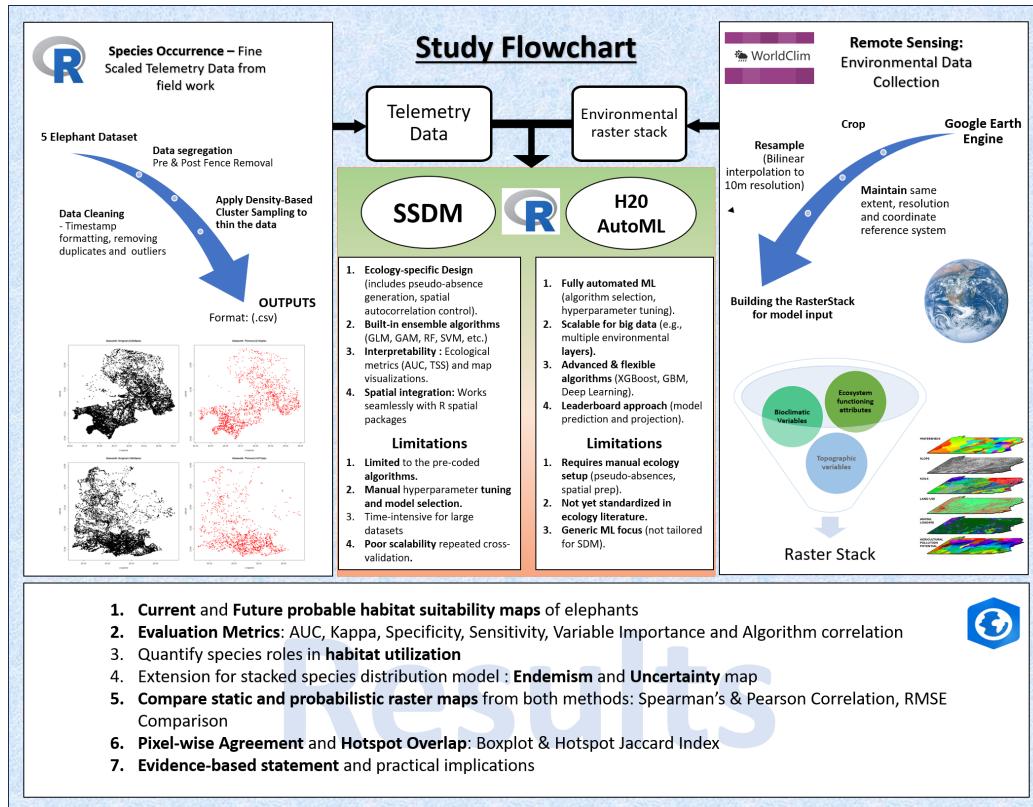


Graphical Abstract

From Ensemble Stacked SDMs to Automated Machine Learning for Predicting Elephant Habitat Suitability

Harin Aiyanna C. R.^{ID}, Brooke Friswold^{ID}, Antoinette van de Water^{ID}, Francesco Pirotti^{ID}



Highlights

From Ensemble Stacked SDMs to Automated Machine Learning for Predicting Elephant Habitat Suitability

Harin Aiyanna C. R., Brooke Friswold, Antoinette van de Water, Francesco Pirotti

- Automated Machine Learning improved accuracy and spatial detail of predictions when compared to traditional SSDMs.
- Habitat suitability shifted markedly (by up to 35%) after fence removal for individual elephants.
- Between-framework differences reveal uncertainty critical for conservation planning emphasizing greater efficiency

From Ensemble Stacked SDMs to Automated Machine Learning for Predicting Elephant Habitat Suitability

Harin Aiyanna C. R.^a, Brooke Friswold^{b,c,d}, Antoinette van de Water^{d,e}, Francesco Pirotti^{a,f}

^a*TESAF Department of Land, Environment, Agriculture, and Forestry, University of Padova, Padova, 35020, Veneto, Italy*

^b*Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, 32611, Florida, USA*

^c*Conservation Ecology Program, King Mongkut's University of Technology Thonburi, Bangkok, 10150, Thailand*

^d*Bring The Elephant Home, Steven Butendiekplateau 130, 3581 GS, Utrecht, The Netherlands*

^e*School of Life Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville, 32009, Pietermaritzburg, South Africa*

^f*CIRGEO Interdepartmental Research Center of Geomatics, University of Padova, Legnaro, 35020, Veneto, Italy*

Abstract

Automated machine learning offers new opportunities to model and predict the ecological trade-offs of fencing, revealing how structures that reduce human–wildlife conflict can also fragment habitats and undermine the connectivity essential for keystone species. This case study analyzes the impact of fencing removal on the suitability and utilization of elephant habitats in the Kariega Game Reserve, South Africa. We utilized fine-scale telemetry data in conjunction with high-resolution environmental covariates to evaluate two modeling methodologies: stacked ensemble species distribution models (SS-SDMs) and the H2O AutoML framework. To mitigate spatial autocorrelation, we implemented a density-based adaptive thinning algorithm (DBSCAN). Results show that the H2O AutoML approach consistently outperformed SS-SDMs ($AUC \approx 0.90$ vs. 0.85) in predictive accuracy, highlighting the potential of automated machine learning for spatial ecological modeling. While both frameworks agreed and captured sharper, spatially explicit habitat gradients at broad scales (Pearson's $r = 0.85\text{--}0.91$), AutoML revealed stronger post-fence expansion dynamics and finer hotspot delineation, underscoring how algorithmic flexibility shapes ecological inference. We also found predicted

range expansion from AutoML expanded by up to 35% than SSDM, reflecting significant increase in elephant movement post fence removal. Pixel-wise difference maps and Jaccard index on the upper quartile ($J_{0.75}$), was moderate (0.22–0.28) indicating potential methodological bias in identifying core areas. Most importantly, we addressed multiple sources of uncertainty, including data-driven (telemetry and thinning), algorithmic (AutoML vs. SSDM) and temporal (prediction stability). Finally, the top-ranked H2O AutoML model, selected cross-validated Area under the Receiver Operating Characteristic Curve (AUC), was projected into future climate scenarios to evaluate shifts in elephant habitat utilization. The findings not only highlight an interdisciplinary machine learning approach to inform landscape delineation for keystone species conservation, but addresses value of precise model with comprehensive benchmarking of AutoML frameworks against established species distribution modelling pipelines.

Keywords: Species Distribution Model, H2O AutoML, Habitat utilization, Species-Environment interactions, Fence removal

1. Introduction

The conservation of keystone species and the restoration of their habitats are essential for preserving ecological stability, biodiversity, and resilience [12]. In an era of accelerating environmental change, with their loss triggering cascading effects that fundamentally destabilize ecological networks and compromise essential ecosystem services [56, 49, 70]. Many keystone species are becoming increasingly vulnerable to climate change, land-use modification, and mounting human pressures [12, 40]. One of the most direct management challenges in southern Africa is the prevalence of wildlife fencing, established to keep wildlife within reserves and to reduce human–wildlife conflict. Fence removal - although central to the continuing debate on land sharing vs land sparing in human-wildlife coexistence, it may as well disrupt ecological connectivity for wide-ranging animals like elephants. [43, 52, 62, 27, 49].

Particularly keystone megafauna such as African elephants, epitomize the tension between the need for conservation and landscape modification for human wellbeing. [73, 28]. They serve as ecosystem engineers, influencing biodiversity and ecological processes via seed dissemination, canopy alteration, soil and water management, and even create micro habitats [40, 69]. Yet these biological functions are increasingly threatened by poaching, habitat

degradation, and climate change [50]. Over the past fifty years, the number of savanna elephants has dropped by around 70%, while the forest elephant populations have dropped by 90%, leading to an estimated continental reduction of 77% [12]. Simultaneously, atmospheric CO₂ concentrations have risen to 420 ppm, amplifying climate extremes and compounding pressures on habitat of keystone taxa [35, 59]. These combined elements make elephants not only a conservation priority but also a sensitive indicator species for assessing how landscape interventions affect ecosystem integrity.

To quantify such changes, species distribution models (SDMs) provide a key means of linking species' niche data with environmental variables. They help policymakers, reserve managers, and ecologists assess habitat suitability, predict spatial movements, and plan conservation strategies [13, 30]. High-resolution remote-sensing products such as Sentinel-2 and MODIS now supply continuous information on habitat data, vegetation, land cover, and human footprint. When integrated into SDMs, these datasets markedly improve predictive performance [1, 36, 14, 7, 44]. Hence leveraging machine learning to integrate environmental variables with species occurrence data yields geographically refined forecasts that can inform proactive conservation and connection strategies [71].

Traditionally, SDMs utilize techniques such generalized linear models (GLMs), generalized additive models (GAMs), random forests, maximum entropy (MaxEnt), or support vector machines (SVMs). Despite their efficacy, these algorithms require substantial model calibration, requires high computational demands, and are prone to collinearity, sampling bias, and overfitting [10, 11, 54, 67, 72]. Ensemble approaches, like those used in the SSDM package, help us understand the underlying problems by integrating multiple algorithms [41, 61]. Yet ensemble SDMs remain difficult to scale and standardize, limiting reproducibility in applied contexts. Several studies [16, 26, 34, 63] highlight the absence of systematic benchmarking or reporting standards, deriving model comparison and scalability challenging.

Here, automated machine learning (AutoML) offers a scalable alternative. By automating algorithm selection, hyperparameter optimization, and ensemble integration, AutoML frameworks enhance reproducibility and reduce user bias while maintaining predictive power [9, 17]. Early applications in engineering [51, 68] indicate that using AutoML can also streamline ecological modeling [24] but systematic benchmarks against established ensemble SDMs are scarce, especially for species-specific conservation problems [78, 39]. Given persistent challenges in SDM reproducibility and the absence of standardized

reporting protocols [78, 16], open-access code and transparent workflows are essential for enabling independent verification, methodological refinement, and collaborative advancement in conservation science [34, 39]. Addressing this gap is critical for assessing the ecological credibility of AutoML-derived outputs.

Relatively few studies have yet combined fine-scale telemetry with high-resolution environmental predictors for elephants, despite the potential of such integration [19, 25, 33]. Telemetry reveals movement and habitat use at the level of individual decisions, while remote sensing provides continuous landscape gradients. Together, they enable fine-grained SDMs that link elephant behavior directly to dynamic environmental conditions. In this study, we benchmark the performance of H2O AutoML [31] against traditional ensemble SDMs using elephant telemetry data from Kariega Game Reserve, South Africa. Specifically, we evaluate differences in predictive accuracy, computational efficiency, and spatial predictions, while examining how targeted fence removal influences habitat suitability and potential range expansion, thereby informing connectivity restoration strategies for elephant populations in fragmented landscapes [22, 48]. Through this approach, we propose a reproducible, scalable machine-learning framework that enhances predictive modeling for keystone species and provides actionable insight for conservation planning.

2. Methodology

2.1. Study Area

Telemetry data were collected at Kariega Game Reserve, Eastern Cape, South Africa, which spans approximately 10,000 hectares (115 km^2) along the Bushmans River. The reserve lies 9.5 km inland from the Indian Ocean and is dominated by dense, woody, and thorny semi-succulent vegetation characteristic of the Albany Thicket biome. Kariega supports a diverse assemblage of wildlife, including African elephant (*Loxodonta africana*), giraffe (*Giraffa camelopardalis*), cheetah (*Acinonyx jubatus*), white and black rhinoceros (*Ceratotherium simum* and *Diceros bicornis*), ostrich (*Struthio camelus*), and numerous avian and reptile species. At the start of this study, the elephant population (75 individuals) was divided by an internal fence into two sectors: Kariega West (KW; 2,792 ha; 55 elephants) and Harvestvale (HV; 4,369 ha; 20 elephants). Fence removal at Kariega occurred in two phases: in September 2022, HV was expanded by 1,031 ha following the first internal fence

removal. A second removal in January 2024 eliminated the primary barrier between KW and HV, creating a contiguous 8,192 ha landscape accessible to the entire elephant population [22] (Fig. ??).(fig 1).

