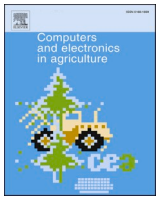
Computers and Electronics in Agriculture 225 (2024) 109283



Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag





Fall armyworm habitat analysis in Africa with multi-source earth observation data

Tonghui Qi [a,b], Yingying Dong [b,c,*], XueLing Li [b], MingXian Zhao [b], Wenjiang Huang [b,c]

[a] Capital Normal University, Beijing 100048, China
[b] Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences. Beijing 100101, China
[c] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Habitat suitability
African fall armyworm
Remote sensing
Exploratory factor analysis
Random forest

ABSTRACT

The fall armyworm (*Spodoptera frugiperda*, FAW) is a significant migratory agricultural pest under the global warning of the Food and Agriculture Organization of the United Nations (FAO). Habitat suitability monitoring and analysis for FAW can help support more scientific management of pest dynamics and decision-making on prevention and control. In this study, we proposed a monthly habitat suitability monitoring model for FAW that integrates multi-source earth observation data such as climate, land use, vegetation, and soil, taking Africa, which is seriously affected by FAW, as the study area. First, exploratory factor analysis (EFA) was employed to reconstruct the climate variables to obtain three factors characterizing temperature, humidity, and wind respectively. Then, habitat suitability monitoring indicators were constructed by combining the factors with other environmental variables, and a monthly habitat suitability monitoring model for FAW in Africa was developed using the random forest algorithm. Finally, based on the model, the distribution of habitat suitability of FAW in Africa by month in 2023 was analyzed, along with the temporal and spatial variation characteristics of habitat suitability. The results indicate that: (1) The exploratory factor analysis effectively extracted the information of raw climate variables and demonstrated good interpretability. The monthly habitat suitability of African FAW was closely associated with Land Use/Land Cover (LULC), Normalized Difference Vegetation Index (NDVI), and humidity factor. (2) Compared with Maxent and SVM, the monthly habitat suitability monitoring model for FAW in Africa constructed using the random forest algorithm showed better performance, with the model being greater than 0.9 in each of the metrics such as Accuracy, Sensitivity, Specificity, F1-score, AUC, Kappa and TSS. The presence points of 2023 verified the validity of the model. (3) In 2023, FAW in Africa is mainly distributed in West Africa and East Africa south of the Sahara Desert, and in the Nile Delta north of the Sahara Desert. With seasonal changes, the suitable and unsuitable areas of FAW shift, and the movement pattern of the center of the presence points is basically consistent with the movement pattern of the precipitation zone. The areas with land use types of grasslands, savannas, and croplands are the hotspots of infestation in the rainy season every year. Partial Dependence Plots (PDPs) for LULC, NDVI, humidity factor and elevation provide deeper insights into how environmental variables drive model predictions. The results of this study demonstrate that the remote sensing analysis approach of FAW habitat suitability in Africa, which integrates multi-source earth observation data, can help to realize the dynamic and high-precision monitoring of FAW, provide references for the prevention and control of insect pests, and promote the sustainable development of agriculture.

1. Introduction

The fall armyworm (*Spodoptera frugiperda*, FAW) is a lepidopteran pest originating from the Americas that feeds on the leaves, stems, and reproductive parts of more than 350 species of plants, causing significant damage to economically important crops such as maize, rice, sorghum, sugarcane, and wheat (Mishra, Sep. 2023); (Rwomushana, 2019) . According to the global spread map of FAW published by the Food and Agriculture Organization of the United Nations(FAO) as of July 2023 (https://www.fao.org/fall-armyworm/monitoring-tools/fa

* Corresponding author at: Aerospace Information Research Institute, Chinese Academy of Sciences No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China.
E-mail address: dongyy@aircas.ac.cn (Y. Dong).

https://doi.org/10.1016/j.compag.2024.109283
Received 3 May 2024; Received in revised form 24 July 2024; Accepted 25 July 2024
Available online 10 August 2024
0168-1699/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

w-map/en/), FAW has spread extensively in African countries except for Western Sahara, Morocco, Algeria, Tunisia, and Libya. Given the serious damage caused by FAW to agriculture, a series of integrated pest management (IPM) measures have been implemented with some effectiveness (G.C et al., Jul. 2020; Tepa-Yotto et al., Aug. 2022). However, to safeguard food security more effectively and reduce overdependence on pesticides, there is an urgent need for dynamic monitoring and analysis of FAW habitat suitability to support more scientific decision-making on pest dynamics management, prevention, and control.

Utilizing species distribution models (SDMs) to monitor the habitat suitability of FAW is a crucial strategy for effective pest management (Day, Oct. 2017); (Early et al., Nov. 2018) ; (Paudel Timilsena, et al., 2022) . SDMs combine species observations with environmental variables to predict the potential distribution of species for habitat suitability assessment and risk zoning for invasive species (Beery et al., 2021); (Ørsted and Ørsted, 2019) ; (Srivastava et al., Apr. 2019) . Currently, research on monitoring the habitat suitability of FAW using SDMs focuses on two aspects: monitoring indicators and monitoring models. In terms of monitoring indicators, bioclimatic variables are often used in existing studies. A commonly used dataset for bioclimatic variables is the WorldClim dataset (Fick and Hijmans, 2017); which provides long-term climatic averages with relatively high spatial resolution. However, it may not capture short-term climatic fluctuations that have a significant impact on FAW behavior and population dynamics (Gardner et al., 2019). Additionally, the life cycle of FAW consists of four stages: egg, larva, pupa, and adult (Mishra, Sep. 2023); (Rwomushana, 2019) ; and its growth, development, reproduction, and migration are also influenced by factors such as host vegetation, soil, topography, and land use (Adan et al., Dec. 2023); (Liu et al., Mar. 2020) ; (Shi et al., Feb. 2021) ; (Sims, Oct. 2008) . Therefore, relying on climatic variables alone may not adequately reflect the habitat requirements of FAW. In terms of monitoring models, species distribution models can be broadly categorized into two groups: correlative models and mechanistic models (Srivastava et al., Apr. 2019). Correlative models are represented by machine learning algorithms such as MaxEnt (Elith et al., 2011). (Abdel-Rahman, et al., 2023) modeled the FAW distribution in Eastern Africa using the bioclimatic, land surface temperature (LST), solar radiation, wind speed, elevation, land use, and maize acreage as monitoring indicators, and the model demonstrated a high prediction accuracy (AUC=0.87). Mechanistic models, on the other hand, are represented by the CLIMEX model (Jung et al., 2016). (Paudel Timilsena, et al., 2022) assessed the global invasion risk of FAW using the CLIMEX model, emphasizing the risk of establishing temporary and permanent populations of FAW in Africa under the current and projected future climates, taking irrigation patterns into account. In summary, correlative models are capable of incorporating almost any environmental condition and objective metrics exist to assess model accuracy. However, the accuracy of correlative models is affected by data availability, and for emerging pests, they typically predict year-round distributions rather than seasonal outbreaks (Early et al., Feb. 2022). Mechanistic models; which combine the mechanisms of biological and environmental factors, adjusting parameters to match simulation with the actual distribution, are capable of predicting both year-round and seasonal distributions but lack objective assessment methods (Yoon and Lee, Feb. 2023). Currently, there is increasing interest in using machine learning algorithms to construct species distribution models for habitat suitability monitoring (Beery et al., 2021).

Nowadays, remote sensing has become an effective tool for monitoring and predicting various crop pests and diseases (Abd El-Ghany et al., 2020); (Ma, 2020) . First, the accuracy of climate datasets based on interpolation of meteorological stations (e.g., WorldClim) is highly dependent on the density of the station networks and the spatiotemporal scale of climate parameters (Amiri et al., May 2020), whereas remotely sensed climate data are continuously observed without interpolation and geographic bias (He, 2015); (Randin, Mar. 2020) . Further, remote sensing can provide richer information on soil, vegetation, land use, etc., which are closely related to pest habitats, thus capturing pest habitat characteristics more adequately (Makori, et al., 2017). Combining remotely sensed features with climate parameters has been proven by some studies to effectively improve the accuracy of pest habitat suitability monitoring models (Zhang, Oct. 2019); (Bhattarai et al., Feb. 2021) utilized Sentinel-1 synthetic aperture radar and Sentinel-2 multispectral imagery in combination with multisite variables using the random forest algorithm, to develop and evaluate the distribution and abundance of host species of the Spruce budworm in northern New Brunswick, Canada, at a spatial resolution of 20 m. The significant increase in temporal and spatial resolution of remote sensing also provides opportunities for dynamic monitoring of pest habitat suitability (Rhodes et al., 2022); (Huang, et al., 2022) constructed a mechanistic model based on remotely sensed data and meteorological assimilation products to dynamically predict the distribution of FAW.

In conclusion, future monitoring models will be more comprehensive by integrating factors such as climate, topography, soil, and vegetation. These models will dynamically reflect changes in pest habitats, enabling more accurate predictions of pest distribution and occurrence. Therefore, on the basis of analyzing the life cycle of FAW, this study will integrate multi-source earth observation data to construct indicators for dynamic monitoring of FAW habitat suitability, and construct a habitat suitability monitoring model of African FAW by employing the random forest algorithm, aiming to more accurately depict the potential distribution range of African FAW in each month, and indicate the hotpots of pests occurrence. By analyzing the characteristics and causes of changes in monthly habitat suitability, decision support will be provided for the formulation of pest control measures, thereby minimizing pest damage and management costs.

2. Materials and methods

2.1. Study areas

This study focuses on Africa (25°22′W-63°30′E, 34°49′S-37°20′N), severely affected by FAW. According to ground survey data from FAO from 2018 to 2022, FAW is widely distributed across Africa, with the densest distribution observed in West and East Africa. Approximately three-quarters of Africa's area lies between the Tropic of Cancer and the Tropic of Capricorn, with the equator passing through it. About 95 % of the continent's area experiences an annual average temperature exceeding 20 °C, which is favorable for the survival and reproduction of FAW that does not have diapause. At the same time, Africa exhibits diverse climatic types, including Mediterranean, subtropical, and tropical deserts, tropical savanna, tropical rainforest, and highland climates, with a highly uneven distribution of precipitation. The diversity in climate and land cover types results in significant differences in habitat suitability for FAW each month in different parts of Africa.

2.2. Data and preprocessing

2.2.1. Fall armyworm occurrence data

The occurrence data of FAW is from the FAO (https://data.apps.fao.org/?lang=en), which provides two types of data: Scout and Trap. The Scout dataset involves systematic inspections of maize plants by farmers, agricultural professionals, or trained scouts to detect signs of infestation, including the presence of egg masses, larvae on the undersides of leaves, or feeding damage on the leaf surface, thereby confirming the presence of FAW. The Trap dataset, on the other hand, was informed of the presence of FAW by trapping male moths into traps through synthesized female moth sex pheromones (Mishra, Sep. 2023). In this study, the two datasets were integrated and 39,941 presence point data (survey parameters include longitude, latitude, and survey time) from 2018 to 2023 were selected, of which 39,239 data from 2018 to 2022 were used for model training and performance evaluation of the model, and 702 data from 2023 were used for further validation of the validity of the

model.

2.2.2. Pseudo-Absence generation

To effectively construct species distribution models using machine learning algorithms, not only species presence data is required but also available pseudo-absences (also known as background data or environmental condition data). The method and quantity of generating pseudo-absence points have a significant impact on the reliability of the model (Hanberry et al., Aug. 2012); (Phillips, 2009) ; Wisz and Guisan, 2009 . Previous studies have found that pseudo-absences selected by the geographic exclusion method can produce predictions with higher sensitivity (Barbet-Massin et al., 2012); while also mitigating issues of class overlap to a certain extent. Therefore, in this study, the geographic exclusion was chosen to generate a buffer of a 1-degree radius around the presence points and randomly generate pseudo-absence points outside the buffer. To ensure a balanced distribution of positive and negative samples (Valavi et al., 2021), this study generated 40,000 pseudo-absence points and randomly assigned them with the survey date from 2018 to 2022. The spatial distribution of FAW presence points and generated pseudo-absence points from 2018 to 2022 is illustrated in Fig. 1.

