Ecological Indicators 184 (2026) 114735

Contents lists available at ScienceDirect

Ecological Indicators

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

UAV-derived absence data improve species distribution model performance for the plateau pika

Yi Sun [a], Bo Huang [a], Yifei Luo [b], Yu Qin [c], Xiong Zhao He [d], Shuhua Yi [a,*]

[a] State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China

[b] Chongqing Institute of Green and Intelligent Technology / Chongqing College, Chinese Academy of Sciences, Chongqing 400714, China

[c] Key Laboratory of Cryospheric Science and Frozen Soil Engineering, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, 320 Donggang West Road, Lanzhou 730000, China

[d] School of Agriculture and Environment, College of Science, Massey University, Private Bag 11-222, Palmerston North 4442, New Zealand

A R T I C L E I N F O

Keywords:
Pseudo-absence data
Unmanned aerial vehicle
Qinghai-Tibetan plateau
BIOMOD
Plateau pika
Ochotona curzoniae
Mammalia Lagomorpha

A B S T R A C T

Reliable absence data remain a major limitation in the development and application of species distribution models (SDMs). Both field-sampled absences and algorithmically generated pseudo-absences are fundamental to model parameterization, yet efficient approaches for obtaining robust absence information and how absence data shape SDM outcomes are still lacking. Here, using Biodiversity Modelling (BIOMOD2), we systematically assessed SDM performance across multiple pseudo-absence generation strategies and sample sizes, leveraging a presence/absence dataset (n = 2261; 1743 presences and 518 absences) for plateau pika (*Ochotona curzoniae*) derived from unmanned aerial vehicle (UAV) surveys on the Qinghai–Tibetan Plateau. UAV-derived absence data consistently produced the highest model accuracy, particularly for Random Forest models (Kappa = 0.566; TSS = 0.611; AUC = 0.863). Among pseudo-absence strategies, the surface range envelope (SRE) strategy performed the best, whereas sample size showed relatively minor influence. Importantly, the spatial configuration of absence samples altered the weighting of environmental predictors, resulting in pronounced differences in model performance. Our findings demonstrate that absence data are pivotal to SDM accuracy and that UAV technology fundamentally advances the acquisition of presence/absence data by enabling efficient, spatially extensive, and reliable sampling. These methodological innovations are essential for improving biodiversity forecasting and informing management strategies under accelerating climate change and intensifying anthropogenic pressures.

1. Introduction

Species distribution models (SDMs) are valuable tools for bridging knowledge gaps in conservation, ecological research, and evolutionary studies, as well as practical applications, by predicting the spatial and temporal distribution of species (Radomski et al., 2022; Broussin et al., 2024). However, a major challenge in SDMs is the lack of reliable, large-scale, and spatially explicit species presence/absence datasets (Schmidt-

Abbreviations: UAV, Unmanned Aerial Vehicle; SDMs, species distribution models; BIOMOD2, Biodiversity Modelling (a R Package); QTP, Qinghai–Tibetan Plateau; AUC, area under the Receiver Operating Characteristic curve; TSS, true skill statistics; MODIS, Moderate-resolution Imaging Spectroradiometer); DISK, pseudo-absences are randomly selected within circles around presence points defined by a minimum and a maximum distance value, a function of BIOMOD2.; SRE, a Surface Range Envelop model is used to randomly select pseudo-absences outside this envelop, a function of BIOMOD2.; FragMAP, Fragmentation Monitoring and Analysis with Aerial Photography System, a developed independently control system for unmanned aerial vehicle by Professor Yi.; GRID, A flighting mode of FragMAP system.; GLM, generalized linear model, a model available in BIOMOD2; GBM, generalized boosted regression model, a model available in BIOMOD2; GAM, generalized additive model, a model available in BIOMOD2; CTA, classification tree analysis, a model available in BIOMOD2; FDA, flexible discriminant analysis, a model available in BIOMOD2; MARS, multiple adaptive regression splines, a model available in BIOMOD2; RF, random forest, a model available in BIOMOD2; MaxEnt, maximum entropy model, a model available in BIOMOD2; XGBOOT, eXtreme Gradient Boosting, a model available in BIOMOD2; AGB, aboveground biomass; DEM, digital elevation model; GP, grassland type; AMT, annual mean temperature; AP, annual precipitation; PCQ, precipitation of coldest quarter.

* Corresponding author.
E-mail address: ysh@lzu.edu.cn (S. Yi).

https://doi.org/10.1016/j.ecolind.2026.114735
Received 20 December 2025; Received in revised form 26 January 2026; Accepted 25 February 2026

Traub, 2022; Kaminski et al., 2025). Obtaining these presence/absence datasets, especially the absence localities, is very difficult due to the high costs, time-intensive efforts, and labor demands associated with traditional field surveys (Zaniewski et al., 2002).

Species occurrence datasets, either presence/absence or presence-only are central to SDM parameterization and directly influence the reliability and accuracy of spatial and temporal predictions of species distributions (Vanderwal et al., 2009; Yu et al., 2024). However, most available data are presence-only and lack accurate, systematically collected absence records; moreover, presence-only data are typically derived from ad hoc or non-stratified sampling efforts (Barbet-Massin et al., 2012). Such datasets present three major limitations: (1) the absence of verifiable absence records, which are essential for most SDMs; (2) unquantifiable sampling biases resulting from non-systematic data collection, often driven by accessibility or environmental conditions rather than stratified sampling design; and (3) undetected detectability biases between rare and common species (Zaniewski et al., 2002; Phillips et al., 2009), with herbarium-based datasets in particular tending to overrepresent rare species. In contrast, presence/absence datasets are collected through standardized and consistent sampling strategies, yielding more interpretable and robust model outputs (Franklin, 1998; Elith et al., 2006). Ultimately, the scarcity of reliable absence data continues to constrain SDM development and application, underscoring its importance as a research priority for ecologists and ecosystem managers.

The SDMs previously used in species distribution studies can be categorized into two groups: (1) presence-only models, which rely solely on presence data only, and (2) presence/absence models, which incorporate both presence and absence data (Brotóns et al., 2004). Presence-only SDMs, e.g., BIOCLIM (Busby, 1991) and Mahalanobis distance (Farber and Kadmon, 2003), are seldom adopted in research and often less accurate due to the lack of true absence data. However, because the absence data are usually unavailable, researchers are often forced to generate artificial absence data (usually termed pseudo-absence or background data) and to construct the presence/absence models. The generation of artificial absence data relies on the interplay between geographical and ecological (environmental or habitat) spaces (Colwell and Rangel, 2009; Sillero and Barbosa, 2021), making the selection of pseudo-absence localities challenging. For instance, geographical biases, such as uneven sampling sites along the main roads or political boundaries (e.g., Huntley et al., 2008), and environmental biases, where sampling fails to cover the full climatic range of a species (e.g., high-altitude areas) (Barbet-Massin et al., 2010), can affect the precisions of SDMs. To mitigate these issues, several selection strategies of pseudo-absence localities have been proposed: (1) random selection of sampling sites within the studied area, excluding the known presence localities (e.g., Random in BIOMOD2), (2) random selection with geographically weighted exclusion (e.g., DISK in BIOMOD2, Hirzel et al., 2001), and (3) random selection with environmentally stratified sampling (e.g., SRE in BIOMOD2, Zaniewski et al., 2002). Although the relative merits of these different strategies have been previously discussed in the context of virtual or some plot datasets within limited ranges (e.g., Chefaoui and Lobo, 2008; Phillips and Dudik, 2008; Broussin et al., 2024), few studies have quantified the effects of absence localities generation strategy and sample size, and the underlying mechanisms have not been properly evaluated, primarily due to the lack of large spatial-scale, field-validated absence datasets.