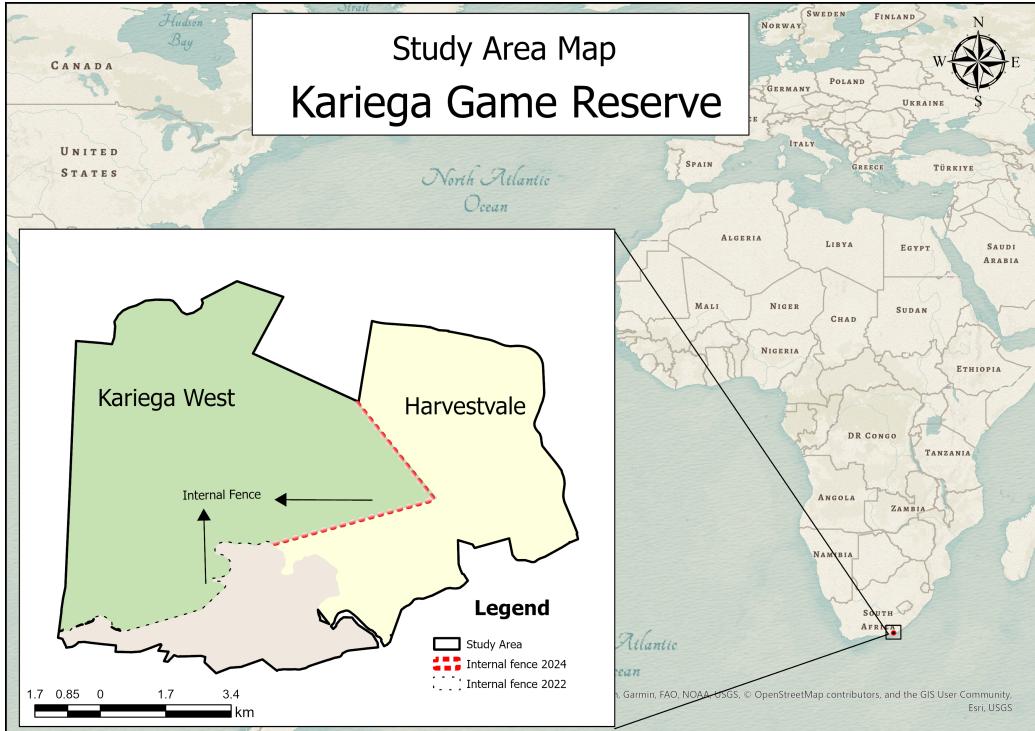


Figure 1: Study Area Map.

2.2. Data Collection and Pre-processing

To monitor spatial responses to habitat expansion, six African elephants—three solitary adult bulls and three matriarchs—were fitted with XL LoRa GPS collars (African Wildlife Tracking) in August 2022. Collars recorded locations every 30 minutes, enabling detailed analysis of daily movement patterns before and after fence removal. Data were transmitted remotely through the LoRa system and managed via EarthRanger software [74].

The telemetry dataset was subsequently cleaned and quality controlled. Duplicate timestamps and erroneous points were removed, and records were standardized across individuals. Encompassing a three-year duration, the conclusive dataset consisted of 127,717 GPS coordinates. Five datasets

documented elephants before the removal of the fence (09-01-2022 to 31-12-2023), whereas the remaining three monitored matriarchs subsequent to the fence’s removal (01-01-2024 to 02-12-2025) (Table 1). Data were exported as CSV files containing species, collar ID, timestamp, longitude, latitude, season, and geometry attributes. For analysis, all locations were grouped by collar ID and partitioned into pre- and post-fence removal periods (cut-off: 1 January 2024).

The datasets of six elephants (E1, E2, E3, E4, E5, E6) contained large record counts (Table 1), which were categorized into two temporal groups: before fence removal (B) and after fence removal (A). Two of the matriarchs (E3 and E4) originated from the Kariega West (KW) section of the reserve, which was smaller and more densely populated (55 elephants), while the matriarch E5 was from the Harvestvale (HV) section, which was larger and less densely populated (20 elephants). E2B contained complete records per timestamp, while the rest of the dataset included 1–2 NA timestamp values (likely due to minor time formatting issues). Additionally, the GPS collar deployed on the matriarch in HV (“E5A: Beauty”) experienced device malfunction four months post-fence removal, producing intermittent fixes that yielded a substantially smaller dataset ($n = 1,121$ points) but still spanning a full year (Jan. 2024–Feb. 2025). We obtained the initial datasets spanning from early 2022 to late 2025, followed by categorizing them into “before 2024” (B files) and “after 2024” (A files). Despite these challenges, each file still contained a consistently large number of records, ensuring robust construction of habitat utilization maps and temporal species distribution models (SDMs).

2.3. Data Thinning

To account for the high density and spatial autocorrelation of GPS fixes, we applied density-based spatial thinning using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [32]. DBSCAN is a widely used clustering method in machine learning and data mining that is well suited for ecological applications because it can identify clusters of arbitrary shape while distinguishing noise and outliers [23]. Unlike traditional approaches such as k -means or hierarchical clustering, which require an *a priori* specification of cluster number, DBSCAN defines clusters based on the density of points within a neighbourhood radius, making it particularly effective for irregular spatial distributions of animal movement data.

In the context of species distribution models (SDMs), retaining local density is crucial because predictive accuracy depends on preserving the spatial

Table 1: Dataset summary showing elephant IDs, record counts, and temporal coverage before and after fence removal.

Dataset	ID	Count	Earliest	Latest
Before Jan 2024				
E1B	Kamva	12,364	09-01-2022 00:07	04-12-2023 23:40
E2B	Kambaku	10,364	09-01-2022 00:03	02-12-2023 23:40
E3B	Bukela	23,816	09-01-2022 00:07	31-12-2023 23:44
E4B	Half Moon	23,808	09-01-2022 00:23	31-12-2023 23:50
E5B	Beauty	18,629	09-01-2022 00:09	31-12-2023 23:56
E6B	Balu	21,259	09-02-2022 09:44	29-11-2023 17:45
After Jan 2024				
E3A	Bukela	25,091	01-01-2024 01:14	02-12-2025 23:54
E4A	Half Moon	24,999	01-01-2024 01:21	02-12-2025 23:56
E5A	Beauty	1,846	01-01-2024 01:57	14-02-2025 23:56

structure of occurrence data relative to environmental covariates. Feeding unthinned GPS data directly into models risks inflating spatial autocorrelation, biasing estimates, and reducing computational efficiency, especially when environmental predictors are available at coarser resolutions than the movement data. This issue was particularly relevant for our telemetry dataset, which contained very high-frequency location fixes (30-minute intervals from September 2022 to February 2025) but was occasionally irregular due to network coverage gaps around the reserve.

We implemented DBSCAN using the `dbSCAN` package in R. Four parameters controlled the clustering and thinning process: (i) the neighbourhood radius (`eps_value` = 0.004 degrees, approximately 400 m); (ii) the minimum number of points required to define a dense cluster (`minPts_value` = 10); (iii) the fraction of points retained from each cluster (`fraction_value` = 0.2); and (iv) the minimum number of samples enforced for small clusters (`min_samples_value` = 4).

The fraction and minimum-sample parameters ensured that thinning preserved representative subsets even in sparsely used areas while reducing the computational burden of dense clusters. This strategy balanced the need to down-sample a very high-density dataset with the requirement to retain spatial heterogeneity and cluster-specific density patterns. The thinning

procedure was applied separately to each collared individual (E1B, E2B, E3B, E4B, E5B, E3A, E4A, and E5A), both before and after fence removal. The resulting reduced datasets maintained the core spatial structure of elephant movement patterns while eliminating redundant records and noise, thereby enhancing the robustness and interpretability of subsequent SDM analyses (Figure A.21).

2.4. Environmental Data Preparation

We assembled a suite of 26 environmental predictors from multiple high-resolution sources to inform the species distribution models (SDMs). Vegetation indices (NDVI, EVI) were derived from Sentinel-2 imagery (10 m resolution), while topographic variables (elevation, slope, aspect) were extracted from the SRTM DEM (30 m). Land cover was characterized using the Global Forest Change and ESA WorldCover products (10–30 m). To capture climatic gradients, we incorporated 19 bioclimatic variables (BIO1–BIO19) from WorldClim at 1 km resolution, representing annual trends, seasonality, and extreme values of temperature and precipitation. Index formulas and topographic derivatives followed standard remote-sensing conventions, ensuring comparability across predictors. To effectively represent the ecological terrain orientation of the Southern Hemisphere, we converted the aspect to radians to produce the cosine-aspect variable (“northness”). Together, these variables provided a multiscale representation of habitat conditions relevant to elephant movement and resource use in the Kariega landscape [7, 59, 76]. To reduce multicollinearity, we evaluated pairwise associations among predictors using Kendall’s τ correlation [8, 64]. A correlation matrix was constructed for all environmental variables, and predictors exhibiting high inter-correlation ($r \geq 0.8$) were excluded. This selection ensured that the retained predictor set was both parsimonious and informative for model development.

2.5. Species Distribution Modeling Frameworks

2.5.1. Stacked Ensemble SDM (SSDM)

The traditional ensemble species distribution modeling (SDM) workflow was implemented using the SSDM R package [61]. This framework constructs habitat suitability models by combining multiple predefined algorithms, including generalized linear models (GLMs), generalized additive models (GAMs), multivariate adaptive regression splines (MARS), generalized boosted regression models (GBMs), classification trees (CTAs), random forests (RFs), artificial neural networks (ANNs), and support vector machines (SVMs).

Pseudoabsence points were generated internally to balance presence-only data, and models were trained using holdout cross-validation (75% training, 25% testing). Predictive performance was evaluated with standard metrics: the Area Under the Receiver Operating Characteristic Curve (AUC), Cohen’s κ [8], sensitivity (true positive rate), and specificity (true negative rate) [15, 21]. By integrating multiple algorithms, SSDM ensembles reduce algorithm-specific biases and provide consensus predictions that are spatially explicit and directly applicable to conservation planning under climate and land-use change [49, 29, 6].

2.5.2. *H2O AutoML Workflow*

To evaluate a fully automated alternative, we applied H2O AutoML using the `h2o` R package [17, 18]. Equal numbers of pseudoabsence points were generated to ensure balanced training data and comparability with SSDM outputs. Environmental predictors were assembled into raster stacks, from which point-level values were extracted for each elephant dataset. Eight individual-specific models were built, capturing fine-scale variation in habitat use. The AutoML framework automatically trained a suite of base learners, including Gradient Boosting Machines (GBMs), eXtreme Gradient Boosting (XGBoost), Deep Learning neural networks, Distributed Random Forest (DRF), and Extremely Randomized Trees (XRT). These base models were integrated into a stacked ensemble via a meta-learner, which weighted contributions by cross-validated AUC to maximize predictive performance [17]. This adaptive ensembling process mitigates algorithm-specific biases and dynamically tailors the model composition to dataset complexity. Variable importance was derived for individual base learners and aggregated across ensembles to quantify the relative contribution of predictors. Spatial predictions were generated in two stages: (i) current habitat suitability maps using present-day environmental conditions, and (ii) future projections under four Shared Socioeconomic Pathways (SSP126, SSP245, SSP370, SSP585) for the periods 2021–2040, 2041–2060, 2061–2080, and 2081–2100. This enabled assessment of both present-day suitability and potential long-term shifts in elephant habitat under climate change.