2.2.3. Earth observation data

In this study, environmental variables such as climate, vegetation, land use, elevation, and soil properties were obtained by Google Earth Engine. Climatic variables such as 2 m temperature, skin temperature, 0 ~ 7 cm soil temperature, wind speed, precipitation, and 0 ~ 7 cm soil moisture obtained from ERA5-Land reflect the climatic information of the FAW habitat. The NDVI provided by MOD13A3 V6.1 reflects the information of FAW host vegetation, serving as an indicator of quantifying damage of pests to the productivity of vegetation (Adan et al., Dec. 2023). Land use, elevation, and slope may affect the occurrence and spread of FAW by altering precipitation patterns and amounts, temperature dynamics, vegetation types, cropping patterns, and the direction, angle, and intensity of solar radiation at the Earth's surface (Abdel-Rahman, et al., 2023). Therefore, in this study, the LULC classified according to the Annual International Geosphere-Biosphere Programme (IGBP) were retrieved from MCD12Q1 v061, and elevation and slope were obtained from the Shuttle Radar Topography Mission (SRTM). Soil properties are involved in influencing pupal survival and adult emergence of FAW (Sims, Oct. 2008). Thus, sand content, clay content, and silt content were obtained from the SoilGrids, a global digital soil mapping system. In this study, a 1 km resampling of the multi-source data was carried out using Google Earth Engine to standardize the spatial resolution of the data. Details about the environmental variables used in this study can be found in Table 1.

2.3. Construction of habitat suitability model for African FAW

The main methods and procedures of this study are illustrated in Fig. 2. Firstly, for the presence and pseudo-absence points from 2018 to 2022 obtained through data preprocessing, the values of environmental variables at the corresponding locations were extracted based on their longitude and latitude. Subsequently, the original climate variables were subjected to variable reconstruction based on exploratory factor analysis to obtain the factors characterizing the information contained in the climate variables. Furthermore, the factor scores of the climate variables, along with other environmental variables were used as monitoring indicators to construct the habitat suitability monitoring model using the random forest algorithm. During the model construction process, the optimal model parameters were determined through 5-fold cross-validation (Cawley and Talbot, 2010), and the model performance was evaluated using multiple metrics. Finally, based on the constructed habitat suitability monitoring model, the monthly habitat suitability of African FAW in 2023 was predicted and analyzed accordingly.

2.3.1. Variable reconstruction based on EFA

The intercorrelation among climate variables may lead to multicollinearity, which in turn causes model overfitting and reduces the

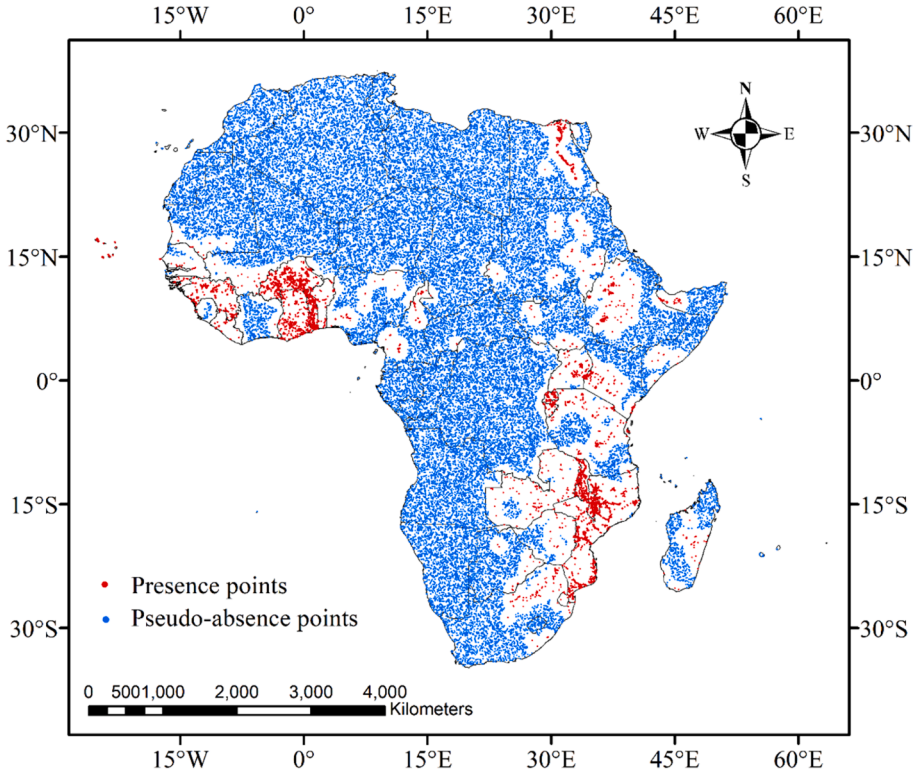


Fig. 1. The spatial distribution of FAW presence points and generated pseudo-absence points from 2018 to 2022.

Table 1
Environmental variables used in FAW habitat suitability studies in Africa.

Variables	Temporal Resolution	Spatial Resolution	Description	Data Source
temperature_2m temperature_2m_min temperature_2m_max	monthly	0.1°	Temperature of air at 2 m above the surface of land, sea or inland waters	(Muñoz Sabater, 2019)
skin_temperature skin_temperature_min skin_temperature_max			Temperature of the surface of the Earth	
soil_temperature_level_1 soil_temperature_level_1_min soil_temperature_level_1_max			Temperature of the soil in layer 1 (0 – 7 cm)	
u_component_of_wind_10m			Eastward component of the 10 m wind	
v_component_of_wind_10m			Northward component of the 10 m wind	
total_precipitation_sum			Accumulated liquid and frozen water that falls to the Earth's surface	
volumetric_soil_water_layer_1 volumetric_soil_water_layer_1_min volumetric_soil_water_layer_1_max			Volume of water in soil layer 1 (0 – 7 cm)	
NDVI	monthly	1 km		(Didan, 2021)
LULC	yearly	500 *m*		(Friedl and Sulla-Menashe, 2022)
elevation slope	/	90 *m*		(Jarvis et al., 2008)
sand_mean	/	250 *m*	Proportion of sand particles in the fine soil part	(Hengl, Feb. 2017)
silt_mean			Proportion of silt particles in the fine soil part	
clay_mean			Proportion of clay particles in the fine soil part	

reliability of predictions (Yoon and Lee, Nov. 2021). Therefore, it is necessary to employ uncorrelated climate variables or reconstruct them through statistical analysis to eliminate multicollinearity and improve model performance. Common methods in previous studies include variable selection based on correlation coefficients, the importance of the variable to the model, the magnitude of eigenvalue, and variance inflation factors (Fan et al., 2020); (Li et al., Jan. 2024) ; (Tepa-Yotto, et al., 2021) ; (Wang et al., 2023) ; (Zhang et al., Apr. 2022) , or variable reconstruction using principal component analysis (Chartois, Feb. 2023); (Huang et al., 2024) .

In this study, climate variables were reconstructed employing exploratory factor analysis (EFA) (Feng et al., 2011).

The purpose of factor analysis is to explain the covariance structure of the observed data through a set of latent variables, known as factors. Suppose there are p observed variables $\mathbf{x} = (x_1, x_2, \cdots, x_p)^T$, and m factors $\mathbf{f} = (f_1, f_2, \cdots, f_m)^T$, where $m < p$. The mathematical model of factor analysis can be expressed as:

$$\mathbf{x} = \mathbf{\Lambda f} + \epsilon \tag{1}$$

where $\mathbf{\Lambda}$ is a $p \times m$ factor loading matrix, and each element λ_{ij} in the matrix denotes the effect of factor f_j on the observed variable x_i; $\mathbf{f}$ is a vector of factor scores, and it is usually assumed that the factors are independent of each other and obey a standard normal distribution, i.e., $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$; ϵ is a vector of errors, and it is usually assumed that the errors are independent of each other and independent of the factors, i.e., $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi})$, $\mathbf{\Psi}$ is a diagonal matrix.

Thus, the covariance matrix of the observed variables $\mathbf{\Sigma}$ can be decomposed as:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}^T + \mathbf{\Psi} \tag{2}$$

EFA is similar to principal component analysis (PCA), but differs in its objective. While PCA aims to construct principal components to explain as much variance as possible, the goal of EFA is to explain the interrelationships among variables as much as possible. Factor analysis achieves better interpretability than principal components through factor rotation (DeCoster, 2024); (Watkins, 2021) .

2.3.2. Random Forest-based FAW habitat suitability monitoring model construction and model evaluation

In this study, a random forest classifier was used to construct a habitat suitability monitoring model using elevation, slope, NDVI, soil properties, LULC, and factors characterizing climatic information, which adequately reflect the habitat information of FAW, as monitoring indicators. The random forest is an algorithm based on ensemble learning that improves the generalization ability and stability of a model by constructing multiple decision trees and combining their predictions. Its core idea is Bagging, a method that constructs multiple training sets with similar features but different sample distributions by randomly drawing samples from the original dataset with put-back. Then, for each training set, a decision tree (Loh, 2011) model is trained and these models are combined to form a forest. In the prediction stage, for a new sample, each decision tree in the forest gives its own prediction, and the final prediction result is obtained by voting (Breiman, Oct. 2001); (Xiang, Apr. 2023) . In species distribution modeling, its main advantages include the simultaneous use of presence and absence data to avoid the introduction of environmental bias; the effective handling of multicollinearity caused by high correlation of variables, avoiding overfitting; the consideration of interactions among variables but without the need to specify interaction terms; the ability to provide estimates of the importance of variables that can help researchers determine which environmental variables have the greatest impact on species distributions (Guo, Apr. 2020); (Lihou and Wall, Nov. 2022) ; (Zhi et al., Dec. 2022) . In addition, each decision tree of the random forest is constructed independently, so it can be processed in parallel, which greatly improves the speed of model training. In this study, the random forest algorithm is implemented through the sklearn package in Python. The parameters of the random forest, the description of the parameters, and the scope of the parameters are shown in Table 2. The optimal parameters of each fold are determined by the grid search algorithm. Furthermore, to better validate the proposed model's performance and demonstrate its relative advantages or disadvantages, Maxent and SVM are chosen as benchmarks for comparison.

There are numerous criteria for evaluating model performance, with common metrics derived from the confusion matrix (Fielding and Bell, Mar. 1997). In this study, 5-fold cross-validation was applied, and model performance was evaluated using Accuracy, Sensitivity, Specificity, F1-score, AUC, Kappa, and the True Skill Statistic (TSS). Among them,

