ACTA BIOLOGICA SIBIRICA
A journal on the biodiversity of Siberia and the adjacent lands

Acta Biologica Sibirica 12: 1693–1710 (2026)
doi: 10.5281/zenodo.24362350
http://journal.asu.ru

RESEARCH ARTICLE

Ensemble habitat suitability modeling of the endemic *Arctium pallidivirens* (Asteraceae) in the Nuratau Mountains, Uzbekistan

Khikmatullo R. Ibragimov[1], Bekhruz Sh. Khabibullaev[1], Khabibullo F. Shomurodov[1], Natalya Yu. Beshko[1], Shukhrat M. Jumayev[2], Giyos X. Buxorov[3], Diyorjon X. Hamrayev[1], Mukhlisa B. Turopova[4], Feruza J. Raimova[5], Gayrat A. Sharifov[6]

1 *Institute of Botany, Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan*
2 *Samarkand State Pedagogical Institute, Samarkand, Uzbekistan*
3 *Turon University, Karshi, Uzbekistan*
4 *Termez State Pedagogical Institute, Termez, Uzbekistan*
5 *Termez State University of Engineering and Agrotechnologies, Termez, Uzbekistan*
6 *State Environmental Research Institution "Nuratau State Reserve", Farish, Uzbekistan*

Corresponding author: Khikmatullo R. Ibragimov (hikmatullai626@gmail.com)

Academic editor: A. Matsyura | Received 31 August 2026 | Accepted 22 October 2026 | Published 21 November 2026

http://zoobank.org/BB15A91C-140B-4400-93E8-219ABA5FDF7A

Citation: Ibragimov KhR, Khabibullaev BSh, Shomurodov KhF, Beshko NYu, Jumayev ShM, Buxorov GX, Hamrayev DX, Turopova MB, Raimova FJ, Sharifov GA (2026) Ensemble habitat suitability modeling of the endemic *Arctium pallidivirens* (Asteraceae) in the Nuratau Mountains, Uzbekistan. Acta Biologica Sibirica 12: 1693–1710. https://doi.org/10.5281/zenodo.24362350

Abstract

The narrow endemic plants of the Central Asian mountains are threatened by grazing and change in land use, but their habitat requirements are rarely quantified. We modeled the habitat of *Arctium pallidivirens*, a vulnerable endemic of the Nuratau Mountains, Uzbekistan. We compiled 48 occurrence records from herbarium specimens (1925–2020), expert field observations, and our own surveys, and fitted an ensemble of four algorithms (GLM, GAM, random forest, and MaxEnt) with eight climatic, vegetation, topographic, and soil predictors, evaluated with spatial block cross-validation. The ensemble reached a mean AUC of 0.93 and a TSS of 0.63 but was less reliable in the west of the range. Precipitation in the warmest quarter was by far the most important predictor, followed by mean NDVI and slope; suitability increased with all three. At the 10th percentile presence threshold, suitable habitat covered 1,184 km² (6.8% of the study area) in a narrow band along the main Nuratau ridge. The

Copyright Khikmatullo R. Ibragimov et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

predictions agreed with evidence not used in the models: 44 of 48 records and 29 of 34 occupied 2 × 2 km cells contained suitable habitat, and no cell below 700 m reached a suitability of 0.5, although elevation was not a predictor. Unsurveyed suitable habitat was predicted on the southern side of the Khobduntau Ridge. Because flowering ends and fruiting takes place in the driest part of the year, summer rain may limit its reproduction. Monitoring should focus on the central ridge, and surveys should target the Khobduntau Ridge and three historical localities where the species was not refound.

Keywords
Bioclimatic predictors, Central Asia, IUCN Red List, narrow endemic, Pamir-Alay, spatial cross-validation, species distribution model, summer precipitation

Introduction

Plants with small ranges are among the most vulnerable to changes in land use and climate. Many depend on particular soils, slopes, or microclimates, and a single change in land use can affect a large part of their range (Kruckeberg and Rabinowitz 1985). Knowing where suitable habitat remains and what limits it is a basic requirement for its protection (Franklin 2010). The mountains of Central Asia are rich in endemic plants, but few studies have measured their habitat requirements (Ma et al. 2024; Abulfayzov et al. 2026). Temperature, moisture, and soils in these mountains change over short distances, and narrow endemics often survive in small, scattered populations. In the Alps, moisture, vegetation cover, and soil properties together explain much of the distribution of mountain plants (Buri et al. 2020; Chauvier et al. 2021). Comparable work in Central Asia is rare: recent studies of narrow endemic and relict plants in Uzbekistan, including several endemics of the Nuratau Mountains, have focused mainly on distribution, population status and extinction risk (Shomurodov et al. 2024; Orsenigo et al. 2025; Fenu et al. 2025) rather than on quantifying habitat requirements. *Arctium pallidivirens* (Kult.) S. López, Romasch., Susanna & N.Garcia (Asteraceae) is endemic to the Nuratau Mountains of Uzbekistan (Beshko 2020; Tojibaev et al. 2020). It was described as *Cousinia pallidivirens* from specimens collected by M.G. Popov in the Aktau Range in 1925 (Kultiassov 1929), later placed in *Anura* (Tscherneva 1962), and transferred to *Arctium* after a molecular revision of the *Arctium–Cousinia* group (Lopez-Vinyallonga et al. 2011). It is a perennial herb of rocky and gravelly slopes at 700–1800 m (Volis and Beshko 2023); its seedlings and a virginile plant are shown in Suppl. material 1: Fig. S1. About 50 populations are known, and the extent of occurrence (EOO) is 3,467 km² and its area of occupancy (AOO) is 136 km² (Volis and Beshko 2023). Under IUCN criteria, it is assessed as Vulnerable in the Red List of Uzbekistan (Shomurodov et al. 2026). The main threats are livestock grazing and the transformation of habitats around villages (Volis and Beshko 2023). Volis and Beshko (2023) modelled the climatically suitable habitat of 27 Nuratau endemics, including *A. pallidivirens*, with bioclimatic MaxEnt models under current and future climates; for this species the

predicted suitable range was largely confined to the Nuratau and Aktau ranges, and translocation to the Surkhan Nature Reserve was proposed as a climate-change adaptation measure. For *Lepidium olgae*, another narrow endemic of the same mountains, an ensemble model showed that wet-season precipitation was the dominant driver of its distribution, followed by vegetation productivity and thermal stability (Abulfayzov et al. 2026). For *A. pallidivirens*, however, non-climatic predictors have not been included and model performance has not been assessed with spatially independent data.