2.6. *Agreement Assessment Between Frameworks*

Following established protocols and recommendations from [13, 39, 42, 46, 57, 78], we evaluated internal consistency, between-method spatial agreement, and temporal change mapping with categorical transitions. Using

individual-based model replicates, we developed a standardized workflow to compare SSDM and H2O AutoML outputs. Predictions were harmonized to a common resolution, extent, and coordinate system prior to comparison. Agreement was quantified using pixel-level similarity metrics, including root mean square error (RMSE), Pearson’s correlation coefficient (r), and Spearman’s rank correlation (ρ). Core hotspot overlap was assessed with the Jaccard index, calculated on the upper quartile (top 25%) of predicted suitability values. In addition, raster differences and agreement maps were generated to visualize areas of concordance and divergence between frameworks. This analysis exposes methodological biases, highlighting how the choice of modeling pipeline influences conservation inferences under dynamic ecological and climatic conditions. For two habitat-suitability rasters, S_{H2O} and S_{SSDM} , aligned on the same grid, pixel-wise correlation, deviation, and overlap metrics were computed.

2.7. Replicates and summary layers

For each individual \times period \times method, three replicate models were fit. We derived per-pixel replicate summaries: mean μ , standard deviation σ , coefficient of variation $\text{CV} = \sigma/(\mu + \varepsilon)$, and stability ϕ_t (fraction of replicates exceeding threshold t). These diagnose internal consistency and inform map interpretation.

2.8. Between-method agreement for all datasets

To evaluate consistency between modeling pipelines, SSDM and H2O AutoML outputs were compared on the same spatial grid. Pixel-wise agreement was quantified using Pearson’s correlation coefficient (r), Spearman’s rank correlation (ρ), root-mean-square error (RMSE), and mean absolute error (MAE).
:

$$r_{\text{Pearson}} = \frac{\sum_i (S_{\text{H2O},i} - \bar{S}_{\text{H2O}})(S_{\text{SSDM},i} - \bar{S}_{\text{SSDM}})}{\sqrt{\sum_i (S_{\text{H2O},i} - \bar{S}_{\text{H2O}})^2} \sqrt{\sum_i (S_{\text{SSDM},i} - \bar{S}_{\text{SSDM}})^2}}, \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_i (S_{\text{H2O},i} - S_{\text{SSDM},i})^2}, \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_i |S_{\text{H2O},i} - S_{\text{SSDM},i}|. \quad (3)$$

Spatial difference maps ($\Delta_{\text{method}} = S_{\text{H2O}} - S_{\text{SSDM}}$) were generated to visualize local divergence between modeling frameworks. Hotspot overlap was quantified using the Jaccard index [58] on the upper-quartile masks (i.e., Intersection over Union).

$$J_{0.75} = \frac{\sum(H_{\text{H2O}} \wedge H_{\text{SSDM}})}{\sum(H_{\text{H2O}} \vee H_{\text{SSDM}})}, \quad H = [S > Q_{0.75}(S)], \quad (4)$$

This approach highlights how pipeline choice influences the delineation of priority cores, defined as the top 25% of suitability values. The upper-quartile threshold ($Q_{0.75}$) isolates high-suitability “core” habitats for Jaccard comparison, minimizing noise from marginal regions. Additional thresholds ($Q_{0.25}, Q_{0.50}, Q_{0.75}$) were applied to evaluate how model agreement and area change vary from broad potential range ($Q_{0.25}$) to optimal core habitat ($Q_{0.75}$).

2.9. Temporal change analysis (E3, E4, E5)

Following established transition matrix methodology [55], temporal change was evaluated within each modeling framework for individuals with data spanning both periods (pre- and post-fence removal)

$$\Delta_{A-B} = S_A - S_B, \quad (5)$$

where S_A denotes the habitat suitability map *after* fence removal, while S_B represents the corresponding map *before* fence removal. The pixel-wise difference $\Delta_{A-B} = S_A - S_B$ thus quantifies the direction and magnitude of temporal change in suitability. Positive values of Δ_{A-B} indicate increased suitability following fence removal, negative values represent declines, and values near zero correspond to spatial stability. This continuous change surface captures both the intensity and spatial distribution of post-fence habitat shifts. Thereby using replicate means for each period, we computed the continuous difference and summarised categorical transitions at thresholds $t \in \{0.25, 0.50, 0.75\}$. Let $A_t = [S_A \geq t]$, $B_t = [S_B \geq t]$. We defined Gain, Stable and Loss classes as

$$\text{Gain}_t = (A_t \wedge \neg B_t), \quad \text{Stable}_t = (A_t \wedge B_t), \quad \text{Loss}_t = (B_t \wedge \neg A_t), \quad (6)$$

where \wedge denotes logical AND (intersection) and \neg denotes logical NOT (complement); thus, Gain, Stable, and Loss classes respectively represent

newly suitable, persistently suitable, and no-longer suitable habitats between periods (GSL maps). Then reported temporal overlap and area change as

$$J_t = \frac{\sum(A_t \wedge B_t)}{\sum(A_t \vee B_t)}, \quad \% \Delta \text{Area}_t = \frac{\sum A_t - \sum B_t}{\sum B_t + \varepsilon} \times 100. \quad (7)$$

Gain/Stable/Loss (GSL) maps, Δ_{A-B} rasters, Jaccard-vs-threshold curves and % area-change summaries provide complementary, threshold-explicit views of post-fence responses. For each dataset we export base suitability maps for both methods, between-method difference maps, and CSV summaries of agreement metrics. For E3, E4 and E5 we additionally export Δ_{A-B} rasters, GSL maps at the three thresholds, and temporal metrics tables. A panel composer assembles per-elephant dashboards (Before/After maps, Δ , GSL strip, Jaccard values J_t and $\% \Delta \text{Area}$) per elephant and method.

2.10. Reproducibility

All analyses were conducted in R (v4.3) using `terra`, `SSDM`, and `h2o`, with parallelisation on a multi-core workstation. Rasters were aligned via reprojection, bilinear resampling for continuous layers and nearest-neighbour for binary outputs. Code paths, thresholds and file conventions are documented in the github repository [2] to ensure end-to-end reproducibility.

3. Results

Across eight telemetry datasets (E1B, E2B, E3B, E4B, E5B; E3A, E4A, E5A), we benchmarked species distribution models (SDMs) derived from H2O AutoML against stacked SDMs (SSDM). For elephants with both before-after coverage (E3, E4, and E5), temporal change was quantified as $\Delta_{A-B} = S_A - S_B$ (After minus Before). Post-fence removal results indicated that differences in reserve size and population density likely influenced habitat use, as two herds originated from the smaller, denser Kariega West (KW) section and one from the larger, less dense Harvestvale (HV) section. Results are presented separately for both KW and HV regions of the game reserve, including continuous suitability surfaces, thresholded maps ($Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$), and composite figures that consolidate spatial predictions and summary metrics.

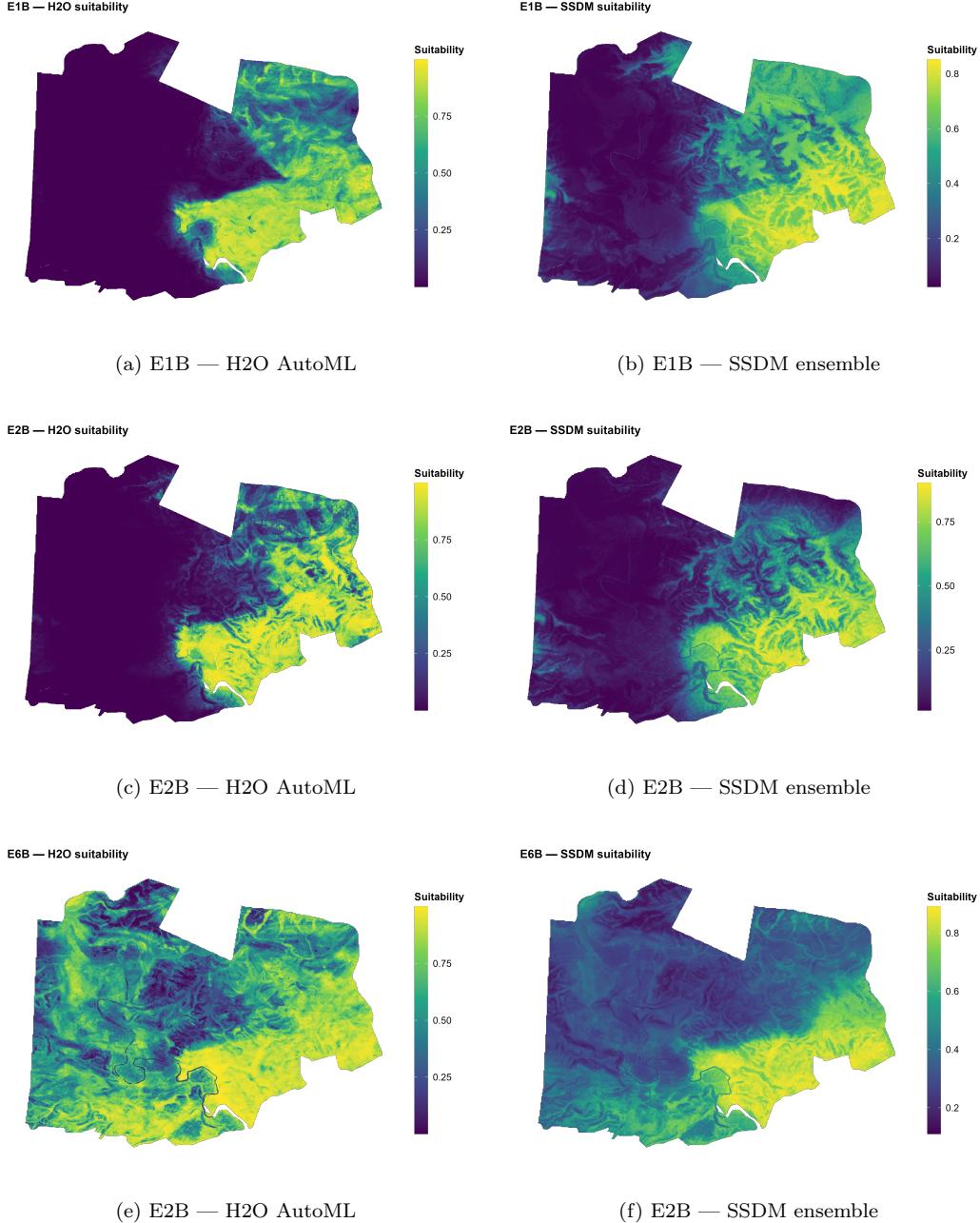


Figure 2: Pre-fence removal habitat suitability maps for male elephants E1B (“Kambaku”), E2B (“Kamva”), and E6B (“Balu”). Side-by-side comparisons display H2O AutoML (left) and SSDM (right) predictions at 1 km resolution. Warmer colors indicate higher suitability (0 = low, 1 = high). Both frameworks identify high-suitability core areas in Harvestvale and Kariega West, though with differences in spatial detail and magnitude (see Table ??)