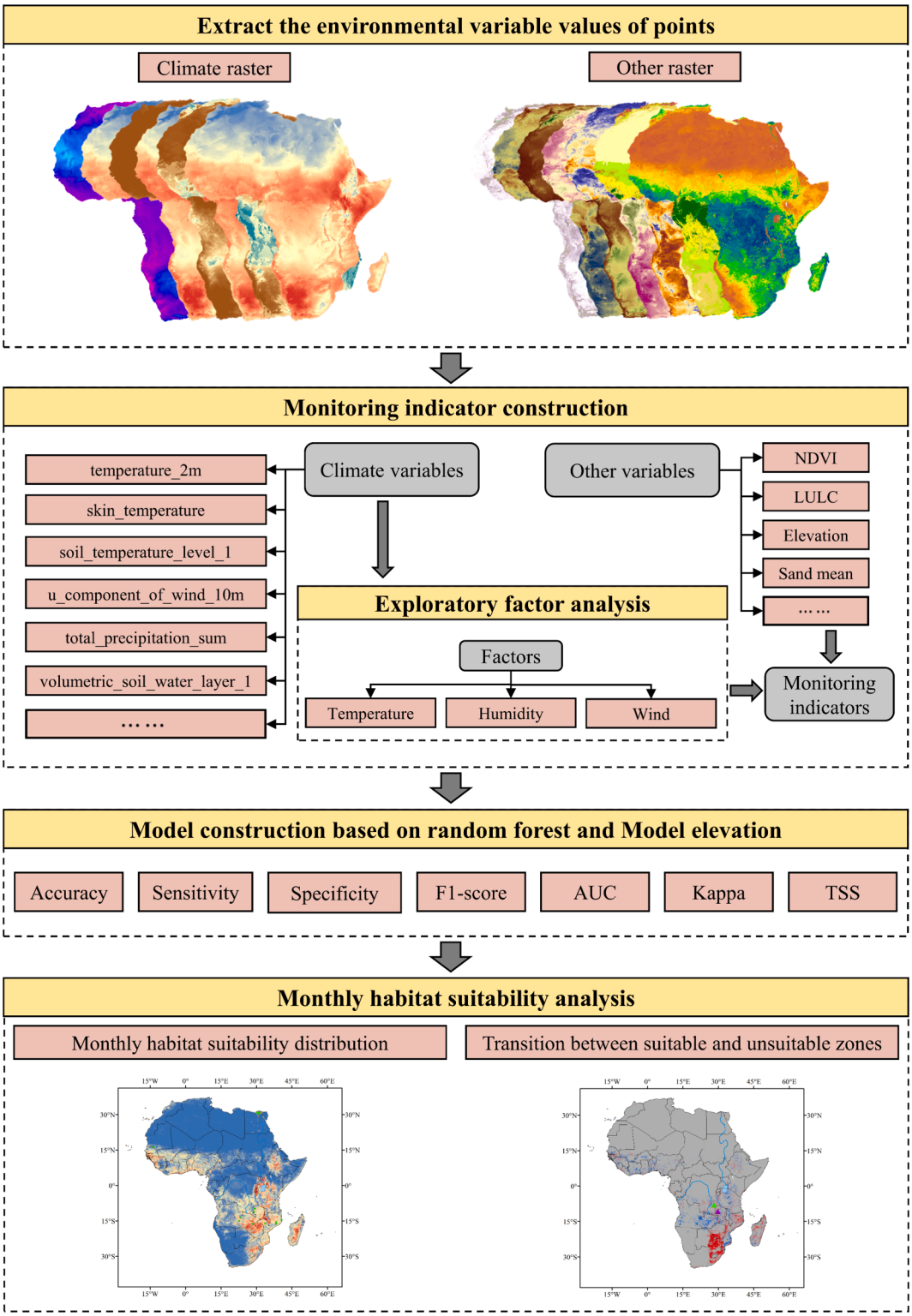


Fig. 2. Methods and procedures.

Accuracy is the simplest and most widely used measure of model performance, reflecting the proportion of presence and absence points that are correctly classified. Sensitivity indicates the proportion of actual presence points correctly identified by the model. Specificity measures the proportion of actual absence points correctly identified by the model. F1-score balances precision, i.e., the proportion of actual presence points among the predicted presence points and recall (sensitivity) (Powell-Romero et al., 2023). AUC measures the model's ability to discriminate between presence and absence across different threshold settings (Jiménez-Valverde, 2012). Kappa measures the consistency of the model's predicted results with the actual situation and takes into account the stochastic classification effects (Allouche et al., 2006). TSS combines sensitivity and specificity and measures the model's comprehensive predictive ability (Shabani et al., 2018). These metrics assess model performance from different perspectives and allow for a more comprehensive comparison of the strengths and weaknesses between

Table 2
Parameters, parameter description, and scope of the parameters of the random forest.

Parameter	Description	Scope
n_estimators	Number of trees in the forest.	10,50,100,200
criterion	The function to measure the quality of a split.	'gini', 'entropy'
max_depth	Maximum depth of the tree.	5,10,20
min_samples_split	Minimum number of samples required to split an internal node.	2,5,10
min_samples_leaf	Minimum number of samples required to be at a leaf node.	1,2,4
max_features	Number of features to consider when looking for the best split.	'sqrt', 'log2'

different models.

3. Results

3.1. Exploratory factor analysis

In this study, the correlation of the original climate variables was examined using Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) method (Feng et al., 2011). The Bartlett sphericity test is used to test the original hypothesis that the correlation matrix is a unit matrix. A unitary matrix implies that the variables are uncorrelated and are not suitable for factor analysis. KMO is a test that examines the strength of partial correlation between variables. KMO values close to 1.0 are considered desirable, while values less than 0.5 are unacceptable. Bartlett's test of sphericity showed $p < 0.05$, while the KMO=0.77, indicating that the original climate variables were suitable for factor analysis. Building upon the correlation test and the Kaiser Criterion (eigenvalue > 1), this study employed the principal component method to extract three common factors. As shown in Table 3, the cumulative variance contribution ratio of the three factors was 0.895 (>0.8), which indicated that the information on the original climate variables was well preserved.

In factor analysis, communality indicates how much of the total variance of each variable is explained by the common factors. As shown in Table 4, the communality of all variables is greater than 0.6, with most variables exceeding 0.9, which indicates that the variables are well explained by the common factors.

In this study, the factor loading matrix was rotated orthogonally using the Varimax method, which enables only a few variables to have higher loadings on a factor and facilitates factor interpretation. The rotated factor loading matrix (Fig. 3) revealed that Factor 1 showed higher loadings on temperature-related variables such as 2 m temperature, skin temperature, and 0 ~ 7 cm soil temperature, so it is named the temperature factor; Factor 2 showed higher loadings on humidity-related variables such as precipitation and 0 ~ 7 cm soil moisture, so it is named the humidity factor; Factor 3 showed higher loadings on variables associated with wind speed, so it was named the wind factor. After factor analysis, the heat map of correlation coefficients of the constructed monitoring indicators (Fig. 4) showed that the correlation coefficients between the indicators are low (<0.8), except for the correlation coefficient between NDVI and the humidity factor, which reaches 0.82.

Table 3
Eigenvalues, variance contribution rates, and cumulative variance contribution rates of common factor.

Common factor	Factor1	Factor2	Factor3
Eigenvalue	6.220	5.426	1.774
Variance contribution rate	0.415	0.362	0.118
cumulative variance contribution rate	0.415	0.776	0.895

Table 4
Variable communality of factor analysis.

Variable	Communality
temperature_2m	0.977
temperature_2m_min	0.956
temperature_2m_max	0.900
skin_temperature	0.988
skin_temperature_min	0.954
skin_temperature_max	0.913
soil_temperature_level_1	0.989
soil_temperature_level_1_min	0.959
soil_temperature_level_1_max	0.941
volumetric_soil_water_layer_1	0.949
volumetric_soil_water_layer_1_min	0.906
volumetric_soil_water_layer_1_max	0.885
total_precipitation_sum	0.630
v_component_of_wind_10m	0.734
u_component_of_wind_10m	0.737

3.2. Model accuracy and the importance of monitoring indicators

In this study, the presence and pseudo-absence data from 2018 to 2022, excluding missing values, were used for modeling, of which 62,976 were used as the training dataset and 15,743 as the validation dataset. The average results of the 5-fold cross-validation (Table 5) demonstrated that the habitat suitability monitoring model constructed using the RF (Random Forest) exhibited high performance and outperformed Maxent and SVM on all evaluation metrics. After considering the performance of the model, computational efficiency, and interpretability of the model, the random forest was selected in this study as the monitoring model for subsequent habitat suitability analysis of FAW in Africa. In the model training, LULC, NDVI, elevation, slope, F1 (temperature factor), F2 (humidity factor), F3 (wind factor), clay content, sand content, and silt content were used as the monitoring indicators, and the entire dataset was used. Considering the stability and generalization performance of the model, the parameter combination with the highest number of occurrences in the 5-fold cross-validation was selected as the optimal parameter (Table 6). For the results of each fold, please see Appendix A.

The importance of monitoring indicators output by the random forest model (Fig. 5) indicates that LULC, NDVI, and F2 (humidity factor) have significant impacts on the presence of FAW. Fig. 6 illustrates the distribution of LULC for FAW presence and pseudo-absence points. FAW presence points are predominantly located in grasslands (48.7 %), savannas (15.2 %), and croplands (15.0 %), which align with the habitat preference of FAW (Ajam et al., May 2024); (Koffi, Jun. 2020) . Pseudo-absence points of FAW are mainly distributed in barren areas (44.7 %), grasslands (19.1 %), and savannas (10.7 %). By observing the locations of pseudo-absence points generated in this study (Fig. 1) and the land use in Africa, it is evident that areas classified as barren are primarily located in the Sahara Desert, which has a hot climate and scarce precipitation, making it unsuitable for the survival of FAW. Due to the vast expanse of the Sahara Desert, a large number of pseudo-absence points are generated there, which has a certain impact on the performance of the model. When these points are excluded, it was found that grasslands and savannas are the two types of land with a high proportion of both presence and pseudo-absence points. This finding has implications for subsequent investigations: grasslands and savannas are marginal to the presence of FAW and can provide FAW with survival and reproduction sites second only to croplands. Therefore, it is necessary to focus on croplands surrounded by grasslands and savannas. There is a significant difference in NDVI between presence points and pseudo-absence points (Fig. 7). The NDVI at presence points is mainly concentrated in the range of 0.5 to 0.7, while that at the pseudo-absence points is concentrated in the range of 0.1 to 0.5. The mean value of NDVI at presence points is significantly higher than that at the pseudo-absence points.

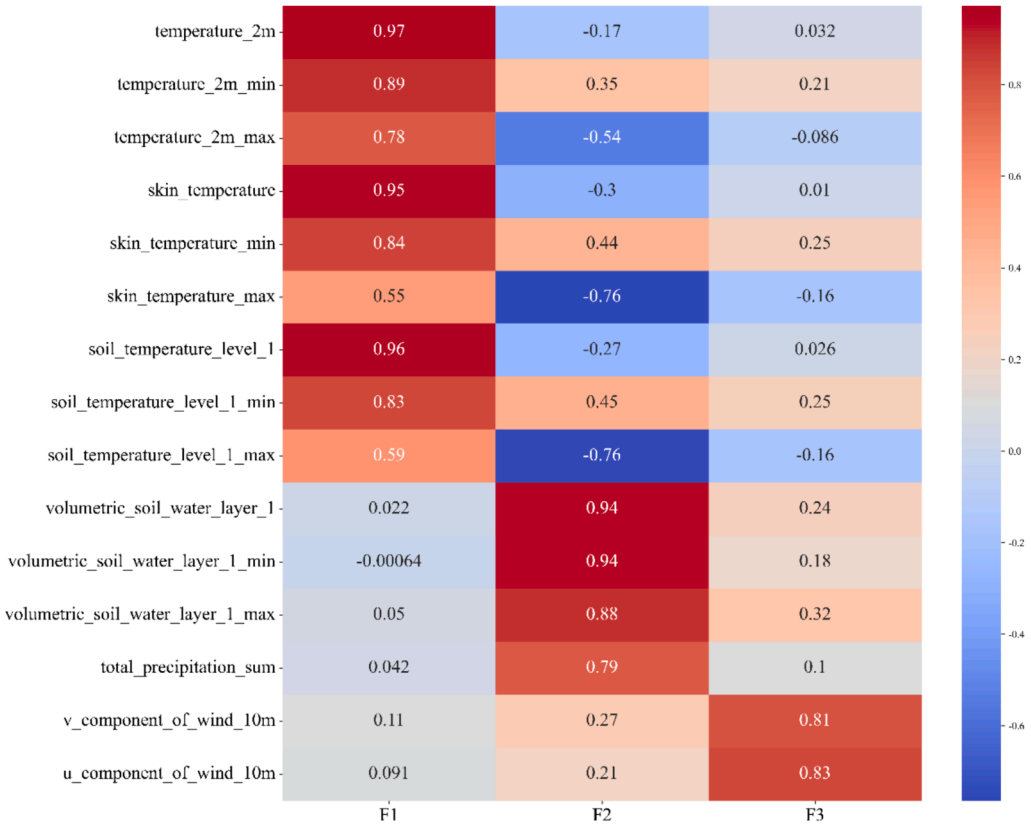


Fig. 3. The rotated factor loading matrix. F1, F2, and F3 represent Factor 1, Factor 2, and Factor 3, respectively.