Herbarium specimens of the species have been collected since the type collection of 1925, most of them before 1960. Since then, settlements, roads, and grazing have expanded in the lower parts of the Nuratau Mountains. Old specimens, therefore, record parts of the range that recent surveys may have missed, and they show where the species should be searched for again. Together with recent field records, they give the most complete picture of its known distribution. Species distribution models relate occurrence records to environmental conditions and show where suitable habitat is likely to occur (Franklin 2010). Ensemble models combine several algorithms, so the results depend less on any single method (Araújo and New 2007; Thuiller et al. 2009). We used this approach to (i) map the potential distribution of *A. pallidivirens* in the Nuratau Mountains, (ii) identify the climatic, vegetation, topographic, and soil variables most closely associated with its occurrence, and (iii) check the predictions against all known records and the elevational range of the species. The results are meant to guide field surveys and the protection of the remaining habitat.

Materials and methods

Study area

The Nuratau Mountains lie in central Uzbekistan, on the right bank of the Zarafshan River (Fig. 1). They stretch for over 250 km, cover nearly 14,000 km² and rise from 300 m to 2169 m above sea level. The mountains are the north-western spurs of the Turkestan Range of the Pamir-Alay system and consist of two parallel ranges, the Nuratau in the north and the Aktau in the south, separated by the Nurata valley and a chain of low ridges (Beshko 2018, 2020; Beshko et al. 2025). The central part is cut by deep gorges and rocky canyons, whereas to the west the mountains become lower and pass into the sandy plains of the Kyzylkum Desert; the northern slopes are much steeper than the southern ones (Beshko 2020). The area lies in the transitional zone between the Pamir-Alay and Tien Shan mountain systems and the arid plains of the Turan Lowland, and belongs to the Nuratau district of the Mountain Central Asian province of the Irano-Turanian floristic region; the district is divided into three phytogeographical regions: the Nuratau, the Aktau and the Nuratau Relic Mountains (Beshko 2020). Its flora is rich in endemic and relict species (Beshko

2018). The study area was the Nuratau phytogeographical district (okrug; Beshko 2020). Besides the mountains themselves, it includes the surrounding foothills and plains, so it is larger (17,374.45 km^2) than the mountain range and its lowest parts lie below 300 m.

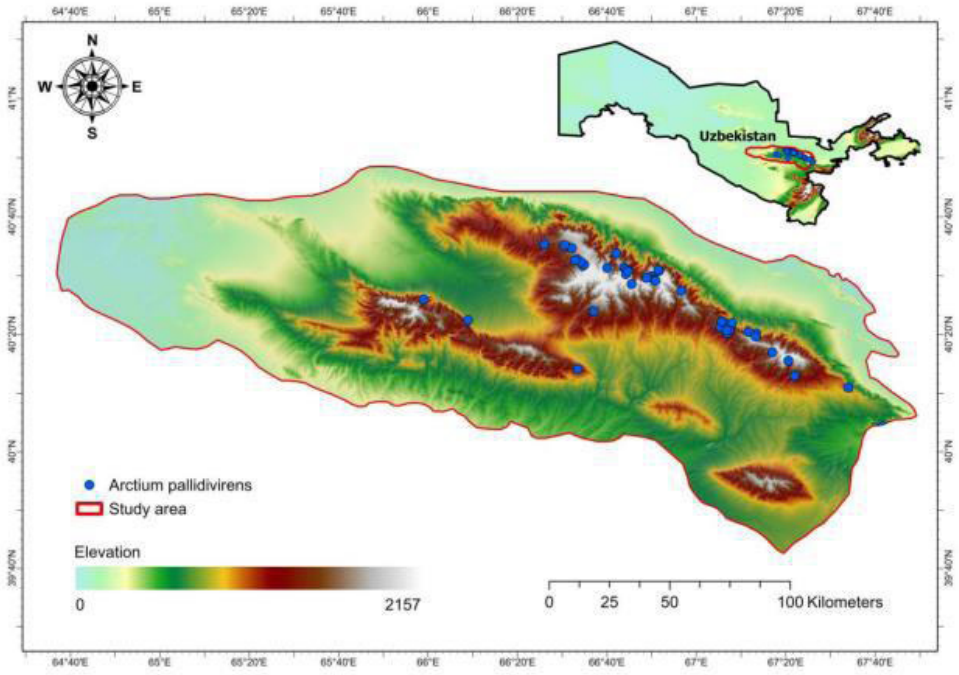

Figure 1. Study area: Nuratau phytogeographical district (Uzbekistan).