Over the past decade, unmanned aerial vehicle (UAV) technology, characterized by its timeliness, high resolution, low cost, and unified standards (Feng et al., 2015), advanced rapidly, offering a promising solution to the limitations of traditional field-based sampling (Sun et al., 2020; Zhang et al., 2023). The integration of specialized software development kits (SDKs) has further enhanced the efficiency and scalability of UAV surveys, enabling coordinated, regional-scale data collection. A notable example is Fragmentation Monitoring and Analysis with Aerial Photography System (FragMAP, Yi, 2017), an aerial photography system designed for standardized route planning and fragmentation analysis, which has demonstrated high precision in detecting plateau pika (*Ochotona curzoniae*, hereafter pika) (Qin et al., 2021). Leveraging this UAV–FragMAP system, researchers have produced the first comprehensive, field-validated dataset of pika burrow presence and absence at a large spatial scale (Zhang J. et al., 2021), effectively overcoming the limitations of presence-only data in species distribution modelling. Building on this technological innovation, the present study applied the BIOMOD2 modelling platform to systematically evaluate SDM performance using UAV-derived pika data on the Qinghai–Tibetan Plateau (QTP). Specifically, the study aimed to: (1) quantify the effects of UAV-based absence data versus pseudo-absence data on SDM performance; (2) assess how different pseudo-absence generation strategies (random vs. environmentally or spatially stratified) and modelling algorithms influence performance, exploring the underlying mechanisms through variable importance and spatial configuration; and (3) examine the feasibility and prospects of using UAV-derived presence/absence datasets for species distribution modelling. This work provides empirical evidence for optimizing SDM workflows and illustrates how emerging UAV technologies can transform ecological field data collection.

2. Material and methods

2.1. UAV-based field sampling and pika burrow detection

From late June to mid-August each year from 2015 to 2023, a total of 2261 UAV-based sampling localities were established across the QTP (Fig. 1). The QTP, with its extensive and varied alpine grassland and broad climatic range provides an ideal study area for predictive modelling on a regional scale (Zhang et al., 2023).

Over 70,000 aerial images were collected using the GRID mode of the FragMAP, a self-developed Android application built on the DJI software development kit (Fig. 2a; Yi, 2017). The GRID mode is an automated flight protocol consisting of 16 fixed waypoints and covering a 40,000 m^2 area (200 m × 200 m) (Fig. 2a–2c). Phantom 3, Mavic 2, and Mavic 3 drones (DJI Innovations, China) were operated in autopilot mode to capture one aerial image at each waypoint. Each images encompassed a comparable ground area of approximately 35 m × 26 m (Fig. 2b–2d).

Three trained volunteers independently inspected each aerial image and manually delineated polygons around all detectable pika burrows following established image-interpretation protocols (Ofli et al., 2016) (Fig. 2d). Following Chen et al. (2008), a sampling site (i.e., GRID route) was classified as a presence locality if any one aerial image interpretation identified nine or more burrows. This threshold was adopted based on Liu et al. (2025), which indicates that approximately nine burrows correspond to the presence of at least one pika. Sampling sites that did not meet this criterion were classified as absence localities. This procedure yielded 1743 presence and 518 absence localities (Fig. 1). The dataset was randomly partitioned into a training subset (80%, $n = 1808$; including 1401 presences and 407 absences, derived from 10 random partitions and averaged) and an evaluation subset (20%, $n = 452$) to ensure robust model calibration and independent validation for SDM construction.

2.2. Generation of pseudo-absence data

In this study, we generated four absence datasets with different sample sizes, i.e., 407, 1401, 2802 and 4203, which correspond to the UAV- derived field absence sample size, and one-, two-, and three-fold the UAV-based presence sample size, respectively (Fig. 3). Depending on the analytical objective, absence information was derived from UAV-based absences (UA), pseudo-absences (PA), or a combination of both (UAV-based absences + pseudo-absences, UPA). Pseudo-absences were generated using the BIOMOD2 package in R (Thuiller et al., 2009) following three strategies: (1) Random – points drawn at random from

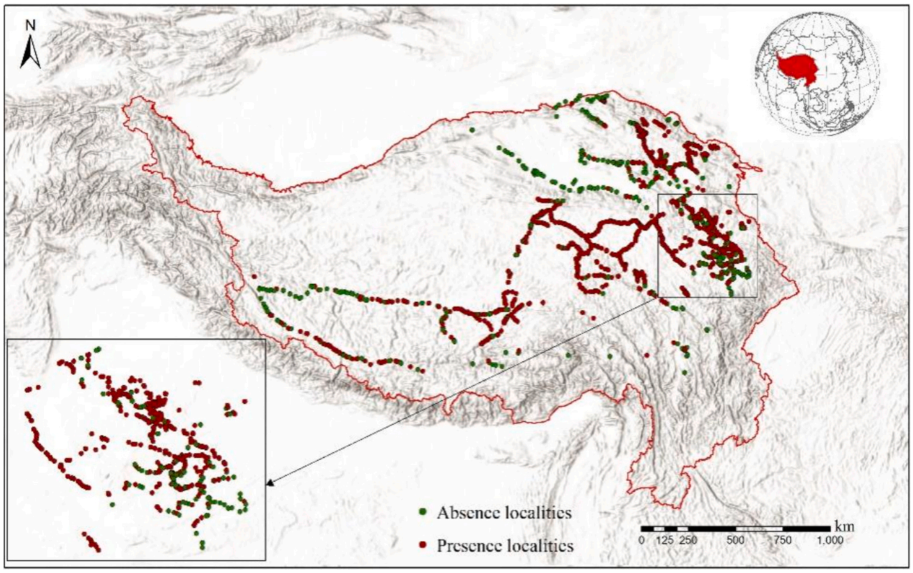

Fig. 1. Sampling localities surveyed using the FragMAP system across the Qinghai–Tibetan Plateau.

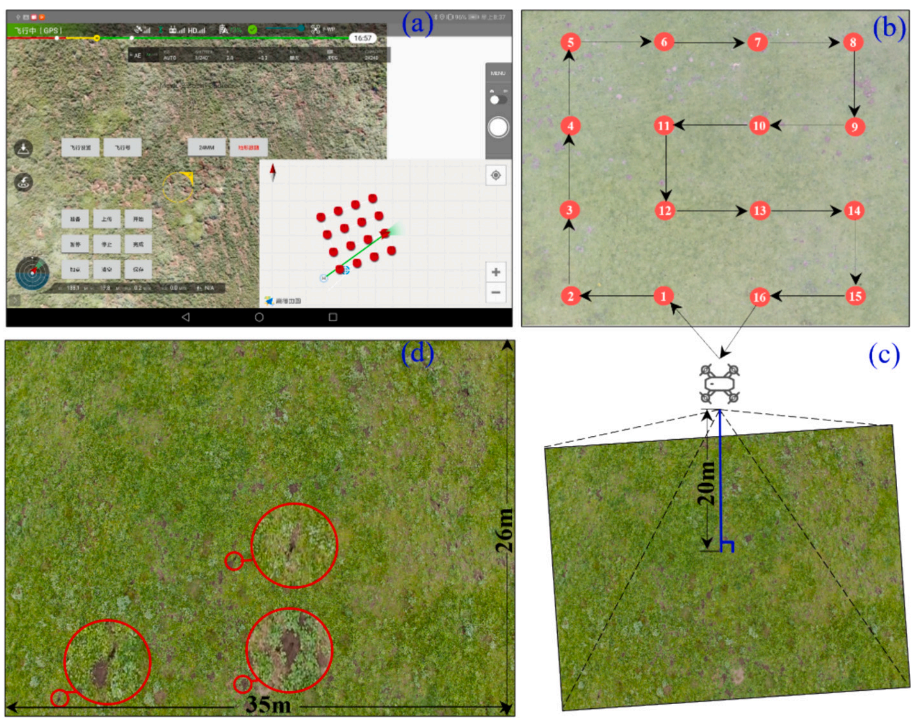

Fig. 2. UAV-based field sampling method and visual detection of pika burrows from high-resolution aerial images.
(a) the FragMAP control interface for automatic flight planning; (b) the GRID route covering a 200 m × 200 m sampling area; (c) a vertical aerial photograph captured at 20 m altitude; and (d) visual identification of pika burrows delineated on an aerial image.

all locations across the QTP, excluding known presence localities; (2) DISK – points randomly sampled from areas located 14.372–43.122 km from any presence locality (Fig. S1); (3) SRE – points selected from areas outside the environmentally suitable range estimated using a rectilinear surface range envelope fitted to presence-only data. In total, the analysis included 216 unique combinations, each subjected to 10 simulation runs for every pseudo-absence treatment, resulting in 2160 SDM modelling scenarios (Fig. 3).

2.3. *Environmental variables for modelling*

2.3.1. *Topography variables*

We downloaded Shuttle Radar Topography Mission (SRTM) data (with a spatial resolution of 90 m × 90 m) derived from the US Geological Survey Administrative (www.usgs.gov). The digital elevation model (DEM), along with derived data for slope and aspect, were used in this study.