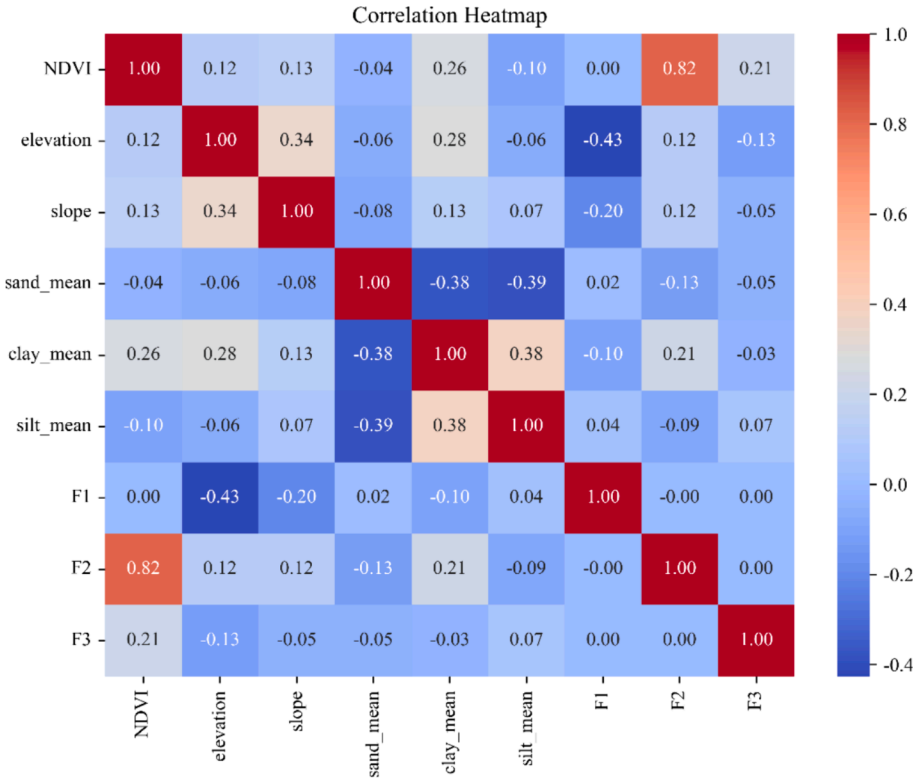


Fig. 4. The heat map of correlation coefficients of monitoring indicators.

Table 5
Model evaluation based on 5-fold cross-validation.

	Accuracy	Sensitivity	Specificity	F1-score	AUC	Kappa	TSS
Maxent	0.900	0.891	0.910	0.898	0.963	0.800	0.800
SVM	0.957	0.972	0.943	0.957	0.958	0.915	0.915
RF	0.975	0.981	0.969	0.975	0.975	0.951	0.951

Table 6
Parameterization of the random forest.

Parameter	Value
n_estimators	100
criterion	'entropy'
max_depth	20
min_samples_split	5
min_samples_leaf	1
max_features	'sqrt'

3.3. Habitat suitability modeling and analysis

The random forest classifier determines the presence of FAW through a voting mechanism, and in this study, the percentage it votes for presence was treated as a habitat suitability index. The monthly habitat suitability of FAW in Africa in 2023 is shown in Fig. 8, where most of the ground survey presence points in 2023 were in areas with high habitat suitability indices, further verifying the validity of the model. The areas with high habitat suitability index for African FAW in all months are mainly located in the western and eastern parts of Africa south of the Sahara Desert. To the north of the Sahara Desert, the Nile Delta region also has a high habitat suitability index in all months.

Although continuous habitat suitability indices can reflect more precise information, for agricultural producers or decision-makers, appropriate zoning is conducive to better decision-making. Therefore, in this study, the habitat suitability index = 0.5 was determined as the threshold for classifying suitable and unsuitable zones based on the distribution of habitat suitability of presence and pseudo-absence points concentrated at the ends of 0 and 1(Fig. 9). The proportion of suitable areas in each month (Fig. 10) indicates that during the spring (March to May), summer (June to August), and autumn (September to November) in the Northern Hemisphere, the area of suitable zones for FAW accounts for more than 10 % of the entire African area, with the highest overall habitat suitability is highest in summer, and relatively low suitability in winter (December to February).

Areas that shifted from unsuitable to suitable zones between months indicate hotspots for FAW control. Fig. 11 depicts the transformation of suitable and unsuitable zones for African FAW from January to December 2023, along with the movement of presence point centers determined by the standard deviation ellipse algorithm. From January to February, a large area of suitable zones appeared at the junction of South Africa, Botswana, and Zimbabwe in southern Africa. From February to March, significant increases in suitable zones were observed in western, central, and eastern Africa, particularly in Ethiopia, Mozambique, Madagascar, and the countries surrounding Lake Victoria, including Kenya, Uganda, Rwanda, Burundi, and Tanzania. At the same time, the large area at the junction of South Africa, Botswana, and Zimbabwe in southern Africa transformed from a suitable zone to an unsuitable zone. From March to April, the area of suitable zones in southern Africa continued to decrease, while those in western, central,

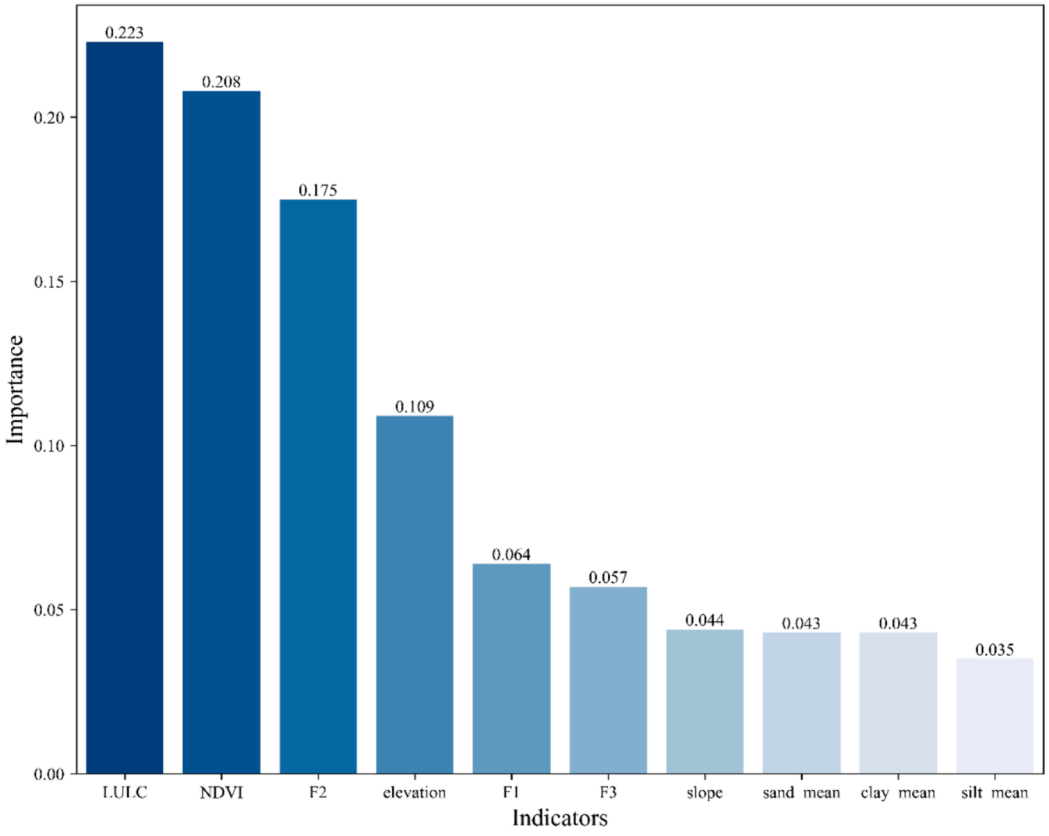


Fig. 5. The importance of monitoring indicators output by the random forest.

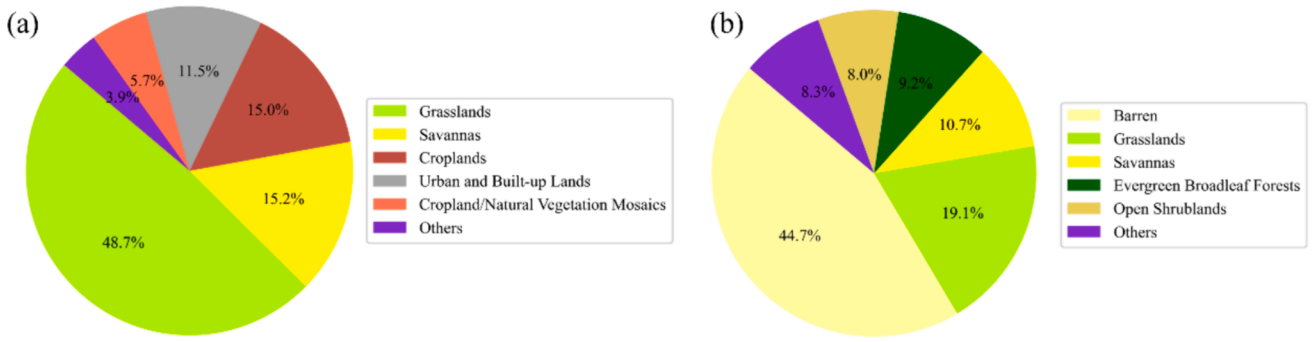


Fig. 6. Percentage of LULC. (a) LULC of presence points; (b) LULC of pseudo-absence points.

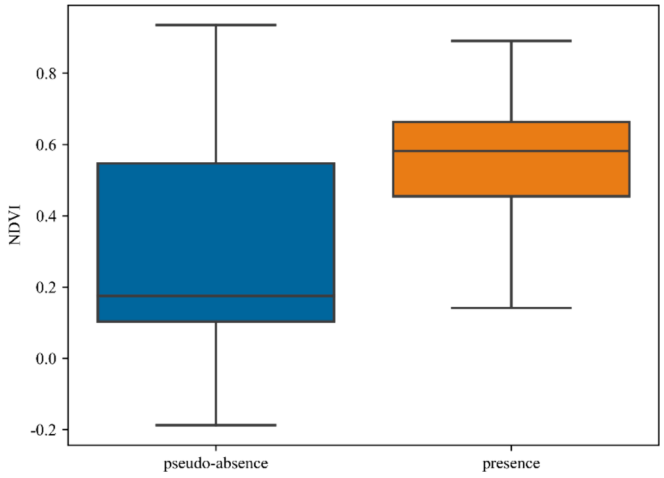


Fig. 7. Distribution of NDVI at presence points and pseudo-absence points.

and eastern Africa continued to increase. From April to August, the area of suitable zones in western Africa expanded continuously, while those in eastern Africa decreased, with the movement of the center of presence points shifting northwestward from eastern Africa to western Africa. From August to December, the area of suitable zones in western Africa gradually decreased, and the center of presence points moved southeastward from western Africa to central Africa. By observing the movement of the center of presence points and the precipitation zone (Fig. 12), it can be found that the movement pattern of the two remains basically the same.

Partial dependence plots (PDPs) provide deeper insights into how environmental variables drive model predictions. PDPs show the dependence between the target response and a set of input features of interest, marginalizing over the values of all other input features. Intuitively, we can interpret the partial dependence as the expected target response as a function of the input features of interest (Greenwell, 2017). As shown in Fig. 5; LULC, NDVI, F2 (humidity factor), and elevation are the top four monitoring indicators in terms of importance. PDPs for these four monitoring indicators are shown in Fig. 13. In terms of LULC, FAW is more likely to be present at number 12 (Croplands), number 13 (Urban and Built-up Lands), and number 14 (Cropland/Natural Vegetation Mosaics). In terms of NDVI, there is a nonlinear relationship between the presence of FAW and NDVI. When NDVI < 0.5, the probability of the presence of FAW increases with increasing NDVI; when $0.5 <$ NDVI < 0.6, the probability of the presence of FAW is the highest; and when NDVI > 0.6, the probability of the presence of FAW decreases with increasing NDVI. In terms of F2 (humidity factor), there is a clear linear relationship between the presence of FAW and humidity, and the probability of presence of FAW increases as humidity increases. In terms of elevation, When elevation $<$ 1000m, the probability of presence of FAW decreases with increasing elevation; and when elevation $>$ 1000m, FAW is less likely to be present.

