



## Article

# Dynamic Forecast of Desert Locust Presence Using Machine Learning with a Multivariate Time Lag Sliding Window Technique

Ruiqi Sun <sup>1,2</sup>, Wenjiang Huang <sup>1,2</sup>, Yingying Dong <sup>1,2,\*</sup>, Longlong Zhao <sup>3</sup>, Biyao Zhang <sup>1</sup>, Huiqin Ma <sup>1</sup>, Yun Geng <sup>1,2</sup>, Chao Ruan <sup>1,2</sup>, Naichen Xing <sup>4</sup>, Xidong Chen <sup>5</sup> and Xueling Li <sup>1</sup>

- <sup>1</sup> State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; sunruiqi19@mailsucas.ac.cn (R.S.); huangwj@aircas.ac.cn (W.H.); zhangby@aircas.ac.cn (B.Z.); mahq@aircas.ac.cn (H.M.); gengyun17@mailsucas.ac.cn (Y.G.); ruanchao@aircas.ac.cn (C.R.); lixl@aircas.ac.cn (X.L.)
- <sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China
- <sup>3</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; ll.zhao@siat.ac.cn
- <sup>4</sup> China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, Beijing 100083, China; xingnc@aircas.ac.cn
- <sup>5</sup> North China University of Water Resources and Electric Power, Zhengzhou 450046, China; chenxd@radi.ac.cn
- \* Correspondence: dongyy@aircas.ac.cn

**Abstract:** Desert locust plagues can easily cause a regional food crisis and thus affect social stability. Preventive control of the disaster highlights the early detection of hopper gregarization before they form devastating swarms. However, the response of hopper band emergence to environmental fluctuation exhibits a time lag. To realize the dynamic forecast of band occurrence with optimal temporal predictors, we proposed an SVM-based model with a temporal sliding window technique by coupling multisource time-series imagery with historical locust ground survey observations from between 2000–2020. The sliding window method was based on a lagging variable importance ranking used to analyze the temporal organization of environmental indicators in band-forming sequences and eventually facilitate the early prediction of band emergence. Statistical results show that hopper bands are more likely to occur within 41–64 days after increased rainfall; soil moisture dynamics increasing by approximately  $0.05 \text{ m}^3/\text{m}^3$  then decreasing may enhance the chance of observing bands after 73–80 days. While sparse vegetation areas with NDVI increasing from 0.18 to 0.25 tend to witness bands after 17–40 days. The forecast model combining the optimal time lags of these dynamic indicators with other static indicators allows for a 16-day extended outlook of band presence in Somalia, Ethiopia, and Kenya. Monthly predictions from February to December 2020 display an overall accuracy of 77.46%, with an average ROC-AUC of 0.767 and a mean F-score close to 0.772. The multivariate forecast framework based on the lagging effect can realize the early warning of band presence in different spatiotemporal scenarios, supporting early decisions and response strategies for desert locust preventive management.

**Keywords:** desert locust; environmental indicator; dynamic forecast; machine learning; time lag



**Citation:** Sun, R.; Huang, W.; Dong, Y.; Zhao, L.; Zhang, B.; Ma, H.; Geng, Y.; Ruan, C.; Xing, N.; Chen, X.; et al. Dynamic Forecast of Desert Locust Presence Using Machine Learning with a Multivariate Time Lag Sliding Window Technique. *Remote Sens.* **2022**, *14*, 747. <https://doi.org/10.3390/rs14030747>

Academic Editor: Tania Stathaki

Received: 20 December 2021

Accepted: 31 January 2022

Published: 5 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The 2019–2021 desert locust (*Schistocerca gregaria*, Forskål) upsurge has posed a major threat to the food security and regional stability of East Africa, the Middle East, and South-West Asia [1]. Countries such as Ethiopia and Somalia have witnessed the worst damage for nearly 25 years, while for Kenya, this is the worst damage for almost 70 years [2]. Desert locust plagues are consequences of changes in population density, triggering solitary individuals to transform into gregarious hopper bands and fledged swarms with booming populations [3–5]. The invasion of swarms is incredibly rapid and destructive, destroying crops and pastures in their path [5]. Due to the mass destruction and high mobility of locust

swarms, it is more cost effective and environmentally friendly to take control measures before the maturity and aggregation of isolated populations [6–9]. Thus, a preventive management strategy (PMS) was adopted by the Food and Agriculture Organization (FAO) of the United Nations [10,11]. PMS relies on early warning of and early reaction to hopper gregarization at the beginning of outbreaks [12–15]. Therefore, it is vital to predict more accurate and timelier band presence of desert locusts dynamically in order to navigate the ground surveyors to the gathering nymphs and treat them before they become another devastating air force [16].

Traditional methods employed by the Desert Locust Information Service (DLIS) of FAO include the desert locust development model (DLDM) and trajectory model (DLTM), which are based on weather forecasts [17]. DLDM is a phenology model that depends on the temperature to estimate the growth rate of eggs and hoppers and thus predicts hopper occurrence, for example, the widely used degree-day model [18]. Whereas DLTM methods, e.g., the HYSPLIT method originating from atmospheric transport and dispersion models, track the displacement of adults with air temperature, pressure, and wind to analyze the invasion risk and simulate the presence of their offspring [10,19,20]. Weather forecasts, primarily the prediction of seasonal rainfall and temperature, facilitated both methods to fill the time extension for long-term predictions [21]. Overall, these meteorological-based models are beneficial for dynamic prediction due to the high temporal resolution of weather products. However, they are also limited by a lack of soil and vegetation information; therefore, they must be managed with intelligence by experienced forecasters [17].

Remote sensing imagery has been increasingly combined with meteorological data and applied in monitoring desert locust breeding habitats where hoppers and bands were produced [5,9,22]. Indeed, the band emergence of desert locusts is influenced by various driving factors, including bioclimatic, edaphic, and host vegetation. Egg pods are laid in sandy soil about 5–10 cm below the surface [12]. Successful hatching requires the uptake of sufficient water [18,23]. After breaking ground, nymphs rely on annual herbs, shrubs, and crops as food and roost [12,23–25]. According to the desert locust biology, Gómez et al. [26] utilized the 1 km MODIS land surface temperature and 9 km Soil Moisture Active Passive (SMAP) root zone estimation to monitor the presence of desert locusts. Escorihuela et al. [27] first attempted to include high-resolution satellite soil moisture in desert locust surveillance based on Sentinel-1 data in synergy with the Soil Moisture and Ocean Salinity (SMOS) microwave observations. Ellenburg et al. [28] evaluated the utility of 3 km land surface model-derived soil moisture products to monitor the nymph distribution. Vegetation parameters mostly derived from medium resolution data fusion products of MODIS and SPOT, such as the vegetation index [29–31], leaf area index [26], green leaf biomass [32], and greenness [33,34], have also had their feasibility tested to map the pest distribution.

Previous studies have confirmed the potential of various near-real-time environmental indicators derived from multisource imagery to detect the presence of desert locusts. Nevertheless, the response of desert locust emergence to environmental fluctuation exhibits a time lag. There is an average of three months between the locust eggs being laid and hatching and the growing hoppers or bands being observed by humans [12,18]. Therefore, the time scale of environmental indicators based on the phenology of desert locusts should be considered in order to extract optimal predictors with temporal variables. In this context, recent research has begun to shift its attention towards the lagging effect of environmental drivers of hopper occurrence during the recession period; soil moisture has received broad interest [28,35–38]. For example, Piou et al. [36] investigated the optimal delays of soil moisture in order to model the hopper emergence through statistical methods.

Although soil water highly influences the incubation of locust eggs, the lagging effect of temperature and vegetation related to hopper growth, distribution, and gregarization still lacks discussion [24,26,39–41]. Other geographical factors, such as topography [42], soil silt or clay content [28,43], and land cover [44–47], have also been proved to bear the potential to map locust biotopes. They play an irreplaceable role but are rarely included in the forecast toolkit. Moreover, most empirical studies on the early warning of hopper

aggregation have focused on monitoring and predicting hopper occurrence in the recession period, whereas few investigated the band presence.

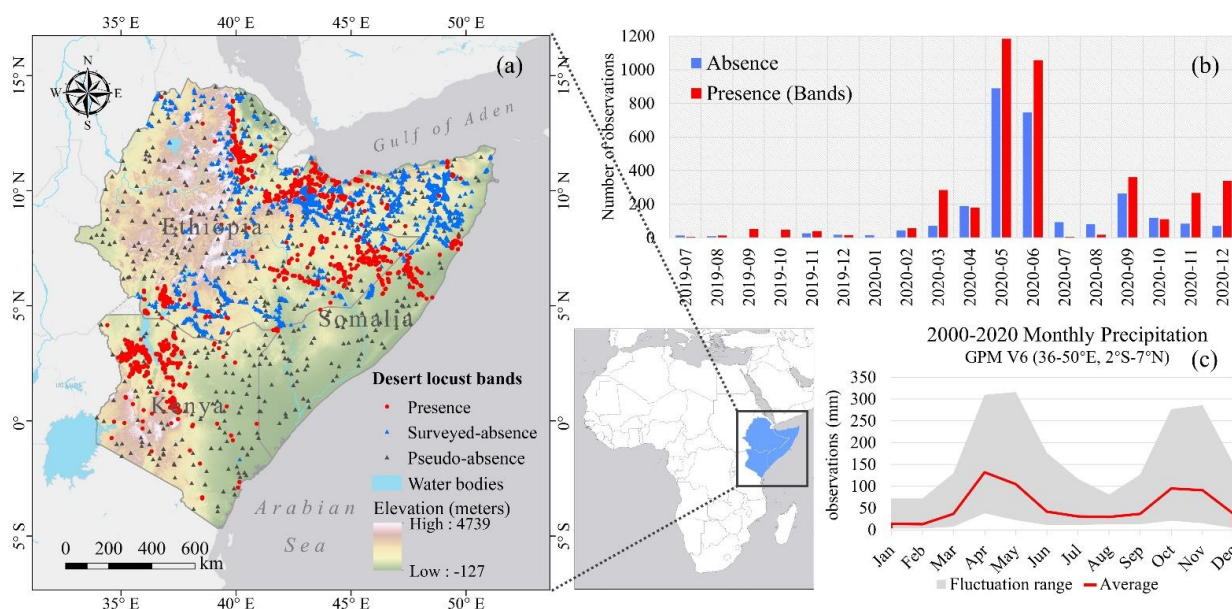
Therefore, we proposed a novel multivariate forecast framework of desert locust band emergence in Somalia, Ethiopia, and Kenya. In addition to the long-term surveillance of solitary individuals, this study is the first attempt to predict hopper band occurrence in order to support the early warning of outbreak risks. The main purposes of this study include (1) coupling historical locust ground survey data and multisource Earth observation data to uncover the lagging effect of critical indicators and to analyze its relationship to desert locust phenology; (2) developing a machine learning-based prediction model with a temporal sliding window selector of multiple indicators to dynamically forecast the band presence; (3) evaluating the accuracy, feasibility and stability of forecast framework to support desert locust PMS.