2.3.2. *Grassland type*

We used grassland type data from the 1:1000000 Chinese Digital

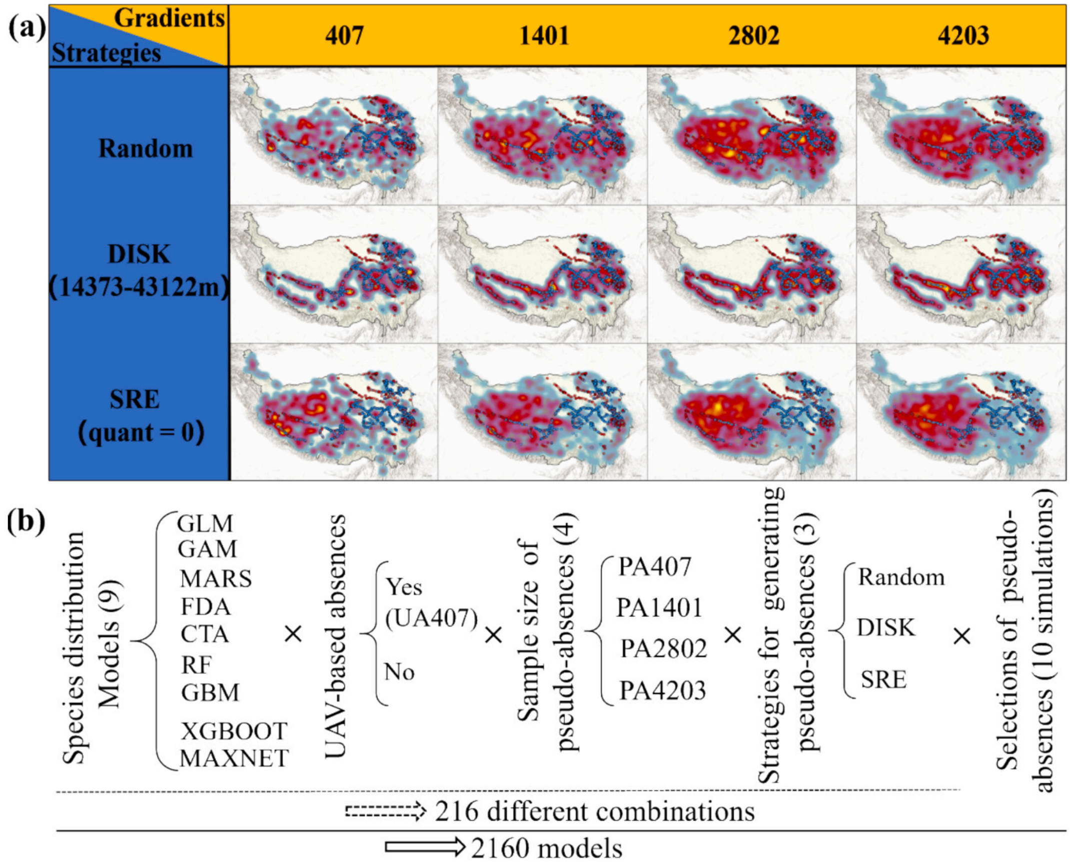

Fig. 3. General framework for data simulation and selection used to evaluate the effects of pseudo-absence generation strategies and sample sizes.
(a) Conceptual illustration of pseudo-absence generation using the Random, DISK, and SRE strategies across four sample-size gradients (PA407, PA1401, PA2802, and PA4203); (b) Overall workflow outlining the data simulation and selection procedures and all factors tested in this study.

Grassland Classification Map provided by the China Resource and Environmental Science and Data Center (https://www.resdc.cn/). The major grassland types include alpine meadow, alpine steppe, alpine steppe meadow, and alpine desert. The total grassland area in this study is 137.54×10^4 km^2.

2.3.3. Climate variables

We downloaded the multi-year mean datasets of monthly minimum and maximum air temperature, mean air temperature, and precipitation from 2010 to 2020, which were downscaled from CRU-TS-4.03 to 2.5 min (about 21 km^2) by the Climatic Research Unit, University of East Anglia, using WorldClim 2.1 for bias correction (https://www.worldclim.org/data/monthlywth.html, Guo et al., 2019). The dataset was further processed to calculate 19 bioclimatic variables of the QTP, with a spatial resolution of 250 m × 250 m.

2.3.4. Above ground biomass

The above ground biomass (AGB) dataset for the QTP was derived by integrating field measurements, UAV-based aerial image, and MODIS (Moderate-resolution Imaging Spectroradiometer) data at a spatial resolution of 250 m (Zhang et al., 2023). The UAV-derived products served as an intermediate scaling layer, substantially mitigating spatial mismatches between ground observations and MODIS spectral indices. Moreover, the AGB estimation model was validated across multiple years, ensuring temporal robustness and confirming the accuracy and reliability of the resulting dataset.

2.3.5. Correlations among environmental variables

Correlations among environmental variables (i.e., vegetation, topography, and climate) were assessed using correlation analysis. Aboveground biomass (AGB) had a significantly positive correlation with bio12 (annual precipitation, AP), bio13 (precipitation of wettest month), bio16 (precipitation of wettest quarter), bio17 (precipitation of driest quarter) and bio18 (precipitation of warmest quarter), while sharply negative correlation with bio14 (precipitation of driest month) and bio15 (precipitation seasonality) (Fig. S2). No significant correlation between topography and climate variables was observed, with exception of those between bio1 (annual mean temperature, AMT), bio5 (max temperature of warmest month), bio8 (mean temperature of wettest quarter), bio10 (mean temperature of warmest quarter), and DEM (Fig. S2). Basing on the correlations among these environmental variables and taking physiological needs (e.g., minimum tolerable temperature and food requirements) and behavior characteristics (e.g., suitability of grasslands for burrow excavation), variables including AGB, DEM, grassland type, AMT, AP, and precipitation of the coldest quarter (bio19, PCQ) were selected for constructing the BIOMOD models.

2.3.6. Construction and evaluation of distribution models

We used the BIOMOD2 package in R (see Thuiller et al., 2009 for further details) to construct SDMs using multiple models, including three regression models (i.e., GLM, GAM, and MARS), two classification models (i.e., FDA and CTA), three machine-learning models (i.e., RF, GBM, and XGBOOST), and Maxent. We split the total localities into 80% for training and 20% for validation, ensuring a robust performance assessment. The model performance was evaluated using the Kappa, area under the Receiver Operating Characteristic curve (AUC), and true skill statistics (TSS). Kappa is a chance-corrected measure of agreement and commonly used in ecological studies with presence/absence data. It requires a threshold to be applied to the predictions, to convert them to presence/absence predictions (Elith et al., 2006). AUC has been used extensively in the species' distribution modelling, and measures the ability of a model to discriminate between sites where a species is present, versus those where it is absent. This provides an indication of the usefulness of the models for prioritizing areas in terms of their relative importance as habitat for the particular species (Hanley and McNeil, 1982). TSS could quantify how well the model distinguishes presence and absence based on omission and commission errors and is unaffected by prevalence (Allouche et al., 2006). The constructed models were then applied to each pixel of grasslands on the QTP using the selected current climate variables, aboveground biomass, and topographic variables. Furthermore, the potential suitable distribution areas could be divided into specific grades, such as: unsuitable area ($P < 22\%$), minimally suitable area ($22\% \leq P < 50\%$), moderately suitable area ($50\% \leq P < 75\%$), and highly suitable area ($75\% \leq P < 1$) (Qi et al., 2022).

2.4. Statistical analysis

We first quantified pairwise correlations among environmental variables and then examined the relationships of individual predictors using the *nlrq* function in the R package quantreg. Normality and homogeneity of variance were assessed using the Shapiro–Wilk test. When these assumptions were satisfied, we applied ANOVA to evaluate differences in model performance metrics (Kappa, AUC, and TSS) across modelling scenarios and to assess variable importance as well as similarity in environmental gradients among UAV-based and pseudo-absence strategies with different sample sizes. If normality or homoscedasticity assumptions were violated, we employed the Wilcoxon signed-rank nonparametric test to compare variables. All calculations and statistical analysis were performed using software R 4.3.1.

