 30 independent replicates. For each replicate, we sampled a new set of 1000 pseudo-absence points from area located outside a 20-km buffer surrounding the presence records using a different random seed. We then extracted environmental variables for all presence and pseudo-absence locations to build the modelling dataset. The selected number of pseudo-absence points provided sufficient environmental contrast while maintaining a moderate presence-to-pseudo-absence ratio.

We parameterized the model using 1500 trees, a learning rate of 0.01, tree complexity of 3 and a bag fraction of 0.7. We selected these settings to balance model flexibility while reducing the risk of

overfitting. Then, we evaluated model performance using spatial block cross-validation implemented through the cv_spatial() function of the blockCV package (Valavi et al. 2019). We partitioned the occurrence dataset into three spatial folds using a predefined block size of 100 km, with blocks randomly allocated to folds (selection = “random”). Folds allocation was optimized over 100 random allocation attempts (iteration = 100) to achieve a balanced spatial partition. The 100 km block size was selected a priori and was not estimated from the spatial autocorrelation range. This cross-validation framework ensured geographically independent model evaluation and reduced overly optimistic estimates of model performance resulting from spatial autocorrelation.

We assessed the model performance for each replicate using the Area Under the Receiver Operating Characteristic Curve (AUC) (Fielding & Bell 1997; Phillips et al. 2006). Although AUC is one of the most widely used metrics in ecological niche modelling, it has been criticized for its sensitivity to prevalence and for its potential to provide overly optimistic estimates of predictive performance (Lobo et al. 2008). Therefore, we additionally calculated the correlation between observed and predicted values, and the prevalence pseudo absence calibrated Sørensen’s index (SIppc) to provide a threshold-dependent assessment of the agreement between observed and predicted presences. The SIppc was calculated as (Ghisbain et al. 2023; Erazo et al. 2024):

$$SIppc = \frac{2TP}{2TP + xFP + FN}$$

Where:

$$x = \left(\frac{P}{A}\right) * \left(\frac{1 - Prevsp}{Prevsp}\right)$$

And

$$Prevsp = \frac{P}{P + A}$$

TP represents true positives, FP false positives, FN false negatives, P the number of presence points, and A the number of pseudo-absence points. The index ranged from 0 to 1, with values closer to 1 indicating better agreement between observed and predicted presences. Because model predictions were continuous probabilities, we converted to binary presence-absence predictions using a fixed threshold of 0.5 prior to calculation of the Sørensen Index. We further calculated evaluation metrics for each spatial fold and subsequently summarized them across the 30 replicates using mean values and 95% confidence intervals.

We averaged suitability predictions from the 30 replicate models to produce the final ensemble suitability map, with continuous suitability predictions ranging from 0 (unsuitable) to 1 (highly suitable) for each 10 arc-minute grid cell using the terra package (Hijmans 2025). We quantified spatial uncertainty by calculating the standard deviation (SD) and standard error (SE) among replicate

projections. Ensemble suitability and uncertainty maps were visualized using the tmap package (Tennekes 2018).

To assess the contribution of individual environmental predictors, we calculated the Relative Importance (RI) of each variable for every replicate model. RI values were estimated based on the frequency with which a variable was selected for tree splitting, weighted by the squared improvement in model fit associated with each split and averaged across all trees in the boosted model. We then summarized RI values as means and 95% confidence intervals across replicates.

We further examined species-environment relationships using Partial Dependence Plots (PDPs). PDPs describe the marginal effect of an environmental predictor on model prediction while holding the remaining predictor variable at their average values. For Boosted Regression Trees, these responses are expressed on the model's logit scale rather than directly habitat suitable values. Positive partial dependence values indicate an increase in the relative likelihood of occurrence, whereas negative values indicate decrease. We then generated PDP for all environmental predictors across the 30 replicate models and displayed all replicates response curves to visualize the stability of predictor effects and the uncertainty associated with model fitting.

MaxEnt

We implemented the MaxEnt modelling using the R package maxnet (Phillips 2021), which estimates species distribution by identifying the probability distribution of maximum entropy constrained by environmental conditions at species occurrence locations. We used the same occurrence records, buffered background (pseudo-absence) generation procedure, environmental predictors, 30-replicate modelling framework and spatial cross-validation procedure described for the BRT analysis.

For each replicate we fitted a MaxEnt model using selected bioclimatic variables and evaluated predictive performance using spatially independent training and testing datasets. We then averaged suitability predictions from 30 replicate models to generate the final ensemble suitability map, with continuous suitability values ranging from 0 (unsuitable) to 1 (highly suitable). Following the same procedure adopted for the BRT model, we quantified spatial uncertainty by calculating the standard deviation (SD) and standard error (SE) among replicate suitability projections. Ensemble suitability and uncertainty maps were visualized using the tmap package (Tennekes 2018).

We assessed model performance using the AUC and True Skill Statistic (TSS). We calculated both metrics for each spatial fold and subsequently summarized them across the 30 replicates using mean values and 95% confidence intervals.

To evaluate the contribution of environmental predictors, we estimated permutation-based RI for each replicate by randomly permuting each predictor and quantifying the resulting change in model predictions (Phillips et al. 2006; Elith et al. 2011). Predictor importance was calculated as one minus the Pearson correlation between the original and permuted predictions and expressed as relative

percentages. RI values were summarized across replicates as mean percentages with corresponding 95% confidence intervals.

Following the same approach adopted for the BRT model, we examined species-environment relationships using MaxEnt response curves. For each predictor, climatic suitability was estimated across the observed environmental gradient while holding all remaining predictors constant at their mean values. Response curves were generated for each of the 30 replicate models and displayed together to assess the stability and variability of predictor effects among replicate runs (Phillips et al. 2006; Elith et al. 2011).

RESULTS

Model Performance

BRT and MaxEnt showed very similar AUC values under spatial block cross validation, indicating comparable discriminatory performance. Both algorithms consistently discriminated between suitable and unsuitable environments, indicating comparable predictive accuracy. The BRT model achieved a mean AUC of 0.924 (95% CI: 0.921–0.926), together with a mean correlation of 0.7 (95% CI: 0.693–0.706) between observed and predicted values and a mean Sørensen similarity index of 0.716 (95% CI: 0.71–0.723) across the 30 replicate models. The MaxEnt model showed comparable predictive performance, with a mean AUC of 0.923 (95% CI: 0.920–0.925) and a mean TSS of 0.746 (95% CI: 0.742–0.750) (Supplementary Tables S1 and S2).

Relative Importance of Environmental Predictors

Both modelling approaches identified a small subset of climatic variables as the primary determinants of *Z. cucurbitae* suitability across Southeast Africa (Fig. 3; Supplementary Tables S3 and S4). In the BRT model, precipitation of the wettest month was the most influential predictor (RI = 35.1%, 95% CI: 33.4–36.8%), followed by isothermality (RI = 23.1%, 95% CI: 21.3–24.8%) and mean temperature of the wettest quarter (RI = 10.7%, 95% CI: 10.0–11.4%). The remaining predictors each contributed less than 8% to the model. In the MaxEnt model, mean temperature of the driest quarter showed the highest relative importance (RI = 35.5%, 95% CI: 33.9–37.1%), followed by mean temperature of the wettest quarter (RI = 24.3%, 95% CI: 23.6–25.0%) and precipitation of the wettest month (RI = 19.1%, 95% CI: 17.5–20.6%). All remaining predictors individually contributed less than 11% to the model.