## 2. Materials and Methods

### 2.1. Study Area and Locust Historical Field Survey Dataset

#### 2.1.1. Study Area

This study concerns Somalia, Ethiopia, and Kenya (SEK) in the Great Horn of Africa, shown in Figure 1a. It extends eastwards to the Arabian Sea, across the Gulf of Aden from Yemen. Complex topography, arid and semi-arid climate, and sea surface temperature anomalies, such as the El Niño Southern Oscillation, shape the bimodal annual rainfall pattern over much of SEK [48]. It includes a long rainy season from March to May and a short rainy season from October to December (see Figure 1c). Hence, SEK can form complementary breeding areas activated by either spring, summer, or winter rains [15,18,49]. Since 2018, the Indian Ocean Dipole has experienced extreme positive phases, with tropical cyclones, Sagar, in May 2018, and Pawan, in late 2019, taking turns to hit coastal regions [2,50]. They brought SEK extraordinary rainfall, providing suitable conditions for desert locust breeding [50,51]. Mass migration of swarms from the Arabian Peninsula since June 2019 has culminated in an outbreak of desert locusts in SEK [51].



**Figure 1.** Spatial and temporal distribution of ground points of the desert locust band used for this study in the SEK region. (a) The geographical location of SEK with band observations in 2000–2020; the red dot represents band presence while the blue triangle refer to surveyed-absence and the grey one indicates pseudo-absence. (b) Monthly count of bands from July 2019 to December 2020. (c) Monthly observations of Global Precipitation Measurement (GPM) V6 in the central SEK region for 20 years (2000–2020); the red line indicates monthly mean rainfall; the grey area indicates the fluctuation interval.

### 2.1.2. Locust Dataset and Data Cleaning

Locust ground points in this study were mainly derived from the DLIS-FAO dataset (<https://locust-hub-hqfao.hub.arcgis.com/> (Accessed: 20 December 2021)). This dataset compiles ground survey observations spanning 25 years, from 1986 to the present, covering some 29 million km<sup>2</sup> spreading range of the species. We focused on records of band presences and absences in SEK between 2000 and 2020, which corresponds to the time range covered by other multisource datasets (see Figure 1a). Band observations are concentrated in May–June and November–December 2020 (see Figure 1b).

Firstly, we rid the records of inaccuracies in the data collection process. Four built-in columns in the dataset, with Boolean values recording whether the sampling location and species are reliable and whether the entry is overall confirmed and credible, were defined as the quality control fields. So, we removed observations with approximate locations, uncertain species, and unconfirmed or dubious reports according to the quality control fields. We also eliminated duplicate records with the same date, latitude, and longitude, which were likely to be repeated observations of the same population.

### 2.1.3. Pseudo-Absence Generation

Nevertheless, some uncertainties remain in the dataset, especially the absence samples caused by the spatial bias and temporal discontinuity of data collection [31]. Surveyed absence (SA) points are usually sampled in potential locust-prone areas of preference [31]. Therefore, we generated pseudo-absence (PA) points by a random generation with exclusion buffer (RGEB) [52]. RGEB uses a buffer to adjust the distance between PA and presence points to avoid both appearing in one grid [53]. Empirical studies suggest that buffer size depends on the data at hand [54–56]. Considering the maximum spatial resolution (~10 km) of our multisource data, we adopted a 10 km radius buffer around each presence site for the PA generation.

Different proportions of absence and presence can affect model performance positively or negatively [57]. We generated different numbers of PA and carried out accuracy and precision tests at different presence-absence ratios (see Supplementary Material 1). In accordance with the results of the experiment, a small percentage of 500 PA points randomly assigned sampling dates in 2019–2020 were added to represent those areas rarely visited by locusts in those years (see Figure 1a). Consequently, a total of 5272 ground points of band occurrence were sampled from the SEK region in 2000–2020 after cleaning, including 2389 presences and 2888 absences (see Figure 1a).

## 2.2. Extraction of Environmental Indicators of Desert Locust Presence

### 2.2.1. Multiple Indicators from Multisource Data

Ten vital influential factors and their most typical indicators were selected based on desert locust biological and applied research [23]. They were extracted from multiple data sources, including meteorological, remote sensing, and basic geographic data and were then divided into dynamic and static categories (see Table 1 for more details) [5]. Dynamic indicators, including precipitation (PREC), soil moisture (SM), normalized difference vegetation index (NDVI), and surface temperature (LST) change rapidly; hence, they are usually observed with high temporal resolutions. Nevertheless, the static indicators are relatively stable, with observations on an annual or even decadal basis, including elevation (DEM), land use and land cover (LULC), soil sand content (SND), soil clay content (SCL), soil silt content (SLT), and soil coarse fragment content (CRF).

Datasets were selected, catering to the desert locust biology and the needs of proposed forecast framework. Stable products with long time series and high temporal resolution were selected to build time series of dynamic drivers and, thus, to investigate the delayed response of band occurrence to these indicators by coupling with the DLIS-FAO dataset. Whereas datasets with higher spatial resolution were selected for a better overall spatial accuracy of prediction. For example, the rainfall assimilation product, which combines satellite imagery with in-situ station data, was chosen for its better performance

in Africa [58–60]. Hourly SM observation converted to daily composite data was used for time series extraction, while a modelled product with a higher spatial resolution was used as SM predictors for the forecast [61–63]. Both layers met the requirement, as much as they could, for monitoring in the root zone where females spawn at a 2–15 cm depth rather than the sediment surface [27]. We also chose the latest version of datasets for static drivers to represent the 2020 scenario for the forecast.

**Table 1.** Multiple indicators and multisource datasets.

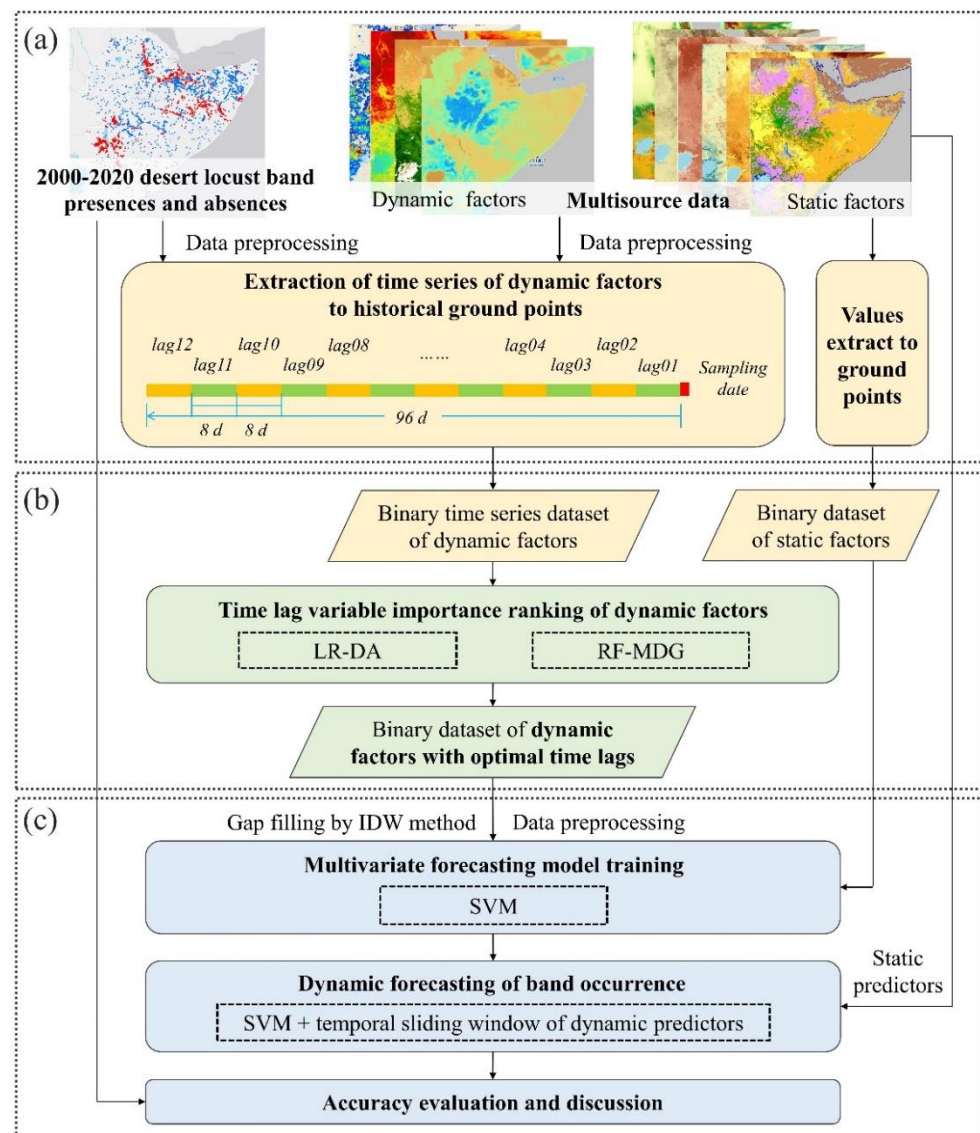
| Indicators         | Data Source | Spatial Resolution                   | Temporal Resolution | Time Range | Description      |   |
|--------------------|-------------|--------------------------------------|---------------------|------------|------------------|---|
| Dynamic indicators | PREC        | CHIRPS <sup>1</sup>                  | 0.05° (~5 km)       | daily      | 1981–2021        | Precipitation (mm/day)  |
|                    | SM          | ERA5-Land <sup>2</sup>               | 0.1° (~10 km)       | hourly     | 1981–2021        | Volumetric soil water layer 2 (7–28 cm below the surface) (m <sup>3</sup> /m <sup>3</sup> ) |
|                    |             | SERVIR <sup>3</sup>                  | 0.03° (~3 km)       | daily      | 2020–2021        | Volumetric soil moisture (0–10 cm below the surface) (%)                                    |
|                    | NDVI        | MOD09GA                              | 1 km                | daily      | 2000–2021        | Red and NIR band normalization  |
|                    | LST         | MOD11A1                              | 1 km                | daily      | 2000–2021        | Units converted from K to °C  |
| Static indicators  | SND         | ISRIC <sup>4</sup><br>SoilGrids [64] | 250 m               | 3-year     | 2014, 2017, 2021 | Sand in depth of 5–15 cm (g/kg)   |
|                    | CLY         |                                      |                     |            |                  | Clay content in depth of 5–15 cm (g/kg)   |
|                    | SLT         |                                      |                     |            |                  | Silt in depth of 5–15 cm (g/kg)   |
|                    | CRF         |                                      |                     |            |                  | Coarse fragments in depth of 5–15 cm (cm <sup>3</sup> /dm <sup>3</sup> )                    |
|                    | LULC        | CGLS-LC100 <sup>5</sup><br>[65]      | 100 m               | 5-year     | 2015–2019        | Discrete classification   |
|                    | DEM         | NASADEM<br>[66]                      | 30 m                | -          | 2020             | Elevation (m)   |

<sup>1</sup> CHIRPS: Climate Hazards Group InfraRed Precipitation with Station. <sup>2</sup> ERA5: European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5. <sup>3</sup> SERVIR: Not an abbreviation but a Spanish word means “to serve”. <sup>4</sup> ISRIC: International Soil Reference and Information Centre. <sup>5</sup> CGLS: Copernicus Global Land Cover Layers.

All data were pre-processed with geometric correction, radiometric calibration, atmospheric correction, and cloud removal. Due to our preference for high temporal resolution data, imagery used in this study was then resampled to a unified spatial resolution of 1 km. Preliminary evidence in previous empirical studies showed that the environmental indicators at 1 km spatial resolution were representative and suitable for regional, large-scale monitoring and forecasts [27,30,36].