3.1. Model discrimination and between-method agreement

Across all datasets, H2O AutoML and SSDM exhibited high overall correspondence in predicted suitability, yet with systematic differences in magnitude and spatial detail (Table ??). Pixel-wise correlations were consistently strong (Pearson’s $r = 0.806\text{--}0.907$; Spearman’s $\rho = 0.856\text{--}0.939$), indicating broad agreement in large-scale suitability gradients. However, local discrepancies remained non-negligible, as reflected by RMSE values ranging from 0.13–0.20 and MAE values between 0.10–0.19. Hotspot overlap, measured by the Jaccard index on the upper quartile ($J_{0.75}$), was moderate (0.515–0.722), indicating variable but generally substantial core-habitat agreement despite high global correlations. Jaccard values showed greater variability among matriarchs (0.515–0.641) compared to bulls (0.625–0.722), potentially reflecting increased habitat heterogeneity following fence removal. Spatial difference maps ($S_{\text{H2O}} - S_{\text{SSDM}}$) revealed that most divergences occurred along transitional zones and mixed-cover mosaics, where subtle environmental gradients amplify model-specific sensitivities.

Table 2: Between-method agreement metrics comparing H2O AutoML and SSDM predictions across all datasets.

Dataset	Pearson r	Spearman ρ	RMSE	MAE	Jaccard$_{0.75}$
E1B	0.866	0.902	0.179	0.133	0.625
E2B	0.896	0.903	0.168	0.116	0.722
E3B	0.838	0.937	0.202	0.138	0.678
E4B	0.849	0.939	0.194	0.132	0.659
E5B	0.907	0.889	0.136	0.106	0.571
E6B	0.806	0.872	0.222	0.190	0.702
E3A	0.852	0.856	0.140	0.114	0.515
E4A	0.883	0.890	0.130	0.106	0.628
E5A	0.898	0.895	0.183	0.120	0.641

3.2. Environmental drivers of suitability

Both modeling frameworks converged on precipitation and temperature as the primary drivers of elephant habitat suitability, particularly high precipitation seasonality (BIO15), precipitation of the warmest quarter (BIO16), and maximum temperature of the warmest month (BIO5). Post-fence datasets

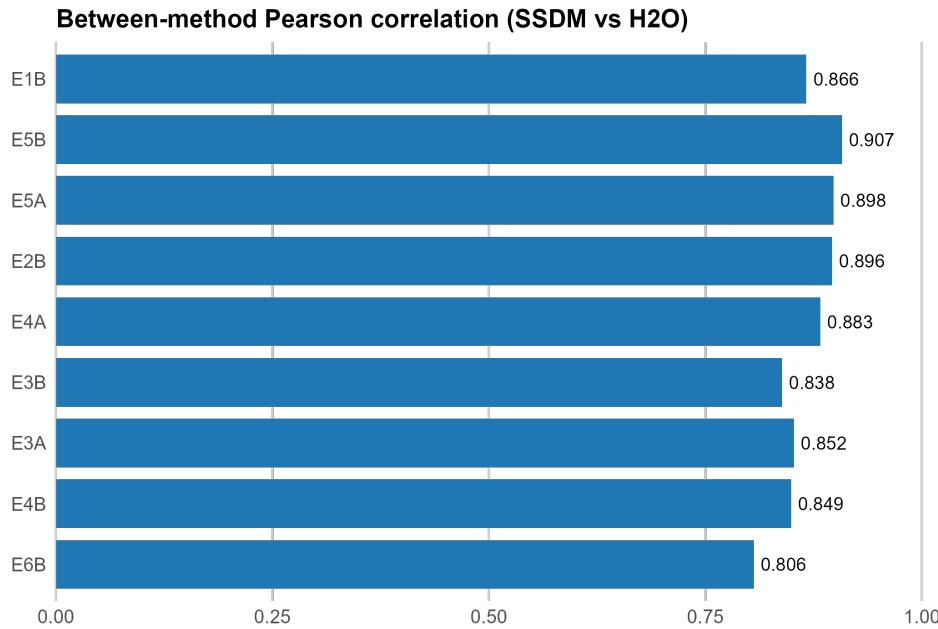


Figure 3: Pearson correlation between H2O AutoML and SSDM predictions across datasets.

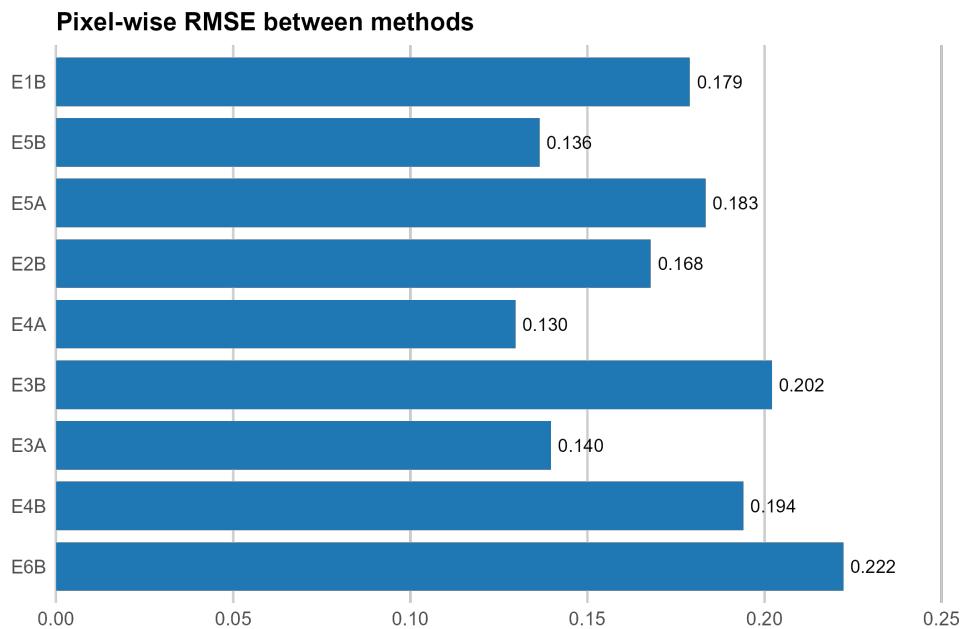


Figure 4: Root Mean Square Error (RMSE) comparing H2O AutoML and SSDM outputs.

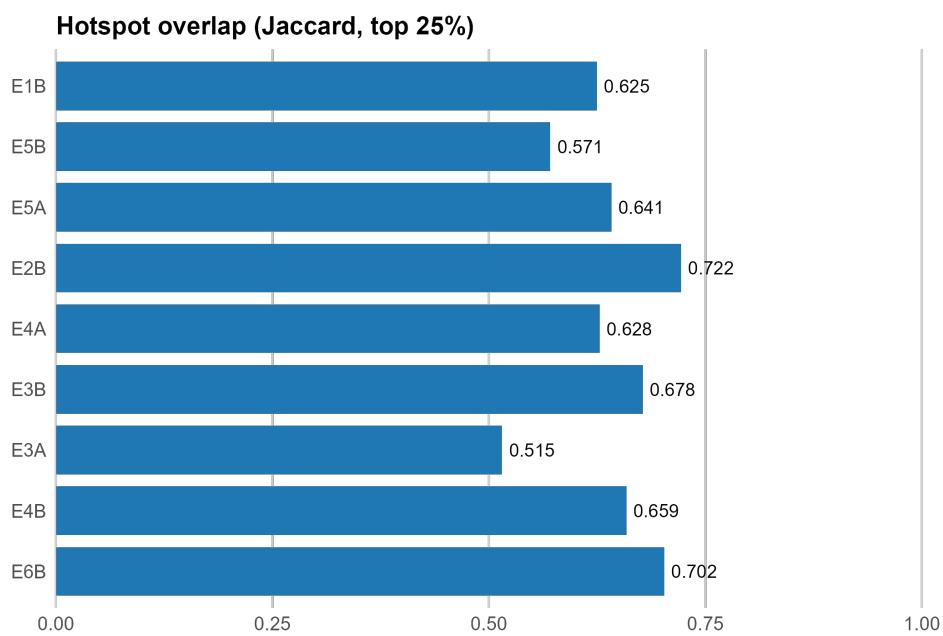


Figure 5: Hotspot Jaccard similarity (upper quartile, Q75) between H2O AutoML and SSDM predictions.

exhibited increased importance of vegetation productivity indices (NDVI and EVI). Rank correlation of variable importance between frameworks exceeded 0.85 across datasets (Fig. A.12), indicating strong consistency in the predictors most influential to habitat suitability, despite spatial divergence in their corresponding predictions.

3.3. Temporal change following fence removal (E3, E4 & E5)

Across elephants E3, E4, and E5, both frameworks revealed strong threshold-dependent differences in habitat area and temporal change (Table 3), though with systematic framework divergence. At the permissive threshold ($t = 0.25$), both H2O AutoML and SSDM predicted substantial habitat expansion following fence removal, with percentage increases of 52–70% for SSDM and 55–64% for H2O, indicating broad agreement on post-removal habitat gain. As thresholds became stricter ($t = 0.50$ and $t = 0.75$), suitable area decreased monotonically and framework divergence increased markedly.

Notably, SSDM predicted complete habitat loss (-100% , $A_{\text{area}} = 0 \text{ km}^2$) at $t = 0.75$ for E3 (“Bukela”) and E4 (“Half-Moon”), with Jaccard indices of 0.00 indicating no temporal overlap of high-suitability cores. In contrast, H2O retained substantial core habitat at the strictest threshold, with Jaccard values ranging from 0.03 (E3) to 0.77 (E5A), indicating greater predicted persistence of high-suitability zones. For E5A (“Beauty”), the frameworks diverged most sharply: SSDM predicted progressive habitat contraction across all thresholds (-25% to -41%), whereas H2O identified net persistence despite modest area reduction. These framework differences highlight how model algorithm, ensemble structure, and extrapolation strategy influence temporal habitat projections - with H2O generally providing more optimistic (higher gain, greater persistence) and SSDM more conservative (sharper decline, lower stability) forecasts.

The Jaccard similarity indices (J_t) quantified spatial overlap between *Before* and *After* habitat maps at three suitability thresholds ($t = 0.25, 0.50, 0.75$) for each elephant and modeling framework. Across all individuals with before–after datasets (E3, E4, and E5), H2O AutoML consistently showed higher temporal overlap than SSDM, particularly at moderate thresholds ($t = 0.25–0.50$). For instance, $J_{0.5}$ averaged 0.41–0.81 for H2O but only 0.11–0.57 for SSDM, indicating that the automated ensemble produced more temporally coherent suitability patterns. At the strictest threshold ($t = 0.75$), overlap declined sharply for SSDM (often $J_t = 0$), reflecting minimal persistence