Although the relative importance of individual predictors differed between modelling approaches, both models consistently identified precipitation of the wettest month and mean temperature of the wettest quarter among the most influential climatic predictors.

Species-environment Relationships

We observed nonlinear relationships between environmental conditions and the predicted climatic suitability of *Z. cucurbitae* (Figs. 4 and 5). Although the two modelling approaches differed in the relative importance assigned to individual predictors and in the shape of the fitted response curves, the overall response patterns were highly consistent across the 30 replicate models, supporting the robustness of the estimated climatic niche.

The BRT model revealed pronounced threshold responses for several climatic predictors. Partial dependence plots showed a strong positive effect of precipitation of the wettest month, with suitability increasing rapidly at approximately 150–250 mm before stabilizing. Suitability also increased with isothermality, precipitation seasonality, mean temperature of the driest quarter, and mean temperature of the wettest quarter, particularly above approximately 15°C. In contrast, mean diurnal range showed a negative relationship with suitability, whereas precipitation of the warmest quarter and precipitation of the coldest quarter exhibited more complex nonlinear responses, with suitability increasing only at relatively high precipitation levels.

The MaxEnt model showed broadly similar response patterns, although the fitted curves were generally smoother. Climatic suitability increased progressively with mean temperature of the driest quarter, reaching its highest values at approximately 15–20°C, before declining slightly at higher temperatures. Suitability also increased with mean temperature of the wettest quarter, peaking at approximately 20–25°C. Among precipitation variables, suitability increased rapidly with precipitation of the wettest month, reaching maximum values at approximately 150–250 mm, and then gradually declined under wetter conditions. Suitability also increased with precipitation seasonality, whereas mean diurnal range showed a negative relationship, with suitability decreasing as daily temperature fluctuations increased.

Predicted Climatic Suitability across Southeast Africa

Both modelling approaches predicted a clear northeast–southwest gradient in climatic suitability across Southeast Africa (Fig. 6). Climatic suitability was highest throughout eastern and northern Mozambique, southern Tanzania, Malawi, and much of Zambia, with similarly high values extending into parts of eastern Zimbabwe, eSwatini and Mpumalanga and Kwazulu Provinces in South Africa. Suitability gradually decreased towards central Mozambique and western Zambia and was lowest across Botswana, southern Namibia, Lesotho, and much of South Africa, where predominantly low suitability values were predicted.

The BRT and MaxEnt models produced highly consistent regional patterns of climatic suitability, identifying the same broad gradients of climatic suitability across the region. However, the BRT model exhibited greater fine-scale spatial heterogeneity, with more localized variation in predicted suitability, whereas the MaxEnt model generated smoother and more spatially continuous suitability gradients across the study area.

Prediction Uncertainty

Prediction uncertainty, expressed as the standard deviation (SD) among the 30 replicate predictions, was generally low across Southeast Africa for both modelling approaches (Fig. 7). The MaxEnt model exhibited consistently lower prediction uncertainty across most of the study area, with uniformly small SD values indicating high agreement among replicate predictions. In contrast, the BRT model showed greater spatial variability in prediction uncertainty, with moderately higher SD values occurring across several regions, particularly in parts of northern and eastern Southeast Africa. Despite these differences, SD values remained relatively low overall for both models, indicating that the predicted patterns of climatic suitability were stable across the 30 independent model replicates.

Discussion

Zeugodacus cucurbitae has become one of the most economically important invasive pests of cucurbit crops throughout sub-Saharan Africa, threatening food production, farmer livelihoods and regional trade. In Tanzania, where cucurbit production is concentrated in favourable agroecological zones, recent studies have confirmed that *Z. cucurbitae* dominates cucurbit fruit communities and remains the principal constraint to sustainable cucurbit production (Tarimo et al. 2025; Bwire et al. 2026). This study provides the first ecological niche modelling assessment specifically developed for Southeast Africa using globally calibrated occurrence data and spatially independent validation. By combining MaxEnt and Boosted Regression Trees, we identified the major climatic constraints governing the establishment of *Z. cucurbitae* while producing robust regional suitability maps that can directly support phytosanitary decision-making.

Our models indicate that climatic suitability is shaped by both moisture availability and seasonal temperature conditions, with the relative importance of these drivers differing between BRT and MaxEnt. Seasonal rainfall influences cucurbit phenology, fruit availability and soil moisture, thereby affecting oviposition success, larval survival and the persistence of adult populations. These findings agree with field studies in Tanzania, where the seasonal abundance of *Z. cucurbitae* closely followed rainfall patterns and host fruit availability, with populations increasing during wetter periods. Likewise, recent surveys in the Morogoro region identified *Z. cucurbitae* as the dominant cucurbit fruit fly across contrasting agroecological zones, although population densities varied markedly between seasons and environments (Mwatawala et al. 2006; Mwatawala et al. 2009; Tarimo et al. 2025; Mwatawala et al. 2026). Our regional models extend these local ecological observations by demonstrating that precipitation not only regulates seasonal population dynamics but also shapes the broad-scale climatic suitability of the melon fly across Southeast Africa. This finding suggests that the same climatic processes operating at local scales also determine regional invasion potential.

Although precipitation primarily determined where suitable habitats occur, temperature further refined the realized climatic niche of *Z. cucurbitae*. Temperature variables further refined the estimated climatic niche, but their relative importance differed between algorithms, with BRT emphasizing isothermality and MaxEnt emphasizing mean temperature of the driest and wettest quarters. Greater thermal variability is likely to impose physiological stress, reducing survival and reproductive performance even under

otherwise favourable environmental conditions. This interpretation is consistent with laboratory studies showing that development, survival, fecundity and longevity of *Z. cucurbitae* are strongly temperature dependent (Vargas et al. 1996; Huang & Chi 2012; Mkiga & Mwatawala 2015; Ghodekar et al. 2025). Together, these findings indicate that precipitation primarily determines where suitable habitats occur, whereas temperature governs the physiological performance, development and reproductive success of populations within those habitats, particularly during the wet season when host availability is greatest.

Our results indicate that large parts of the region currently provide favourable climatic conditions for the establishment and persistence of the melon fly, with a pronounced northeast–southwest gradient. The highest suitability occurred in eastern Mozambique, Southern Tanzania and Northern Malawi, regions characterized by relatively high precipitation, stable thermal conditions and intensive cucurbit production. This spatial pattern is biologically plausible because it coincides with the principal cucurbit-producing landscapes of eastern Africa, where *Z. cucurbitae* has consistently been reported as the dominant pest of cucurbit crops (Mwatawala et al. 2006; Mwatawala et al. 2009; Tarimo et al. 2025; Bwire et al. 2026). Conversely, suitability declined towards the drier inland areas and southern part of the region, reflecting the combined effects of reduced precipitation and greater temperature variability (Mwatawala et al. 2006; Mwatawala et al. 2009; De Meyer et al. 2015).

Our models also identified climatically suitable areas beyond the current documented distribution of *Z. cucurbitae*, particularly in southern Mozambique, eSwatini and parts of Mpumalanga and Kwazulu Natal Provinces in South Africa. Such discrepancies between predicted suitability and known occurrence records are common in ecological niche modelling and may reflect incomplete surveillance, recent dynamics, dispersal limitations or anthropogenic barriers rather than unsuitable environmental conditions (Broennimann et al. 2007; Kumar et al. 2014). Alternatively, these areas may represent climatically suitable regions that have not yet been colonized because of dispersal constraints or limited human-mediated introduction. This is particularly relevant because southern Mozambique remains free of the pest despite extensive cucurbit production and active trade with infested regions. Considering the strategic importance of Southern Mozambique and parts of Mpumalanga and Kwazulu Natal Provinces in South Africa for agricultural production and regional trade, these areas should be prioritized for surveillance, monitoring and early detection programmes.