### 2.2.2. Extraction of Time Series of Dynamic Indicators

We realized a bulk automated extraction of time series of dynamic indicators to ground observation sites. We assumed that it took an average of around 96 days before the hopper bands were observed, based on the locust biology and empirical studies (see Figure 2a) [23]. Imagery of four dynamic indicators were dynamically selected by a sliding window of the given 96-day band development cycle prior to the sampling date, and daily mean values were extracted to each point (noted as  $d_{96}$ ,  $d_{95}$ ,  $d_{94}$ ,  $d_{93}$  . . . ,  $d_{02}$ ,  $d_{01}$ ). Based on historical ground survey sites in SEK between 2000 and 2020, we generated over 5000 daily time series of each dynamic indicator at each point.



**Figure 2.** Methods and process. (a) Extraction of dynamic and static indicators of desert locust presence. (b) Time lag variable importance ranking for dynamic indicators. (c) Forecast of desert locust presence based on multivariate machine learning model with a temporal sliding window.

For actual preventive control operations, daily to seasonal forecasts to detect band emergence grounds are required. The frequency of reporting varies with the situation and scale. A forecast of 8 to 10 days is usually used on a national scale. Thus, an 8-day time step was adopted to cut the time series. Meanwhile, the mean values of each eight-day period were calculated as a secondary time slice noted as 12 time-lagged variables ( $lag12, lag11, lag10 \dots, lag02, lag01$ ). A time-series dataset of four dynamic indicators for the development sequence of bands was eventually constructed (see Figure 2a).

We then used the dataset to analyze and evaluate the variable importance of the time lag in order to select the best temporal window for each dynamic indicator for forecast in Section 2.3. Meanwhile, we exposed and discussed the lagging response of band presence to each indicator buried in these statistical results.

### 2.2.3. Extraction of Static Indicators

Static indicators extracted from multisource Earth observation data such as topography, soil texture, and land use, which represent inherent ecological conditions of the study area, are also indispensable for the forecast of locust bands. Thus, we also overlaid ground

points on the imagery of six static indicators in order to extract the static values to these locations (see Figure 2a). These values were then added to the time-series dataset created in Section 2.2.2 to construct a complete dichotomous dataset (see Figure 2b). This dataset with static indicators and other dynamic indicators was finally used for the dynamic forecast of band presence in Section 2.4.

### 2.3. Time Lag Variable Importance Ranking for Dynamic Indicators

We constructed two variable importance evaluation models for each dynamic indicator with all their time lag variables based on the 2000–2020 band occurrence classified dataset with multivariate time-series indicators established in Section 2.2.3 (see Figure 2b). Logistic Regression-Dominance Analysis (LR-DA) and Random Forest-Mean Decrease Gini (RF-MDG) were classic methods widely used for feature importance ranking and feature selection [67,68]. LR is simple but efficient and interpretable. In contrast, RF makes up for the inability of LR to solve non-linear problems whilst being able to quantify the relative contribution of features. Therefore, statistical results of the two methods were comprehensively analyzed to draw reliable conclusions on the most significant time lag variables (see Figure 2b). LR described the conditional probability of desert locust hopper band occurring (Equation (1)) or not occurring (Equation (2)) by the sequential variables of dynamic indicators at each ground observation site [69]:

$$P(Y = 1|X) = \frac{e^{(WX+B)}}{1+e^{(WX+B)}} \quad (1)$$

$$P(Y = 0|X) = \frac{1}{1+e^{(WX+B)}} \quad (2)$$

where the binary state of band absence or presence at each ground observation point is  $Y = (Y_1, Y_2, \dots, Y_k) \in \{0, 1\}$ ;  $k$  denotes the ordinal number of 5272 ground points,  $k \in N$ . The time lag variable for each dynamic indicator is  $X = (X_{1,j,k}, X_{2,j,k}, \dots, X_{i,j,k})$ ;  $i$  identifies the time lag variables  $\{lag01, lag02, lag03 \dots, lag12\}$ ,  $i \in [1, 12]$ ,  $i \in N$ ;  $j$  is the ordinal number of four dynamic indicators,  $j \in \{1, 2, 3, 4\}$ , corresponding to PREC, SM, NDVI, and LST, respectively. The weight coefficients corresponding to the time lag variables of each environmental indicator are  $W = (w_{1,j}, w_{2,j}, \dots, w_{i,j})^T$ ; whereas  $B = (b_1, b_2, \dots, b_j)^T$  refers to the bias constants of each environmental indicator.

LR-DA calculated the relative importance scores for lag features [70,71]. McFadden  $R^2$  was selected to measure the conditional and additional contributions for all possible sets of variable groups and finally to determine the combined contribution ranking of all lagging indicators [72]. At the same time, RF-MDG reported the importance of variables through evaluating the improvement in error at each node for each randomly selected variable and the ratio for all nodes in the forest [67,73]. The normalized results of the relative importance ranking of LR and RF were combined to identify one to three optimal lagging variables for each dynamic indicator.

### 2.4. Model for Forecast Based on a Temporal Sliding Window

The Support Vector Machine (SVM) has been widely used in species distribution simulation for its simplicity and robustness in maximizing separation margins [31,74]. RF has also been proven to perform well in modeling the presence-absence of species, so we conducted a comparative experiment of both machine learning methods (see Supplementary Material 2). We finally decided to use SVM as the fundamental model for the forecast for a better overall accuracy throughout the year. For dynamic indicators, a temporal sliding window selector was invented to choose trainers and predictors dynamically based on the time lag information mining from the historical ground survey information and long time series of satellite imagery in Section 2.3. Lagging variables of dynamic indicators with lower significance were removed and those that contributed highly survived. We then combined other static indicators for model training and prediction. Thus, we proposed a

data-driven multivariate approach combining machine learning and a temporal sliding window to predict band occurrence for early warning (see Figures 2c and 3).

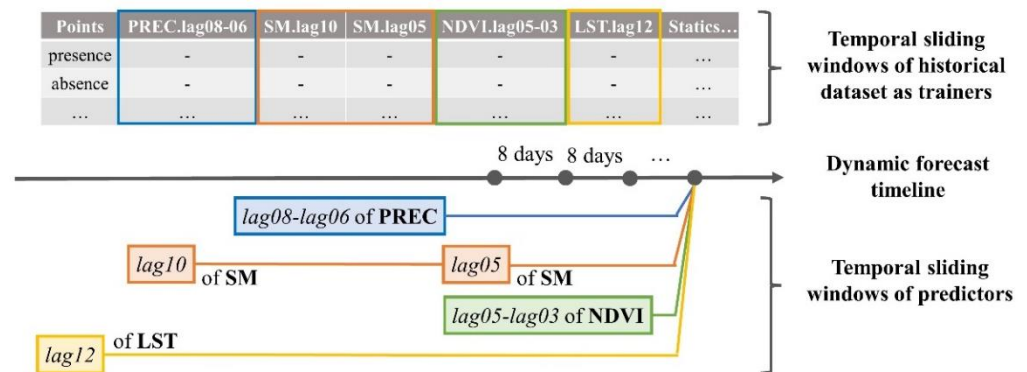


Figure 3. Illustration of time lag sliding window technique.

Theoretically, this method allowed for dynamic forecasts of band occurrence every 8 days. In this paper, monthly forecasts from February to December 2020 (20th of each month) were implemented as an example to validate the model performance and robustness. The 2000–2020 historical dataset constructed in Section 2.2, with all of the static indicators and optimal time lag variables of each dynamic indicator retained, was used as training sets. Records in the target month were set aside as ground truth for each training set in order to evaluate this prediction afterwards. A filtered subset then trained an optimal SVM regression model with a radial-based function by tuning hyperparameters.

Moreover, images in the optimal time lag windows of each dynamic indicator and all static images were prepared as a batch of predictors. We resampled the multisource heterogeneous data to the spatial resolution of 1 km. Alongside other elementary preprocessing, we also used the Inverse Distance Weighted (IDW) method to fill in the blanks for images with vast missing pixels after cloud masking. Predictors were input into the training models that we built. Raw values of regressions were normalized to represent the likelihood of desert locust bands occurring.

### 2.5. Model Evaluation and Accuracy Assessment

The model was validated by both presences and absences of the 2020 ground observations set apart from the historical datasets (see Figure 2c). We divided the regression results into three categories, low, average, and high probability of band presence. The probability threshold, when maximizing the sensitivity and specificity or true positive rate and true negative rate (max TPR + TNR) derived from the training model, was defined as the optimal splitting point to divide the regression results into low and average probability of band presence for model evaluation (see Figure 4) [75,76]. Projected results with average probability of band presence were then partitioned by the first quartile, while the high probability was merely used for better visualization and demonstration.

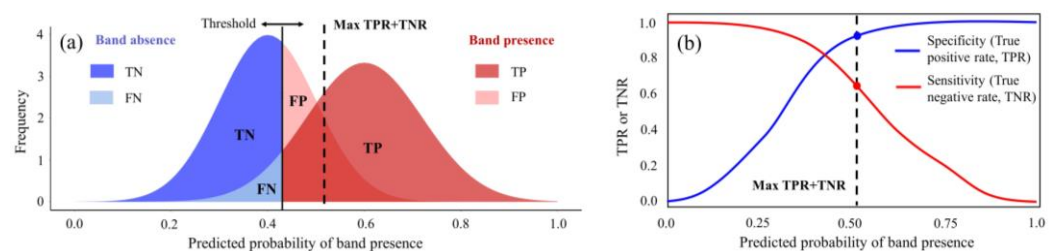


Figure 4. Illustration of max TPR + TNR threshold for binary classification. (a) Frequency of band presence-absence distribution [77]; (b) sensitivity-specificity curve.