3. Results

3.1. Effects of absence generation strategies and model algorithms on SDM performance

Using a fixed set of 1401 presence localities, models incorporating 407 UAV-derived absences consistently outperformed those relying the pseudo-absence data. This superiority was robust across all modelling algorithms and pseudo-absence generation strategies. Among models using 407 pseudo-absences generated via the Random, DISK, and SRE strategies, the SRE strategy produced the highest predictive accuracy across performance metrics, with the exception of TSS in the FDA model. In contrast, pseudo-absences generated using the Random and DISK strategies yielded significantly lower model accuracy ($P < 0.05$, Fig. 4). Across all algorithms implemented in BIOMOD2, the Random Forest (RF) model performed markedly better than the others (Kappa = 0.566, TSS = 0.611, AUC = 0.863; Fig. 4).

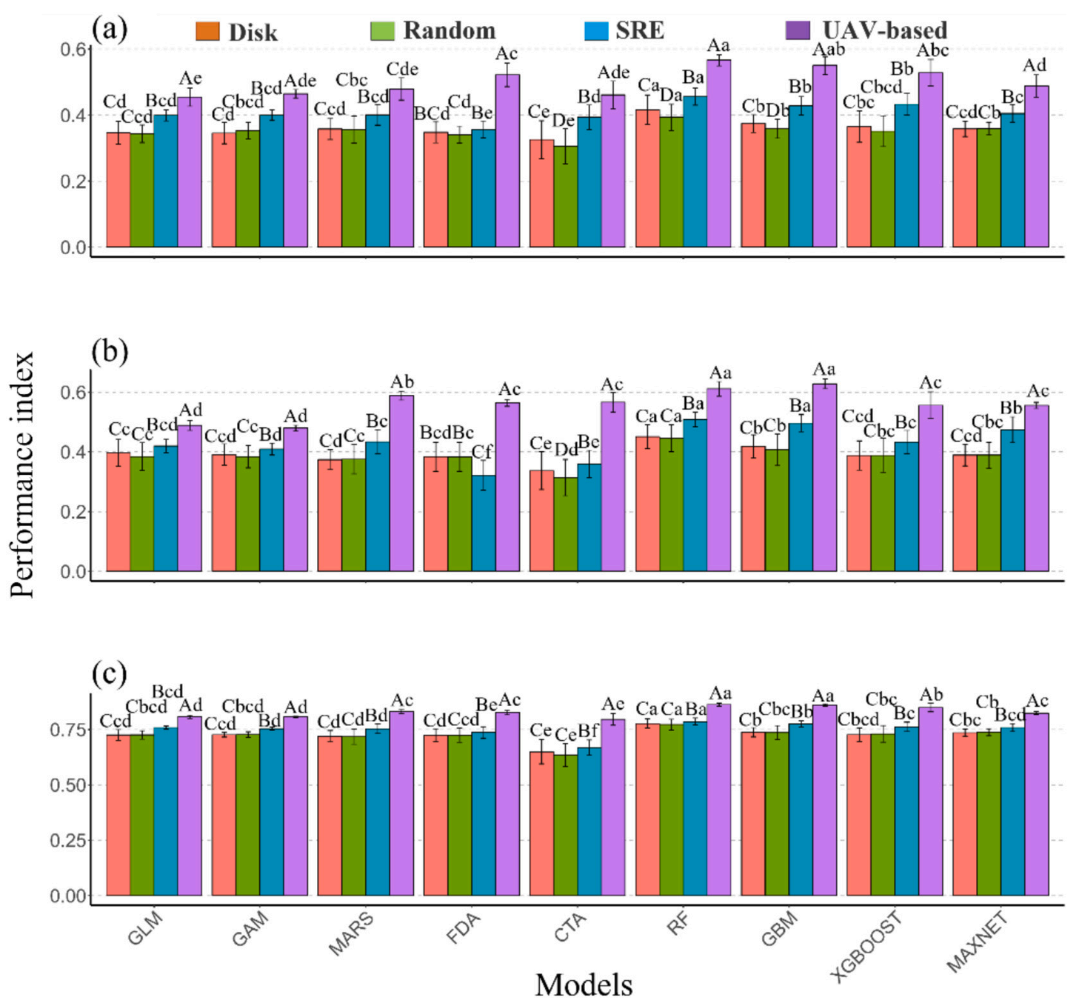

Fig. 4. Comparative evaluation of model performance using 1401 presence localities and either 407 UAV-based absences or 407 pseudo-absence localities. (a) Kappa, (b) TSS, and (c) AUC. Different capital letters above the bars indicate significant differences among pseudo-absence generation strategies ($P < 0.05$), while different lowercase letters denote significant differences among BIOMOD modelling algorithms ($P < 0.05$).

3.2. *Impact of absence data type and sample size on RF model performance*

Based on the top-performing Random Forest (RF) model, the inclusion of 407 UAV-derived absence localities consistently yielded the highest predictive performance (mean Kappa = 0.566, TSS = 0.611, AUC = 0.863), outperforming all pseudo-absence strategies (Random, DISK, and SRE) as well as their combinations with UAV-derived absence data (Fig. 5).

Across all pseudo-absence strategies, models incorporating both UAV-derived absence data and pseudo-absences generally achieved significantly higher performance than models constructed using pseudo-absences alone, particularly for the Random and DISK strategies ($P < 0.05$). When comparing pseudo-absence strategies overall, Random models showed higher average performance than DISK and SRE models under combined scenarios. In contrast, when models were constructed using pseudo-absences alone, the SRE strategy consistently produced the highest performance, followed by Random and then DISK (Fig. 5).

Model performance further varied with pseudo-absence sample size and strategy. Under the combined scenarios, the model performance for the DISK strategy exhibited a unimodal pattern, increasing initially with pseudo-absence sample size and then declining as sample size increased further. In contrast, performance under the Random and SRE strategies showed a decline with increasing pseudo-absence numbers. For models constructed with solely on pseudo-absence data, the SRE strategy consistently outperformed Random and DISK across all evaluated sample sizes ($P < 0.05$). In addition, pseudo-absence datasets with sample sizes equal to the number of presence records generally achieved higher overall performance in terms of Kappa and TSS, although AUC values were slightly lower than those obtained using larger pseudo-absence datasets (Fig. 5).

3.3. *Relative importance of environment variables*

For models incorporating UAV-derived absence localities, above-ground biomass (AGB) was identified as the most influential