Our findings agree with previous ecological niche modelling studies conducted for other economically important tephritid fruit flies, suggesting broadly similar climatic controls across invasive species despite differences in host range and invasion history. Similar climatic patterns have been reported for *Bactrocera dorsalis*, in which precipitation and temperature jointly determined the species' potential distribution (De Meyer et al. 2009). These similarities suggest that precipitation and temperature are the principal climatic drivers governing fruit fly establishment and invasion across tropical regions.

Despite minor differences in spatial detail, BRT and MaxEnt converged on remarkably similar regional suitability patterns, increasing confidence in the robustness of the predictions. BRT produced more heterogeneous suitability surfaces and captured finer-scale environmental variation, whereas MaxEnt

generated smoother climatic suitability surfaces. These differences are consistent with greater flexibility of BRT in modelling complex nonlinear responses and interactions among predictors (Elith et al. 2008), while MaxEnt generally produces broader suitability gradients from presence-background data (Phillips et al. 2006). Nevertheless, both algorithms consistently identified the same principal hotspots and the same dominant climatic drivers.

Both models also showed good predictive performance under spatial block cross-validation, with BRT performing marginally better than MaxEnt. The use of spatially structured validation is particularly important because it reduces the inflation of predictive accuracy commonly associated with random partitioning and provides a more realistic assessment of model performance (Dormann et al. 2007; Roberts et al. 2017; Valavi et al. 2019; Erazo et al. 2024). Likewise, the use of global occurrence records enabled the models to capture a broader representation of the realized climatic niche represented across both the native and invaded ranges of *Z. cucurbitae*, thereby increasing the robustness of predictions within Southeast Africa.

From a phytosanitary perspective, the predicted suitability maps provide a practical framework for prioritizing surveillance and monitoring efforts across Southeast Africa. High-risk areas identified consistently by both modelling approaches should receive intensified trapping, early detection activities and awareness campaigns to facilitate rapid detection and prevent establishment. Conversely, areas that are currently free of the pest but predicted to be climatically suitable should be prioritized for contingency planning, quarantine preparedness and the protection of Pest Free Areas (PFAs). The suitability maps may also support pest risk analysis, delimiting surveys following new detections, and the regional harmonization of surveillance programmes among neighbouring countries (FAO 2017a; FAO 2017b; FAO 2017c).

Our study focused exclusively on climatic suitability and therefore predicts the potential climatic distribution of *Z. cucurbitae*. Other ecological and anthropogenic factors that influence establishment and spread were not considered. Future studies should integrate these other factors such as host availability, competition with other cucurbit feeders which could offer significant insight into the ecological suitability of the melon fly, land use practices, irrigation, trade pathways and human mediated dispersal, as they can also influence establishment and spread. These ecological and anthropogenic drivers, together with projected climatic change scenarios, would provide a more comprehensive assessment of invasion risk and better support regional surveillance, integrated pest management and phytosanitary planning across Southeast Africa.

Conclusion

This study provides the first regional assessment of the climatic suitability of *Z. cucurbitae* across Southeast Africa using globally calibrated ecological niche models. Both BRT and MaxEnt identified precipitation of the wettest month and seasonal temperature variables as key contributors to climatic suitability, although their relative importance differed between algorithms. The strong agreement

between modelling approaches increases confidence in the predicted suitability patterns and the identification of areas climatically favourable for establishment. The resulting suitability maps provide a valuable basis for pest risk assessment by identifying priority areas for surveillance, early detection and protection of pest-free areas. Although the present study considered only climatic factors, the integration of host distribution, land use, trade pathways, human-mediated dispersal and future climate change scenarios would further improve invasion risk assessments and strengthen evidence-based phytosanitary planning across the region.

Declarations

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

This research was supported by the Belgian Directorate-General for Development Cooperation and Humanitarian Aid (DGD) through the Royal Museum for Central Africa (RMCA), Belgium, under DISPEST II Project. The funding body had no role in the design of the study, data collection, analysis, interpretation of the results, or writing of the manuscript.

Author Contribution

LB, DC, LC and MV conceived and designed the study. LB conducted fruit fly trapping in some regions, compiled the occurrence database, performed all ecological niche modelling analyses, prepared the figures and tables, and wrote the first draft of the manuscript. DE provided methodological guidance on ecological niche modelling and spatial validation and critically revised the manuscript. MV contributed to the conceptual development of the study, methodological refinement, interpretation of the results, and critical revision of the manuscript. DC supervised the research and contributed to the interpretation of the results and manuscript revision. EC conducted trapping for data collection in Malawi. LC, LE, DR, JT, JR, TC and MM contributed to data interpretation and critically reviewed the manuscript. All authors read, revised, and approved the final manuscript.

Acknowledgement

This research was supported and financed by the Directorate General for Development Cooperation and Humanitarian Aid through the DISPEST project under the 2024–2028 framework agreements with RMCA. The authors are grateful to Dr. Marc De Meyer for his long-standing support to fruit fly research in Africa and for his valuable comments on an earlier version of this manuscript, and for providing access

to the major part of the occurrence records used in this study, and all colleagues who contributed to field surveys, fruit fly monitoring, and constructive discussions during the development of this research.

Data Availability

The datasets generated and analyzed during the current study, including the modelling workflow and derived model outputs, are available from the corresponding author upon reasonable request. Occurrence records obtained from the CABI Invasive Species Compendium and the Royal Museum for Central Africa (RMCA) are not publicly redistributable by the authors and are subject to the data-sharing policies and usage conditions of their respective providers.

References

1. Beery S, Cole E, Parker E, Perona P, Winner K (2021) Species Distribution Modeling for Machine Learning Practitioners: A Review. In ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '21), June 28-July 2,2021, Virtual Event, Australia. ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/3460112.3471966
2. Boontop Y, Schutze MK, Clarke AR, Cameron SL, Krosch MN (2017) Signatures of invasion: using an integrative approach to infer the spread of melon fly, *Zeugodacus cucurbitae* (Diptera: Tephritidae), across Southeast Asia and the West Pacific. Biol Invasions (2017) 19:1597–1619. http://doi:10.1007/s10530-017-1382-8
3. Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson AT, Guisan A (2007) Evidence of climatic niche shift during biological invasion. Ecol Lett 10(8):701–709. https://doi.org/10.1111/j.1461-0248.2007.01060
4. Bwire PM, Bakengesa JA, De Meyer M, Virgilio M, Mwatawala MW, Majubwa MO, Deguine JP, Kudra AB (2026) Economic benefits of GAMOUR Agroecology in smallholder cucurbit production in Tanzania. European Journal of Horticultural Science (2026) 91:1 https://doi.org/10.1079/ejhs.2026.0004
5. CABI (2024) https://www.cabidigitallibrary.org/doi/10.1079/cabicompendium.17683#REF-DDB-206236. Accessed 12 August 2025
6. Cugala D, Canhanga L (2023) Monitoring report of *Zeugodacus cucurbitae* in Mozambique. Technical Report. UEM
7. De Meyer M, Robertson MP, Mansell MW, Ekesi S, Tsuruta K, Mwaiko W, Vayssières J-F, Peterson AT (2009) Ecological niche and potential geographic distribution of the invasive fruit fly *Bactrocera invadens* (Diptera: Tephritidae). Bull Entomol Res 100(1):35–48. https://doi.org/10.1017/S0007485309006713
8. De Meyer M, Robertson MP, Mansell MW, Ekesi S, Tsuruta K, Mwaiko W, Vayssières JF, De Meyer M, Delatte H, Mwatawala M, Quilici S, Vayssières J-F, Virgilio M (2015) A review of the current

knowledge on *Zeugodacus cucurbitae* (Coquillett) (Diptera, Tephritidae) in Africa, with a list of species included in *Zeugodacus*. ZooKeys. 2015;539