Species occurrence data

Plant names follow the Catalogue of Life (Roskov et al. 2018). The species in its habitat is shown in Fig. 2. The records of *A. pallidivirens* came from three sources. The first was 14 herbarium specimens collected between 1925 and 2020 and deposited in the National Herbarium of Uzbekistan (TASH). The second was 13 field observations without vouchers made by N.Yu. Beshko between 2002 and 2022; three of them are documented with photographs in the Plantarium database (www.plantarium.ru). The third was 24 records from our own field surveys, located with a handheld GPS receiver. The records without vouchers were made by N.Yu. Beshko and by us during our own field surveys. N.Yu. Beshko, a specialist on the flora of the Nuratau Mountains, checked the identification of all herbarium specimens and verified the field records; none of the records are crowdsourced. Historical specimens were georeferenced from their label data. Another specimen, collected in the Malguzar Mountains in 1947 (Korotkova 61, TASH), was excluded because the

Malguzar Range lies beyond the Tamerlane's Gate gorge, outside the study area, and the species has not been found there since. Of the remaining 51 records, three were duplicates, leaving 48 records in 40 cells of the 1-km predictor grid. To reduce spatial clustering, we kept only records at least 1 km apart (Veloz 2009). This distance equals the resolution of the predictors: records closer together fall in the same or adjacent cells and carry nearly identical predictor values, so retaining them would add weight to a few localities without adding environmental information. Thinning left 32 records for model calibration. To test the effect of this choice, we repeated the cross-validation with a 500 m threshold, which retained 38 records (Suppl. material 2: Table S1). The 48 records were used to evaluate the final predictions.

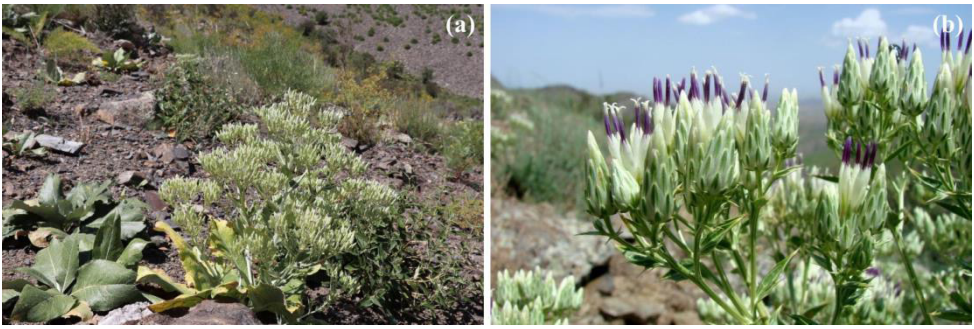

Figure 2. *Arctium pallidivirens* in the Nuratau Mountains: (a) flowering plant in its habitat on a gravelly slope; (b) capitula.

Environmental predictors

Nineteen bioclimatic variables were taken from WorldClim v2.1 at 30 arc seconds resolution (Fick and Hijmans 2017). The slope and TPI were derived from the SRTM digital elevation model (Farr et al. 2007), the mean NDVI of MODIS MOD13A2 (Didan et al. 2015), the mean daily solar radiation (MJ m^{-2} day^{-1}) from GLDAS (Rodell et al. 2004), and the content of sand and clay (g kg^{-1}) from SoilGrids (Hengl et al. 2017). All layers were resampled into the WorldClim grid. To verify the layers, we compared the bioclimatic values used for modeling with the original WorldClim layers in the 26,328 cells of the study area that had values in the resampled layers; they agreed in all cells. Collinearity was checked in the calibration data with Pearson correlations and the variance inflation factor (VIF; Dormann et al. 2013). Of the strongly correlated variables, only one was kept; for example, precipitation of the wettest quarter (bio16) was excluded because it was almost identical to precipitation of the wettest month (bio13; r = 0.99). Eight predictors were retained: precipitation of the warmest quarter (bio18, mm), precipitation of the wettest month (bio13, mm), mean NDVI, slope (°), TPI, sand and clay content and mean solar radiation. All had VIF < 5 (maximum 4.55). Two pairs had correlations slightly above 0.7 (bio13–bio18, r = 0.73; bio18–slope, r = 0.71); we kept them because they describe

different seasons and processes, but the effects of bio18 and slope cannot be fully separated. The full correlation matrix is given in Suppl. material 1: Fig. S2. Elevation was not used as a predictor, so that it could serve as an independent check of the predictions.

Model calibration and evaluation

Presence and pseudo-absence data were prepared with the data-formatting function of biomod2 (Thuiller et al. 2009). We generated ten sets of 320 pseudo-absences (ten per presence) at random within the boundary of the study area only; all pseudo-absences lay at least 1 km from the nearest presence record (Barbet-Massin et al. 2012). Four algorithms were then fitted to each set in R, giving 40 models: a generalized linear model (GLM) with linear terms; a generalized additive model (GAM; package mgcv, smooth terms with k = 3, REML estimation and shrinkage of unimportant terms; Wood 2017); random forest (RF; package ranger, 500 trees, probability output; Wright and Ziegler 2017); and MaxEnt (package maxnet, linear and quadratic features, regularization multiplier 1, cloglog output; Phillips et al. 2017). Because random splitting of spatially clustered records overestimates model performance (Veloz 2009), the models were evaluated with spatial block cross-validation. The 32 calibration records were ordered by longitude and divided into four blocks of eight records each (W1, 65.98–66.57° E; W2, 66.58–66.76° E; E1, 66.81–67.11° E; E2, 67.13–67.57° E), and pseudo-absences were assigned to the same blocks according to their longitude, with the two outer blocks extending to the edges of the study area. To reduce spatial dependence between training and test data, training records within 5 km of the withheld block were excluded (Roberts et al. 2017). Each block was withheld in turn, the models were fitted to the other three, and performance on the withheld block was measured with AUC, TSS, sensitivity, and specificity (Allouche et al. 2006). The threshold for TSS was chosen on the training data only (maximum Youden index; package pROC; Robin et al. 2011). The final prediction was the unweighted mean of the 40 models fitted to all calibration records (maps of the individual algorithms in Suppl. material 1: Fig. S3), and spatial uncertainty was expressed as the standard deviation among them. Variable importance was measured by permutation: for each model, the values of one predictor in the calibration data were randomly shuffled, and importance was calculated as one minus the correlation between the original and the new predictions (Thuiller et al. 2009). Scores were averaged over the 40 models. Response curves were obtained by varying one predictor across its range while holding the others at their median values in the 32 calibration records, so that each curve shows the effect of one predictor under the conditions where the species occurs; they are shown as the mean and standard deviation of the 40 models.