4. Discussion

In this study, habitat suitability monitoring indicators for FAW in Africa were constructed by integrating multi-source earth observation data. To eliminate multicollinearity in variables, we used exploratory factor analysis (EFA) to reconstruct raw climate variables. Compared with the traditional correlation-based variable selection or variable reconstruction based on principal component analysis, the factors extracted by EFA retained the main information of the original climate variables while having better interpretability. The monthly habitat suitability monitoring model for FAW in Africa constructed using the random forest algorithm exhibited high performance. The importance of monitoring indicators output by the model showed that the presence of FAW was closely related to LULC, NDVI, and humidity factor. The presence points and pseudo-absence points of FAW differed considerably in LULC and NDVI. Differences in LULC reflected that grasslands and savannas were important temporary refuges for FAW and that FAW might migrate from croplands to these areas when the population density of FAW is high. NDVI reflects variations in vegetation productivity, and a higher NDVI means sufficient food and good shelter for FAW. However, areas with higher NDVI are not necessarily more suitable for FAW, e.g. the Congo Basin in the tropical rainforest climate zone, which is not the optimal habitat for FAW (Gómez-Undiano, Sep. 2022). The humidity factor ranked 3rd in the importance of monitoring indicators. This is due to the significance of atmospheric relative humidity on the development, survival, and reproduction of FAW, as well as soil moisture on the pupation and emergence of FAW (He et al., Jun. 2021). The F1 (temperature factor) was only ranked 5th in the importance of monitoring indicators, probably because most of Africa is located between the Tropics of Cancer and Capricorn, with high temperatures in all months, which can meet the survival needs of FAW.

Habitat suitability maps of FAW in Africa by month (Fig. 8) indicated that year-round suitable habitats for FAW exist in savannah climate and plateau alpine climate regions such as West Africa and East Africa south of the Sahara Desert. These findings are in general agreement with the results of the CLIMEX model (Paudel Timilsena, et al., 2022); which takes into account climate change and irrigation patterns. However, a notable difference arises regarding the predictions for the Central African countries of the Democratic Republic of the Congo, Congo, and Gabon. According to the CLIMEX model, these countries exhibit a high Ecoclimatic Index (EI) that supports the establishment of year-round populations of FAW. In contrast, this study shows that FAW only has small seasonal suitable areas in these three countries. Two main reasons contribute to this disparity: First, survey data from these three countries were missing from the FAW ground survey data in this study, resulting in numerous pseudo-absence points within their borders. Second, this study considers not only the climate but also land use when assessing FAW habitat suitability. Therefore, the Democratic Republic of the

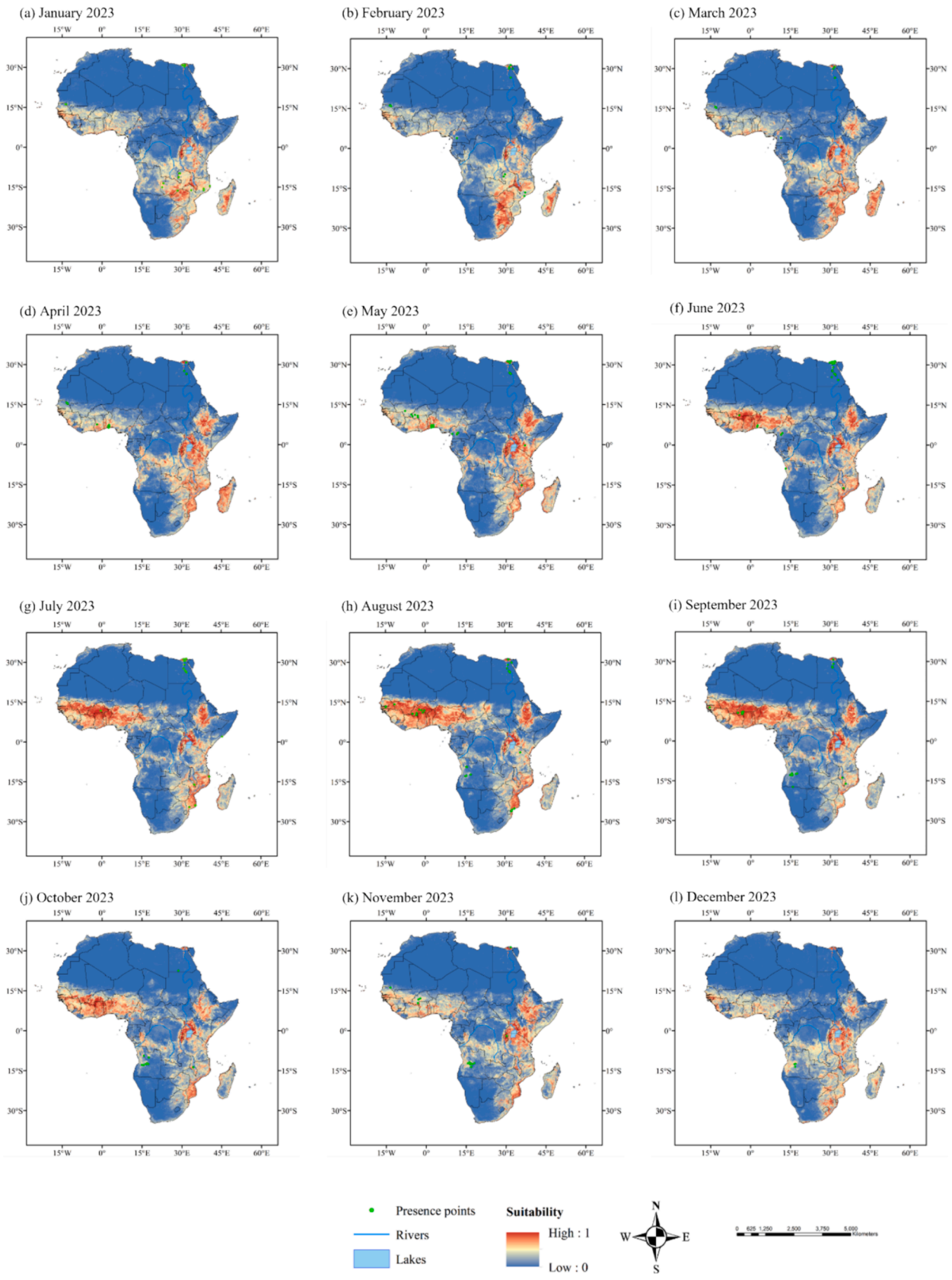


Fig. 8. African FAW Monthly Habitat Suitability in 2023.

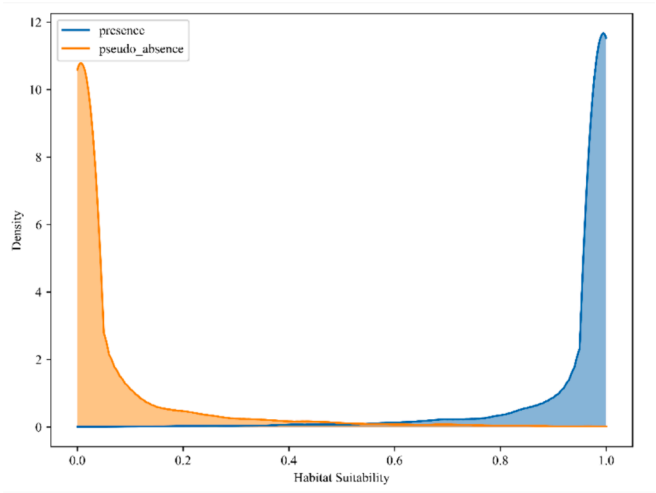


Fig. 9. Kernel density estimation of habitat suitability index for presence points and pseudo-absence points.

Congo, Congo, and Gabon, which are located in tropical rainforest climate zones with predominant land use as evergreen broadleaf forests, are not optimal habitats for FAW. Similarly, the Nile Delta in the tropical desert climate zone is also suitable for FAW all year round, contrary to the findings of (Guimapi, Jun. 2022); which suggest that the Nile Delta region is unsuitable for FAW survival due to ignoring the favorable irrigation conditions along the Nile River. The results of this study, based on actual survey data, better reflect the real situation. Regarding the source of infestation in the Nile River Basin, FAW might migrate from Ethiopia, northward along the Nile to Sudan and Egypt (Niassy et al., 2024). The Mediterranean coasts of Morocco, Algeria, and Tunisia in North Africa have high habitat suitability indices in some months, indicating seasonal suitable areas. However, the likelihood of FAW migrating to these areas through wind dispersal is low (Early et al., Nov. 2018). Although FAW has not been observed along the Mediterranean coast of North Africa according to the data served by FAO, in the study by (Paudel Timilsena, et al., 2022); the presence of FAW has been recorded in this region, suggesting a possibility of FAW migrating from North Africa to Southern Europe by monsoon (Wang, Apr. 2023).

To better assist stakeholders in developing and implementing FAW control measures, we partitioned the habitat suitability index and analyzed the transition of unsuitable and suitable areas each month in Africa. The results showed that there was a seasonal transition in the hotspots of FAW infestation in Africa, and the pattern of movement of the center of the presence points was basically consistent with that of the precipitation zone. According to the relative contributions of climatic variables in predicting the geographic distribution of FAW as indicated by the MaxEnt model, precipitation in the driest months and precipitation in the coldest season were the most important variables in explaining the potential distribution of FAW (Dessie et al., Apr. 2024). Combined with previous analyses on land use, we concluded that areas characterized by grasslands, savannas, and croplands are at higher risk of infestation during the annual rainy season, and the monitoring, prevention, and control of FAW should be strengthened accordingly.

In this study, the validity of the model is based on the following premise assumptions. Although the Trap and Scout datasets are different for determining the presence of FAW, we assume that the results of both methods are true and reliable. The Random Forest-based FAW habitat suitability monitoring model is encouraging in terms of its performance, but the premise assumptions are that the FAW's response to the environment stays relatively stable and that discriminatory accuracy is informative only if the model is similar to the real ecological niche (Warren et al., 2020). Moreover, to ensure accurate implementation by practitioners and researchers, it is essential to describe the conditions under which the model can be effectively applied. Although the FAW habitat suitability analysis methodology for integrating multi-source data proposed in this study is generalizable, the model parameters need to be recalibrated when the model is applied to regions or countries outside Africa. Specifically, the climate varies considerably from region to region and from country to country, so the factor analysis needs to be re-performed based on local monthly climate data. Changes in the input data will cause corresponding changes in the Random Forest parameters, and the most relevant environments may also change.

Despite the results, this study has several limitations. First, the study area of this study covered the entire African region, which limited the use of high spatiotemporal resolution environmental variables because of severe missing data. Second, the habitat suitability monitoring model constructed in this study relies heavily on existing ground survey data on FAW, which may suffer from sampling bias (Hirzel and Guisan, Nov. 2002), potentially leading to the positioning of pseudo-absences within suitable areas during pseudo-absence generation. Current methods for pseudo-absence generation could be classified into two categories: geospatial-based generation and feature-space-based generation, and the geospatial-based method was used in this study. However, there is

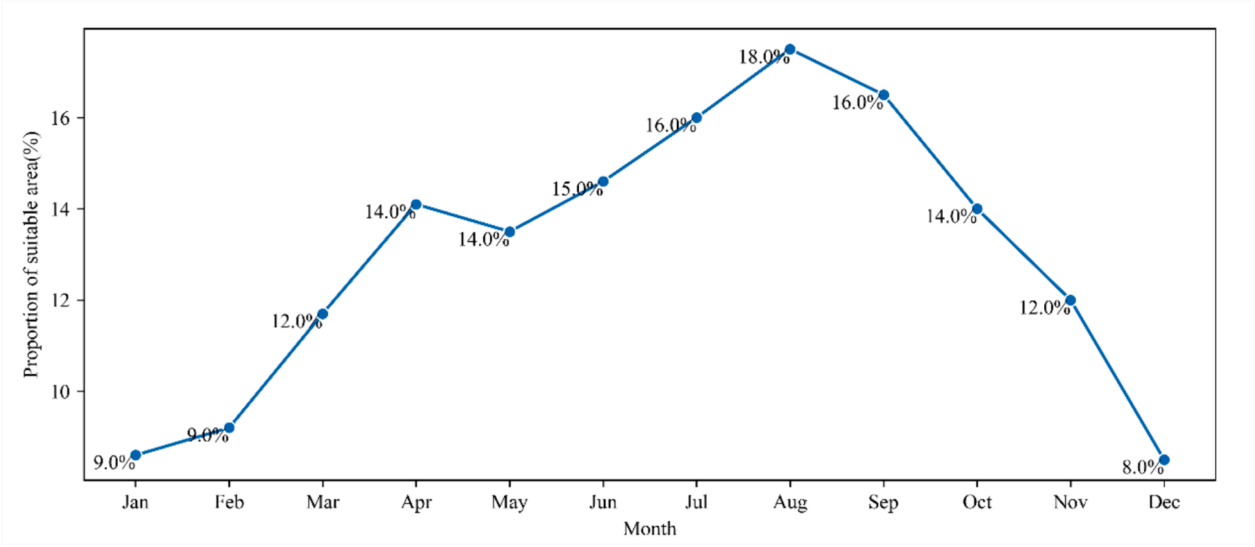


Fig. 10. The proportion of suitable areas for FAW to the area of Africa in each month.