9. Dhillon MK, Singh R, Naresh JS, Sharma HC (2005) The melon fruit fly, *Bactrocera cucurbitae*: A review of its biology and management. Journal of Insect Science. ISSN: 1536–2442
10. Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: A Review. Ecography, Oct., 2007, Vol. 30, No. 5 (Oct., 2007), pp. 609–628
11. Duan RY, Kong XQ, Huang MY, Varela S, Ji X (2014) The predictive performance and stability of six species distribution models. PLoS ONE 9(11):e112764. https://doi.org/10.1371/journal.pone.0112764
12. Elith J, Leathwick JR, Hastie J (2008) A working guide to Boosted Regression Trees. *Journal of Animal Ecology* 2008, 77, 802–813. http://doi:10.1111/j.1365-2656.2008.01390
13. Elith J, Leathwick JR (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Rev Ecol Evol Syst 40:677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159
14. Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17:43–57. https://doi.org/10.1111/j.1472-4642.2010.00725
15. Erazo D, Grant L, Ghisbain G, Marini G, Colón-González FJ, Wint W, Rizzoli A, Van Bortel W, Vogels BF, Grubaugh ND, Mengel M, Frieler K, Thiery W, Dellicour S (2024) Contribution of climate change to the spatial expansion of West Nile virus in Europe. https://doi.org/10.1038/s41467-024-45290-3
16. FAO (2017a) International Standard for Phytosanitary Measures (ISPM) No. 4. *Requirements for the establishment of pest free areas*. Food and Agriculture Organization of the United Nations (FAO), International Plant Protection Convention, Rome
17. FAO (2017b) International Standard for Phytosanitary Measures (ISPM) No. 6. *Surveillance*. Food and Agriculture Organization of the United Nations (FAO), International Plant Protection Convention, Rome
18. FAO (2017c) International Standard for Phytosanitary Measures (ISPM) No. 11. Pest risk analysis for quarantine pests. Food and Agriculture Organization of the United Nations (FAO), International Plant Protection Convention, Rome
19. FAOSTAT (2024) Statistical database. http://www.fao.org/faostat/en/#data/QC Accessed in 15 August 2025
20. Fick SE, Hijmans RJ (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
21. Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24(1):38–49. 10.1017/S0376892997000088

22. Ghisbain G, Thiery W, François M, Erazo D, Pierre R, Michez D, Dellicour S (2023) Projected decline in European bumblebee populations in the twenty-first century. Nature 628. 10.1038/s41586-023-06471-0
23. Ghodekar KS, Yendrembam KD, Sonawane VK, Longkumer IY, Ibrahim MM (2025) Biology and development of the melon fruit fly, Zeugodacus cucurbitae, on various cucurbit hosts. N Z J Crop Hortic Sci. 2025;53:2603–18
24. He J, Sun P, Wang C, Jiang F, Zhang Y, Qin Y (2025) Ecological niche comparisons of three polyphagous fruit fly (Diptera: Tephritidae) species using ensemble models with soil variables. J Econ Entomol. https://doi.org/10.1093/jee/toaf143
25. Hijmans R (2025) terra: Spatial Data Analysis_. R package version 1.8–29. https://CRAN.R-project.org/package=terra
26. Hijmans RJ, Barbosa M, Ghosh A, Mandel A (2024) geodata: Download Geographic Data_. R package version 0.6-2. https://CRAN.R-project.org/package=geodata
27. Huang Y-B, Chi H (2012) Age-stage, two-sex life tables of *Bactrocera cucurbitae* (Coquillett) (Diptera: Tephritidae). Insect Sci. 2012;19:263–73
28. Jacquard C, Virgilio M, David P, Quilici S, De Meyer M, Delatte H (2013) Population structure of the melon fly, *Bactrocera cucurbitae*. Reun Island Biol Invasions 15:759–773. https://doi.org/10.1007/s10530-012-0324-8
29. Kumar S, Neven LG, Yee WL (2014) Evaluating correlative and mechanistic niche models for assessing the risk of pest establishment. Ecosphere 5(7). https://doi.org/10.1890/ES14-00050.1
30. Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17(2):145–151. https://doi.org/10.1111/j.1466-8238.2007.00358
31. Mkiga AM, Mwatawala MW (2015) Developmental biology of *Zeugodacus cucurbitae* in three cucurbitaceous hosts at different temperature regimes. J Insect Sci. 2015;15:160
32. Mwatawala MW, Maerere AP, Makundi R, De Meyer M (2010) Incidence and host range of the melon fruit fly *Bactrocera cucurbitae* (Coquillett) (Diptera: Tephritidae) in Central Tanzania. Int J Pest Manag. 2010;56:265–73
33. Mwatawala MW, De Meyer M, Makundi RH, Maerere AP (2009) Host range and distribution of fruit-infesting pestiferous fruit flies (Diptera, Tephritidae) in selected areas of Central Tanzania. Bull Entomol Res. 2009;99(6):629–41. https://doi.org/10.1017/S0007485309006695
34. Mwatawala MW, Meyer MD, Makundi RH, Maerere AP (2006) Biodiversity of fruit flies (Diptera, Tephritidae) in orchards in different agro-ecological zones of the Morogoro region, Tanzania. Fruits. 2006;61:321– 32. https://doi.org/10.1051/fruits:2006031
35. Mwatawala MW, De Meyer M, Bakengesa J, Zinga AM, Runoha JO (2026) Long-term monitoring reveals climate-driven seasonal dynamics of *Zeugodacus cucurbitae* (Coquillett) in Tanzania and prospects for pest control. BMC Ecol Evo. https://doi.org/10.1186/s12862-026-02536-6