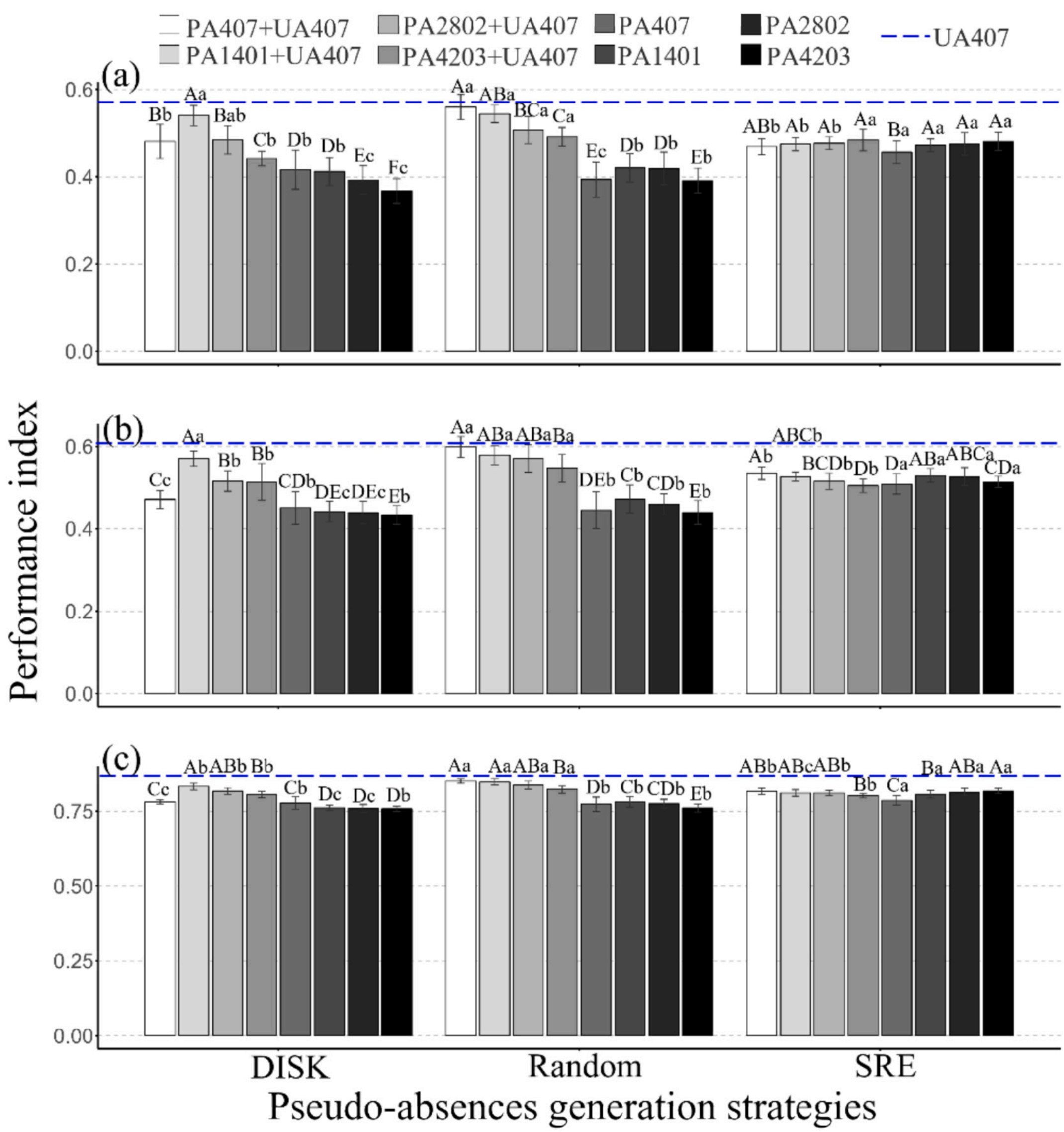

Fig. 5. Performance comparison of Random Forest (RF) models constructed using different combinations of absence-data sampling strategies and sample sizes (the presence sample size was 1401).
(a) Kappa, (b) TSS and (c) AUC. PA, pseudo-absences; UA, UAV-based absences. Different capital letters above the bars indicate significant differences among absence-data combination patterns ($P < 0.05$), while different lowercase letters indicate significant differences among pseudo-absence generation strategies ($P < 0.05$).

environmental predictor, followed by annual mean precipitation (AMP), annual precipitation (AP), and precipitation of the coldest quarter (PCQ). In contrast, for models constructed using pseudo-absence localities generated by different strategies, the digital elevation model (DEM) consistently ranked as the dominant predictor. Across models moreover, no uniform trend in variable importance was observed with increasing sample size (Fig. 6).

In addition to these general patterns, variable importance differed systematically among pseudo-absence generation strategies, reflecting differences in the spatial configuration of absence localities. Under the Random strategy, pseudo-absence localities were unconstrained spatially and became progressively more evenly distributed as sample size increased. Correspondingly, the relative importance of AGB and PCQ declined, whereas the contribution of DEM increased. For the DISK strategy, pseudo-absence localities remained spatially constrained by the distribution of presence records. As pseudo-absence sample size increased, the importance of DEM and PCQ remained relatively stable, while AGB reached its highest contribution when the number of pseudo-absence localities matched that of presence records. Under the SRE strategy, pseudo-absences were selected primarily in environmental space. With increasing pseudo-absence sample size, the importance of DEM showed little variation, whereas the relative contributions of AGB increased and PCQ showed a moderate increase (Fig. 6).

3.4. Environment divergence between pseudo-absence and presence localities

The interaction between sample size and pseudo-absence generation strategy had no significant effect on the main environmental variables, except for grassland type (Table S1). Among the pseudo-absence generation strategies, the SRE produced localities most divergent from the presence sites, exhibiting the lowest overlap across all environmental variables ($P < 0.05$), followed by the Random and DISK strategies (Fig. 7). Specifically, compared to the presence localities, the localities generated by SRE strategy were more frequently associated with lower AGB, AMT, AP and PCQ, but occurred at higher elevations (Fig. 7).

Under the Random strategy, sample size had little effect on divergence, except that divergence in PCQ increased with larger sample sizes. For the DISK strategy, increasing the number of pseudo-absence points amplified divergence in AGB, AP, and PCQ. In contrast, for the SRE strategy, sample size had no significant effect on environmental divergence ($P < 0.05$; Fig. 7).

4. Discussion

Drawing on a large UAV-derived presence/absence dataset of pika burrows ($n = 2261$) across the QTP, our study provides a unique broad geographic basis for evaluating SDMs. By applying multiple modelling algorithms to both presence-only and presence/absence datasets, we were able to conduct a comprehensive assessment of SDM performance and determine the conditions under which reliable predictions can be achieved.

4.1. Incorporation of absence data substantially improves SDM performance

Across all modelling scenarios, models integrating UAV-based absence data consistently outperformed those relying on pseudo-absence data, demonstrating the critical importance of collecting

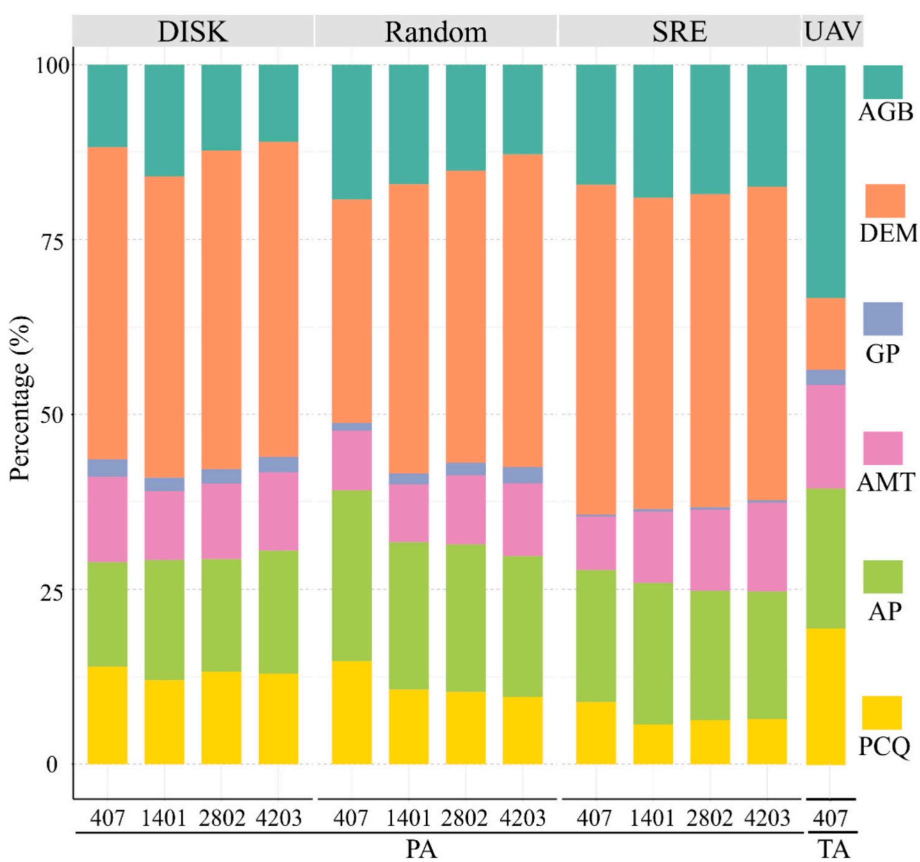

Fig. 6. The importance of the environment variables to the RF model's performance. AGB, aboveground biomass; DEM, digital elevation model; GP, grassland type; AMT, annual mean temperature; AP, annual precipitation; PCQ, precipitation of coldest quarter. Presence sample size = 1401, the UAV-based absences (UA) sample size = 407, and there were 407, 1401, 2802 and 4203 pseudo-absence (PA) samples were generated by DISK, Random and SRE strategies, respectively.