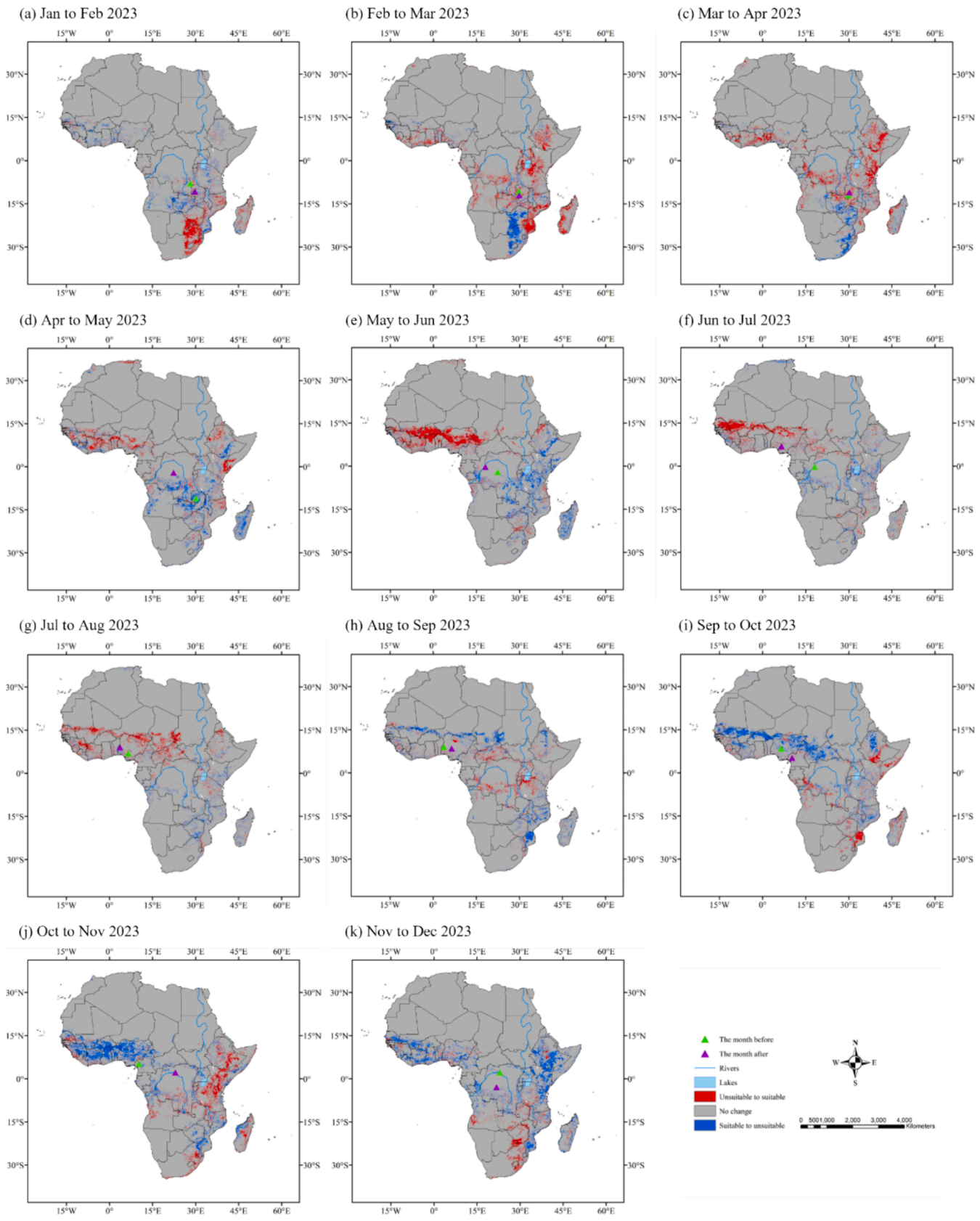


Fig. 11. The transformation between suitable and unsuitable zones and the movement of the center of presence points in each month of 2023.

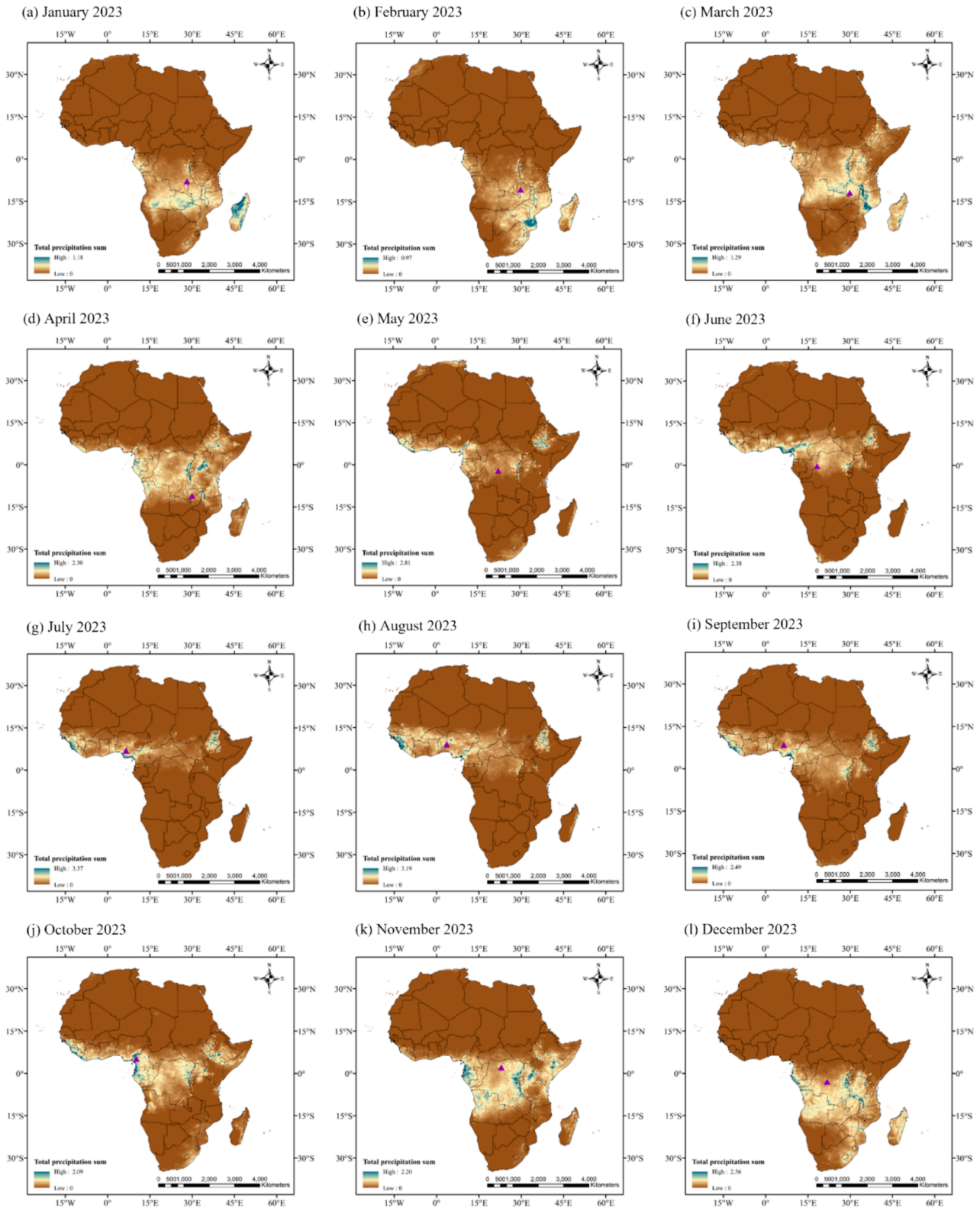


Fig. 12. Precipitation distribution in Africa and the center of FAW presence points in each month of 2023.

no standard rule for the buffer size of pseudo-absence generation, and the reliability has not been determined (Hong et al., May 2019). Third, several environmental variables were selected in this study, but the presence of FAW is also influenced by biological interactions (e.g., natural enemies) and human activities (e.g., pest control measures), which were not taken into account. Moreover, the habitat suitability index provided by the model acts solely as a relative indicator of the probability of presence. Higher values indicate a higher probability of the presence of FAW, while lower values do not definitively rule out the presence of FAW. When FAW has a high population density, it is likely to migrate from suitable to unsuitable areas (Abdel-Rahman, et al., 2023). Further, in the partitioning of habitat suitability, we divided suitable

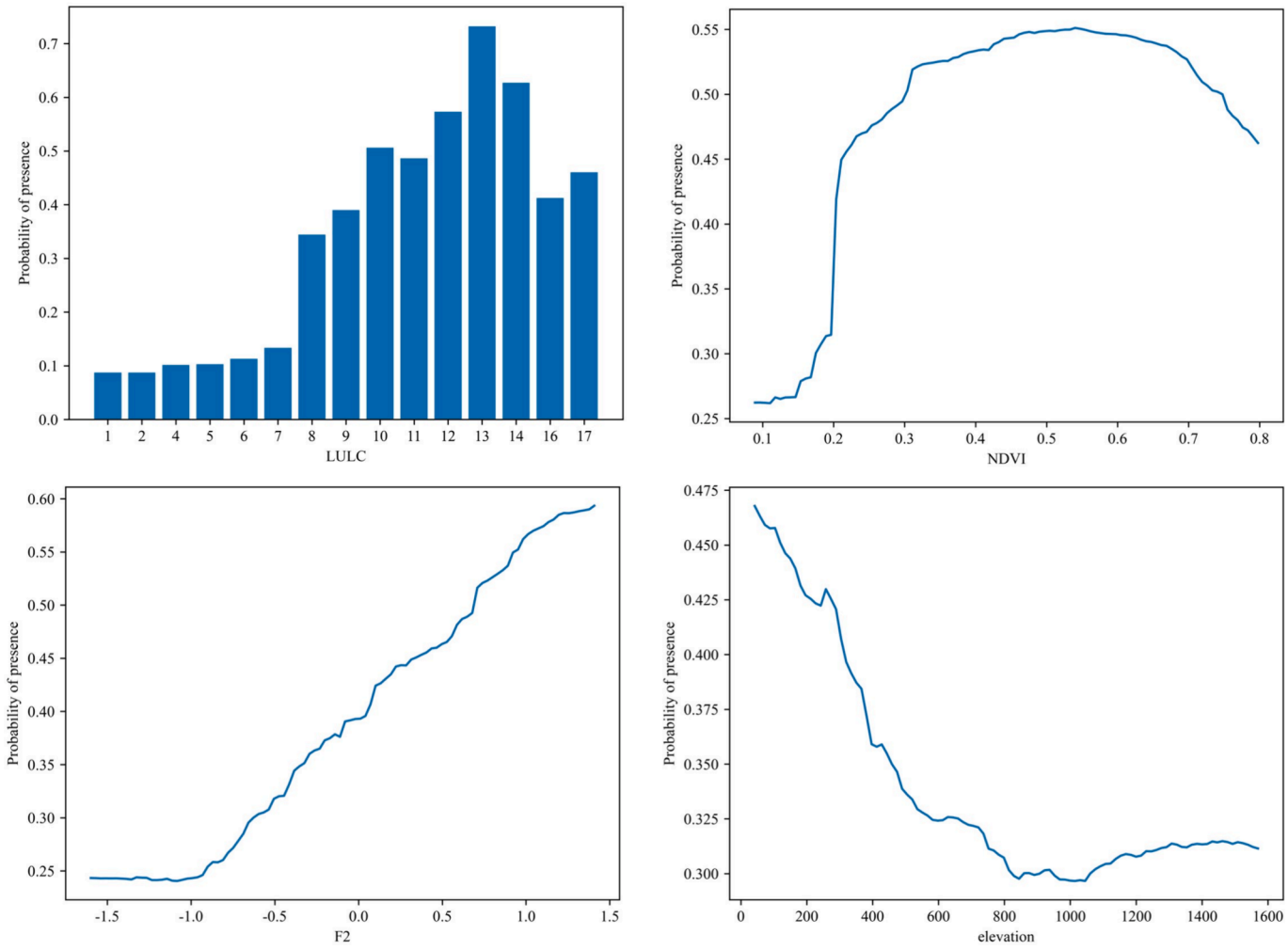


Fig. 13. Partial dependence plots for LULC, NDVI, F2 (humidity factor), and elevation.