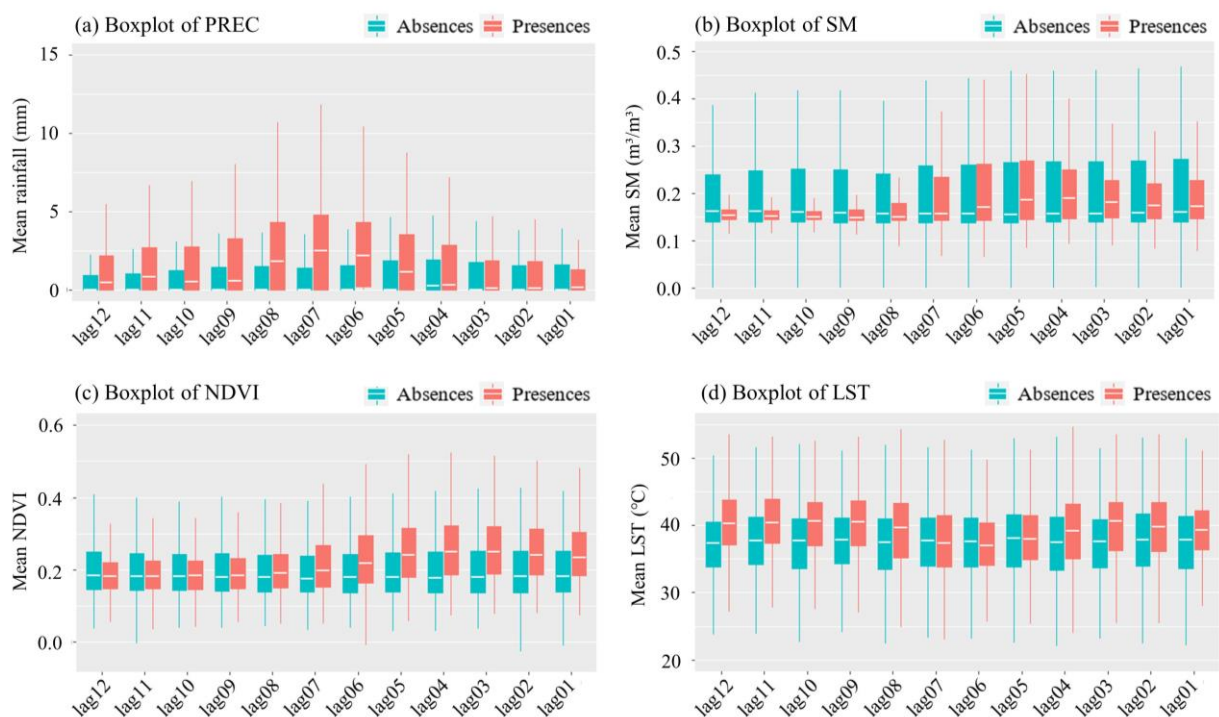


The monthly reserved band occurrence as ground truths were overlaid on the classified results to build confusion matrixes. If a presence point fell within the predicted area with average or high probability of band presence, it was recorded as a true positive (TP); otherwise, it was a false negative (FN). If an absence point fell within the predicted low probability area for band presence, it was recorded as a true negative (TN); otherwise, it was a false negative (FN). Nevertheless, absence observation indicates only a failure to observe any band in sight in this location; it does not make it unsuitable for locust breeding or band developing. Thus, F-score was selected to focus on the positive performance by a combined consideration of precision and recall. Ultimately, five evaluation metrics including the overall accuracy, sensitivity, specificity, ROC-AUC, and F-score were calculated based on the confusion matrix.

### 3. Results

#### 3.1. Lagging Response of Hopper Band Occurrence to Environmental Drivers

Figure 5 depicts the value distribution of four environmental drivers over the desert locust band-forming sequences. PREC, SM, and NDVI showed a similar trend in the band-forming period. Instead, LST probably exhibited an opposite trend or lacked an apparent sequential characteristic. All indicators where hopper bands were absent remained roughly stable, but where hopper bands were present, indicators experienced a rise at different time lags.

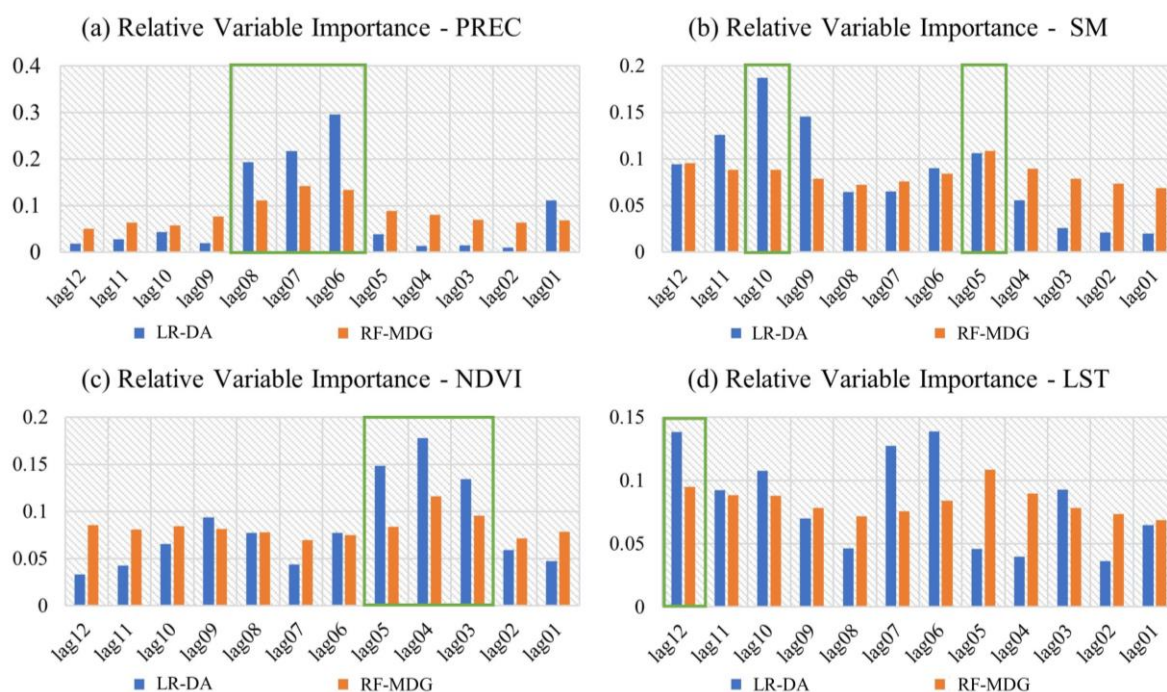


**Figure 5.** Value distributions of dynamic indicators at different time-lagged variables. (a) Boxplot of PREC. (b) Boxplot of SM. (c) Boxplot of NDVI. (d) Boxplot of LST. The red box represents the value distributions of dynamic indicators in band presences while the green box refers to that in band absences.

PREC (see Figure 5a) at band occurrence consistently obtained higher medians over the breeding and developing period of hopper bands and gradually increased until it peaked at *lag07*. The differences between value distributions of presences and absences reached their maximum from *lag08* to *lag06*. The SM factor (see Figure 5b) at band attendance fell behind band absence for the first 32 days from *lag12* to *lag08*. They then elevated to a peak at around 0.2 ( $\text{m}^3/\text{m}^3$ ) from 40 (*lag05*) to 17 (*lag03*) days before bands were observed. The trend of SM is similar to that of PREC, however, its peak was delayed by 16 days compared to PREC. NDVI (see Figure 5c) followed the same trend as PREC and SM, peaking at about 0.25 from 32 (*lag04*) to 9 (*lag02*) days ahead of the observation, with a further delay of

about 8 to 16 days compared to SM. Additionally, the median LST (see Figure 5d) of band presence revealed possibly opposite volatility in the propagated region compared to other factors. LST of band presence remained high from *lag12* to *lag09* and then decreased and rose again to its highest point.

Figure 6 demonstrates the relative importance rankings of time lag variables for each dynamic indicator. Larger contributions appeared essentially at time lags with the most significant difference of values between presences and absences in Figure 5. Due to the spatial heterogeneity of the phenology of desert locusts on the large scale of the study area, multiple consecutive lags were preferred in this study to provide the sliding window with a temporal tolerance. Results of LR-DA and RF-MDG agreed that the PREC factor had the highest relative importance in the temporal window from 64 days (*lag08*) to 41 days (*lag06*) before band observation (see Figure 6a). NDVI also made a higher contribution from 40 days (*lag05*) to 17 days (*lag03*) ahead of band detection without any doubt (see Figure 6c).



**Figure 6.** The normalized relative importance of time-lagged variables of each dynamic indicator. The green box refers to the optimal temporal sliding window. (a) Relative variable importance of PREC. (b) Relative variable importance of SM. (c) Relative variable importance of NDVI. (d) Relative variable importance of LST. The blue bar represents the relative variable importance of LR-DA while the orange box refers to that of RF-MDG.

However, the two approaches had slightly different views on the optimal time lag of the SM factor. One indicated that the time lag that contributed most to the model was *lag05* (40 to 33 days before band emergence), whereas the other also detected a sub-peak at *lag05* but ultimately supported *lag10* (80 to 73 days before band occurrence). Combined with the boxplot of SM (see Figure 6b), both votes of the optimal lagging window should be accepted. Moreover, the results of RF concluded that the contribution of *lag05* was the most considerable, with the next largest contribution from *lag12*. Whereas the rankings of LR displayed a discontinuity, tending to suggest that *lag12* made the highest contribution, that *lag06* took second place, and that *lag05* made a meagre contribution. However, directly using near-real-time LST indicators with no delay as predictors was still not a good choice compared to other lagged variables. We finally identified *lag12* as the most important variable of LST (see Figure 6d). To sum up, we eventually adopted mean PREC from

*lag08* to *lag06*, SM at *lag10* and *lag05*, mean NDVI from *lag05* to *lag03*, and LST at *lag12* to participate in the forecast (see the green boxes in Figure 6).

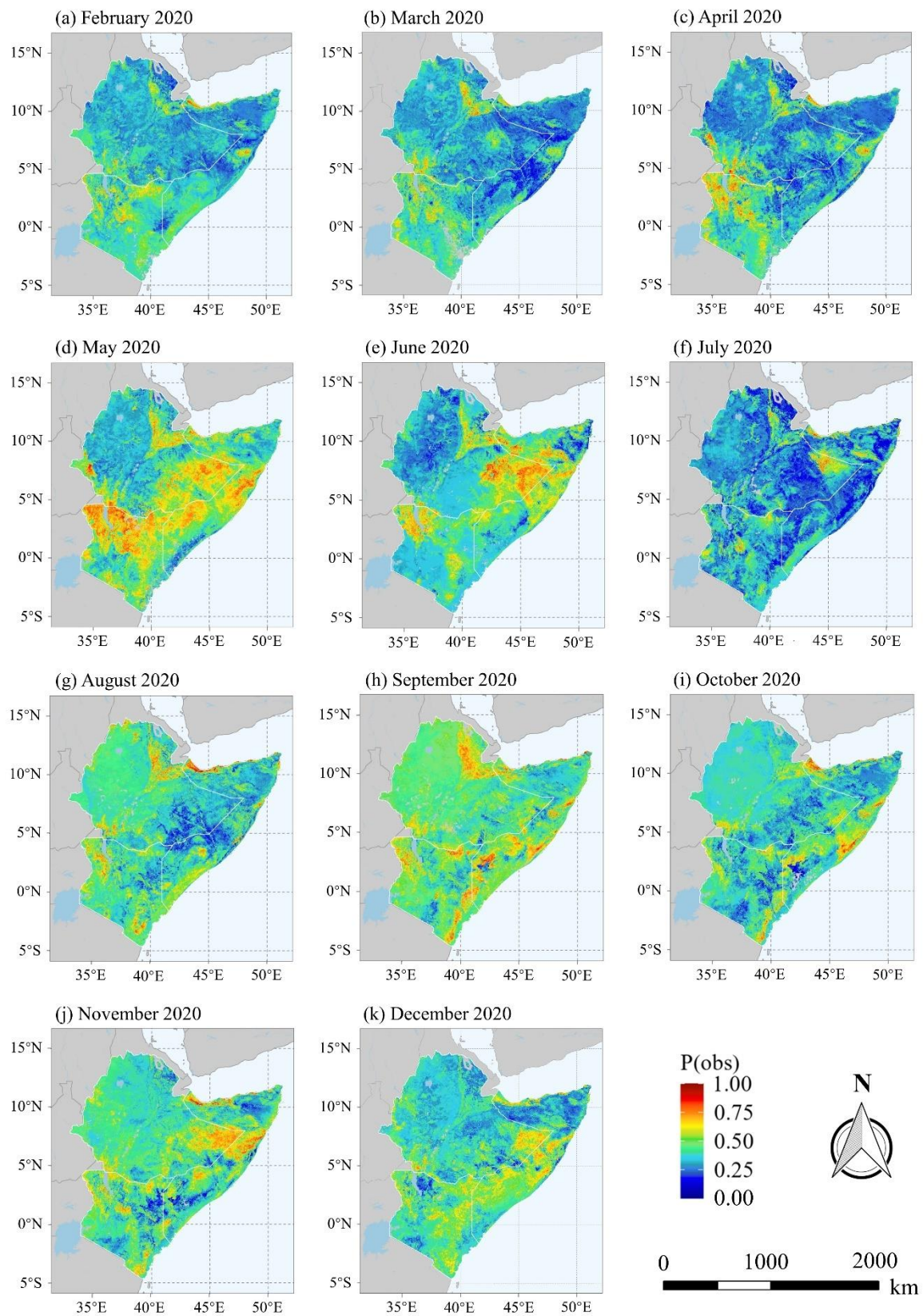
### 3.2. Dynamic Forecast of Hopper Band Presence in SEK

Eleven forecast experiments from February to December 2020 demonstrated satisfactory overall performance with an average accuracy of over 77%, a ROC-AUC value of 0.7666, and an F-score close to 0.7715 (see Table 2). The forecast accuracies for March, April, May, and June were exceptionally high, above 0.8. Predicted band-forming areas in other months shared an accuracy of about 0.75.