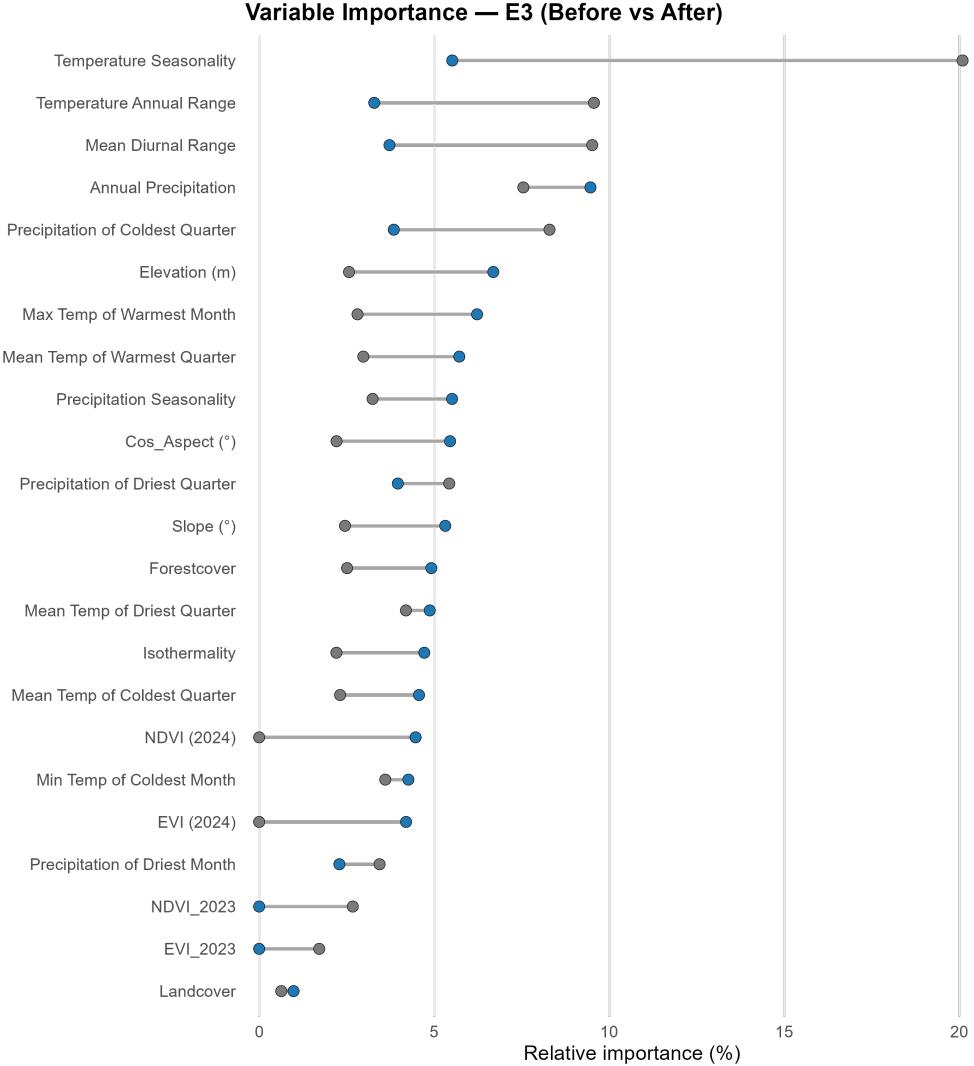


Figure 6: Elephant E3 (“Bukela”). Dumbbell plots display the relative contribution (%) of each environmental predictor to H2O AutoML and SSDM habitat suitability models in the pre-fence removal period (gray circles; before 2024) and post-fence removal period (blue circles; after 2024). Horizontal line spans indicate the magnitude of change in predictor importance between periods, with leftward shifts representing decreased importance and rightward shifts representing increased importance. The distance between circles illustrates how fence removal altered the relative weight of environmental drivers in explaining habitat-use patterns. Variable importance rankings remained strongly correlated (Spearman’s $\rho > 0.85$) between frameworks, indicating agreement on dominant predictors despite differences in spatial detail. Figures for Elephant E4 (“Half-Moon”) and E5A (“Beauty”) are provided in the Appendix.

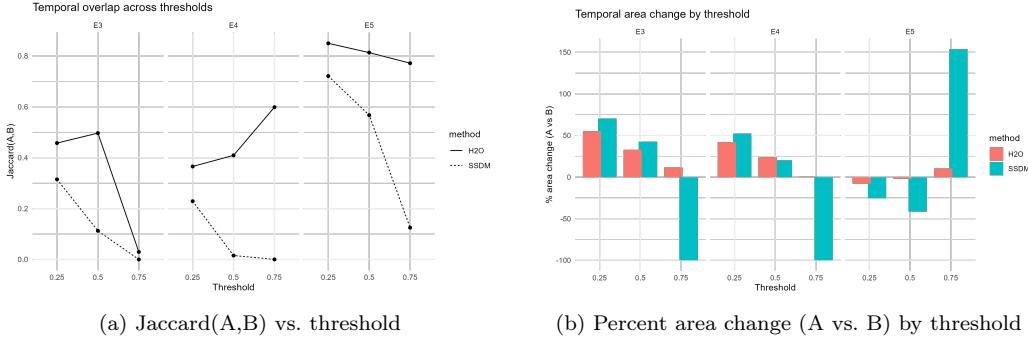


Figure 7: Temporal robustness: overlap and area change across thresholds (0.25, 0.50, 0.75) for E3, E4 and E5 within each framework.

of high-suitability zones through time. In contrast, H2O retained moderate persistence (e.g., $J_{0.75} = 0.03$ for E3; E4: strong persistence, $J = 0.60$, $A = 41.43 \text{ km}^2$; E5: very strong persistence, $J = 0.77$, $A = 52.67 \text{ km}^2$), suggesting greater stability of predicted core habitats. Composite panels integrate base maps, Δ maps, gain–stability–loss (GSL) strips, and metric tables (Figs. 8–A.10–A.11).

Overall, we observe three consistent patterns. (i) *Discrimination*: H2O AutoML outperformed SSDM across all datasets, achieving lower error and higher AUC. (ii) *Spatial agreement*: Global correlations were high, but hotspot overlap remained moderate and spatial difference maps revealed fine-scale divergence, particularly in transitional zones. (iii) *Temporal change*: Both frameworks detected post-fence habitat expansion, with H2O offering a more sensitive lens (sharper Δ , larger gains) and SSDM providing a conservative baseline (smoother cores, higher stability). Together, these perspectives delineate the plausible uncertainty envelope for conservation interpretation and highlight the complementary insights of automated versus ensemble SDM approaches.

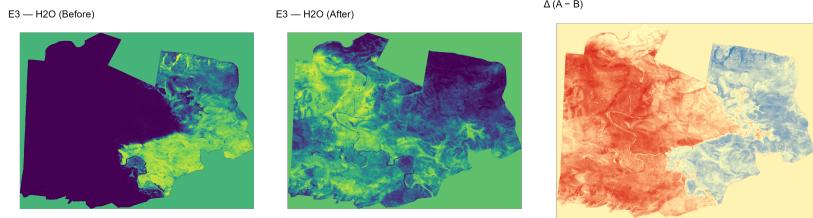
3.4. Synthesis

We observe three patterns in our results. First, *discrimination*: H2O outperformed SSDM across all datasets, with lower error and higher AUC. Second, *spatial agreement*: global correlations are high, but hotspot overlap is modest and difference maps reveal fine-scale divergence, especially in transitional zones. Third, *temporal change*: both frameworks detect post-fence expansion, with H2O offering a more sensitive lens (sharper Δ , larger

Table 3: Temporal change metrics (Before–After) for elephants E3, E4, and E5 across three suitability thresholds ($t \in \{0.25, 0.5, 0.75\}$) and two modelling frameworks (SSDM and H2O AutoML). Threshold t is the suitability cutoff. B_{area} and A_{area} are suitable habitat areas *Before* and *After* fence removal, respectively; $\% \Delta \text{ Area} = \frac{A_{\text{area}} - B_{\text{area}}}{B_{\text{area}}} \times 100\%$. J_t is the Jaccard similarity between Before and After maps (0–1). All areas are in km^2 .

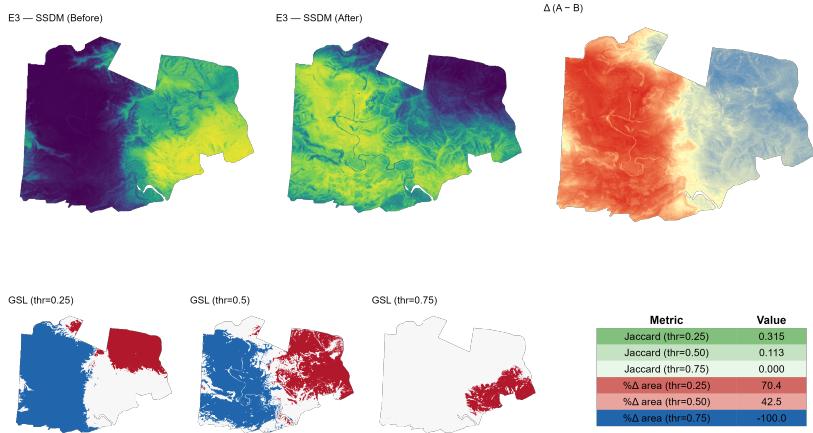
ID	Method	t	B_{area} (km^2)	A_{area} (km^2)	Loss (km^2)	Stable (km^2)	Gain (km^2)	$\% \Delta$ (%)	J_t
E3	SSDM	0.25	45.66	77.82	16.09	29.57	48.26	70.44	0.31
E3	SSDM	0.50	32.53	46.37	24.55	7.98	38.39	42.54	0.11
E3	SSDM	0.75	10.80	0.00	10.80	0.00	0.00	-100.00	0.00
E3	H2O	0.25	71.34	110.74	14.18	57.16	53.58	55.23	0.46
E3	H2O	0.50	62.90	83.54	14.31	48.59	34.95	32.82	0.50
E3	H2O	0.75	11.43	12.76	10.74	0.70	12.06	11.58	0.03
E4	SSDM	0.25	45.89	69.94	24.32	21.57	48.37	52.41	0.23
E4	SSDM	0.50	33.89	40.74	32.76	1.14	39.60	20.21	0.02
E4	SSDM	0.75	8.60	0.00	8.60	0.00	0.00	-100.00	0.00
E4	H2O	0.25	72.94	103.57	25.69	47.25	56.32	42.00	0.37
E4	H2O	0.50	64.79	80.28	22.67	42.13	38.16	23.91	0.41
E4	H2O	0.75	55.12	55.53	13.69	41.43	14.10	0.74	0.60
E5	SSDM	0.25	49.05	36.56	13.19	35.86	0.70	-25.46	0.72
E5	SSDM	0.50	38.71	22.68	16.49	22.22	0.46	-41.40	0.57
E5	SSDM	0.75	2.31	5.87	1.40	0.91	4.96	153.58	0.12
E5	H2O	0.25	89.48	82.50	10.48	79.00	3.50	-7.80	0.85
E5	H2O	0.50	74.11	72.65	8.29	65.82	6.83	-1.96	0.81
E5	H2O	0.75	57.45	63.48	4.78	52.67	10.81	10.50	0.77

Elephant E3 — H₂O | Before vs After: base maps, Δ (A–B), GSL strip, metrics



(a) E3 — H₂O AutoML

Elephant E3 — SSDM | Before vs After: base maps, Δ (A–B), GSL strip, metrics



(b) E3 — SSDM

Figure 8: Per-elephant comparison for Bukela (E3). The upper panel shows habitat suitability maps before and after fence removal, with a difference map $\Delta = A - B$ indicating increases (red) and decreases (blue) in suitability. The lower panel presents threshold-specific spatial changes: blue denotes areas that became suitable (Gain), red indicates areas that became unsuitable (Loss), and white shows areas that remained suitable (Stable) after the intervention. The accompanying table summarizes model metrics—Jaccard similarity (spatial overlap at each threshold) and percentage change in suitable area (% Δ area). All areas are in km²

gains) and SSDM a conservative baseline (smoother cores, higher stability). Together these perspectives bracket the plausible uncertainty envelope for conservation interpretation.