and unsuitable areas based on the distribution of habitat suitability at presence points and pseudo-absence points with a threshold of 0.5, but this threshold is somewhat subjective. Last but not least, there are some limitations related to the model's applicability. The environmental data used in this study were all from Google Earth Engine, which has undergone pre-processing steps such as radiometric correction, geometric correction, filling of voids, and smoothing treatment, but may still have some deviation from the actual situation. In addition, high-performance computing resources are required to run the model, and it is recommended to use a multi-core processor to ensure the computational efficiency and accuracy of the results.

In response to these shortcomings, this study argues for several improvements. In cases where the study area is relatively small, increasing the temporal and spatial resolution of environmental variables can be beneficial. Higher temporal resolution enables the acquisition of richer temporal information, while higher spatial resolution allows for the inclusion of landscape structure information in the model. Additionally, besides relying solely on the data provided by the FAO, data from multiple databases can be combined to obtain a more comprehensive distribution of FAW. Regarding the generation method of pseudo-absence points, (Senay et al., 2024) proposed an approach that considers both geographic extent and environmental requirements to generate spatially and ecologically balanced pseudo-absence points. Subsequent studies could attempt to compare the effects of different pseudo-absence generation methods on model performance. In Africa, areas with intensive agro-economic activities are more vulnerable to FAW, and socio-economic factors such as farmers' education level, farm size, and government support policies have a significant impact on the prevention and control of FAW (Makale, Nov. 2021; Kansiime, Jul. 2019; Tambo, Oct. 2020). Therefore, follow-up studies could consider combining environmental data with socioeconomic data to construct a comprehensive risk assessment model, which could help allocate resources more precisely. In the next step, based on this study, the seasonal migration and population dynamics of FAW in Africa will be investigated using graph neural networks with high-resolution remote sensing imagery as the data source to achieve high spatiotemporal accuracy of migration prediction.

5. Conclusions

In this study, a remote sensing analysis method for habitat suitability of FAW in Africa that integrates multi-source earth observation data was proposed to realize high-precision monitoring of monthly habitat suitability of FAW in Africa. Firstly, based on exploratory factor analysis, the climate variables were reconstructed, and three factors namely, temperature, humidity, and wind, were extracted, with a cumulative variance contribution ratio of 0.895, which effectively reflected the information of the original climate variables. Subsequently, combining the extracted factors and environmental variables such as land use, vegetation, and soil as monitoring indicators, a random forest algorithm was used to construct a monthly habitat suitability monitoring model for FAW in Africa, which exhibited high performance in several metrics such as Accuracy, Sensitivity, Specificity, F1-score, AUC, Kappa, and TSS. The importance of monitoring indicators revealed that LULC

(0.223), NDVI (0.208), and humidity factor (0.175) were closely related to the presence of FAW. FAW presence points were mainly distributed in grasslands (48.7 %), savannas (15.2 %), and croplands (15.0 %), with NDVI concentrations ranging from 0.5 to 0.7 at presence points. Further analysis of the monthly habitat suitability distribution of FAW in Africa in 2023 showed that FAW was predominantly distributed in the western and eastern of Africa south of the Sahara Desert, as well as the Nile Delta north of the Sahara Desert. With seasonal changes, there were transitions between suitable and unsuitable areas for FAW, with the overall habitat suitability being higher from June to August and lower from December to February of the following year. The center of presence points moves between western and eastern Africa, and its movement pattern is basically consistent with that of the precipitation zone. Therefore, we inferred that areas with grasslands, savannas, and croplands have a higher risk of pest infestation during the rainy season each year and monitoring and control of FAW should be strengthened. Partial dependence plots for LULC, NDVI, humidity factor, and elevation provided deeper insights into how these four most important monitoring indicators drive model predictions. Future work will focus on the seasonal migration and population dynamics of FAW in Africa to realize migration prediction with high spatiotemporal precision.

CRediT authorship contribution statement

Tonghui Qi: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Yingying Dong:** Writing – review & editing, Resources, Project administration, Methodology, Data curation. **XueLing Li:** Software, Methodology, Data curation. **MingXian Zhao:** Visualization, Methodology. **Wenjiang Huang:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was funded by the Alliance of International Science Organizations (ANSO-CR-KP-2021-06), National Natural Science Foundation of China (42071320, 42071423), GEO-PDRS: Global Vegetation Pest and Disease Dynamic Remote Sensing Monitoring and Forecasting, 2023-2025.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2024.109283.

References

N. M. Abd El-Ghany, S. E. Abd El-Aziz, and S. S. Marei, "A review: application of remote sensing as a promising strategy for insect pests and diseases management," *Environ. Sci. Pollut. Res.*, vol. 27, no. 27, pp. 33503–33515, Sep. 2020, doi: 10.1007/s11356-020-09517-2.

E. M. Abdel-Rahman *et al.*, "Computational biogeographic distribution of the fall armyworm (*Spodoptera frugiperda* J.E. Smith) moth in eastern Africa," *Heliyon*, vol. 9, no. 6, p. e16144, Jun. 2023, doi: 10.1016/j.heliyon.2023.e16144.

Adan, M., Tonnang, H.E.Z., Greve, K., Borgemeister, C., Goergen, G., Dec. 2023. Use of time series normalized difference vegetation index (NDVI) to monitor fall armyworm (Spodoptera frugiperda) damage on maize production systems in Africa. Geocarto Int. 38 (1), 2186492. https://doi.org/10.1080/10106049.2023.2186492.

Ajam, A.L., Karungi, J., Ogwal, G., Adumo, S.A., Paparu, P., Otim, M.H., May 2024. Population dynamics of Fall Armyworm (Lepidoptera: Noctuidae) in Maize Fields in uganda. Insects vol. 15, no. 5, Art. no. 5. https://doi.org/10.3390/insects15050301.

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 (6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x.

Amiri, M., Tarkesh, M., Jafari, R., Jetschke, G., May 2020. Bioclimatic variables from precipitation and temperature records vs. remote sensing-based bioclimatic variables: Which side can perform better in species distribution modeling? Ecol. Inform. 57, 101060 https://doi.org/10.1016/j.ecoinf.2020.101060.

B. Greenwell M., "pdp: An R Package for Constructing Partial Dependence Plots," *R J.*, vol. 9, no. 1, p. 421, 2017, doi: 10.32614/RJ-2017-016.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3 (2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x.

S. Beery, E. Cole, J. Parker, P. Perona, and K. Winner, "Species Distribution Modeling for Machine Learning Practitioners: A Review," in *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies*, in COMPASS '21. New York, NY, USA: Association for Computing Machinery, Sep. 2021, pp. 329–348. doi: 10.1145/3460112.3471966.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., Meneghini, A., MacLean, D.A., Feb. 2021. Spruce budworm tree host species distribution and abundance mapping using multi-temporal Sentinel-1 and Sentinel-2 satellite imagery. ISPRS J. Photogramm. Remote Sens. 172, 28–40. https://doi.org/10.1016/j.isprsjprs.2020.11.023.

Breiman, L., Oct. 2001. Random Forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Cawley, G.C., Talbot, N.L., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Chartois, M., et al., Feb. 2023. Environmental factors driving the abundance of Philaenus spumarius in mesomediterranean habitats of Corsica (France). Sci. Rep. 13 (1), 1901. https://doi.org/10.1038/s41598-023-28601-4.

Day, R., et al., Oct. 2017. Fall Armyworm: Impacts and Implications for Africa. Outlooks Pest Manag. 28 (5), 196–201. https://doi.org/10.1564/v28_oct_02.

J. DeCoster, "Overview of Factor Analysis." Accessed: Apr. 07, 2024. [Online]. Available: http://www.stat-help.com/notes.html.

Dessie, B., Ferede, B., Taye, W., Shimelash, D., Apr. 2024. Field infestation of the invasive fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae) on maize in Southern Ethiopia. Crop Prot. 178, 106587 https://doi.org/10.1016/j.cropro.2024.106587.

K. Didan "MODIS/Terra Vegetation Indices Monthly L3 Global 1km SIN Grid V061. NASA EOSDIS Land processes distributed active archive 2021 Center" 10.5067/MODIS/MOD13A3.061.

Early, R., González-Moreno, P., Murphy, S.T., Day, R., Nov. 2018. Forecasting the global extent of invasion of the cereal pest Spodoptera frugiperda, the fall armyworm. NeoBiota 40, 25–50. https://doi.org/10.3897/neobiota.40.28165.

Early, R., Rwomushana, I., Chipabika, G., Day, R., Feb. 2022. Comparing, evaluating and combining statistical species distribution models and CLIMEX to forecast the distributions of emerging crop pests. Pest Manag. Sci. 78 (2), 671–683. https://doi.org/10.1002/ps.6677.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17 (1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x.

S. Fan, C. Chen, Q. Zhao, J. Wei, and H. Zhang, "Identifying Potentially Climatic Suitability Areas for Arma custos (Hemiptera: Pentatomidae) in China under Climate Change," *Insects*, vol. 11, no. 10, Art. no. 10, Oct. 2020, doi: 10.3390/insects11100674.

Feng, D., Nan, X., Wei, S., Xuhong, R., Zhenyu, J., Qing, G., 2011. Study on exploratory factor analysis in measurement information extraction and reconstruction. In: In *IEEE 2011 10th International Conference on Electronic Measurement & Instruments*, pp. 76–79. https://doi.org/10.1109/ICEMI.2011.6037769.

S. E. Fick and R. J. Hijmans, "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas - Fick - 2017 - International Journal of Climatology - Wiley Online Library." Accessed: Apr. 17, 2024. [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086.

Fielding, A.H., Bell, J.F., Mar. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24 (1), 38–49. https://doi.org/10.1017/S0376892997000088.

Friedl, M., Sulla-Menashe, D., 2022. "MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061. NASA EOSDIS Land Processes Distributed Active Archive Center". https://doi.org/10.5067/MODIS/MCD12Q1.061.

Gardner, A.S., Maclean, I.M.D., Gaston, K.J., 2019. Climatic predictors of species distributions neglect biophysiologically meaningful variables. Divers. Distrib. 25 (8), 1318–1333. https://doi.org/10.1111/ddi.12939.

Gómez-Undiano, I., et al., Sep. 2022. Predicting potential global and future distributions of the African armyworm (Spodoptera exempta) using species distribution models. Sci. Rep. 12 (1), 16234. https://doi.org/10.1038/s41598-022-19983-y.

Guimapi, R.A., et al., Jun. 2022. Harnessing data science to improve integrated management of invasive pest species across Africa: an application to Fall armyworm (*Spodoptera frugiperda*) (J.E. Smith) (Lepidoptera: Noctuidae). Glob. Ecol. Conserv. 35, e02056.

Guo, X., et al., Apr. 2020. Prediction of the spatial distribution of soil arthropods using a random forest model: a case study in Changtu County, Northeast China. Agric. Ecosyst. Environ. 292, 106818 https://doi.org/10.1016/j.agee.2020.106818.