Suitable area and validation

To convert continuous predictions into suitable and unsuitable cells, we used the 10th percentile training presence threshold (P10; 0.332), which classifies 90% of the calibration records as suitable and is recommended for small samples (Liu et al. 2005; Pearson et al. 2007). For comparison, we also report the threshold that maximized TSS on the withheld blocks (0.10) and a fixed threshold of 0.5. Two hundred cells (131 km²) within the study area lacked values in one or more resampled predictor layers; for these cells, predictor values were taken from the original layers before resampling and predictions were made with the same 40 models. The areas were calculated from the 26,528 cells of the study area (17,374.45 km²). The predictions were then checked in three ways: by the proportion of the 48 occurrence records that fell in suitable cells; by the proportion of occupied 2 × 2 km cells, the grid used to calculate the area of occupancy (IUCN Standards and Petitions Committee 2024), which contained suitable habitat; and by the distribution of predicted suitability across elevation bands, which we compared with the known elevational range of the species (700–1800 m; Volis and Beshko 2023).

Results

Model performance and predictors

In spatial block cross-validation, the four algorithms reached similar AUC values (0.91) but differed in TSS, from 0.60 for GAM to 0.38 for RF, the latter mainly due to low sensitivity (0.46). The ensemble achieved a mean AUC of 0.93 ± 0.03 and a TSS of 0.63 ± 0.22 across the four blocks (Fig. 3A, B). The performance was highest in block E1 (TSS 0.85) and lowest in W1 (TSS 0.34, sensitivity 0.38), so the predictions for the western part of the range are less reliable.

Precipitation in the warmest quarter (bio18) was by far the most important predictor (permutation score 0.72 ± 0.23), followed by mean NDVI (0.19 ± 0.12) and slope (0.09 ± 0.07). The other five predictors each scored below 0.04 (Fig. 3C). Within the range of the occurrence records, suitability rose steeply with bio18, from about 0.15 at 18 mm to 0.85 at 45 mm, and leveled off above this value (Fig. 4A). It also increased with mean NDVI, from about 0.3 to 0.8 in the occupied range (0.14–0.29; Fig. 4B), and with slope, from approximately 0.4 at 5° to 0.7 at 30° (Fig. 4C). Suitability changed little with bio13 (Fig. 4D); all records fell between 60 and 72 mm.

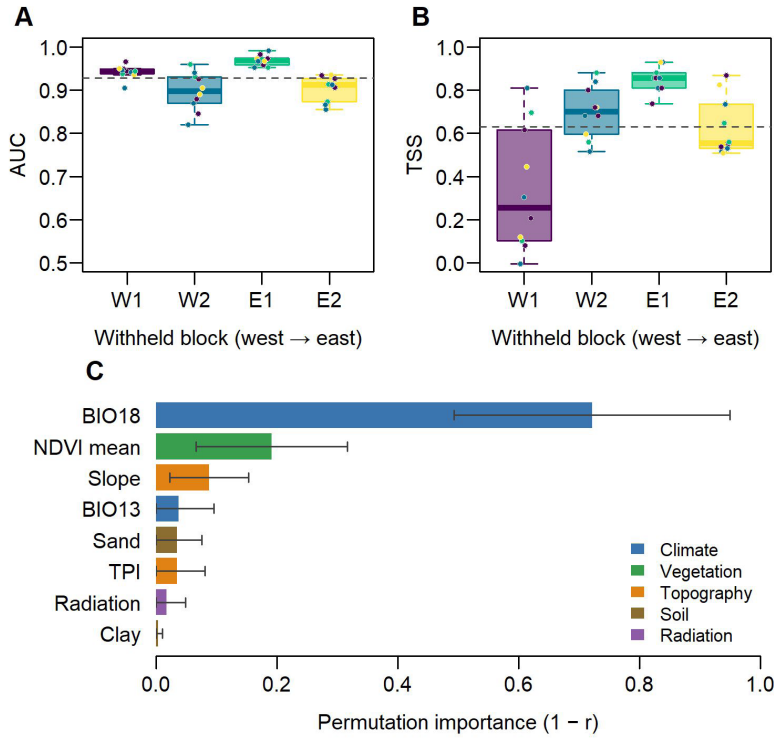

Figure 3. Performance of the ensemble model under spatial block cross-validation and importance of the predictors. (A) AUC and (B) TSS on each block withheld (W1–E2) in the ten pseudo-absence sets; the dashed line shows the mean. (C) Permutation importance of the eight predictors, averaged over the 40 models. Colors in (A) and (B) indicate blocks from west to east; colors in (C) indicate predictor groups.

Predicted distribution

Most of the study area had very low predicted suitability (Fig. 5A). The suitability of 0.5 to 0.8 was predicted for 596.85 km² (3.4%) and 0.8 or more for only 51.70 km² (0.3%), almost all along the main Nuratau ridge between about 66.4° and 67.4° E. A smaller, separate area of suitable habitat was predicted on the southern side of the Khobduntau Ridge, about 34 km from the nearest known record. At the P10 threshold, 1,183.55 km² (6.8%) was classified as suitable (Table 1). Uncertainty was generally low (SD < 0.18) and highest in the central part of the ridge, where suitable and unsuitable cells alternate over short distances (Fig. 5B).