4. Discussion

5. Discussion and Conservation Implications

This study demonstrates that H2O AutoML consistently outperformed the traditional SSDM (Stacked Species Distribution Model) framework across multiple performance metrics in elephant habitat suitability modeling. AutoML achieved superior discrimination (higher ROC–AUC values), calibration (lower RMSE and log-loss), and computational efficiency, revealing a stronger capacity to capture nonlinear environmental responses than traditional ensemble SDMs. AutoML reflects its ability to dynamically optimize model architecture and hyperparameters for dataset-specific complexity [18, 31]. Unlike traditional ensemble approaches [61] that use fixed weighting schemes, AutoML adaptively reweighted base learners to match the inherent structure of each elephant’s movement data. GBM- and XRT-dominated ensembles best captured Bukela’s habitat patterns, while Half-Moon’s variability required deeper neural network components, and Beauty’s smaller sample was adequately represented by simpler GBM architecture. This adaptive capacity, combined with clear variable importance outputs, maintained both predictive accuracy and ecological interpretability [26]. These findings align with emerging evidence that machine learning automation enhances predictive power and scalability in wildlife conservation applications [71, 24], while its parallelized architecture [9] substantially reduced training time, highlighting the practical utility of automated pipelines for large-scale or multispecies conservation studies.

Addressing collinearity among environmental predictors is fundamental for reliable inference in habitat suitability modeling [10, 11]. While conventional approaches such as variance inflation factor (VIF) analysis or principal component analysis (PCA) are widely used, both have significant drawbacks: VIF is sensitive to sample size and scaling, whereas PCA often obscures ecological meaning by collapsing variables into abstract components that lack biological interpretability [11]. Therefore, this study employed Kendall’s τ rank correlation as a robust, non-parametric method for variable screening [16]. By retaining ecologically interpretable predictors rather than statistical surrogates, we preserved the biological interpretability of model outputs while

maintaining predictive performance (see Appendix Fig. 8). Spatial block cross-validation [60, 72] confirmed that this variable selection approach maintained predictive stability across spatially independent partitions (Figs.A.14, A.16), demonstrating that model reliability was not compromised by the reduced predictor set. This methodological synergy proved particularly valuable: by pre-screening variables for ecological interpretability, AutoML could focus its automated feature selection on optimizing among biologically meaningful predictors rather than navigating statistically transformed surrogates [18, 71]. This integration resulted in models that balanced predictive accuracy with transparency—ensuring a critical requirement for conservation decision-making where stakeholder trust depends on interpretable outputs [26].

The contrast between pre- and post-fence datasets provides compelling evidence of the conservation value of restoring landscape connectivity. Before fence removal, elephants occupied restricted, fragmented habitats, whereas the post-removal period revealed clear expansions into newly accessible terrain. This pattern mirrors broader evidence that fences constrain movement and erode long-term viability of wide-ranging species [40, 48], while providing quantitative support for habitat reconnection efforts. Site-specific differences in reserve configuration (pre-fence removal: KW and HV) likely shaped initial patterns of habitat use, with localized population densities (KW: 55 elephants; HV: 20 elephants) and spatial heterogeneity strongly mediating connectivity gains and behavioral shifts following fence removal [62]. The post-fence habitat gains identified here (Table 3) provide quantitative evidence for the conservation value of reconnecting fragmented landscapes. Individual elephant responses to fence removal revealed distinct patterns of habitat expansion and exploration. Matriarch Bukela (E3A) demonstrated the most pronounced expansion into newly accessible terrain (Fig. 8), suggesting that experienced individuals may be more effective at identifying and utilizing restored connectivity. AutoML models maintained stronger temporal consistency for Bukela, preserving continuous high-suitability patches even under strict thresholds, while SSDM tended to fragment predicted suitable areas.

Half-Moon (E4) displayed a more heterogeneous pattern, with sharp mosaics of gain and loss immediately after fence removal, reflecting exploratory use of newly available space typical of initial post-barrier investigations in large mammals [52, 48]. Beauty (E5) showed modest but spatially coherent expansion along existing corridors, retaining stable habitat cores across thresholds. This behavior suggests that Beauty’s herd had already established optimal habitat selection within suitable areas of the reserve, with expansion

primarily following established movement patterns rather than exploring entirely new areas, consistent with documented corridor fidelity in elephant populations [75, 4]. Notably, habitat suitability expansion for both Bukela and Half-Moon was directed toward Harvestvale, indicating that restored connectivity enabled access to previously inaccessible high-quality habitat patches. These patterns collectively highlight the importance of assessing habitat change across multiple probability thresholds to effectively identify both expansion fronts and persistent refugia [60, 72].

Despite AutoML’s superior predictive accuracy, it exhibited limited spatial congruence with SSDM, sharing only 17–28% of high-suitability areas. This divergence is consistent with previous benchmarking studies showing that algorithmic ensembles often differ in spatial predictions even when their discrimination scores are comparable [20, 27]. Such divergence reflects differences in model architecture and sensitivity to environmental gradients rather than outright predictive error, where each framework highlights different aspects of habitat suitability based on its underlying assumptions and algorithmic approach. The difference maps ($\Delta = S_{\text{AutoML}} - S_{\text{SSDM}}$; Fig. 3.7) reveal that H2O AutoML precisely delineated fence boundaries, showing sharp transitions between red and blue areas that correspond to the physical barrier location, while SSDM provided more moderate spatial separation. This pattern highlights AutoML’s enhanced sensitivity to complex ecological discontinuities and its ability to capture fine-scale habitat boundaries that may be critical for conservation planning. Relying on a single modeling framework risks over- or underestimating critical habitat areas, whereas cross-framework comparisons expose the boundaries of predictive confidence. The spatial divergence observed here underscores the value of ensemble approaches that incorporate multiple algorithmic perspectives, particularly when making high-stakes conservation decisions about habitat protection or corridor placement [3, 61].

Across all models, precipitation and temperature variables (Figs. A.12 and A.12a,A.12b) emerged as dominant predictors of elephant habitat suitability, consistent with the ecology of water-dependent megafauna that require reliable access to surface water and thermally suitable foraging areas [45, 4]. This finding aligns with established understanding of elephant resource requirements and validates the biological realism of both modeling approaches. Following fence removal, vegetation indices such as NDVI and EVI gained prominence in variable importance rankings, indicating that elephants rapidly exploited newly accessible, high-productivity areas. This shift suggests that access to quality forage became a more influential factor in habitat selection

once movement barriers were removed, allowing elephants to optimize resource acquisition across a broader landscape [75].

Forward projections under different climate scenarios revealed contrasting habitat trajectories through 2100. Under low-emission scenarios (SSP1–2.6), suitable habitat remained relatively stable, suggesting that current conservation areas and connectivity corridors may continue to support elephant populations under moderate climate change. However, high-emission pathways (SSP5–8.5) predicted substantial habitat loss, consistent with projected climate refugia contractions for African elephants [49] and broader patterns of climate-driven range loss for large mammals [53]. These projections emphasize the critical importance of maintaining and expanding habitat connectivity now, as future climate conditions may further constrain suitable areas. The connectivity gains achieved through fence removal in this study may prove essential for enabling climate-driven range shifts and maintaining population viability under changing environmental conditions.

While this study demonstrates AutoML’s potential for ecological modeling and provides clear evidence of connectivity restoration benefits, several limitations constrain broader generalization and highlight areas for future research. After the first phase of fence removal, collared bulls E1B and E2B concentrated within Harvestvale while E6B explored widely (Fig. 3.1), and each suitability map revealed persistent patterns along remaining internal fence boundaries, indicating incomplete connectivity restoration. Only the complete 2024 removal enabled the substantial habitat expansions observed in matriarchs Bukela and Half-Moon, demonstrating that incremental connectivity improvements may be insufficient to restore landscape-scale movement for elephant populations. For instance, individual-level variation in responses for Bukela’s extensive habitat utilization compared to Beauty’s corridor-following behavior—suggests that conservation strategies must account for behavioral diversity within populations, as average population responses may not serve all individuals. Operational challenges further limited data completeness: all three bulls experienced collar twisting within six months of deployment, attributed to dense vegetation and design flaws, ultimately necessitating collar removal and resulting in loss of male movement data post-2024 [22]. Combined with the small sample size (six collared elephants, reduced to three post-fence), these constraints restrict understanding of population-level responses to connectivity restoration and limit broader generalization. The short monitoring period may not capture longer-term behavioral adaptations or seasonal variations in habitat use patterns. Although, regardless of algo-

rithmic sophistication, correlative models cannot fully capture the behavioral or physiological processes underlying habitat selection [5, 47]. Short-term disturbances such as droughts, fires, or anthropogenic pressures may significantly alter habitat suitability patterns but remain inadequately represented with only abiotic environmental predictors. Hence, reliance on WorldClim data provides global consistency but may also limit spatial precision relative to higher-resolution climate products such as CHELSA [37] or Copernicus datasets.

Future research should address these limitations by integrating mechanistic approaches such as resource selection functions (RSFs) or step-selection functions (SSFs) with AutoML frameworks to capture biological realism of process-based models [5, 19, 33]. Dynamic connectivity models [26, 77] and energetically informed dispersal kernels [66] could provide more realistic representations of movement costs and corridor functionality. Use of multimodal datasets [65] as well as incorporating remote sensing time series, acoustic monitoring, genetic sampling, and human–wildlife conflict data could enable more comprehensive, policy-relevant predictions [71, 29, 38]. The computational efficiency demonstrated by AutoML in this study suggests significant potential for scaling to landscape, regional, or continental analyses. Future work can explore transferability across different keystone taxa, habitat types, and geographic regions to develop more generalizable conservation tools. Critical to this advancement is embracing open science principles: the automated nature of AutoML frameworks offers opportunities for improved reproducibility through open-source code sharing, benchmarked validation protocols, and transparent reporting of hyperparameter optimization processes [78, 34].

The practical conservation implications of our findings are immediate and urgent. The quantified habitat gains following fence removal (Table 3) provide strong empirical support for connectivity restoration as a conservation strategy. Having our code open-sourced [2] leverages AutoML’s precise delineation of expansion areas to inform targeted protection efforts, anti-poaching patrol routes, and human–wildlife conflict mitigation measures in newly utilized habitats. This becomes particularly critical given climate change projections (Fig. 4.1) indicating substantial habitat loss under high-emission scenarios (SSP5–8.5), which emphasize the urgency of current connectivity restoration efforts [49, 53]. Habitat corridors established through fence removal may prove essential for facilitating climate-driven range shifts and maintaining population viability under future environmental conditions. Finally, the spatial divergence between AutoML and SSDM modeling frameworks highlights the importance

of using multiple approaches for high-stakes conservation decisions, with management plans incorporating uncertainty estimates and considering the range of predictions across different modeling approaches rather than relying on single-model outputs. By combining AutoML’s scalability with mechanistic ecological understanding and multimodal data integration, future research can advance toward adaptive, data-rich conservation strategies that reflect both ecological complexity and practical management constraints while maintaining the transparency and reproducibility essential for science-based conservation policy.