Hanberry, B.B., He, H.S., Palik, B.J., Aug. 2012. Pseudoabsence generation strategies for species distribution models. PLoS One 7 (8), e44486.

He, K.S., et al., 2015. Will remote sensing shape the next generation of species distribution models? Remote Sens. Ecol. Conserv. 1 (1), 4–18. https://doi.org/10.1002/rse2.7.

He, L., Zhao, S., Ali, A., Ge, S., Wu, K., Jun. 2021. Ambient Humidity affects development, survival, and reproduction of the invasive fall armyworm, spodoptera frugiperda (Lepidoptera: Noctuidae), in China. J. Econ. Entomol. 114 (3), 1145–1158. https://doi.org/10.1093/jee/toab056.

Hengl, T., et al., Feb. 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12 (2), e0169748.

Hirzel, A., Guisan, A., Nov. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecol. Model. 157 (2), 331–341. https://doi.org/10.1016/S0304-3800(02)00203-X.

Hong, H., Miao, Y., Liu, J., Zhu, A.-X., May 2019. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. Catena 176, 45–64. https://doi.org/10.1016/j.catena.2018.12.035.

T. Huang, T. Yang, K. Wang, and W. Huang, "Assessing the Current and Future Potential Distribution of Solanum rostratum Dunal in China Using Multisource Remote Sensing Data and Principal Component Analysis," *Remote Sens.*, vol. 16, no. 2, Art. no. 2, Jan. 2024, doi: 10.3390/rs16020271.

Y. Huang *et al.*, "Mapping the Spatio-Temporal Distribution of Fall Armyworm in China by Coupling Multi-Factors," *Remote Sens.*, vol. 14, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/rs14174415.

A. Jarvis, H. I. Reuter, A. Nelson, and E. Guevara, "Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m." 2008. doi: https://srtm.csi.cgiar.org.

Jiménez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Glob. Ecol. Biogeogr. 21 (4), 498–507. https://doi.org/10.1111/j.1466-8238.2011.00683.x.

Jung, J.-M., Lee, W.-H., Jung, S., 2016. Insect distribution in response to climate change based on a model: Review of function and use of CLIMEX. Entomol. Res. 46 (4), 223–235. https://doi.org/10.1111/1748-5967.12171.

M. K. Kansiime *et al.*, "Farmer perception of fall armyworm (Spodoptera frugiderda J.E. Smith) and farm-level management practices in Zambia", doi: 10.1002/ps.5504.

Koffi, D., et al., Jun. 2020. Maize Infestation of fall armyworm (Lepidoptera: Noctuidae) within agro-ecological zones of togo and ghana in west africa 3 Yr After Its invasion. Environ. Entomol. 49 (3), 645–650. https://doi.org/10.1093/ee/nvaa048.

Li, J., Zhang, B., Mao, Y., Jiang, J., Li, K., You, S., Jan. 2024. Comparative analysis of habitat suitability for a crop and its primary insect herbivore: providing insights for crop planting and pest management strategies. Front. Ecol. Evol. 11 https://doi.org/10.3389/fevo.2023.1305369.

Lihou, K., Wall, R., Nov. 2022. Predicting the current and future risk of ticks on livestock farms in Britain using random forest models. Vet. Parasitol. 311, 109806 https://doi.org/10.1016/j.vetpar.2022.109806.

Liu, T., Wang, J., Hu, X., Feng, J., Mar. 2020. Land-use change drives present and future distributions of Fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae). Sci. Total Environ. 706, 135872 https://doi.org/10.1016/j.scitotenv.2019.135872.

Loh, W.-Y., 2011. Classification and regression trees. Wires Data Min. Knowl. Discov. 1 (1), 14–23. https://doi.org/10.1002/widm.8.

Ma, L.B., et al., 2020. Multi-Scale Remote Sensing for Fall Armyworm Monitoring and Early Warning Systems. In: In IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Sep, pp. 4886–4889. https://doi.org/10.1109/IGARSS39084.2020.9323181.

F. Makale *et al.*, "Fall armyworm in Botswana: impacts, farmer management practices and implications for sustainable pest management", doi: 10.1002/ps.6717.

D. M. Makori *et al.*, "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.*, vol. 6, no. 3, Art. no. 3, Mar. 2017, doi: 10.3390/ijgi6030066.

Mishra, P.K., et al., Sep. 2023. A Comprehensive review on seasonal incidence of fall army worm, spodoptera frugiperda (J.E. Smith) and their Management. Int. J. Environ. Clim. Change 13 (10), 2458–2467. https://doi.org/10.9734/ijecc/2023/v13i102912.

Muñoz Sabater, J., 2019. "ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)". https://doi.org/10.24381/cds.68d2bb30.

S. Niassy, M. K. Agbodzavu, and E. Kimathi, "Bioecology of fall armyworm Spodoptera frugiperda (J. E. Smith), its management and potential patterns of seasonal spread in Africa | PLOS ONE." Accessed: Apr. 07, 2024. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249042.

Ørsted, I.V., Ørsted, M., 2019. Species distribution models of the Spotted Wing Drosophila (Drosophila suzukii, Diptera: Drosophilidae) in its native and invasive range reveal an ecological niche shift. J. Appl. Ecol. 56 (2), 423–435. https://doi.org/10.1111/1365-2664.13285.

B. Paudel Timilsena *et al.*, "Potential distribution of fall armyworm in Africa and beyond, considering climate change and irrigation patterns," *Sci. Rep.*, vol. 12, no. 1, p. 539, Jan. 2022, doi: 10.1038/s41598-021-04369-3.

Phillips, S.J., et al., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19 (1), 181–197. https://doi.org/10.1890/07-2153.1.

Powell-Romero, F., Fountain-Jones, N.M., Norberg, A., Clark, N.J., 2023. Improving the predictability and interpretability of co-occurrence modelling through feature-based joint species distribution ensembles. Methods Ecol. Evol. 14 (1), 146–161. https://doi.org/10.1111/2041-210X.13915.

Randin, C.F., et al., Mar. 2020. Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. Remote Sens. Environ. 239, 111626 https://doi.org/10.1016/j.rse.2019.111626.

M. W. Rhodes, J. J. Bennie, A. Spalding, R. H. ffrench-Constant, and I. M. D. Maclean, "Recent advances in the remote sensing of insects," *Biol. Rev.*, vol. 97, no. 1, pp. 343–360, Feb. 2022, doi: 10.1111/brv.12802.

I. Rwomushana, "Spodoptera frugiperda (fall armyworm)," *CABI Compend.*, vol. CABI Compendium, p. 29810, May 2019, doi: 10.1079/cabicompendium.29810.

S. G.C, A. Bhusal, and L. Khatri, "An introduction of fall armyworm (Spodoptera frugiperda) with management strategies: a review paper," *Nippon J. Environ. Sci.*, vol. 1, p. 1010, Jul. 2020, doi: 10.46266/njes.1010.

Senay, S.D., Worner, S.P., Ikeda, T., 2024. Novel Three-Step Pseudo-absence selection technique for improved species distribution modelling | PLOS ONE [Online]. Available: Accessed: Mar. 16 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0071218.

Shabani, F., Kumar, L., Ahmadi, M., 2018. Assessing accuracy methods of species distribution models: auc, specificity, sensitivity and the true skill statistic. Glob. J. Hum.-Soc. Sci. B Geogr. Geo-Sci. Environ. Sci. Disaster Manag. 18 (1), pp.

Shi, Y., Li, L.-Y., Shahid, S., Smagghe, G., Liu, T.-X., Feb. 2021. Effect of soil moisture on pupation behavior and inhabitation of Spodoptera frugiperda (Lepidoptera: Noctuidae). Appl. Entomol. Zool. 56 (1), 69–74. https://doi.org/10.1007/s13355-020-00709-3.

Sims, S.R., Oct. 2008. Influence of Soil Type and Rainfall on Pupal Survival and Adult Emergence of the Fall Armyworm (Lepidoptera: Noctuidae) in southern florida. J. Entomol. Sci. 43 (4), 373–380. https://doi.org/10.18474/0749-8004-43.4.373.

Srivastava, V., Lafond, V., Griess, V.C., Apr. 2019. Species distribution models (SDM): applications, benefits and challenges in invasive species management. CABI Rev. 2019, 1–13. https://doi.org/10.1079/PAVSNNR201914020.

Tambo, J.A., et al., Oct. 2020. Tackling fall armyworm (Spodoptera frugiperda) outbreak in Africa: an analysis of farmers' control actions. Accessed: Jun. 23, 2024. [Online]. Available: Int. J. Pest Manag. https://www.tandfonline.com/doi/abs/10.1080/09670874.2019.1646942.

Tepa-Yotto, G.T., Chinwada, P., Rwomushana, I., Goergen, G., Subramanian, S., Aug. 2022. Integrated management of Spodoptera frugiperda 6 years post detection in Africa: a review. Curr. Opin. Insect Sci. 52, 100928 https://doi.org/10.1016/j.cois.2022.100928.

G. T. Tepa-Yotto *et al.*, "Global Habitat Suitability of Spodoptera frugiperda (JE Smith) (Lepidoptera, Noctuidae): Key Parasitoids Considered for Its Biological Control," *Insects*, vol. 12, no. 4, Art. no. 4, Apr. 2021, doi: 10.3390/insects12040273.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2021. Modelling species presence-only data with random forests. Ecography 44 (12), 1731–1742. https://doi.org/10.1111/ecog.05615.

Wang, J., et al., Apr. 2023. Migration risk of fall armyworm (Spodoptera frugiperda) from North Africa to Southern Europe. Front. Plant Sci. 14 https://doi.org/10.3389/fpls.2023.1141470.

J. Wang, J. Deng, W. Yan, and Y. Zheng, "Habitat Suitability of Pine Wilt Disease in Northeast China under Climate Change Scenario," *Forests*, vol. 14, no. 8, Art. no. 8, Aug. 2023, doi: 10.3390/f14081687.

Warren, D.L., Matzke, N.J., Iglesias, T.L., 2020. Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. J. Biogeogr. 47 (1), 167–180. https://doi.org/10.1111/jbi.13705.

M. W. Watkins, *A Step-by-Step Guide to Exploratory Factor Analysis with SPSS*. 2021. Accessed: Apr. 07, 2024. [Online]. Available: https://www.taylorfrancis.com/books/mono/10.4324/9781003149347/step-step-guide-exploratory-factor-analysis-spss-marley-watkins.

Wisz, M.S., Guisan, A., Apr. 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol. 9 (1), 8. https://doi.org/10.1186/1472-6785-9-8.

Xiang, H., et al., Apr. 2023. Mapping potential wetlands by a new framework method using random forest algorithm and big Earth data: a case study in China's Yangtze River Basin. Glob. Ecol. Conserv. 42, e02397.

Yoon, S., Lee, W.-H., Nov. 2021. Methodological analysis of bioclimatic variable selection in species distribution modeling with application to agricultural pests (*Metcalfa pruinosa* and *Spodoptera litura*). Comput. Electron. Agric. 190, 106430 https://doi.org/10.1016/j.compag.2021.106430.

Yoon, S., Lee, W.-H., Feb. 2023. Application of true skill statistics as a practical method for quantitatively assessing CLIMEX performance. Ecol. Indic. 146, 109830 https://doi.org/10.1016/j.ecolind.2022.109830.

Zhang, J., et al., Oct. 2019. Monitoring plant diseases and pests through remote sensing technology: A review. Comput. Electron. Agric. 165, 104943 https://doi.org/10.1016/j.compag.2019.104943.

Zhang, Y., Hughes, A.C., Zhao, Z., Li, Z., Qin, Y., Apr. 2022. Including climate change to predict the global suitable area of an invasive pest: *Bactrocera correcta* (Diptera: Tephritidae). Glob. Ecol. Conserv. 34, e02021.

Zhi, X., Du, H., Zhang, M., Long, Z., Zhong, L., Sun, X., Dec. 2022. Mapping the habitat for the moose population in Northeast China by combining remote sensing products and random forests. Glob. Ecol. Conserv. 40, e02347.