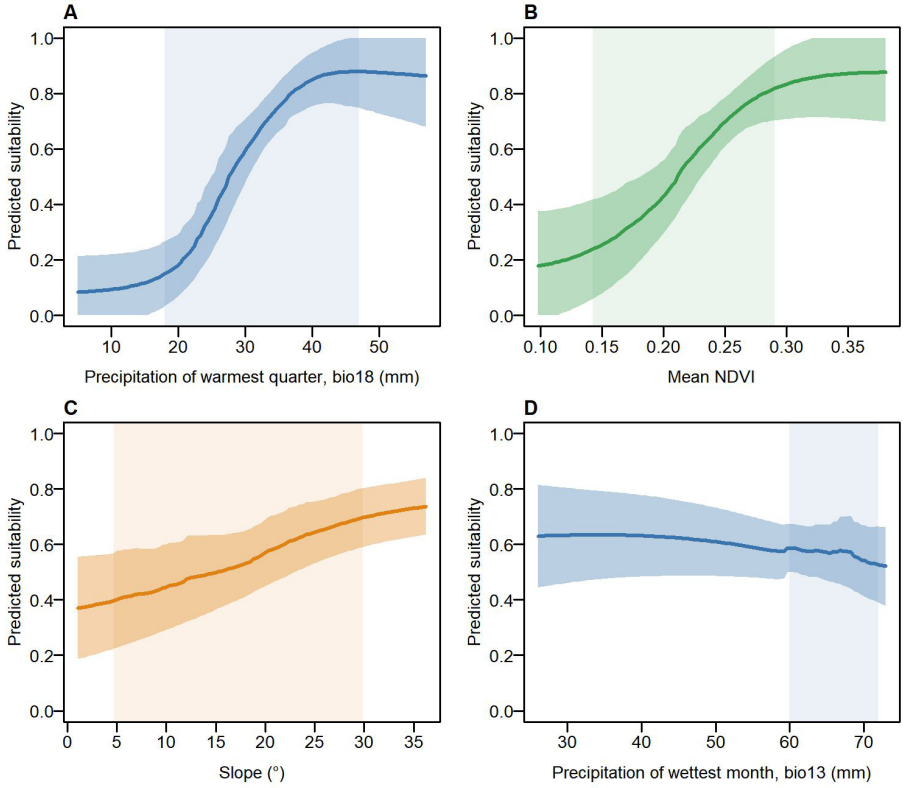

Figure 4. Response curves of *Arctium pallidivirens* to (A) precipitation of the warmest quarter (bio18), (B) mean NDVI, (C) slope, and (D) precipitation of the wettest month (bio13). Each predictor was varied across its range, while the others were held at their median values in the 32 calibration records. Lines show the mean and shaded bands the standard deviation of the 40 models; the light shaded area marks the range of the occurrence records. The colors follow the predictor groups in Fig. 3C.

Table 1. Suitable habitat area and occurrence records within suitable cells across three thresholds

Threshold	Value	Suitable area (km²)	% of study area	Records in suitable cells (of 48)
Maximum TSS (spatial CV)	0.10	3,206.01	18.5	48 (100%)
10th-percentile presence (P10)	0.33	1,183.55	6.8	44 (92%)
Fixed	0.50	648.55	3.7	34 (71%)

Note: The total study area comprises 17,374.45 km².

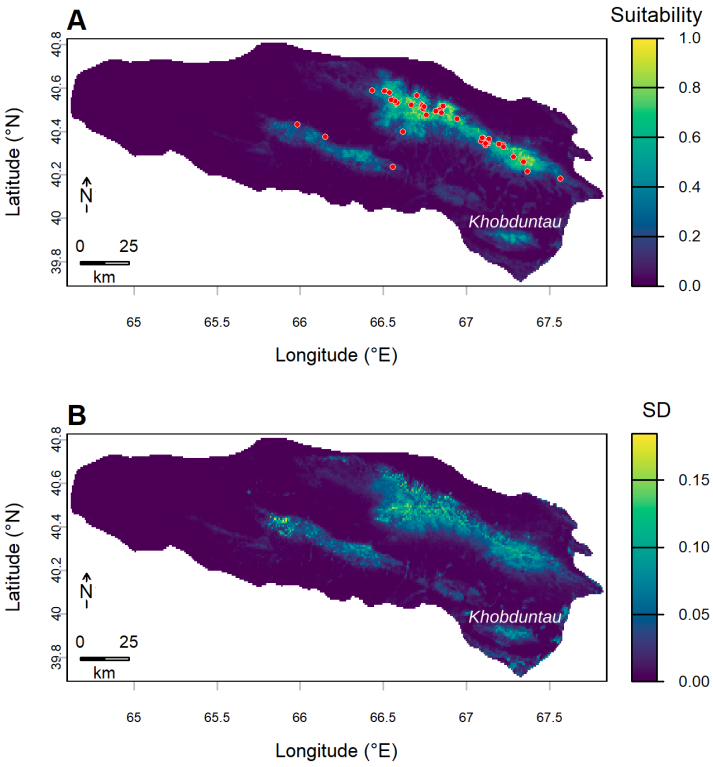

Figure 5. (A) Predicted habitat suitability of *Arctium pallidivirens* (mean of 40 models) with the occurrence records (red dots); (B) spatial uncertainty, expressed as the standard deviation among the 40 models.

Spatial and elevational model validation

At the P10 threshold, 44 of the 48 records (92%) were in suitable cells. The four exceptions were the three western records (65.98–66.56° E) and one record at the eastern margin of the range (67.57° E). The 48 records occupied 34 cells in a 2 × 2 km grid (AOO = 136 km², the same value as reported by Volis and Beshko 2023); 29 of these cells (85%) contained suitable habitat at the P10 threshold and in 26 (76%) suitable habitat covered more than half of the cell. The mean suitability was 0.32 above 1000 m and below 0.06 in all lower elevation bands; no cell below 700 m reached a suitability of 0.5 (Table 2). All records above 1500 m were classified as suitable at both P10 and 0.5, while records below 1000 m were often missed at the stricter threshold.