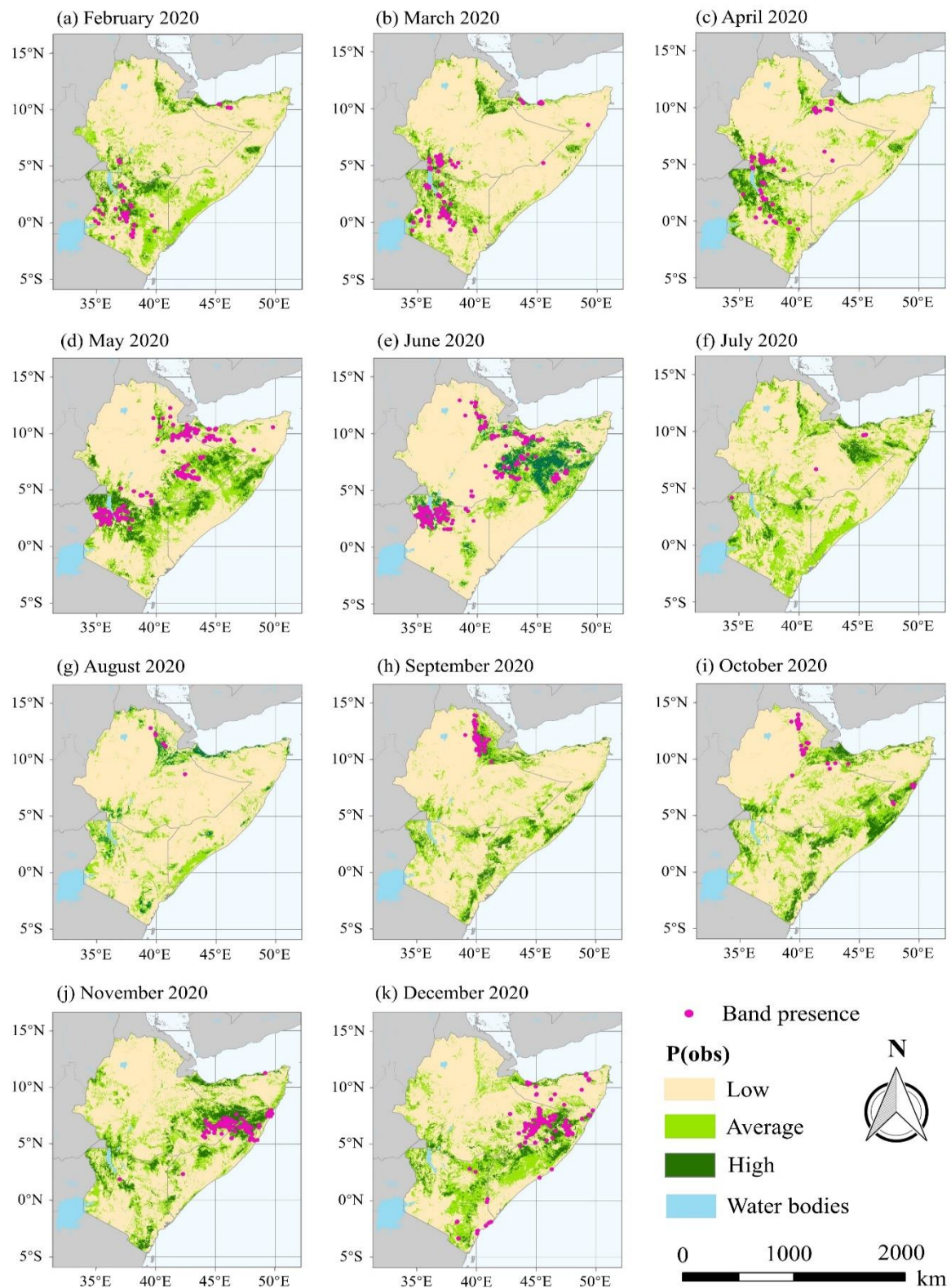
**Table 2.** Evaluation of dynamic predicting results.

| Date           | Evaluation Metrics |             |             |         |         |
|----------------|--------------------|-------------|-------------|---------|---------|
|                | Accuracy (%)       | Sensitivity | Specificity | ROC-AUC | F-Score |
| February 2020  | 74.44              | 0.6047      | 0.8759      | 0.7403  | 0.7792  |
| March 2020     | 80.15              | 0.6934      | 0.9329      | 0.8131  | 0.7930  |
| April 2020     | 82.59              | 0.7264      | 0.9002      | 0.8133  | 0.7811  |
| May 2020       | 88.68              | 0.8886      | 0.8814      | 0.8850  | 0.9218  |
| June 2020      | 85.31              | 0.8971      | 0.6667      | 0.7819  | 0.9081  |
| July 2020      | 70.00              | 0.6167      | 0.7714      | 0.6940  | 0.6549  |
| August 2020    | 76.99              | 0.6238      | 0.9412      | 0.7825  | 0.7453  |
| September 2020 | 79.81              | 0.6314      | 0.9203      | 0.7759  | 0.7258  |
| October 2020   | 66.77              | 0.5988      | 0.7419      | 0.6704  | 0.6515  |
| November 2020  | 73.41              | 0.7500      | 0.7116      | 0.7308  | 0.7673  |
| December 2020  | 73.95              | 0.7087      | 0.7816      | 0.7451  | 0.7586  |
| Average        | 77.46              | 0.7036      | 0.8296      | 0.7666  | 0.7715  |

The forecast model integrating all dynamic and static factors can eventually report up to 16 days in advance. Predicted probabilities of band occurrence, classified results, and corresponding ground observations in SEK 2020 were displayed in Figures 7 and 8. The spatial distribution of band presence varied from season to season. Desert locusts preferred spring and summer to propagate in SEK, while the suitable range was significantly reduced from July to October. Aggregated band present areas from February to April were mainly located along the Ethiopian Rift, the Eastern Branch of the East African Rift Valley, and a broad area of Kenya, including the central area of the country and the areas surrounding Lake Victoria in the southwestern corner. May and June were clearly the best months for desert locust reproduction. The expanded range and suitability of summer band present areas were concentrated in the bare ground covered by sparse scrubs and herbs around Lake Turkana in northwestern Kenya, in the Somali State of Ethiopia and in the Togdheer, Mudug, and Galguduud Regions of Somalia. From July onwards, local ecology was less suitable for desert locust, leading to adults and swarms migrating to their usual winter band present areas along the Red Sea and the Gulf of Aden. However, winter's short rainy season brought back the breeding season. The parcel suitable for local breeding from November to December was distributed in the eastern horn of Ethiopia and central Somalia regions, where locusts landed in warmer and wetter areas along the southeastern coast.



**Figure 7.** Dynamic predicted probability of observing desert locust band presence in SEK from February 2020 to December 2020. (a) February 2020, (b) March 2020, (c) April 2020, (d) May 2020, (e) June 2020, (f) July 2020, (g) August 2020, (h) September 2020, (i) October 2020, (j) November 2021, (k) December 2020.



**Figure 8.** Dynamic projected areas with classified probabilities of band presence and monthly ground truths of band presence in SEK from February 2020 to December 2020. Areas were divided into low and average probability of band presence by statistics at max TPR + TNR of training models as dichotomous thresholds for model evaluation, while the high probability area partitioned by the first quartile was only used for better visualization and demonstration. (a) February 2020, (b) March 2020, (c) April 2020, (d) May 2020, (e) June 2020, (f) July 2020, (g) August 2020, (h) September 2020, (i) October 2020, (j) November 2020, (k) December 2020.

## 4. Discussion

### 4.1. Lag Effect of Environmental Drivers in Desert Locust Ecology

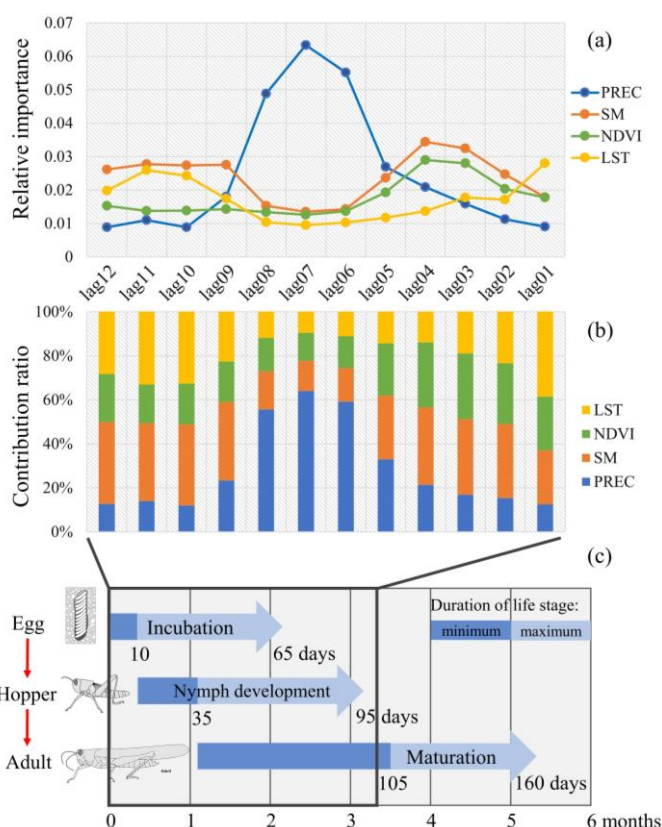
PREC is the initial external trigger for the reproductive events of desert locusts; thus, it possesses a longer lag. Females generally lay eggs after consecutive rainfall surges [78–82]. In SEK, a total median PREC of almost 60 mm for about 24 days is likely to observe hopper bands after 41 to 64 days. This result is similar to the predicted temporal gain of rainfall on hopper emergence in Mauritania, whose PREC could forecast band emergence up to 50 days in advance [36]. Tratalos et al. [83], however, suggested that the previous generation reproduces and forms bands about 4 months after the arrival of rainfall. This is because his indicator came from the recession zone within the Sahara. Eggs and hoppers developed much more slowly there, meaning that the time required before sighting band occurrence was greatly extended.