36. Naimi B, Hamm Na, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling. Ecography 37:191–203. https://doi:10.1111/j.1600-587.2013.00205
37. Olarewaju OO, Fajinmi OO, Arthur GD, Coopoosamy RM, Naidoo KK (2021) Food and medicinal relevance of Cucurbitaceae species in Eastern and Southern Africa. Bull Natl Res Centre 45(1). https://doi.org/10.1186/s42269-021-00659-y
38. Phillips S (2021) maxnet: Fitting 'Maxent' Species Distribution Models with 'glmnet'_. R package version 0.1.4. https://CRAN.R-project.org/package=maxnet
39. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190(3–4):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
40. Qin Y, Paini DR, Wang C, Fang Y, Li Z (2015) Global establishment risk of economically important fruit fly species (Tephritidae). PLoS ONE 10(1):e0116424. https://doi.org/10.1371/journal.pone.0116424
41. Ridgeway G, Developers G (2024) gbm: Generalized Boosted Regression Models_. R package version 2.2.2,<https://CRAN.R-project.org/package=gbm
42. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913–929. https://doi.org/10.1111/ecog.02881
43. Tarimo P, Kabota S, Majubwa R, Kudra A, Virgilio M, Jordaens K, De Meyer M, Mwatawala M (2025) Fruit fly infestation of cucurbitaceous vegetables in Morogoro—Eastern Central Tanzania. PLoS ONE 20(4):e0322277. https://doi.org/10.1371/journal.pone.0322277
44. Tennekes M (2018) tmap: Thematic Maps in R. J Stat Softw 84(6):1–39. 10.18637/jss.v084.i06
45. R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing. www.R-project.org, Vienna, Austria
46. Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2019) blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol Evol. 2019; 10:225–232. https://doi.org/10.1111/2041-210X.13107
47. Vargas RI, Walsh WA, Nishida T (1996) Survival and development of immature melon fly, *Bactrocera cucurbitae* (Coquillett) (Diptera: Tephritidae), at low temperatures. J Econ Entomol 89(3):716–723. https://doi.org/10.1093/jee/89.3.716
48. Vayssières J-F, Rey J-Y, Traoré L (2007) Distribution and host plants of *Bactrocera cucurbitae* in West and Central Africa. Fruits. 2007;62(6):391–6. https://doi.org/10.1051/fruits:2007037
49. Vayssières J-F, Carel Y, Coubes M, Duyck PF (2008) Development of immature stages and comparative demography of two cucurbit-attacking fruit flies in Reunion Island: *Bactrocera cucurbitae* and *Dacus ciliatus* (Diptera Tephritidae). Environ Entomol 37(2):307–314. https://doi.org/10.1603/0046-225X(2008)37[307:DOISAC]2.0.CO;2

50. Virgilio M, Delatte H, Backeljau T, De Meyer M (2010) Macrogeographic population structuring in the cosmopolitan agricultural pest Bactrocera cucurbitae (Diptera: Tephritidae). Mol Ecol. 2010;19:2713–24
51. Wickham H, François R, Henry L, Müller K, Vaughan D (2023) dplyr: A Grammar of Data Manipulation. R package version 1.1.4. https://CRAN.R-project.org/package=dplyr
52. Supplementary Table S Summary of the performance metrics of the 30 BRT Replicates. The final row reports the mean ± Standard Error (SE) across replicates

Figures

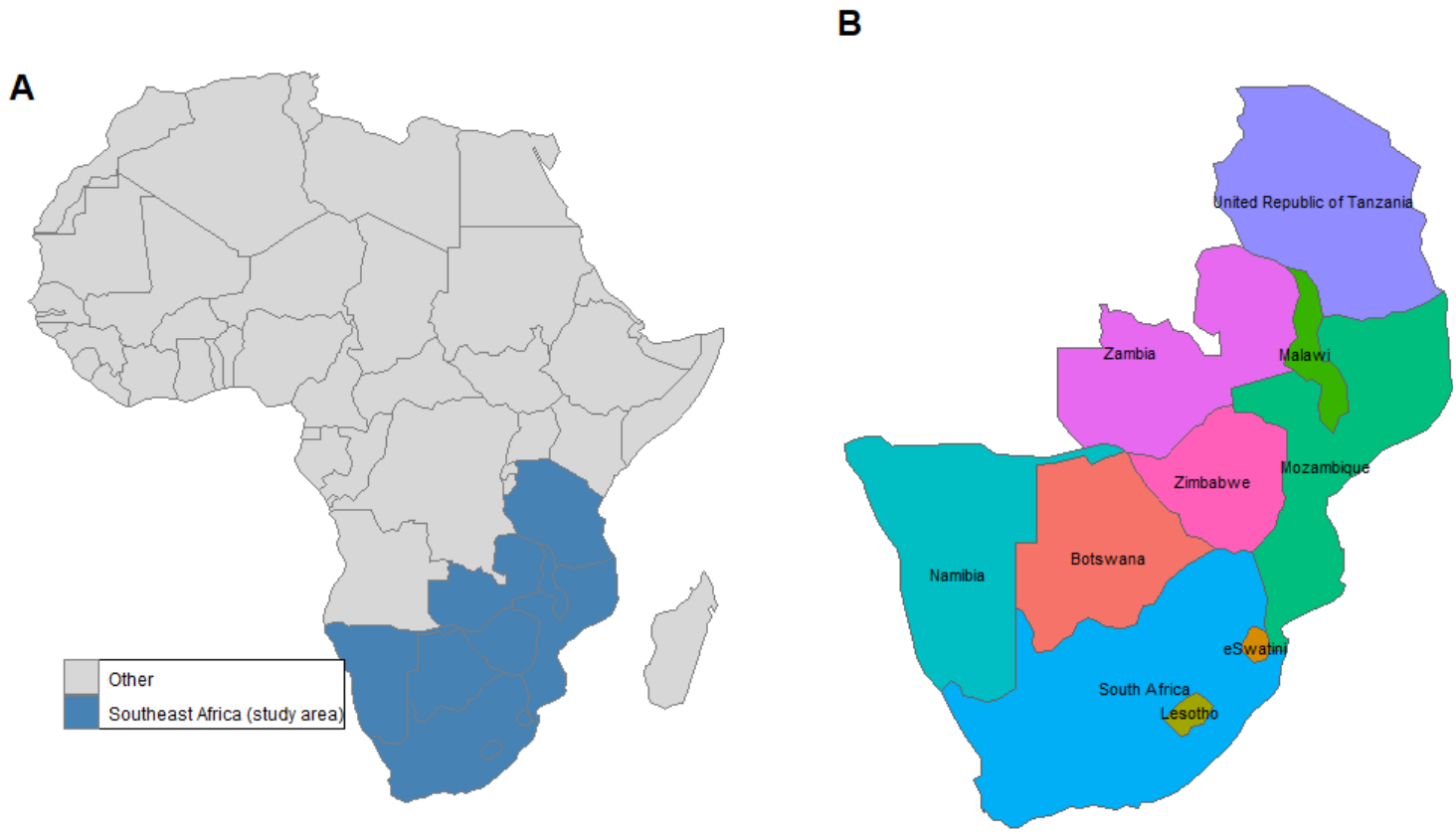

Figure 1

Geographic extent of the study area. (A) Location of Southeast Africa within the African continent. (B) Countries included in the ecological niche modelling of the melon fly.