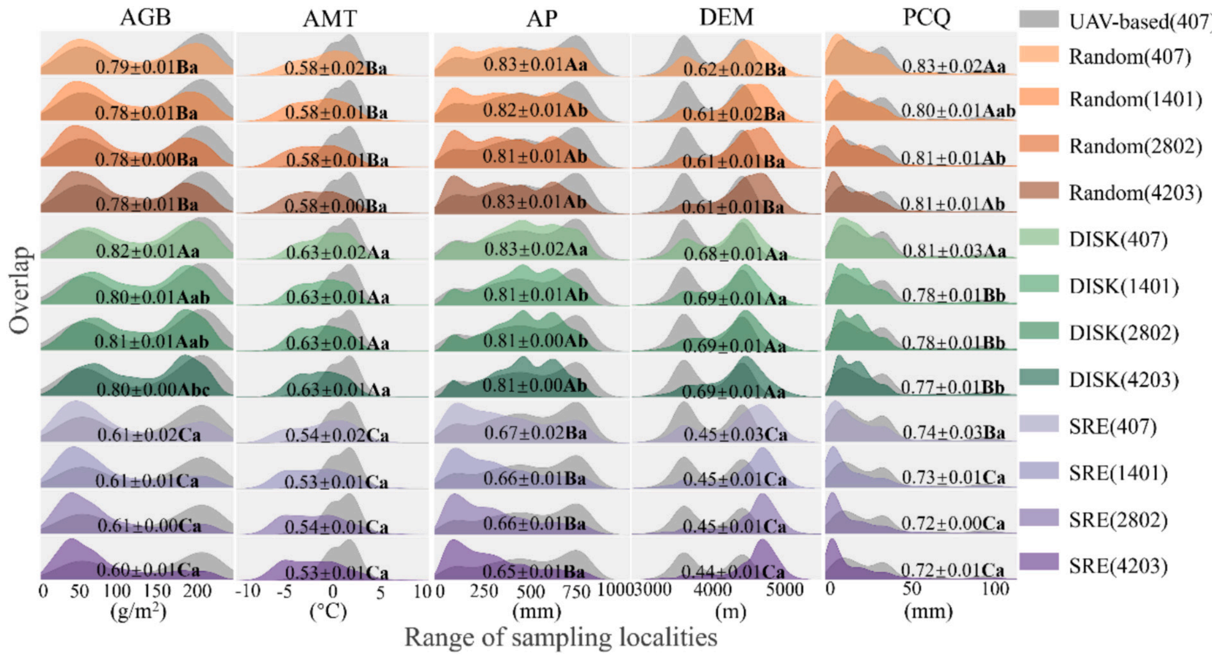

Fig. 7. Environmental divergence between presence localities and pseudo-absence localities generated using different strategies and sample sizes.
AGB, aboveground biomass (g/m^2); AMT, annual mean temperature (°C); AP, annual precipitation (mm); DEM, digital elevation model (m); PCQ, precipitation of coldest quarter (mm). The different capital letters in a row indicate the statistical difference between generating strategies ($P < 0.05$), and the different lowercase letters in a column denote statistical difference between pseudo-absence sample sizes ($P < 0.05$).

robust absence information for improving SDM accuracy (Zhang et al., 2021a, 2021b). This finding, based on extensive standardized field observations, provides the first direct evidence that field-validated absences significantly enhance SDM performance in the alpine grasslands.

Among the BIOMOD2 models, the RF achieved the highest predictive accuracy. Nonetheless, RF is not universally superior for all taxa, as model performance often varies with species-specific ecological constraints and environmental contexts (Elith et al., 2006; Barbet-Massin et al., 2012). In this study, RF-predicted distributions of the plateau pika identified highly suitable habitats mainly in southern Gansu, eastern Qinghai, and northwestern Sichuan (Fig. 8a), covering approximately 11.0% of the QTP. Moderately suitable habitats occurred along the

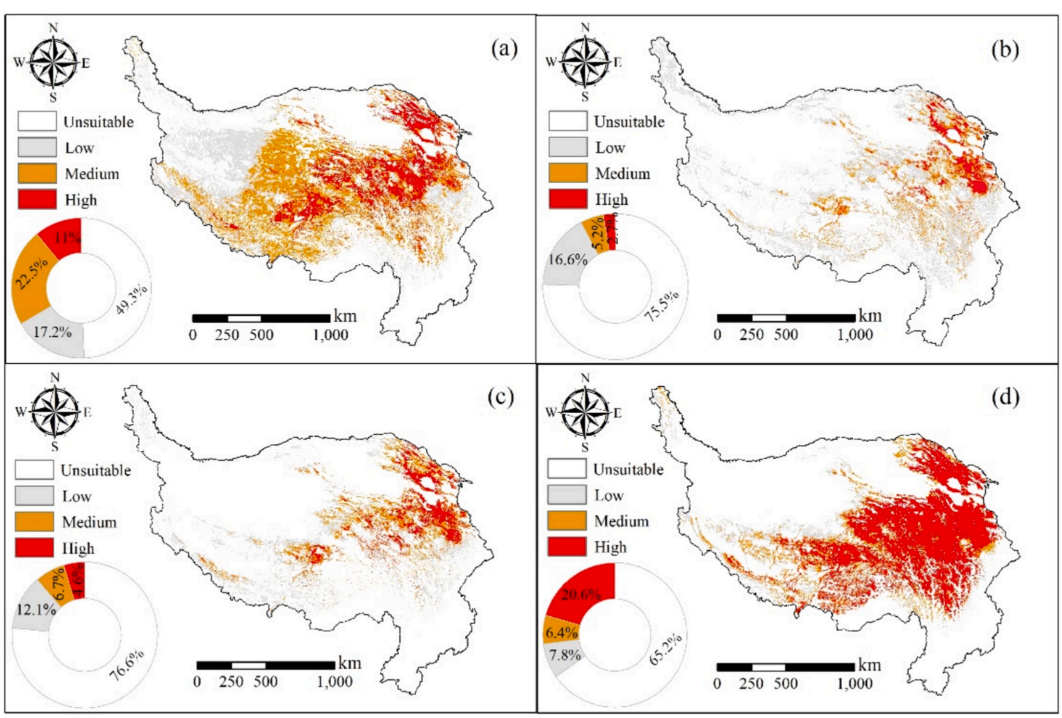

Fig. 8. Predicted habitat suitability for the plateau pika based on Random Forest (RF) models.
(a) The map derived from RF using UAV-based absence localities; (b) the map derived from RF using pseudo-absences generated by DISK strategy; (c) the map derived from RF using pseudo-absences generated by Random strategy; (d) the map derived from RF using pseudo-absences generated by SRE strategy. Suitability classification: **0**–0.22 = Unsuitable habitat; **0.22**–0.5 = Low suitability; **0.5**–0.75 = Medium suitability; **0.75**–**1** = High suitability.

periphery of these areas (22.5%), whereas unsuitable habitats dominated the northwest and southeast (49.3%). All models using pseudo-absence data consistently underestimated suitable habitat extent, with SRE showing the smallest deviation relative to Random and DISK strategies (Fig. 8b–d). This systematic underestimation aligns with patterns observed in previous presence-only SDM studies (Table S2; Hua et al., 2023; Qi et al., 2022).

Plateau pika has long been recognized as a keystone species in the alpine rangeland ecosystem on the QTP (Cui et al., 2024; Smith and Foggin, 1999). However, high-density outbreaks are often associated with grassland degradation, leading to their classification as pests in certain contexts (Hua et al., 2023). Recent research indicates that pika impacts on alpine grasslands are minimal at moderate densities and closely tied to activity intensity (Qin et al., 2025). Since species distribution and density form the basis for effective prevention and management (Huang et al., 2025), our spatially explicit habitat suitability assessment provides essential ecological insights and supports the development of evidence-based management strategies tailored to regional conditions.

4.2. Optimizing pseudo-absence strategies and sample sizes in presence-only SDMs

The incorporation of UAV-based absence localities consistently enhanced model performance across all pseudo-absence generation strategies and sample sizes, likely because field-verified absences improve the environmental representation of unsuitable conditions and reduce ambiguity in species–environment relationships (Figs. 6 and 7). Nevertheless, due to logistical, financial, and temporal constraints, most species datasets in practice remain presence-only, making pseudo-absence generation an indispensable component of SDMs (Zaniewski et al., 2002).

Our results demonstrate that the effects of pseudo-absence generation strategies on model performance are contingent on pseudo-absence sample size. When pseudo-absence numbers were limited and comparable to the number of presence records, SDM performance exhibited pronounced strategy-specific sensitivities. For the Random strategy, moderate pseudo-absence sample sizes yielded the highest predictive accuracy. In contrast, excessively large random pseudo-absence sets increasingly sampled environmentally marginal yet ecologically irrelevant conditions, thereby weakening the contrast between suitable and unsuitable environments and reducing model discrimination (Figs. 5 and 3a). The DISK strategy showed a different response: model performance declined as pseudo-absence sample size increased. The potential reason could be that DISK restricts pseudo-absence selection to geographically defined buffers around presence localities, increasing sample density intensifies spatial dependence and local autocorrelation. This overrepresentation of local environmental gradients limits the model's ability to generalize beyond the immediate vicinity of observed presences, an effect that is particularly evident at small to intermediate sample sizes (Figs. 5 and 3a). In contrast, the SRE strategy displayed improved performance with increasing pseudo-absence sample size. As an environment-based approach (Hao et al., 2019), SRE increasingly delineates unsuitable environmental space as sample size grows, even when presence data are limited, resulting in more robust discrimination of species' environmental boundaries (Figs. 5 and 3a). This pattern suggests that SRE is comparatively resilient to sample size limitations, provided that pseudo-absences adequately span the environmental gradients of the study region.