Suitable SM content and duration is not only a booster of oviposition and incubation but is also an early signal of vegetation growth. Locust egg hatching and survival rates are limited by a moisture content of 5–20% per unit volume of soil [84,85]. Hence, bands are likely to be observed after 33 to 40 days, when the median SM value rises to approximately  $0.2 \text{ m}^3/\text{m}^3$  after a rainfall upsurge in SEK. Another scenario is when the median SM value remains extremely low at  $0.15 \text{ m}^3/\text{m}^3$  and increases after one month; bands probably appear after 73 to 80 days. These findings agreed with those of Gómez et al. [38], who found that a median SM of more than  $0.11 \text{ m}^3/\text{m}^3$  for at least 16 days increased the probability of hoppers occurring after 79 days. Meanwhile, Piou et al. [36] also observed that soil moisture dynamics with similar trends increased the chance of hopper sightings after 72 days.

Sparsely vegetated areas where the median NDVI slightly rises to around 0.25 were more likely to track successful band development after 17 to 40 days. The shortening of the time lag is due to the fact that the vegetation mainly affects the hopper development period by providing them with nutrition and shelter. Piou et al. [35,36] found that vegetation growth around one and a half months before the survey was one of the best predictors of hopper presence during the recession period in Mauritania. It is a mildly shorter response time than in this study. Both results are reliable given that SEK has a more humid climate than Mauritania in West Africa.

LST did not show a significant optimal time window for observation, mainly due to the fact that the hatching and developing phases of the reproduction process are temperature dependent [18,86,87]. Laboratory studies revealed that most eggs incubate at low thermal cycles ( $30^\circ\text{C}$  or  $25^\circ\text{C}$ ) of desert locust biotopes [88]. Temperatures below  $15^\circ\text{C}$  and above  $35^\circ\text{C}$  might raise egg mortality [23]. Gómez et al. [26] introduced LST into the desert locust monitoring system and found that eggs were sensitive to an average LST of around  $25^\circ\text{C}$ , while hoppers preferred an LST of around  $35^\circ\text{C}$ . Similarly, prosperous band-forming habitats of SEK exhibited a high median LST in the early stage due to their locations in and around the deserts. It then declined, partly because of the increased rainfall required for propagation and partly in response to the temperature needs of the pests to reproduce.

To validate these lagging results for single factors, an integrated model was built combining the time lag variables of all dynamic indicators (see Figure 9). An 8 to 16-day delay of the optimal predictive temporal window between PREC-SM-NDVI indicators was detected by previous analysis and the integrated model. Although the precise mechanism driving desert locust reproduction requires further research in biology and ecology, results have confirmed that PREC, SM, and NDVI successively influence the oviposition, incubation, and nymph development in distinct ways. In general, the surge in PREC allows for an appropriate increase in SM in the root zone, which facilitates the hatching of eggs and promotes the annual vegetation springing up, leading to an increased NDVI. During this process, LST also affects soil evapotranspiration and vegetation development. Ultimately, the weather, together with the soil and vegetation, determines the success of band formation.



**Figure 9.** (a) Relative importance and (b) contribution ratio of dynamic indicators in (c) desert locust life sequences of desert locust [12].

#### 4.2. Strengths and Weaknesses of Environmental Drivers for Forecast Framework

PREC remains an essential predictor of desert locust hopper band emergence. It maintains egg incubation by moistening the soil while promoting vegetation growth and thus supporting hopper development. As the most critical external trigger of desert locust outbreaks, the high temporal resolution of satellite or in-situ rainfall estimates meets the demand for near-real-time forecasting with a robust performance. The disadvantage is that rainfall indirectly affects desert locust breeding through the surface environment, where evaporation, desert aerosols, or runoff may lead to a spatially inaccurate final prediction [58].

The potential of SM as a proxy for PREC to predict desert locust reproduction has been widely considered in recent years [27,28,36,37,40]. The greatest strength of SM compared to other indicators is that it reflects the immediate need for locust eggs to hatch, effectively combining remote sensing techniques with biological mechanisms. It can advance the prediction by almost three months, buying ample time for early ground control decisions and actions. With microwave remote sensing, adaptive scaling algorithms and data assimilation models developed, soil moisture products with higher spatial resolutions will better facilitate desert locust early warning framework as an essential member of the indicator system [89].

Vegetation has always been one of the critical predictors of nymph occurrence [32,90]. Although moisture indicators offer longer predictive time gains than NDVI, vegetation provides larva development with nutrition and refuge [82]. At the same time, patches of thicket and meadow with low coverage and irregularities are more likely to concentrate solitary hoppers, increasing the likelihood of gregarization into social bands [24,39]. Therefore, vegetation is crucial to the later stages of reproduction and should not be excluded in any reproductive habitat indicator system.

LST continuously regulates the breeding behavior, especially at early stages and near band emergence. It must be included in the indicator system. However, in the context

of early warning for preventive management strategy of desert locust, we suggest using an early predictor (in this study we use variable lag12 instead of lag01) in the forecast framework to earn adequate time to employ ground control.

#### 4.3. Feasibility and Robustness of Forecast Model for Early Warning

The sound overall accuracy of the predictions for SEK, mainly thanks to the multivariate model, integrates various explanatory indicators containing bioclimatic, soil, vegetation, topography, land cover, and other information. Recently, experiments and evaluations of different indicators, primarily some soil physical attributes, have emerged to predict the reproduction area of desert locusts [22,28,36]. We found that introducing more factors, particularly critical surface hydrothermal conditions required for propagation through the time-series variables of the optimal time lag windows, improves the performance of band presence predictions compared to existing studies.

Preliminary evidence of the applicability and stability of the proposed model has been demonstrated in time and space and in combination with data availability in this study. The data-driven model is applicable to species with tremendous ground tracking information to quantify the phenology of desert locusts by statistical methods and data mining approaches. In terms of temporal feasibility, the framework is capable of a year-round forecast for different months and seasons. However, accuracy is more prominent in summer. This result is inevitable due mainly to the uneven seasonal distribution of the historical ground survey data used for training. Most of the band points in the SEK region were collected in March, April, May, and June. In other words, the time-lagged response pattern of dynamic indicators obtained in this study can be applied to the prediction of band emergence areas in SEK all year round but is best suited to the forecast needs of summer band emergence areas. While in terms of spatial feasibility, it is reasonable to adopt different lagging indicators for different scenarios due to the changing phenology of desert locust caused by different ecological contexts [91]. The temporal sliding window technique requires calibration and localization when applied to different areas. However, the indicator system derived from multisource imagery and the predictive framework can be promoted to other regions to assist in the local ground control of desert locusts.

Furthermore, the time span that the predictive framework can report in advance varies with different combinations of environmental factors. It ultimately depends on the nearest optimal time lag of a specific factor. For example, the single-factor model based on SM allows for the longest-term prediction, achieving a 72-day extended forecast. Whereas the prediction model combining all the four dynamic indicators demonstrated in this study will eventually be able to report the band presence up to 16 days in advance. Different lengths of terms can be adapted to the needs of different ground applications.

## 5. Conclusions

This study proposed a data-driven forecast framework based on machine learning and a temporal sliding window technique to realize the early warning of the presence of desert locust. Long time series satellite imagery and other Earth observation data were combined with historical locust ground survey data to construct the sliding window approach based on the statistical relationship between environmental dynamics and desert locust phenology. Forecasts can be carried out every 8 days to provide dynamic maps of prosperous band habitats 2–11 weeks in advance, saving time for ground decisions and actions and making them more timely, effective, and environmentally friendly. Compared to meteorological-based models or remote sensing models using near-real-time metrics, this framework embeds the ecological mechanism on the time scale of hopper band emergence. It requires little expert knowledge or experiences and is appropriate for other spatial or temporal scenarios.

A noteworthy shortcoming of the proposed model is that the population dynamics and migratory behavior of desert locusts have not been taken into account [92,93]. It provides habitat suitability or potential risks of band presence while residents and invaders can also



further affect the final distribution of band risks. This partly led to a larger scope of the predicted zone of band occurrence in September and October 2020 than that which actually occurred (see Figure 6c). Some successfully breeding and maturing summer swarms had already migrated with the prevailing winds towards the Red Sea coast and northeast Somalia for overwintering [94,95]. Further research should focus on coupling the habitat-based model with population dynamics, embedding the gregarization mechanism into the data-driven methods.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/rs14030747/s1>, Figure S1: Spatial distribution of the different numbers of pseudo-absence points generated, Table S1: Evaluation of predicting results with different numbers of pseudo-absence, Figure S2: Predicting possibility of band presence with different numbers of pseudo-absence, Table S2: Comparison of evaluation metrics of SVM and RF, Figure S3: Dynamic predicted probability of observing desert locust band presence with monthly ground truths of band presence in SEK from February 2020 to December 2020.

**Author Contributions:** Conceptualization, R.S. and Y.D.; methodology, R.S., L.Z. and B.Z.; data curation, R.S.; formal analysis, R.S.; software, R.S., B.Z., C.R. and X.C.; validation, R.S. and H.M.; writing—original draft preparation, R.S.; writing—review and editing, W.H., Y.D., L.Z., B.Z., H.M., Y.G., N.X. and X.L.; visualization, R.S.; supervision, W.H. and Y.D.; project administration, Y.D. and W.H.; funding acquisition, W.H., Y.D., and L.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China (Grant No. 2021YFE0194800), the National Natural Science Foundation of China (Grant No. 42171323, 41801338), the Alliance of International Science Organizations (Grant No. ANSO-CR-KP-2021-06), the SINO-EU, Dragon 5 proposal: Application of Sino-Eu Optical Data into Agronomic Models to Predict Crop Performance and to Monitor and Forecast Crop Pests and Diseases (Grant No. 57457), and the External Cooperation Program of the Chinese Academy of Sciences (Grant No. 183611KYSB20200080).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** All the data used in this study are available publicly. The desert locust ground survey data are available from FAO DLIS (<https://locust-hub-hqfao.hub.arcgis.com/> (accessed on 20 December 2021)). The CHIRPS data are available from the Climate Hazards Center (CHC) (<https://data.chc.ucsb.edu/products/CHIRPS-2.0/> (accessed on 20 December 2021)). ERA5-Land hourly data is available from the Climate Data Store of Copernicus (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview> (accessed on 20 December 2021)). The modeled soil moisture data are available from SERVIR ([https://gis1.servirglobal.net/data/locusts/east\\_africa/](https://gis1.servirglobal.net/data/locusts/east_africa/) (accessed on 20 December 2021)) and information on the other soil physical attributes comes from ISRIC SoilGrids (<https://soilgrids.org/> (accessed on 20 December 2021)). Land cover data is available from CGLS (<https://land.copernicus.eu/global/products/lc> (accessed on 20 December 2021)). NASADEM data is available from USGS ([https://lpdaac.usgs.gov/products/nasadem\\_hgtv001/](https://lpdaac.usgs.gov/products/nasadem_hgtv001/) (accessed on 20 December 2021)).