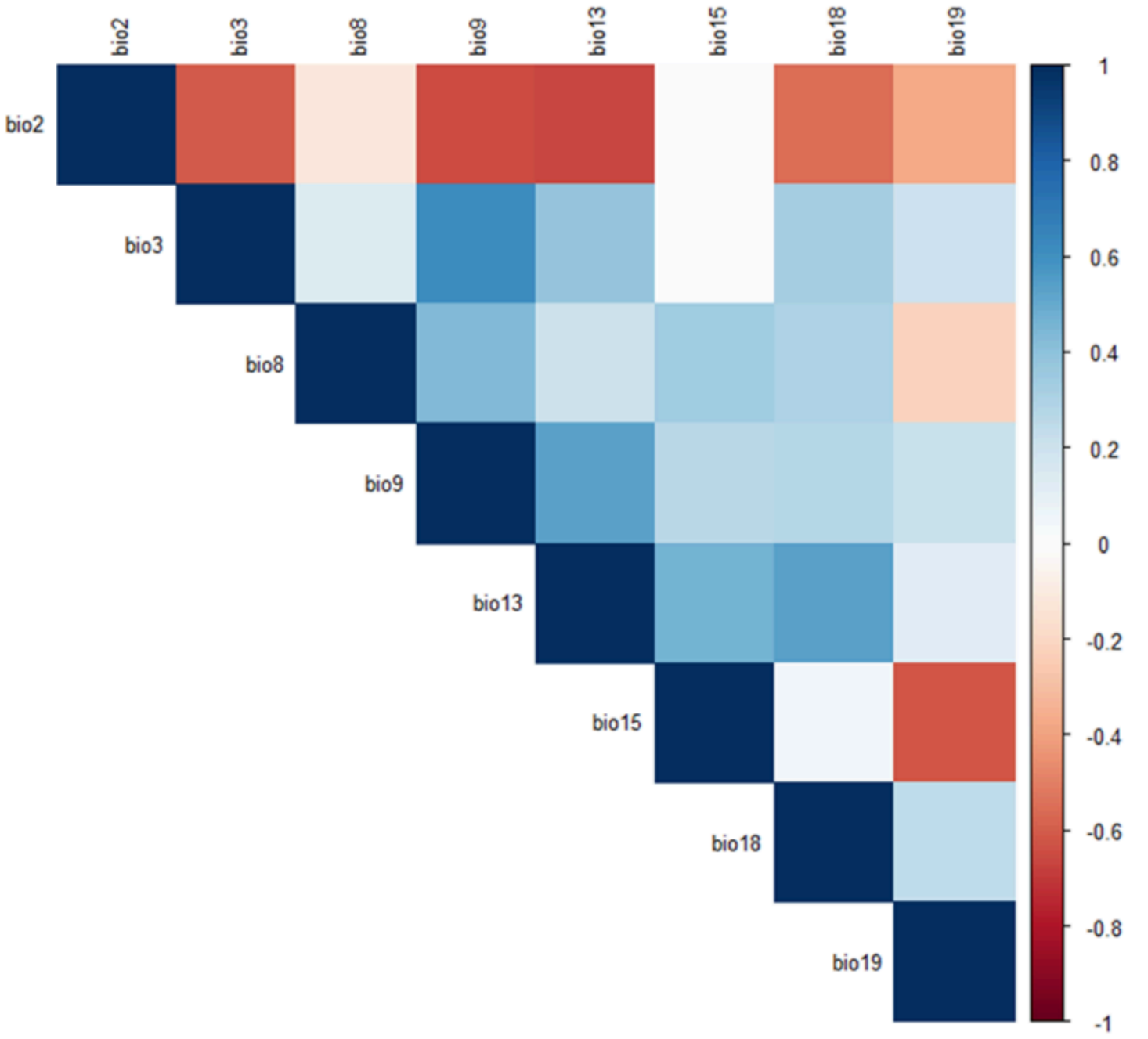

Figure 2

Pairwise Pearson correlation matrix of the selected environmental variables used for ecological niche modelling. Blue colors indicate positive correlations, red colors indicate negative correlations, and colors intensity reflects the magnitude of the correlation coefficient.

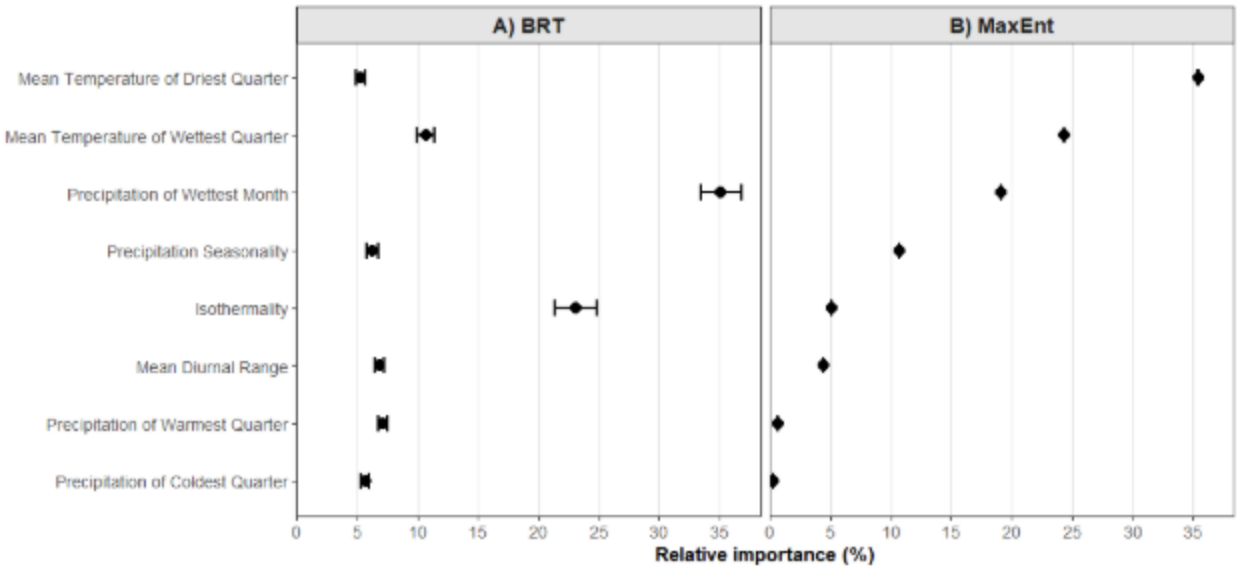

Figure 3

Relative importance of bioclimatic variables in ecological niche models for the melon fly in Southeast Africa. Relative importance values represent the mean contribution of each predictor across 30 independent model replicates and error bars indicate the 95% confidence intervals.