Beyond overall model performance, our findings further reveal that pseudo-absence strategies influence SDM outcomes by reshaping the relative importance of environmental predictors through their spatial configuration. Under the Random strategy, the unconstrained spatial distribution of pseudo-absences increasingly approximates uniform coverage as sample size increases (Fig. 3). This spatial homogenization amplifies the influence of broad-scale variables such as elevation (DEM), while progressively reducing the relative importance of ecologically proximal predictors, including aboveground biomass (AGB) and precipitation of the coldest quarter (PCQ). For the DISK strategy, the spatial coupling between pseudo-absences and presence records stabilizes the importance of DEM and PCQ across sample sizes, while allowing AGB to exert its strongest influence when pseudo-absence and presence samples are numerically balanced (Figs. 3 and 6). This pattern indicates an optimal contrast between suitable and marginal habitats when spatial dependence is maintained without excessive sampling density. In contrast, the SRE strategy operates primarily in environmental space rather than geographic space. As pseudo-absence sample size increases, this environmentally stratified configuration enhances contrasts along productivity and climatic gradients, resulting in increasing contributions of AGB and PCQ while maintaining a relatively stable influence of DEM (Figs. 3 and 6). This mechanism is consistent with the “reverse niche” concept, in which unsuitable environmental conditions are explicitly emphasized to sharpen species–environment relationships (Broussin et al., 2024).

Taken together, these results indicate that the effects of pseudo-absence sample size are conditional on the spatial and environmental structure imposed by different pseudo-absence generation strategies. Consequently, pseudo-absence sample size should not be optimized according to a universal “more-is-better” principle (Wisz and Guisan, 2009; Phillips and Dudík, 2008; Barbet-Massin et al., 2012). Rather, sample size and generation strategy should be jointly considered in light of the target species' ecology and the intended modelling objective to achieve robust, generalizable, and ecologically interpretable SDM outcomes.

4.3. UAV-based monitoring enhances data structure and representativeness for SDMs

SDMs have become indispensable tools in ecology, conservation, and ecosystem management (Broussin et al., 2024). However, their performance depends heavily on the availability of high-quality presence/absence datasets collected via consistent sampling designs, which are often costly, labor-intensive, and logistically challenging (Austin et al., 1994; Franklin, 1998). UAV technology offers a transformative solution. Over the past decade, UAV-based surveys have emerged as an efficient, scalable, and non-destructive method for collecting presence/absence data at broad spatial extents (Qin et al., 2020; Surasinghe et al., 2025). Compared to the traditional field visual-based surveys and literature method, UAV-based datasets provide clear advantages: (1) ability to collect absence data, addressing a fundamental limitation of presence-only dataset; (2) greater spatial coverage and systematic sampling, spanning diverse grassland types and climatic zones despite partial accessibility constraints; (3) larger, standardized sampling units (e.g., 200 m × 200 m), implemented through FragMAP's GRID mode, which improves representativeness and reduces heterogeneity in plot sizes (Yi, 2017); (4) more reliable detection of pika presence/absence, as burrow features are easily identified in aerial image, abandoned burrows degrade quickly, helping distinguish historical from current activity (Yi et al., 2016); (5) Compatibility with long-term monitoring, enabling fixed-point observations across large spatial and temporal scales to capture population dynamics and habitat responses (Sun et al., 2018). Collectively, these advantages greatly enhance the reliability and ecological interpretability of SDMs and support deeper insights into pika behavioral patterns, dispersal dynamics, and habitat responses.

Despite these advantages, UAV-based acquisition of presence/absence data is subject to several practical limitations that warrant consideration. Data collection remains sensitive to adverse weather conditions, including heavy or prolonged rainfall, strong winds, and fog, which could reduce flight efficiency and compromise image quality. In addition, UAV operations are governed by airspace regulations and data-security policies that differ among countries and regions, potentially restricting flight permissions, operational flexibility, and spatial coverage. Such regulatory constraints may limit the scalability and transferability of UAV-based monitoring frameworks. Furthermore, although the cost of UAV hardware has declined in recent years, financial limitations faced by early-career researchers, together with the limited availability of widely adopted standardized survey and processing systems (e.g., FragMAP, Yi, 2017), continue to constrain the widespread implementation of UAV-derived presence/absence datasets. As a result, UAV-based datasets remain relatively sparse at regional to continental scales. Nevertheless, ongoing technological advances in UAV platforms, increasing regulatory clarity, and the growing accessibility of standardized monitoring and analytical frameworks are expected to progressively alleviate these constraints. As these developments mature, UAV-based monitoring is likely to play an increasingly central role in structuring high-quality presence/absence datasets, thereby enhancing the robustness, scalability, and practical applicability of SDMs for biodiversity assessment and ecosystem management.

4.4. Limitations and future prospects

Pseudo-absence quality remains a central challenge in SDMs. Model performance decreases when pseudo-absences are sampled from overly narrow or excessively broad environmental spaces (Van Der Wal et al., 2009). Localities too close or too distant from presence points may contribute limited ecological information and reduce model reliability. As study extent expands, the probability of selecting environmentally uninformative pseudo-absences increases. This likely explains why models incorporating composite samples (UAV-based absences + pseudo-absences) performed better than those using pseudo-absences alone, yet still did not exceed the performance of models based solely on UAV-confirmed absences (Fig. 5). Further research is required to explore the mechanisms underlying pseudo-absence generation and the effects of sample size on model performance.

Accurate identification of true absence localities remains a major challenge in species distribution modelling due to issues related to imperfect detectability, observer bias, and inconsistent sampling designs (Zaniewski et al., 2002). The high-precision, fixed-point monitoring framework implemented by FragMAP offers a promising solution by enabling repeated surveys of standardized 200 m × 200 m plots. Pika burrows are readily detectable in high-resolution UAV image (Qin et al., 2021), and abandoned burrows tend to collapse or become obscured by vegetation over time, making them reliable indicators of species absence. Despite these advantages, the UAV-based datasets used in this study are spatially constrained, as surveys were conducted primarily along accessible road networks (Kadmon et al., 2004; Zhang et al., 2021b). Future research should therefore expand sampling efforts into more remote regions and across broader climatic and environmental gradients to construct more comprehensive and representative presence–absence datasets. In addition, although presence/absence data are effective for delineating species distributions, they offer limited insight into population density. Given the high efficiency and fine spatial resolution of UAV-based surveys, future studies could exploit this technology to generate quantitative density estimates. Integrating density information with distribution models would provide deeper ecological understanding and enhance the development of evidence-based, regionally tailored management strategies in the face of accelerating climate change and intensifying anthropogenic pressures.

5. Conclusions

Using an extensive UAV-derived presence/absence dataset for plateau pika across the QTP, we evaluated SDM performance under multiple pseudo-absence generation strategies and sample sizes. Our results show that incorporating UAV-based absence data markedly enhances model accuracy, with Random Forest achieving the best predictive performance. Among the pseudo-absence strategies, SRE consistently outperformed DISK and Random, while sample size had only a limited effect on model outcomes. All pseudo-absence–based models systematically underestimated the extent of suitable pika habitat, a pattern likely driven by shifts in the relative importance of key environmental variables and by the environmental divergence between pseudo-absence and true presence locations.

CRediT authorship contribution statement

Yi Sun: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Bo Huang:** Software, Methodology, Investigation, Data curation. **Yifei Luo:** Writing – original draft, Software, Investigation, Formal analysis. **Yu Qin:** Software, Methodology, Investigation, Formal analysis, Data curation. **Xiong Zhao He:** Writing – review & editing, Visualization, Validation, Methodology, Data curation. **Shuhua Yi:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

This research was jointly funded by the grants from the National Natural Science Foundation of China (42071056) and the CAS "Light of West China" Program (xbzglzb2022022).

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2026.114735.

Data availability

Data will be made available on request.

References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232.

Austin, M.P., Nicholls, A.O., Doherty, M.D., Meyers, J.A., 1994. Determining species response functions to an environmental gradient by means of a β-function. J. Veg. Sc. 5, 215–228.