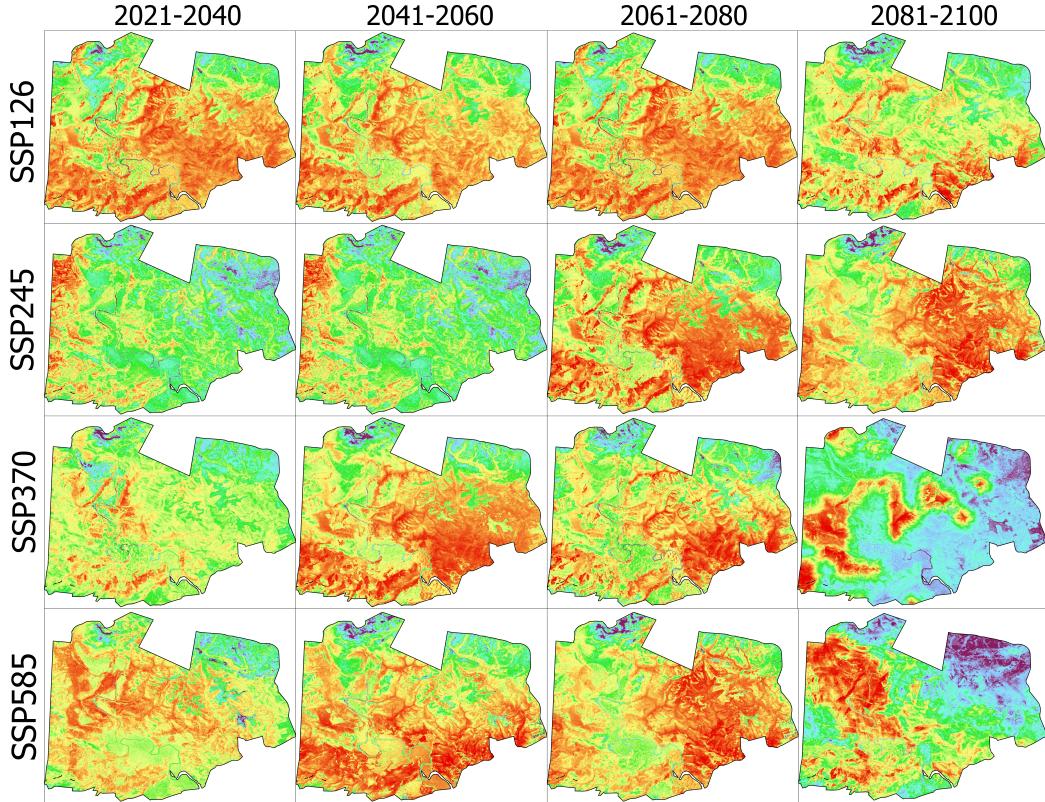


Figure 9: Future projection for Bukela-E3A.

6. Conclusion

This study demonstrates that AutoML represents a significant advancement in habitat suitability modeling for conservation applications, offering

superior predictive performance, computational efficiency, and ecological interpretability compared to traditional ensemble approaches. The framework’s ability to adaptively optimize model architecture while maintaining biological interpretability makes it particularly valuable for species-specific conservation planning. The quantified benefits of fence removal provide compelling evidence for landscape connectivity restoration as an effective conservation strategy, with individual elephants rapidly exploiting newly accessible, high-quality habitats. However, the spatial divergence between modeling frameworks and the projected impacts of climate change underscore the importance of adaptive management approaches that incorporate multiple sources of uncertainty. Looking forward, the integration of AutoML with mechanistic models, multimodal datasets, and open-source frameworks offers a pathway toward more robust, scalable, and policy-relevant conservation tools. As automated machine learning capabilities continue to evolve, their application to conservation challenges promises to enhance our ability to protect biodiversity in an increasingly complex and rapidly changing world. The success of fence removal in this system provides a template for similar connectivity restoration efforts globally, while the methodological advances demonstrated here offer tools for more effective conservation planning across species and ecosystems. Most importantly, this work illustrates how cutting-edge computational approaches can be harnessed to address urgent conservation challenges, providing both immediate management guidance and a foundation for adaptive, evidence-based conservation strategies.

References

- [1] Emiliano Agrillo, Federico Filippini, Alice Pezzarossa, Laura Casella, Daniela Smiraglia, Arianna Orasi, Fabio Attorre, and Andrea Taramelli. “Earth Observation and Biodiversity Big Data for Forest Habitat Types Classification and Mapping”. In: *Remote Sensing* 13.7 (Mar. 2021), p. 1231. ISSN: 2072-4292. DOI: 10.3390/rs13071231. URL: <http://dx.doi.org/10.3390/RS13071231>.
- [2] Harin Aiyanna. *MySpeciesSDM: R code for species distribution modeling with H2O AutoML*. R scripts. Accessed: 2025-10-28. 2025. URL: https://github.com/Caspian-Aiyanna/ems_paper.git.
- [3] Miguel Araújo and Mark New. “Ensemble forecasting of species distributions”. In: *TREE* 22 (Jan. 2007), pp. 42–47.

- [4] Gil Bohrer, Pieter SA Beck, Shadrack M Ngene, Andrew K Skidmore, and Ian Douglas-Hamilton. “Elephant movement closely tracks precipitation-driven vegetation dynamics in a Kenyan forest-savanna landscape”. In: *Movement Ecology* 2.1 (Jan. 2014). ISSN: 2051-3933. DOI: 10.1186/2051-3933-2-2. URL: <http://dx.doi.org/10.1186/2051-3933-2-2>.
- [5] Mark S Boyce, Pierre R Vernier, Scott E Nielsen, and Fiona K.A Schmiegelow. “Evaluating resource selection functions”. In: *Ecological Modelling* 157.2–3 (Nov. 2002), pp. 281–300. ISSN: 0304-3800. DOI: 10.1016/S0304-3800(02)00200-4. URL: [http://dx.doi.org/10.1016/S0304-3800\(02\)00200-4](http://dx.doi.org/10.1016/S0304-3800(02)00200-4).
- [6] Harin Aiyanna Cheriyanda Raveendra, Lorenzo Picco, and Francesco Pirotti. “Using Earth observation to support the predictive accuracy of species distribution models in ecological restoration: a case study of Poland beavers (*Castor fiber*)”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-3-2024 (Nov. 2024), pp. 63–70. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-xlviii-3-2024-63-2024. URL: <http://dx.doi.org/10.5194/ISPRS-ARCHIVES-XLVIII-3-2024-63-2024>.
- [7] Hyeyeong Choe, Junhwa Chi, and James H. Thorne. “Mapping Potential Plant Species Richness over Large Areas with Deep Learning, MODIS, and Species Distribution Models”. In: *Remote Sensing* 13.13 (June 2021), p. 2490. ISSN: 2072-4292. DOI: 10.3390/rs13132490. URL: <http://dx.doi.org/10.3390/rs13132490>.
- [8] Jacob Cohen. “A Coefficient of Agreement for Nominal Scales”. In: *Educational and Psychological Measurement* 20.1 (Apr. 1960), pp. 37–46. ISSN: 1552-3888. DOI: 10.1177/001316446002000104. URL: <http://dx.doi.org/10.1177/001316446002000104>.
- [9] Felix Conrad, Mauritz Mälzer, Michael Schwarzenberger, Hajo Wiemer, and Steffen Ihlenfeldt. “Benchmarking AutoML for regression tasks on small tabular data in materials design”. In: *Scientific Reports* 12.1 (Nov. 2022). ISSN: 2045-2322. DOI: 10.1038/s41598-022-23327-1. URL: <http://dx.doi.org/10.1038/s41598-022-23327-1>.
- [10] Paulo De Marco and Caroline Corrêa Nóbrega. “Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation”. In: *PLOS ONE* 13.9 (Sept. 2018). Ed. by

- Luciano Bosso, e0202403. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0202403. URL: <http://dx.doi.org/10.1371/journal.pone.0202403>.
- [11] Carsten F. Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J. Leitão, Tamara Münkemüller, Colin McClean, Patrick E. Osborne, Björn Reineking, Boris Schröder, Andrew K. Skidmore, Damaris Zurell, and Sven Lautenbach. “Collinearity: a review of methods to deal with it and a simulation study evaluating their performance”. In: *Ecography* 36.1 (May 2012), pp. 27–46. ISSN: 1600-0587. DOI: 10.1111/j.1600-0587.2012.07348.x. URL: <http://dx.doi.org/10.1111/j.1600-0587.2012.07348.x>.
 - [12] Charles T. T. Edwards, Kathleen S. Gobush, Fiona Maisels, Dave Balfour, Russell Taylor, and George Wittemyer. “Survey-based inference of continental African elephant decline”. In: *Proceedings of the National Academy of Sciences* 121.48 (Nov. 2024). ISSN: 1091-6490. DOI: 10.1073/pnas.2403816121. URL: <http://dx.doi.org/10.1073/pnas.2403816121>.
 - [13] Jane Elith and John R. Leathwick. “Species Distribution Models: Ecological Explanation and Prediction Across Space and Time”. In: *Annual Review of Ecology, Evolution, and Systematics* 40.1 (Dec. 2009), pp. 677–697. ISSN: 1545-2069. DOI: 10.1146/annurev.ecolsys.110308.120159. URL: <http://dx.doi.org/10.1146/ANNUREV.ECOLSYS.110308.120159>.
 - [14] Joaquim Estopinan, Maximilien Servajean, Pierre Bonnet, François Munoz, and Alexis Joly. “Deep Species Distribution Modeling From Sentinel-2 Image Time-Series: A Global Scale Analysis on the Orchid Family”. In: *Frontiers in Plant Science* 13 (Apr. 2022). ISSN: 1664-462X. DOI: 10.3389/fpls.2022.839327. URL: <http://dx.doi.org/10.3389/fpls.2022.839327>.
 - [15] Tom Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8 (June 2006), pp. 861–874. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2005.10.010. URL: <http://dx.doi.org/10.1016/j.patrec.2005.10.010>.

- [16] Xiao Feng, Daniel S. Park, Ye Liang, Ranjit Pandey, and Monica Papes. “Collinearity in ecological niche modeling: Confusions and challenges”. In: *Ecology and Evolution* 9.18 (Aug. 2019), pp. 10365–10376. ISSN: 2045-7758. DOI: 10.1002/ece3.5555. URL: <http://dx.doi.org/10.1002/ece3.5555>.
- [17] Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. “Auto-sklearn 2.0: The Next Generation”. In: *Proceedings of the AutoML Conference 2020*. 2020. URL: https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf.
- [18] Matthias Feurer and Frank Hutter. “Hyperparameter Optimization”. In: *Automated Machine Learning*. Springer International Publishing, 2019, pp. 3–33. ISBN: 9783030053185. DOI: 10.1007/978-3-030-05318-5_1. URL: http://dx.doi.org/10.1007/978-3-030-05318-5_1.
- [19] John Fieberg, Johannes Signer, Brian Smith, and Tal Avgar. “A ‘How to’ guide for interpreting parameters in habitat-selection analyses”. In: *Journal of Animal Ecology* 90.5 (Mar. 2021), pp. 1027–1043. ISSN: 1365-2656. DOI: 10.1111/1365-2656.13441. URL: <http://dx.doi.org/10.1111/1365-2656.13441>.
- [20] Joern Fischer and David B. Lindenmayer. “Landscape modification and habitat fragmentation: a synthesis”. In: *Global Ecology and Biogeography* 16.3 (Feb. 2007), pp. 265–280. ISSN: 1466-8238. DOI: 10.1111/j.1466-8238.2007.00287.x. URL: <http://dx.doi.org/10.1111/j.1466-8238.2007.00287.x>.
- [21] Joseph L. Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical Methods for Rates and Proportions*. Wiley, Sept. 2003. ISBN: 9780471445425. DOI: 10.1002/0471445428. URL: <http://dx.doi.org/10.1002/0471445428>.
- [22] Brooke Friswold, Brett Mitchell, George Gale, and Antoinette Van de Water. “Twisting collars on male elephants in shrub terrain: animal welfare considerations for researchers, managers and manufacturers”. In: *Pachyderm* 64 (Nov. 2023), pp. 78–91. ISSN: 1026-2881. DOI: 10.69649/pachyderm.v64i.534. URL: <http://dx.doi.org/10.69649/pachyderm.v64i.534>.