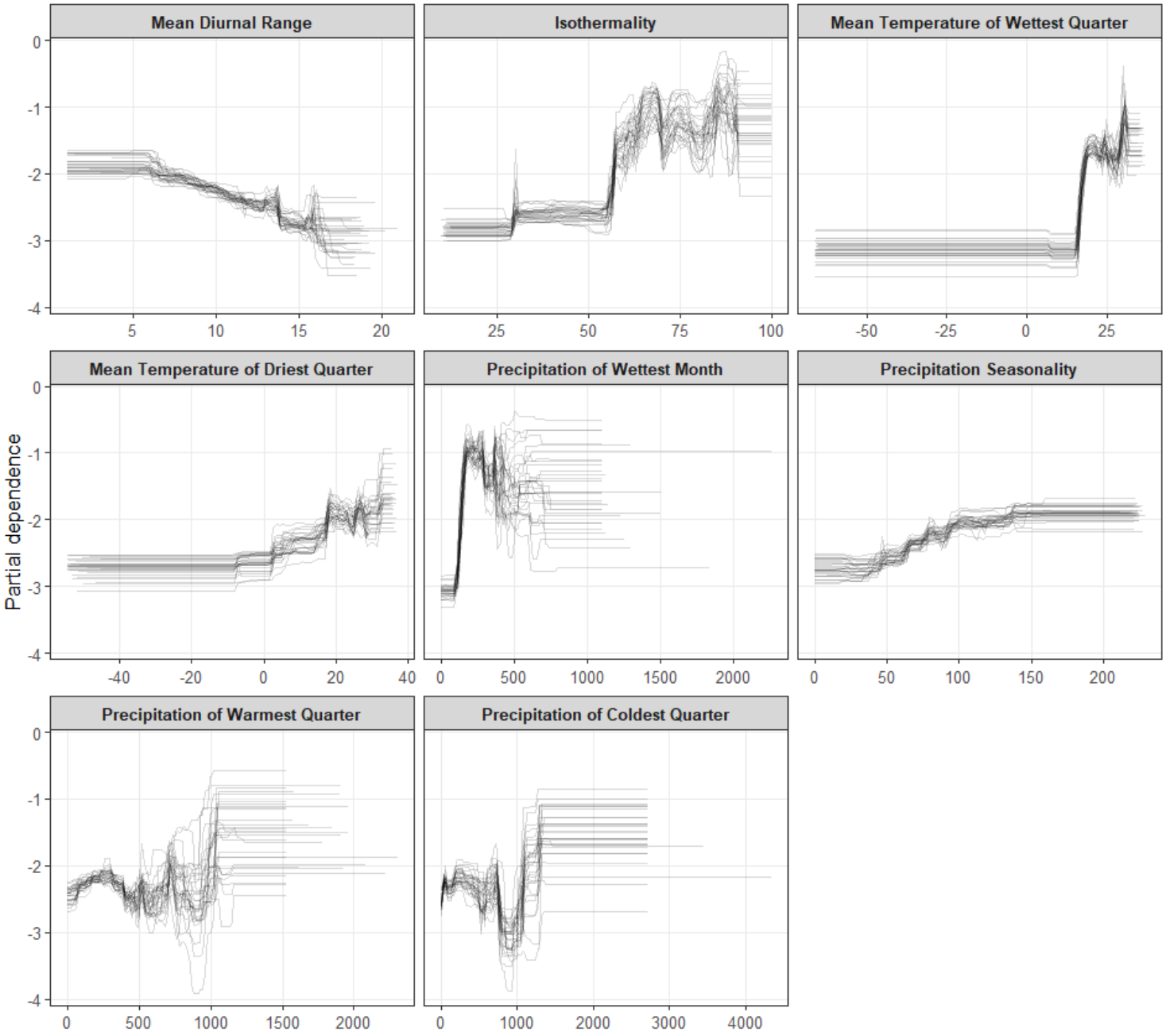

Figure 4

Partial dependence plots showing the relationships between individual bioclimatic predictors and the predicted climatic suitability of *Zeugodacus cucurbitae*derived from the Boosted Regression Tree (BRT) models. Each curve represents one of the 30 independent replicate models fitted using different buffered pseudo-absence samples. Partial dependence values illustrate the marginal effect of each predictor, expressed on the logit scale, while all remaining predictors are held at their average values.

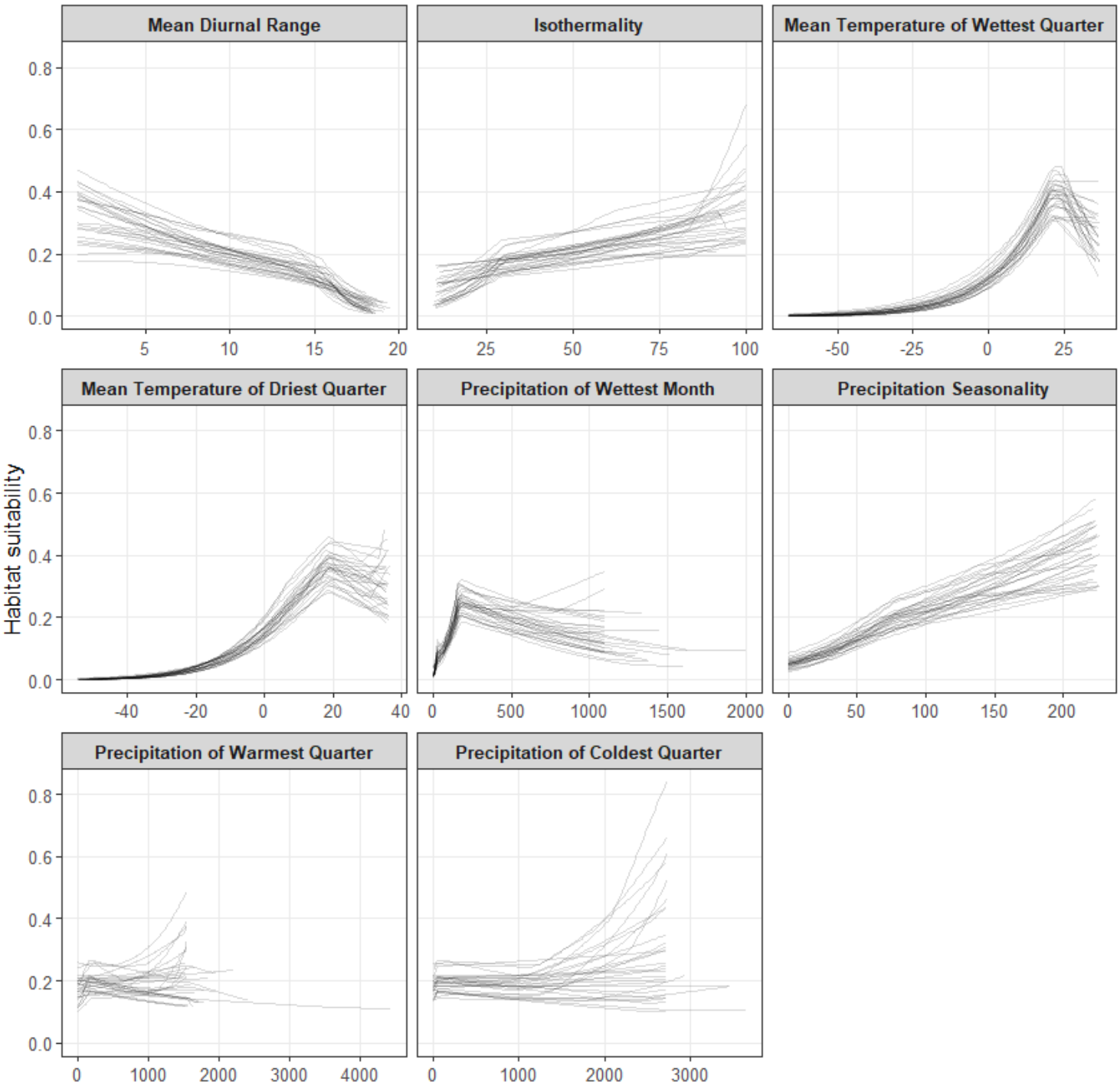

Figure 5

Response curves showing the relationships between individual bioclimatic predictors and the predicted climatic suitability of *Zeugodacus cucurbitae* derived from the MaxEnt models. Each curve represents one of the 30 independent replicate models generated using different buffered pseudo-absence samples. Curves show the predicted climatic suitability across the environmental gradient of each predictor while all remaining predictors are held constant at their mean values. climatic suitability is expressed as cloglog output ranging from 0 to 1.