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, 327–338.

Barbet-Massin, M., Thuiller, W., Jiguet, F., 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? Ecography 33, 878–886.

Brotóns, L., Thuiller, W., Araújo, M.B., Hirzel, A.H., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27, 437–448.

Broussin, J., Mouchet, M., Goberville, E., 2024. Generating pseudo-absences in the ecological space improves the biological relevance of response curves in species distribution models. Ecol. Model. 498, 110865.

Busby, J.R., 1991. BIOCLIM: a bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds.), nature conservation: cost effective biological surveys and data analysis. CSIRO, pp. 64–68.

Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecol. Model. 210, 478–486.

Chen, J., Wang, Z., Wang, Y., Li, B., Zhaxi, Xiaozhaxi, Luosangdawa, Zhang, M., 2008. Methods for investigating the density of the plateau pika in northern Tibetan plateau. Plant Prot. 34, 114–117.

Colwell, R.K., Rangel, T.F., 2009. Hutchinson's duality: the once and future niche. P. Natl Acad. Sci. USA 106, 19651–19658.

Cui, H., Wang, Y., Zhou, X., Li, W., 2024. Positive role of plateau pika (*Ochotona coronae*) on environmental quality at low and moderate density on the Tibetan plateau: evidence from a meta-analysis. J. Environ. Manage. 361, 121239.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Farber, O., Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecol. Model. 160, 115–130.

Feng, Q., Liu, J., Gong, J., 2015. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. Remote Sens–BASEL 7, 1074–1094.

Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. J. Veg. Sci. 9, 733–748.

Guo, Y., Li, X., Zhao, Z.F., Nawaz, Z., 2019. Predicting the impacts of climate change, soils and vegetation types on the geographic distribution of *Polyporus umbellatus* in China. Sci. Total Environ. 648, 1–11.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Hao, T., Elith, J., Guillera-Arroita, G., Lahoz-Monfort, J.J., 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Divers. Distrib. 25, 839–852.

Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual species. Ecol. Model. 145, 111–121.

Hua, R., Hua, L., Tang, Z., Dong, R., Bao, D., Ye, G., La, M., Sun, W., Zhang, Z., Wang, L., Dong, L., Cai, B., Chu, B., Hao, Y., 2023. Maxent modeling for predicting habitat suitability and potential distribution of plateau pika (*Ochotona curzoniae*) on the Qinghai-Tibet plateau, China. Rangel. Ecol. Manag. 87, 34–43.

Huang, T., Yu, H.Y., Liu, X.Y., Luo, S.H., Meng, B.P., Wang, X.Y., Sun, Yi, Yi, S., Qin, Y., 2025. The influencing mechanism and changing evaluation of plateau pika hazard: implications for grassland management. Land Degrad. Dev. 0, 1–16.

Huntley, B., Collingham, Y., Willis, S., Green, R., 2008. Potential impacts of climatic change on European breeding birds. PloS One 3, e1439.

Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of perdictive maps produced by bioclimatic models. Ecol. Appl. 14, 401–413.

Kaminski, D.J., Poole, K.E., Harms, T.M., Andress, A.J., 2025. A Bayesian method for estimating multiscale resource selection using presence-only data: a case study predicting winter distributions for little brown, northern long-eared, and tricolored bats. Landsc. Ecol. 40, 70.

Liu, X., Qin, Y., Sun, Y., Yi, S., 2025. Monitoring plateau pika and revealing the associated influencing mechanisms in the alpine grasslands using unmanned aerial vehicles. Drones 9, 298.

Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., Parkan, M., Joost, S., 2016. Combining human computing and machine learning to make sense of big (aerial) data for disaster response. Big Data 4, 47–59.

Phillips, S., Dudík, M., 2008. Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197.

Qi, Y., Pu, X., Li, Y., Li, D., Huang, M., Zheng, X., Guo, J., Chen, Z., 2022. Prediction of suitable distribution area of plateau pika (*Ochotona curzoniae*) in the Qinghai–Tibet Plateau under shared socioeconomic pathways (SSPs). Sustainability 14, 12114.

Qin, Y., Huang, T., Yu, H., Wang, X., Luo, S., Liu, X., 2025. Moderate plateau pika activity had minor effect on ecosystem C and N in alpine grasslands. Glob. Ecol. Conserv. 63, e03890.

Qin, Y., Sun, Y., Zhang, W., Qin, Y., Chen, J., Wang, Z., Zhou, Z., 2020. Species monitoring using unmanned aerial vehicle to reveal the ecological role of plateau pika in maintaining vegetation diversity on the northeastern Qinghai-Tibetan plateau. Remote Sens. 12, 2480.

Qin, Y., Yi, S., Ding, Y., Qin, Y., Zhang, W., Sun, Y., Hou, X., Yu, H., Meng, B., Zhang, H., Chen, J., Wang, Z., 2021. Effects of plateau pikas' foraging and burrowing activities on vegetation biomass and soil organic carbon of alpine grasslands. Plant and Soil 458, 201–216.

Radomski, T., Beamer, D., Babineau, A., Wilson, C., Pechmann, J., Kozak, K.H., 2022. Finding what you don't know: testing SDM methods for poorly known species. Divers. Distrib. 28, 1769–1780.

Schmidt-Traub, G., 2022. National climate and biodiversity strategies are hamstrung by a lack of maps. Nat. Ecol. Evol. 5, 1325–1327.

Sillero, N., Barbosa, A.M., 2021. Common mistakes in ecological niche models. Int. J. Geogr. Inf. Sci. 35, 213–226.

Smith, A., Foggin, M., 1999. The plateau pika (*Ochotona curzoniae*) is a keystone species for biodiversity on the Tibetan plateau. Anim. Conserv. 2, 235–240.

Sun, Y., Yi, S., Hou, F., 2018. Unmanned aerial vehicle methods makes species composition monitoring easier in grasslands. Ecol. Indic. 95, 825–830.

Sun, Y., Yi, S., Hou, F., Luo, D., Hu, J., Zhou, Z., 2020. Quantifying the dynamics of livestock distribution by unmanned aerial vehicles (UAVs): a case study of yak grazing at the household scale. Rangeland Ecol. Manag. 73, 642–648.

Surasinghe, T., Singh, K., Frazier, A., 2025. Harnessing the full potential of drones for fieldwork. BioScience 75, 1–9.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.

Van Der Wal, J., Shoo, L., Graham, C., Williams, S., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol. Model. 220, 589–594.

Wisz, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol. 9, 8.

Yi, S., 2017. FragMAP: a tool for long-term and cooperative monitoring and analysis of small-scale habitat fragmentation using an unmanned aerial vehicle. Int. J. Remote Sens. 38, 2686–2697.

Yi, S., Chen, J., Qin, Y., Xu, G., 2016. The burying and grazing effects of plateau pika on alpine grassland are small: a pilot study in a semiarid basin on the Qinghai-Tibet plateau. Biogeosciences 13, 6273–6284.

Yu, B., Dai, W., Li, S., Wu, Z., Wang, J., 2024. A new threshold selection method for species distribution models with presence-only data: extracting the mutation point of the P/E curve by threshold regression. Ecol. Evol. 14, e11208.

Zaniewski, A.E., Lehmann, A., Overton, J.M., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 157, 261–280.

Zhang, H., Tang, Z., Wang, B., Kan, H., Sun, Y., Qin, Y., Meng, B., Li, M., Chen, J., Lv, Y., Zhang, J., Niu, S., Yi, S., 2023. A 250 m annual alpine grassland AGB dataset over the Qinghai–Tibet Plateau (2000–2019) in China based on in situ measurements, UAV photos, and MODIS data. Earth Syst. Sci. Data 15, 821–846.

Zhang, J., Liu, D., Meng, B., Jianjun, C., Wang, X., Jiang, H., Yu, Y., Yi, S., 2021a. Using UAVs to assess the relationship between alpine meadow bare patches and disturbance by pikas in the source region of Yellow River on the Qinghai-Tibetan plateau. Glob. Ecol. Conserv. 26, e01517.

Zhang, X., Yuan, Y., Zhu, Z., Ma, Q., Yu, H., Li, M., Ma, J., Yi, S., He, X., Sun, Y., 2021b. Predicting the distribution of *Oxytropis ochrocephala* Bunge in the source region of the Yellow River (China) based on UAV sampling data and species distribution model. Remote Sens–BASEL 13, 5129.