**Acknowledgments:** The authors would like to sincerely thank Cyril Piou, an ecologist from CIRAD (French Agricultural Research Centre for International Development) for his valuable guidance and advice. We also appreciate the academic editor and reviewers for their insightful comments which greatly helped us improve the quality of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Huang, W.; Dong, Y.; Zhao, L.; Geng, Y.; Ruan, C.; Zhang, B.; Sun, Z.; Zhang, H.; Ye, H.; Wang, K. Review of locust remote sensing monitoring and early warning (Chinese). *J. Remote Sens.* **2020**, *24*, 1270–1279.
2. Salih, A.A.M.; Baraibar, M.; Mwangi, K.K.; Artan, G. Climate change and locust outbreak in East Africa. *Nat. Clim. Change* **2020**, *10*, 584–585. [CrossRef]
3. Hemming, C.F.; Popov, G.B.; Roffey, J.; Waloff, Z. Characteristics of desert locust plague upsurges. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **1979**, *287*, 375–386.

4. Hassanali, A.; Bashir, M.O.; Njagi, P.G.N.; Ely, S.O. Desert locust gregarization: A conceptual kinetic model. *J. Orthoptera Res.* **2005**, *14*, 223–226. [[CrossRef](#)]
5. Latchininsky, A.V. Locusts and remote sensing: A review. *J. Appl. Remote Sens.* **2013**, *7*, 75099. [[CrossRef](#)]
6. Michel, L. Recent progress in Desert and Migratory Locust management in Africa. Are preventative actions possible? *J. Orthoptera Res.* **2001**, *10*, 277–291.
7. Lecoq, M. Desert locust threat to agricultural development and food security and FAO/international role in its control. In Proceedings of the Eighth Arab Congress of Plant Protection, El-Beida, Libya, 12–16 October 2003; Arab Society for Plant Protection: Beyrouth, Lebanon, 2003.
8. Magor, J.I.; Lecoq, M.; Hunter, D.M. Preventive control and Desert Locust plagues. *Crop. Prot.* **2008**, *27*, 1527–1533. [[CrossRef](#)]
9. Sword, G.A.; Lecoq, M.; Simpson, S.J. Phase polyphenism and preventative locust management. *J. Insect Physiol.* **2010**, *56*, 949–957. [[CrossRef](#)]
10. Cressman, K. Role of remote sensing in desert locust early warning. *J. Appl. Remote Sens.* **2013**, *7*, 75098. [[CrossRef](#)]
11. Zhang, L.; Lecoq, M.; Latchininsky, A.; Hunter, D. Locust and grasshopper management. *Annu. Rev. Entomol.* **2019**, *64*, 15–34. [[CrossRef](#)]
12. Symmons, P.M.; Cressman, K. *Desert Locust Guidelines: Biology and Behaviour*; FAO: Rome, Italy, 2001.
13. Dobson, H.M. *Desert Locust Guidelines: Control*; FAO: Rome, Italy, 2001.
14. Symmons, P.M. A computer simulation of Desert Locust band target block detection and demarcation. *Crop. Prot.* **2004**, *23*, 183–190. [[CrossRef](#)]
15. van Huis, A.; Cressman, K.; Magor, J.I. Preventing desert locust plagues: Optimizing management interventions. *Entomol. Exp. Appl.* **2007**, *122*, 191–214. [[CrossRef](#)]
16. Gay, P.E.; Lecoq, M.; Piou, C. The limitations of locust preventive management faced with spatial uncertainty: Exploration with a multi-agent model. *Pest. Manag. Sci.* **2019**, *76*, 1094–1102. [[CrossRef](#)]
17. Cressman, K. The use of new technologies in desert locust early warning. *Outlooks Pest Manag.* **2008**, *19*, 55–59. [[CrossRef](#)]
18. Roffey, J.; Magor, J.I. *Desert Locust Technical Series—Desert Locust Population Dynamics Parameters*; Desert Locust Field Research Stations (FAO): Rome, Italy, 2003; p. 35.
19. Wang, Y.; Wu, M.; Lin, P.; Wang, Y.; Chen, A.; Jiang, Y.; Zhai, B.; Chapman, J.W.; Hu, G. Plagues of desert locusts: Very low invasion risk to china. *Insects* **2020**, *11*, 628. [[CrossRef](#)] [[PubMed](#)]
20. Amarjyothi, K.; Kumar, D.P.; Saikrishnan, K.C. Identification and tracking of locust swarms by indian doppler weather radar. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–4. [[CrossRef](#)]
21. Ceccato, P.; Cressman, K.; Giannini, A.; Trzaska, S. The desert locust upsurge in West Africa (2003–2005): Information on the desert locust early warning system and the prospects for seasonal climate forecasting. *Int. J. Pest Manag.* **2007**, *53*, 7–13. [[CrossRef](#)]
22. Kimathi, E.; Tonnang, H.E.Z.; Subramanian, S.; Cressman, K.; Abdel-Rahman, E.M.; Tesfayohannes, M.; Niassy, S.; Torto, B.; Dubois, T.; Tanga, C.M.; et al. Prediction of breeding regions for the desert locust *Schistocerca gregaria* in East Africa. *Sci. Rep. UK* **2020**, *10*, 11937. [[CrossRef](#)]
23. Cressman, K.; Stefanski, R. *Weather and Desert Locusts*; World Meteorological Organization and Food and Agriculture Organization of the United Nations: Rome, Italy, 2016.
24. Despland, E.; Collett, M.; Simpson, S.J. Small-scale processes in desert locust swarm formation: How vegetation patterns influence gregarization. *Oikos* **2000**, *88*, 652–662. [[CrossRef](#)]
25. Culmsee, H. The habitat functions of vegetation in relation to the behaviour of the desert locust *Schistocerca gregaria* (Forsk.) (*Acrididae: Orthoptera*)—A study in Mauritania (West Africa). *Phytocoenologia* **2002**, *32*, 645–664. [[CrossRef](#)]
26. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, C.; Taratiel, D.; Casanova, J.L. Desert locust detection using Earth observation satellite data in Mauritania. *J. Arid Environ.* **2019**, *164*, 29–37. [[CrossRef](#)]
27. Escorihuela, M.J.; Merlin, O.; Stefan, V.; Moyano, G.; Eweys, O.A.; Zribi, M.; Kamara, S.; Benahi, A.S.; Ebbe, M.A.B.; Chihrane, J.; et al. SMOS based high resolution soil moisture estimates for desert locust preventive management. *Remote Sens. Appl. Soc. Environ.* **2018**, *11*, 140–150.
28. Ellenburg, W.L.; Mishra, V.; Roberts, J.B.; Limaye, A.S.; Case, J.L.; Blankenship, C.B.; Cressman, K. Detecting desert locust breeding grounds: A Satellite-Assisted modeling approach. *Remote Sens.* **2021**, *13*, 1276. [[CrossRef](#)]
29. Hielkema, J.U.; Roffey, J.; Tucker, C.J. Assessment of ecological conditions associated with the 1980/81 desert locust plague upsurge in West Africa using environmental satellite data. *Int. J. Remote Sens.* **1986**, *7*, 1609–1622. [[CrossRef](#)]
30. Ceccato, P. Operational early warning system using SPOT-vegetation and Terra-MODIS to predict Desert Locust outbreaks. In Proceedings of the 2nd International Vegetation User Conference: 1998–2004: 6 Years of Operational Activities, European Commission, Antwerp, Belgium, 24–26 March 2004.
31. Renier, C.; Waldner, F.; Jacques, D.; Babah Ebbe, M.; Cressman, K.; Defourny, P. A dynamic vegetation senescence indicator for Near-Real-Time desert locust habitat monitoring with MODIS. *Remote Sens.* **2015**, *7*, 7545–7570. [[CrossRef](#)]
32. Tucker, C.J.; Hielkema, J.U.; Roffey, J. The potential of satellite remote sensing of ecological conditions for survey and forecasting desert-locust activity. *Int. J. Remote Sens.* **1985**, *6*, 127–138. [[CrossRef](#)]
33. Pekel, J.; Ceccato, P.; Vancutsem, C.; Cressman, K.; Vanbogaert, E.; Defourny, P. Development and application of Multi-Temporal colorimetric transformation to monitor vegetation in the desert locust habitat. *IEEE J. Stars* **2011**, *4*, 318–326. [[CrossRef](#)]