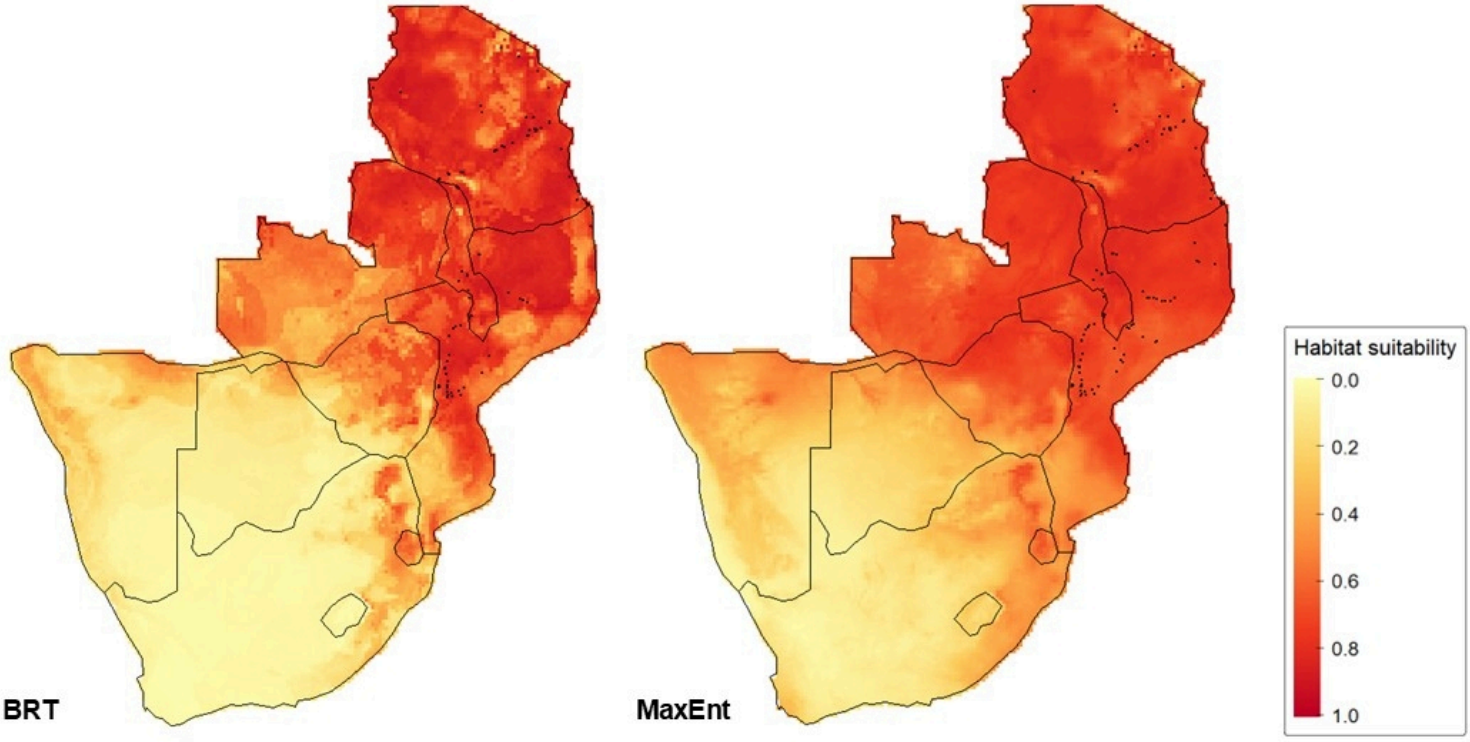

Figure 6

Predicted climatic suitability for *Zeugodacus cucurbitae* across Southeast Africa derived from the Boosted Regression Tree (BRT) and MaxEnt models. Maps show the predicted climatic suitability, ranging from 0 to 1, obtained from 30 replicate models generated using different buffered pseudo-absence samples. Higher values indicate greater predicted climatic suitability.

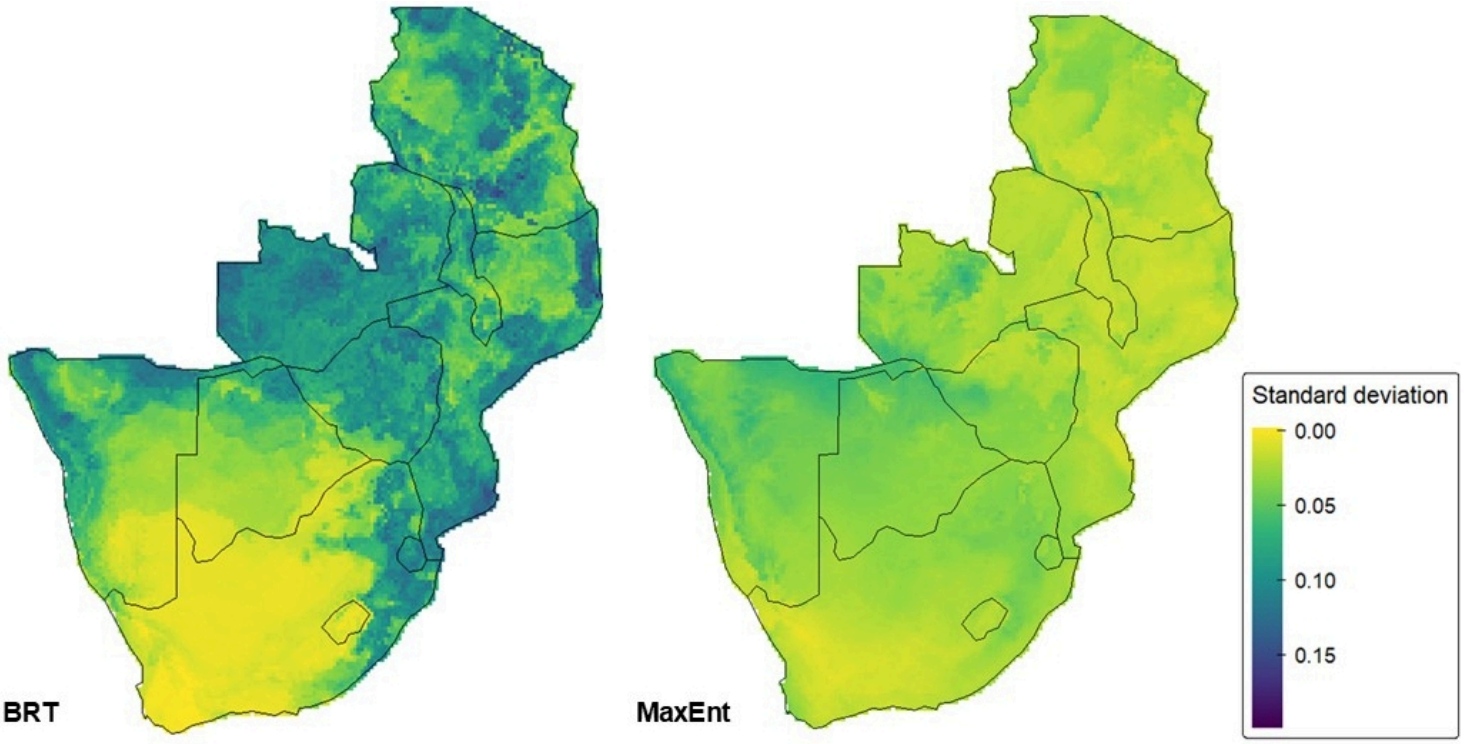

Figure 7

Prediction uncertainty for *Zeugodacus cucurbitae* across Southeast Africa derived from the BRT and MaxEnt models. Maps show the standard deviation among the 30 replicate predictions generated using different buffered pseudo-absence samples. Lower SD values indicate greater agreement among replicate models, whereas higher values indicate greater prediction uncertainty.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.docx