Table 2. Predicted suitability and occurrence records by elevation band

Elevation (m)	Cells	Mean suitability	Cells ≥ 0.5	Records	Records suitable at P10	Records suitable at 0.5
< 300	2,427	0.004	0	0	–	–
300–500	6,117	0.010	0	0	–	–
500–700	5,769	0.023	0	0	–	–
700–1000	8,002	0.056	23	6	4	2
> 1000	4,213	0.317	968	42	40	32

Discussion

The suitable habitat for *A. pallidivirens* is confined to a narrow band along the main Nuratau ridge. With the 10th percentile training presence threshold (0.33), 1,184 km² (6.8% of the study area) was classified as suitable; with a stricter threshold of 0.5 this decreased to 649 km² (3.7%) and only 51.7 km² exceeded 0.8. Even the larger estimate is about one third of the extent of occurrence of the species (3,467 km²), but it is almost nine times its area of occupancy (136 km²; Volis and Beshko 2023). Therefore, part of the suitable habitat is either unoccupied or has not yet been surveyed. Precipitation in the warmest quarter (bio18) was by far the strongest predictor, and suitability increased with it from about 18 mm to 45 mm before levelling off. In the Nuratau Mountains, rainfall occurs mainly in winter and spring, with a maximum in March–April, and the dry period lasts from June to September (Beshko 2020). Summer rain is scarce in the study area, but the main ridge receives more of it than the foothills and surrounding plains. *A. pallidivirens* flowers from May to July and its fruits ripen from July to September (Beshko 2020; N.Yu. Beshko, published data and personal field observations in the Nuratau State Reserve), so flowering ends and the entire fruiting phase falls in the driest part of the year. Thus, small differences in summer rain can decide whether plants complete reproduction and whether seedlings survive their first year, in line with the strong dependence of plants in dry mountains on water supply (Körner 2021). Summer precipitation in the study area also increases with elevation, so bio18 probably captures part of the wider contrast between the mountains and the lowlands, including lower temperatures and evaporative demand on the ridge. The two effects cannot be separated with our data. For *Lepidium olgae* (Brassicaceae), another narrow endemic of the same mountains, wet-season precipitation was the main climatic predictor (Abulfayzov et al. 2026). Both species appear to be limited by water but in the season that matches their reproductive period: *L. olgae* flowers in April and May and sets fruit in May–June, when the wet season ends, whereas *A. pallidivirens* fruits from July to September, in the driest part of the year. The suitability also increased with slope throughout the range occupied by the species (about 5–30°), which agrees with our

field observation that it grows mainly on rocky and gravelly slopes. Steep slopes are less accessible to livestock and carry sparser herbaceous cover, so they can offer the species both protection from grazing and reduced competition. The suitability also increased with the mean NDVI in the occupied range (0.14–0.29). Therefore, the species is more likely where vegetation cover is somewhat denser than on the bare surrounding slopes, which may mark sites where soil moisture lasts longer into the summer.

Soil texture contributed little to the model (permutation scores ≤ 0.04). This contrasts with studies in the Alps, where soil variables greatly improved plant distribution models (Dubuis et al. 2013; Buri et al. 2020). A likely reason is that the SoilGrids estimates for Central Asia are based on few soil profiles and cannot resolve the shallow, stony substrates on which the species grows. The weak soil effect therefore says more about the available soil data than about the role of substrate for *A. pallidivirens*. The predictions agree reasonably well with evidence that was not used to fit the model. Although elevation was not a predictor, the mean suitability was 0.32 in cells above 1000 m and below 0.06 in every lower elevation band, and no cell below 700 m reached a suitability of 0.5, which matches the known elevation range of the species (700–1800 m). Of the 48 occurrence records, 44 (92%) were in suitable cells at the P10 threshold and 34 (71%) at the stricter threshold of 0.5; all records above 1500 m were suitable under both. The model was less reliable at the edges of the range. The four records outside the suitable habitat at P10 were the three westernmost records (65.98–66.56° E) and one record at the eastern margin (67.57° E), and at the 0.5 threshold only 2 of the six records below 1000 m were classified as suitable. This matches the weak performance of the western block in cross-validation (TSS 0.34) and probably reflects the small number of records from the lower and western parts of the range, where suitability is likely underestimated. The climate of the Nuratau Mountains is also significantly more arid in the west than in the east (Beshko 2020), so the western records lie closer to the dry limit of the species, where predicted suitability is more sensitive to small differences in summer rainfall. The model was calibrated with 32 records after spatial thinning. Such small samples increase the uncertainty of distribution models (Wisz et al. 2008; van Proosdij et al. 2016), and spatial block cross-validation gave lower, but more realistic, performance estimates than random splitting would have (Veloz 2009). Seven older herbarium records were collected before GPS was available and were georeferenced from the locality descriptions on their labels, only to the nearest arc minute (about 1.9 km), which is coarser than the 1-km resolution of the predictors, and their median suitability (0.37) was much lower than that of records with GPS coordinates (0.64). With the available data we cannot tell whether this reflects coordinate error or loss of habitat since the specimens were collected. For conservation, the location of the remaining suitable habitat is more useful than its exact area. The central ridge between about 66.4° and 67.4° E holds most of the highly suitable cells and most of the current records, and monitoring should focus there. Outside strictly protected areas, populations are exposed to livestock grazing and trampling; cattle

and horses browse the plants readily, so the damage is not limited to trampling (Fig. 6b), and at some sites, orthopterans consumed a large share of the developing seeds during our surveys in 2026 (Fig. 6a). Both bush-crickets and locusts (*Dociostaurus maroccanus*, *Calliptamus italicus*) feed on the species, and locust outbreaks recur in the region at intervals of about three to four years (N.Yu. Beshko, personal field observations). The model also points to suitable habitat without records, most clearly a separate area on the southern side of the Khobduntau Ridge (centered at about 39.92° N, 67.29° E), some 34 km south of the nearest known record. These areas, together with the lower slopes where the model is least reliable, should be prioritized in targeted surveys (Raxworthy et al. 2003). At three localities where the species was collected in 1947 and 2012 – one of them in the Malguzar Mountains, a separate range beyond the Tamerlane's Gate gorge (Beshko 2020), just outside the study area – the habitat is now heavily modified by settlements, roads, and grazing, and we did not find the species there during our surveys. These sites should be revisited before the species is considered lost from them.