34. Waldner, F.; Ebbe, M.; Cressman, K.; Defourny, P. Operational monitoring of the desert locust habitat with earth observation: An assessment. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2379–2400. [[CrossRef](#)]
35. Piou, C.; Lebourgeois, V.; Benahi, A.S.; Bonnal, V.; Jaavar, M.E.H.; Lecoq, M.; Vassal, J. Coupling historical prospection data and a remotely-sensed vegetation index for the preventative control of Desert locusts. *Basic Appl. Ecol.* **2013**, *14*, 593–604. [[CrossRef](#)]
36. Piou, C.; Gay, P.E.; Benahi, A.S.; Babah Ebbe, M.A.O.; Chihrane, J.; Ghaout, S.; Cisse, S.; Diakite, F.; Lazar, M.; Cressman, K.; et al. Soil moisture from remote sensing to forecast desert locust presence. *J. Appl. Ecol.* **2019**, *56*, 966–975. [[CrossRef](#)]
37. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, C.; Taratiel, D.; Casanova, J.L. Machine learning approach to locate desert locust breeding areas based on ESA CCI soil moisture. *J. Appl. Remote Sens.* **2018**, *12*, 036011. [[CrossRef](#)]
38. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Modelling desert locust presences using 32-year soil moisture data on a large-scale. *Ecol. Indic.* **2020**, *117*, 106655. [[CrossRef](#)]
39. Cisse, S.; Ghaout, S.; Mazih, A.; Babah Ebbe, M.A.O.; Benahi, A.S.; Piou, C. Effect of vegetation on density thresholds of adult desert locust gregarization from survey data in Mauritania. *Entomol. Exp. Appl.* **2013**, *149*, 159–165. [[CrossRef](#)]
40. Lazar, M.; Piou, C.; Doumandji-Mitiche, B.; Lecoq, M. Importance of solitary desert locust population dynamics: Lessons from historical survey data in Algeria. *Entomol. Exp. Appl.* **2016**, *161*, 168–180. [[CrossRef](#)]
41. Huang, J. Remote sensing of locust and grasshopper plague in China: A review. In Proceedings of the 5th International Conference on Agro-Geoinformatics, Tianjin, China, 18–20 July 2016.
42. Amichay, G.; Ariel, G.; Ayali, A. The effect of changing topography on the coordinated marching of locust nymphs. *Peer J.* **2016**, *4*, e2742. [[CrossRef](#)]
43. Steedman, A. *Locust Handbook*, 2nd ed.; Overseas Development Natural Resource Institute: London, UK, 1988.
44. Sivanpillai, R.; Latchinsky, A.V.; Driese, K.L.; Kambulin, V.E. Mapping locust habitats in River Ili Delta, Kazakhstan, using LANDSAT imagery. *Agric. Ecosyst. Environ.* **2006**, *117*, 128–134. [[CrossRef](#)]
45. Zhao, L.; Huang, W.; Chen, J.; Dong, Y.; Ren, B.; Geng, Y. Land use/cover changes in the Oriental migratory locust area of China: Implications for ecological control and monitoring of locust area. *Agric. Ecosyst. Environ.* **2020**, *303*, 107110. [[CrossRef](#)]
46. Shi, Y.; Huang, W.; Dong, Y.; Peng, D.; Zheng, Q.; Yang, P. The influence of landscape's dynamics on the Oriental Migratory Locust habitat change based on the time-series satellite data. *J. Environ. Manag.* **2018**, *218*, 280–290. [[CrossRef](#)]
47. Geng, Y.; Zhao, L.; Dong, Y.; Huang, W.; Shi, Y.; Ren, Y.; Ren, B. Migratory locust habitat analysis with PB-AHP model using Time-Series satellite images. *IEEE Access* **2020**, *8*, 166813–166823. [[CrossRef](#)]
48. Tierney, J.E.; Ummenhofer, C.C.; DeMenocal, P.B. Past and future rainfall in the Horn of Africa. *Sci. Adv.* **2015**, *1*, e1500682. [[CrossRef](#)]
49. FAO-DLIS; WCS. *Seasonal precipitation predictions in Desert Locust spring/summer/winter breeding areas (June–November 2021)*; FAO: Rome, Italy, 2021.
50. Kalakkal, J.; Singh, A. *Desert Locusts' Upsurges: A Harbinger of Emerging Climate Change-Induced Crises?* The United Nations Environment Programme (UNEP): Nairobi, Kenya, 2021.
51. Stone, M. *A Plague of Locusts has Descended on East Africa. Climate Change May be to Blame*; National Geographic: Washington, DC, USA, 2020; Volume 2021.
52. Aiello-Lammens, M.E.; Boria, R.A.; Radosavljevic, A.; Vilela, B.; Anderson, R.P. SpThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **2015**, *38*, 541–545. [[CrossRef](#)]
53. Bedia, J.; Iturbide, M.; Gutiérrez, J. Tackling Uncertainties of Species Distribution Model Projections with *Package mopa*. *R J.* **2018**, *10*, 122–135.
54. Rew, J.; Cho, Y.; Hwang, E. A robust prediction model for species distribution using bagging ensembles with deep neural networks. *Remote Sens.* **2021**, *13*, 1495. [[CrossRef](#)]
55. Chefaoui, R.M.; Lobo, J.M. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol. Model.* **2008**, *210*, 478–486. [[CrossRef](#)]
56. Graham, C.H.; Hijmans, R.J. A comparison of methods for mapping species ranges and species richness. *Glob. Ecol. Biogeogr.* **2006**, *15*, 578–587. [[CrossRef](#)]
57. Iturbide, M.; Bedia, J.; Herrera, S.; Del Hierro, O.; Pinto, M.; Gutiérrez, J.M. A framework for species distribution modelling with improved pseudo-absence generation. *Ecol. Model.* **2015**, *312*, 166–174. [[CrossRef](#)]
58. Dinku, T.; Ceccato, P.; Cressman, K.; Connor, S.J. Evaluating detection skills of satellite rainfall estimates over desert locust recession regions. *J. Appl. Meteorol. Clim.* **2010**, *49*, 1322–1332. [[CrossRef](#)]
59. Dinku, T.; Funk, C.; Peterson, P.; Maidment, R.; Tadesse, T.; Gadain, H.; Ceccato, P. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 292–312. [[CrossRef](#)]
60. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [[CrossRef](#)]
61. Kumar, S.V.; Peters-Lidard, C.D.; Tian, Y.; Houser, P.R.; Geiger, J.; Olden, S.; Lighty, L.; Eastman, J.L.; Doty, B.; Dirmeyer, P. Land information system: An interoperable framework for high resolution land surface modeling. *Environ. Model. Softw.* **2006**, *21*, 1402–1415. [[CrossRef](#)]
62. McNally, A.; Arsenault, K.; Kumar, S.; Shukla, S.; Peterson, P.; Wang, S.; Funk, C.; Peters-Lidard, C.D.; Verdin, J.P. A land data assimilation system for sub-Saharan Africa food and water security applications. *Sci. Data* **2017**, *4*, 1–19. [[CrossRef](#)] [[PubMed](#)]

63. Case, J.L.; Mungai, J.; Sakwa, V.; Zavodsky, B.T.; Srikishen, J.; Limaye, A.S.; Blankenship, C.B. *Transitioning Enhanced Land Surface Initialization and Model Verification Capabilities to the Kenya Meteorological Service*; American Meteorological Society Fall Meeting: New Orleans, LA, USA, 2016.
64. Hengl, T.; Mendes De Jesus, J.; Heuvelink, G.B.M.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e169748. [[CrossRef](#)] [[PubMed](#)]
65. Buchhorn, M.; Lesiv, M.; Tsendbazar, N.; Herold, M.; Bertels, L.; Smets, B. Copernicus global land cover layers—Collection 2. *Remote Sens.* **2020**, *12*, 1044. [[CrossRef](#)]
66. NASA JPL. NASADEM Merged DEM Global 1 arc second V001. 2020, distributed by NASA EOSDIS Land Processes DAAC. Available online: [https://doi.org/10.5067/MEaSURES/NASADEM/NASADEM\\_HGT.001](https://doi.org/10.5067/MEaSURES/NASADEM/NASADEM_HGT.001) (accessed on 20 December 2021).
67. Evans, J.S.; Murphy, M.A.; Holden, Z.A.; Cushman, S.A. *Modeling Species Distribution and Change Using Random Forest*; Springer: New York, NY, USA, 2011; pp. 139–159.
68. Pearce, J.; Ferrier, S. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **2000**, *133*, 225–245. [[CrossRef](#)]
69. Keating, K.A.; Cherry, S. Use and interpretation of logistic regression in habitat-selection studies. *J. Wildl. Manag.* **2004**, *68*, 774–789. [[CrossRef](#)]
70. Azen, R.; Traxel, N. Using dominance analysis to determine predictor importance in logistic regression. *J. Educ. Behav. Stat.* **2009**, *34*, 319–347. [[CrossRef](#)]
71. Luo, W.; Azen, R. Determining predictor importance in hierarchical linear models using dominance analysis. *J. Educ. Behav. Stat.* **2013**, *38*, 3–31. [[CrossRef](#)]
72. Soares, F.M.C.D. *Modelling the Distribution of São Tomé Bird Species: Ecological Determinants and Conservation Prioritization*. Doctoral dissertation, University of Lisbon, Lisbon, Portugal, 2017.
73. Han, H.; Guo, X.; Yu, H. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016.
74. Zhang, J.; Li, S. A Review of Machine Learning Based Species' Distribution Modelling. In Proceedings of the 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, China, 2–3 December 2017.
75. Chang, T.; Hansen, A.J.; Piekielek, N. Patterns and variability of projected bioclimatic habitat for *pinus albicaulis* in the greater yellowstone area. *PLoS ONE* **2014**, *9*, e111669. [[CrossRef](#)]
76. Hong, C.S.; Oh, T.G. TPR-TNR plot for confusion matrix. *Commun. Stat. Appl. Methods* **2021**, *28*, 161–169. [[CrossRef](#)]
77. Sterr, U. Analysis of 2019 Kaggle ML & DS Survey. Available online: <https://uwesterr.github.io/KaggleMISurvey2019/> (accessed on 20 December 2021).
78. Magor, J.I. *Rainfall as a factor in the geographical distribution of the desert locust breeding areas, with particular reference to the summer breeding area of India and Pakistan*. Doctoral dissertation, University of Edinburgh, Edinburgh, UK, 1963.
79. Roffey, J.; Popov, G.; Hemming, C.F. Outbreaks and recession populations of the desert locust *Schistocerca gregaria* (Forsk.). *Bull. Entomol. Res.* **1970**, *59*, 675–680. [[CrossRef](#)]
80. Bennett, L.V. The development and termination of the 1968 plague of the Desert locust, *Schistocerca gregaria* (Forskål) (Orthoptera, Acrididae). *Bull. Entomol. Res.* **1976**, *66*, 511–552. [[CrossRef](#)]
81. Dempster, J.P. The population dynamics of grasshoppers and locusts. *Biol. Rev.* **1963**, *38*, 490–529. [[CrossRef](#)]
82. Latchininsky, A.V.; Sivanpillai, R. *Locust Habitat Monitoring and Risk Assessment Using Remote Sensing and GIS Technologies*; Springer: Dordrecht, The Netherlands, 2010; pp. 163–188.
83. Tratalos, J.A.; Cheke, R.A.; Healey, R.G.; Stenseth, N.C. Desert Locust Populations, Rainfall and Climate Change: Insights from Phenomenological Models Using Gridded Monthly data. *Clim. Res.* **2010**, *43*, 229–239. [[CrossRef](#)]
84. Hunter, J.P. *Egg Development in the Desert Locust (Schistocerca gregaria Forsk.) in Relation to the Availability of Water*; Wiley Online Library: Hoboken, NJ, USA, 1964; pp. 25–33.
85. Popov, G.B. *Ecological Studies on Oviposition by Swarms of the Desert Locust (Schistocerca gregaria Forskal) in Eastern Africa*. Anti-Locust Research Centre: London, UK, 1958.
86. Symmons, P.M. A model to predict the incubation and nymphal development periods of the desert locust, *Schistocerca gregaria* (Orthoptera: Acrididae) 1. *Bull. Entomol. Res.* **1992**, *82*, 517–520.
87. Wardhaugh, K.; Ashour, Y.; Ibrahim, A.O.; Khan, A.M.; Bassonbol, M. Experiments on the incubation and hopper development periods of the desert locust (*Schistocerca gregaria* Forskål) in Saudi Arabia. *Lond. Anti locust Res. Cent Anti locust Bull.* **1969**, *46*, 35–38.
88. Nishide, Y.; Suzuki, T.; Tanaka, S. Synchrony in the hatching of eggs in the desert locust *Schistocerca gregaria* (Orthoptera: Acrididae): Egg condition influences hatching time in the laboratory and under simulated field temperatures. *Appl. Entomol. Zool.* **2017**, *52*, 599–604. [[CrossRef](#)]
89. Mohanty, B.P.; Cosh, M.H.; Lakshmi, V.; Montzka, C. Soil moisture remote sensing: State-of-the-science. *Vadose Zone J.* **2017**, *16*, 1–9. [[CrossRef](#)]

90. Deveson, E.D. Satellite normalized difference vegetation index data used in managing Australian plague locusts. *J. Appl. Remote Sens.* **2013**, *7*, 75096. [[CrossRef](#)]
91. Gómez, D.; Salvador, P.; Sanz, J.; Rodrigo, J.F.; Gil, J.; Casanova, J.L. Prediction of desert locust breeding areas using machine learning methods and SMOS (MIR\_SMNRT2) Near Real Time product. *J. Arid Environ.* **2021**, *194*, 104599. [[CrossRef](#)]
92. Roffey, J.; Popov, G. Environmental and behavioural processes in a desert locust outbreak. *Nature* **1968**, *219*, 446–450. [[CrossRef](#)]
93. Veran, S.; Simpson, S.J.; Sword, G.A.; Deveson, E.; Piry, S.; Hines, J.E.; Berthier, K. Modeling spatiotemporal dynamics of outbreaking species: Influence of environment and migration in a locust. *Ecology* **2015**, *96*, 737–748. [[CrossRef](#)] [[PubMed](#)]
94. FAO. *Desert Locust Bulletin—General Situation during August 2020 and Forecast until Mid-October 2020*; FAO: Rome, Italy, 2020.
95. FAO. *Desert Locust Bulletin—General Situation during September 2020 and Forecast until Mid-November 2020*; FAO: Rome, Italy, 2020.