Figure 6. Field evidence of threats to *Arctium pallidivirens* in the Nuratau Mountains: (a) seed and inflorescence predation by orthopterans; (b) habitat degradation and trampling caused by unmanaged livestock grazing in nonprotected areas.

Conclusion

Suitable habitat for *A. pallidivirens* is limited to a narrow band along the main Nuratau ridge, mostly above 1000 m. Summer precipitation was the strongest predictor of its occurrence, which fits the timing of its fruiting in the driest part of the year. Steep slopes and somewhat denser vegetation also favored the species. The model agrees well with the known records and the elevation range, but was less reliable

at the western edge and on the lower slopes, where more surveys are needed. For conservation, we recommend monitoring the populations on the central ridge and reducing grazing pressure there, surveying the predicted habitat on the southern side of the Khobduntau Ridge, and revisiting the three historical localities where the species was not refound before treating these populations as lost.

Acknowledgements

This study was supported by the state research program "DIGITAL NATURE" (2025–2029) implemented by the Institute of Botany, Academy of Sciences of the Republic of Uzbekistan. We thank the staff of the Nuratau State Reserve for their help during the field surveys.

References

Abulfayzov K, Khabibullaev B, Shomurodov K, Beshko NYu, Sullieva S, Li Y, Fan L (2026) Ecological niche modeling of the narrow-range endangered endemic *Lepidium olgae* in Uzbekistan. Plants 15(7): 1125. https://doi.org/10.3390/plants15071125

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43(6): 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22(1): 42–47. https://doi.org/10.1016/j.tree.2006.09.010

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

Beshko NYu, Shomurodov K, Khabibullaev B (2025) Floristic diversity of the Nuratau Mountains: a 150-year retrospective analysis. Biogeographia – The Journal of Integrative Biogeography 40(2): a053. https://doi.org/10.21426/B6.47012

Beshko NYu (2018) Endemic plants of the Nuratau Mountains (North-Western Pamir-Alay). Uzbek Biological Journal 4: 13–18. [In Russian]

Beshko NYu (2020) Flora of Central Asia: The Nuratau Mountains. Korea National Arboretum, Pocheon, Republic of Korea, 432 pp.

Buri A, Grand S, Yashiro E, Adatte T, Spangenberg JE, Pinto-Figueroa E, Verrecchia E, Guisan A (2020) What are the most crucial soil variables for predicting the distribution of mountain plant species? A comprehensive study in the Swiss Alps. Journal of Biogeography 47: 1143–1153. https://doi.org/10.1111/jbi.13803

Chauvier Y, Thuiller W, Brun P, Lavergne S, Descombes P, Karger DN, Renaud J, Zimmermann NE (2021) Influence of climate, soil, and land cover on plant species distribution in the European Alps. Ecological Monographs 91(2): e01433. https://doi.org/10.1002/ecm.1433

Didan K, Barreto Munoz A, Solano R, Huete A (2015) MODIS Vegetation Index User's Guide (MOD13 Series). Version 3.00, Collection 6. Vegetation Index and Phenology Lab, The University of Arizona, Tucson, AZ, USA.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Marquéz JR, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1): 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dubuis A, Giovanettina S, Pellissier L, Pottier J, Vittoz P, Guisan A (2013) Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. Journal of Vegetation Science 24(4): 593–606. https://doi.org/10.1111/jvs.12002

Fenu G, Beshko NYu, Branca F, Khabibullaev BS, Ilardi V, Maggioni L, Shomurodov KF, Orsenigo S (2025) Global and regional IUCN Red List Assessments: 19. Italian Botanist 20: 187–200. https://doi.org/10.3897/italianbotanist.20.177927

Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf D (2007) The Shuttle Radar Topography Mission. Reviews of Geophysics 45(2): RG2004. https://doi.org/10.1029/2005RG000183

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37(12): 4302–4315. https://doi.org/10.1002/joc.5086

Franklin J (2010) Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511810602

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017) SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE 12(2): e0169748. https://doi.org/10.1371/journal.pone.0169748

IUCN Standards and Petitions Committee (2024) Guidelines for using the IUCN Red List Categories and Criteria. Version 16. Prepared by the Standards and Petitions Committee. Available from: https://www.iucnredlist.org/documents/RedListGuidelines.pdf

Körner C (2021) Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems. 3rd ed. Springer, Cham, Switzerland, 500 pp. https://doi.org/10.1007/978-3-030-59538-8

Kruckeberg AR, Rabinowitz D (1985) Biological aspects of endemism in higher plants. Annual Review of Ecology and Systematics 16: 447–479. https://doi.org/10.1146/annurev.es.16.110185.002311

Kultiassov MV (1929) Materialien zur Kenntnis der Turkestanischen Arten der Gattung Cousinia. Acta Universitatis Asiae Mediae, Series VIII-b, Botanica 6: 1–28.

Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28(3): 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Lopez-Vinyallonga S, Romaschenko K, Susanna A, Garcia-Jacas N (2011) Systematics of the Arctioid group: disentangling *Arctium* and *Cousinia* (Cardueae, Carduinae). Taxon 60(2): 539–554. https://doi.org/10.1002/tax.602020

Ma S, Li W, Tojibaev KSh, Turginov O, Yang W, Ma K (2024) Regionwide and nationwide floristic richness reveal vascular plant diversity in Central Asia. Plants 13(16): 2275. https://doi.org/10.3390/plants13162275

Orsenigo S, Beshko NYu, Casolo V, Castello M, Khabibullaev BS, Petrussa E, Shomurodov KF, Uliana LR, Fenu G (2025) Global and Regional IUCN Red List Assessments 18. Italian Botanist 19: 185–203. https://doi.org/10.3897/italianbotanist.19.156865

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 34(1): 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography 40(7): 887–893. https://doi.org/10.1111/ecog.03049

Raxworthy CJ, Martinez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA, Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. Nature 426(6968): 837–841. https://doi.org/10.1038/nature02205

Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40(8): 913–929. https://doi.org/10.1111/ecog.02881

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77. https://doi.org/10.1186/1471-2105-12-77

Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng CJ, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The Global Land Data Assimilation System. Bulletin of the American Meteorological Society 85(3): 381–394. https://doi.org/10.1175/BAMS-85-3-381

Roskov Y, Ower G, Orrell T, Nicolson D, Bailly N, Kirk PM, Bourgoin T, DeWalt RE, Decock W, van Nieukerken E, Zarucchi J, Penev L (Eds) (2018) Species 2000 & ITIS Catalogue of Life, 2018 Annual Checklist. Species 2000: Naturalis, Leiden, the Netherlands. Available from: www.catalogueoflife.org

Shomurodov K, Adilov B, Rudov A, Sharipova V, Abduraimov O, Khayitov R, Khabibullaev B (2024) Ecological aspects of distribution and population status assessment of *Rhamnus erythroxyloides* subsp. *sintenisii* (Rich.f) Mabb., a relict species in the Kyzylkum Desert of Uzbekistan. Plants 13(22): 3154. https://doi.org/10.3390/plants13223154

Shomurodov K, Fenu G, Khabibullaev B, Beshko NYu, Adilov B, Karimov F, Turginov O, Abduraimov O, Dekhkonov D, Kadirov U, Makhmudjanov D, Juramurodov I, Makhmudov A, Esanov K, Kosimov Z, Allamuratov A, Akbarov F, Gulomov R, Hoshimov H, Akhmedov A, Saribaeva S, Daminova N, Turdiev D, Pulatov S, Abulfayzov K, Sadinov J, Isomov J, Fayziev K, Ibragimov K, Tojibaev K (2026) Red list of the threatened vascular

plants of Uzbekistan. Biological Conservation 322: 112024. https://doi.org/10.1016/j.biocon.2026.112024

Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32(3): 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Tojibaev KSh, Beshko NYu, Batoshov AR, Azimova DE, Yusupov ZT, Deng T, Sun H (2020) Flora of the Dzhizak Province, Uzbekistan. China Forestry Publishing House, Beijing, 523 pp.

Tscherneva OV (1962) *Anura* (Juz.) Tschern. In: Shishkin BK, Bobrov EG (Eds) Flora SSSR. Vol. 27. Izdatelstvo Akademii Nauk SSSR, Moscow–Leningrad, 107–108. [In Russian]

van Proosdij ASJ, Sosef MSM, Wieringa JJ, Raes N (2016) Minimum required number of specimen records to develop accurate species distribution models. Ecography 39(6): 542–552. https://doi.org/10.1111/ecog.01509

Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. Journal of Biogeography 36(12): 2290–2299. https://doi.org/10.1111/j.1365-2699.2009.02174.x

Volis S, Beshko NYu (2023) How to preserve narrow endemics in view of climate change? The Nuratau Mountains as the case. Plant Diversity of Central Asia 2(2): 82–101. https://doi.org/10.54981/PDCA/vol2_iss2/a3

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, NCEAS Predicting Species Distributions Working Group (2008) Effects of sample size on the performance of species distribution models. Diversity and Distributions 14(5): 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Wood SN (2017) Generalized Additive Models: An Introduction with R. 2nd ed. Chapman and Hall/CRC, New York, 496 pp. https://doi.org/10.1201/9781315370279

Wright MN, Ziegler A (2017) ranger: a fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77(1): 1–17. https://doi.org/10.18637/jss.v077.i01

Supplementary material 1

Figure S1. *Arctium pallidivirens* in the Nuratau Mountains: (a) seedlings; (b) virginile plant

Figure S2. Pearson correlation coefficients between the eight predictors used in the final model, calculated from the occurrence and pseudo-absence points of the calibration data

Figure S3. Habitat suitability for *Arctium pallidivirens* predicted by the four algorithms, each averaged over the ten pseudo-absence sets: (A) GLM; (B) GAM; (C) random forest; (D) MaxEnt

Authors: Khikmatullo R. Ibragimov, Bekhruz Sh. Khabibullaev, Khabibullo F. Shomurodov, Natalya Yu. Beshko, Shukhrat M. Jumayev, Giyos X. Buxorov, Diyorjon X. Hamrayev, Mukhlisa B. Turopova, Feruza J. Raimova, Gayrat A. Sharifov

Data type: images

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://journal.asu.ru/biol/article/view/20769/17685

Supplementary material 2

Table S1. Spatial block cross-validation performance with 1-km (main analysis, 32 records) and 500-m (38 records) thinning thresholds

Authors: Khikmatullo R. Ibragimov, Bekhruz Sh. Khabibullaev, Khabibullo F. Shomurodov, Natalya Yu. Beshko, Shukhrat M. Jumayev, Giyos X. Buxorov, Diyorjon X. Hamrayev, Mukhlisa B. Turopova, Feruza J. Raimova, Gayrat A. Sharifov

Data type: table

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://journal.asu.ru/biol/article/view/20769/17686