- [23] Matthias Fuchs and Wolfram Höpken. “Clustering: Hierarchical, k-Means, DBSCAN”. In: *Applied Data Science in Tourism*. Springer International Publishing, 2022, pp. 129–149. ISBN: 9783030883898. DOI: 10.1007/978-3-030-88389-8_8. URL: http://dx.doi.org/10.1007/978-3-030-88389-8_8.
- [24] Max Gaber, Yanghui Kang, Guy Schurges, and Trevor Keenan. “Using automated machine learning for the upscaling of gross primary productivity”. In: *Biogeosciences* 21.10 (May 2024), pp. 2447–2472. ISSN: 1726-4189. DOI: 10.5194/bg-21-2447-2024. URL: <http://dx.doi.org/10.5194/bg-21-2447-2024>.
- [25] Richard A. Giliba, Christian Kiffner, Pascal Fust, and Jacqueline Loos. “Modelling elephant corridors over two decades reveals opportunities for conserving connectivity across a large protected area network”. In: *PLOS ONE* 18.10 (Oct. 2023). Ed. by Daniel de Paiva Silva, e0292918. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0292918. URL: <http://dx.doi.org/10.1371/journal.pone.0292918>.
- [26] Eric Gilman and Milani Chaloupka. “Evidence from interpretable machine learning to inform spatial management of Palau’s tuna fisheries”. In: *Ecosphere* 15.2 (Feb. 2024). ISSN: 2150-8925. DOI: 10.1002/ecs2.4751. URL: <http://dx.doi.org/10.1002/ecs2.4751>.
- [27] Ingo Grass, Jacqueline Loos, Svenja Baensch, Péter Batáry, Felipe Librán-Embid, Anoush Ficiciyan, Felix Klaus, Maraja Riechers, Julia Rosa, Julia Tiede, Kristy Udy, Catrin Westphal, Annemarie Wurz, and Teja Tscharntke. “Land-sharing/-sparing connectivity landscapes for ecosystem services and biodiversity conservation”. In: *People and Nature* 1.2 (Apr. 2019), pp. 262–272. ISSN: 2575-8314. DOI: 10.1002/pan3.21. URL: <http://dx.doi.org/10.1002/pan3.21>.
- [28] Michael Gross. “Of elephants and men”. In: *Current Biology* 34.2 (Jan. 2024), R37–R40. ISSN: 0960-9822. DOI: 10.1016/j.cub.2024.01.001. URL: <http://dx.doi.org/10.1016/j.cub.2024.01.001>.
- [29] Antoine Guisan, Reid Tingley, John B. Baumgartner, Ilona Naujokaitis-Lewis, Patricia R. Sutcliffe, Ayesha I. T. Tulloch, Tracey J. Regan, Lluis Brotons, Eve McDonald-Madden, Chrystal Mantyka-Pringle, Tara G. Martin, Jonathan R. Rhodes, Ramona Maggini, Samantha A. Setterfield, Jane Elith, Mark W. Schwartz, Brendan A. Wintle, Olivier Broennimann, Mike Austin, Simon Ferrier, Michael R. Kearney, Hugh P.

- Possingham, and Yvonne M. Buckley. “Predicting species distributions for conservation decisions”. In: *Ecology Letters* 16.12 (Oct. 2013). Ed. by Hector Arita, pp. 1424–1435. ISSN: 1461-0248. DOI: 10.1111/ele.12189. URL: <http://dx.doi.org/10.1111/ele.12189>.
- [30] Antoine Guisan and Niklaus E. Zimmermann. “Predictive habitat distribution models in ecology”. In: *Ecological Modelling* 135.2–3 (Dec. 2000), pp. 147–186. ISSN: 0304-3800. DOI: 10.1016/S0304-3800(00)00354-9. URL: [http://dx.doi.org/10.1016/S0304-3800\(00\)00354-9](http://dx.doi.org/10.1016/S0304-3800(00)00354-9).
- [31] H2O.ai. *H2O AutoML User Guide*. Accessed: 2024-11-30. 2024. URL: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html>.
- [32] Michael Hahsler, Matthew Piekenbrock, and Derek Doran. “dbSCAN: Fast Density-Based Clustering with R”. In: *Journal of Statistical Software* 91.1 (2019). ISSN: 1548-7660. DOI: 10.18637/jss.v091.i01. URL: <http://dx.doi.org/10.18637/JSS.V091.I01>.
- [33] Mark Hebblewhite and Daniel T. Haydon. “Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1550 (July 2010), pp. 2303–2312. ISSN: 1471-2970. DOI: 10.1098/rstb.2010.0087. URL: <http://dx.doi.org/10.1098/rstb.2010.0087>.
- [34] Maximilian H. K. Hesselbarth, Jakub Nowosad, Alida de Flamingh, Craig E. Simpkins, Martin Jung, Gemma Gerber, and Martí Bosch. “Computational Methods in Landscape Ecology”. In: *Current Landscape Ecology Reports* 10.1 (Dec. 2024). ISSN: 2364-494X. DOI: 10.1007/s40823-024-00104-6. URL: <http://dx.doi.org/10.1007/s40823-024-00104-6>.
- [35] Bärbel Hönißch, Dana L. Royer, Daniel O. Breecker, Pratigya J. Polissar, Gabriel J. Bowen, Michael J. Henehan, Ying Cui, Margret Steinthorsdóttir, Jennifer C. McElwain, Matthew J. Kohn, Ann Pearson, Samuel R. Phelps, Kevin T. Uno, Andy Ridgwell, Eleni Anagnostou, Jacqueline Austermann, Marcus P. S. Badger, Richard S. Barclay, Peter K. Bijl, Thomas B. Chalk, Christopher R. Scotese, Elwyn de la Vega, Robert M. DeConto, Kelsey A. Dyez, Vicki Ferrini, Peter J. Franks, Claudia F. Giulivi, Marcus Gutjahr, Dustin T. Harper, Laura L. Haynes, Matthew Huber, Kathryn E. Snell, Benjamin A. Keisling, Wilfried Konrad, Tim K.

- Lowenstein, Alberto Malinverno, Maxence Guillermic, Luz María Mejía, Joseph N. Milligan, John J. Morton, Lee Nordt, Ross Whiteford, Anita Roth-Nebelsick, Jeremy K. C. Rugenstein, Morgan F. Schaller, Nathan D. Sheldon, Sindia Sosdian, Elise B. Wilkes, Caitlyn R. Witkowski, Yi Ge Zhang, Lloyd Anderson, David J. Beerling, Clara Bolton, Thure E. Cerling, Jennifer M. Cotton, Jiawei Da, Douglas D. Ekart, Gavin L. Foster, David R. Greenwood, Ethan G. Hyland, Elliot A. Jagiecki, John P. Jasper, Jennifer B. Kowalczyk, Lutz Kunzmann, Wolfram M. Kürschner, Charles E. Lawrence, Caroline H. Lear, Miguel A. Martínez-Botí, Daniel P. Maxbauer, Paolo Montagna, B. David A. Naafs, James W. B. Rae, Markus Raitzsch, Gregory J. Retallack, Simon J. Ring, Osamu Seki, Julio Sepúlveda, Ashish Sinha, Tekie F. Tesfamichael, Aradhna Tripati, Johan van der Burgh, Jimin Yu, James C. Zachos, and Laiming Zhang. “Toward a Cenozoic history of atmospheric CO₂”. In: *Science* 382.6675 (Dec. 2023). ISSN: 1095-9203. DOI: 10.1126/science.adi5177. URL: <http://dx.doi.org/10.1126/science.adi5177>.
- [36] Louis Jochems, Jodi Brandt, Clayton Kingdon, Samuel J. Schurkamp, Andrew Monks, and Shane C. Lishawa. “Active remote sensing data and dispersal processes improve predictions for an invasive aquatic plant during a climatic extreme in Great Lakes coastal wetlands”. In: *Journal of Environmental Management* 370 (Nov. 2024), p. 122610. ISSN: 0301-4797. DOI: 10.1016/j.jenvman.2024.122610. URL: <http://dx.doi.org/10.1016/j.jenvman.2024.122610>.
- [37] Dirk Nikolaus Karger, Olaf Conrad, Jürgen Böhner, Tobias Kawohl, Holger Kreft, Rodrigo Wilber Soria-Auza, Niklaus E. Zimmermann, H. Peter Linder, and Michael Kessler. “Climatologies at high resolution for the earth’s land surface areas”. In: *Scientific Data* 4.1 (Sept. 2017). ISSN: 2052-4463. DOI: 10.1038/sdata.2017.122. URL: <http://dx.doi.org/10.1038/sdata.2017.122>.
- [38] Melissa A Karp, Megan Cimino, J Kevin Craig, Daniel P Crear, Christopher Haak, Elliott L Hazen, Isaac Kaplan, Donald R Kobayashi, Hassan Moustahfid, Barbara Muhling, Malin L Pinsky, Laurel A Smith, James T Thorson, and Phoebe A Woodworth-Jefcoats. “Applications of species distribution modeling and future needs to support marine resource management”. In: *ICES Journal of Marine Science* 82.3 (Mar. 2025). Ed. by Anita Gilles. ISSN: 1095-9289. DOI: 10.1093/icesjms/fsaf024. URL: <http://dx.doi.org/10.1093/icesjms/fsaf024>.

- [39] Jamie M. Kass, Adam B. Smith, Dan L. Warren, Sergio Vignali, Sylvain Schmitt, Matthew E. Aiello-Lammens, Eduardo Arlé, Ana Márcia Barbosa, Olivier Broennimann, Marlon E. Cobos, Maya Guéguen, Antoine Guisan, Cory Merow, Babak Naimi, Michael P. Nobis, Ian Ondo, Luis Osorio-Olvera, Hannah L. Owens, Gonzalo E. Pinilla-Buitrago, Andrea Sánchez-Tapia, Wilfried Thuiller, Rozbeh Valavi, Santiago José Elías Velazco, Alexander Zizka, and Damaris Zurell. “Achieving higher standards in species distribution modeling by leveraging the diversity of available software”. In: *Ecography* 2025.2 (Nov. 2024). ISSN: 1600-0587. DOI: 10.1111/ecog.07346. URL: <http://dx.doi.org/10.1111/ecog.07346>.
- [40] Minyi Kau, Byron V. Weckworth, Sheng Li, Mathias M. Pires, Daiying Jin, Michela Pacifici, Carlo Rondinini, Luigi Boitani, Thomas M. McCarthy, Zhi Lu, George B. Schaller, Steven R. Beissinger, and Juan Li. “Umbrella, keystone, or flagship? An integrated framework for identifying effective surrogate species”. In: *Biological Conservation* 303 (Mar. 2025), p. 111025. ISSN: 0006-3207. DOI: 10.1016/j.biocon.2025.111025. URL: <http://dx.doi.org/10.1016/j.biocon.2025.111025>.
- [41] R. Kindt. “Ensemble species distribution modelling with transformed suitability values”. In: *Environmental Modelling & Software* 100 (Feb. 2018), pp. 136–145. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2017.11.009. URL: <http://dx.doi.org/10.1016/J.ENVSOFT.2017.11.009>.
- [42] Kamil Konowalik and Agata Nosol. “Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage”. In: *Scientific Reports* 11.1 (Jan. 2021). ISSN: 2045-2322. DOI: 10.1038/s41598-020-80062-1. URL: <http://dx.doi.org/10.1038/s41598-020-80062-1>.
- [43] Claire Kremen. “Reframing the land-sparing/land-sharing debate for biodiversity conservation”. In: *Annals of the New York Academy of Sciences* 1355.1 (July 2015), pp. 52–76. ISSN: 1749-6632. DOI: 10.1111/nyas.12845. URL: <http://dx.doi.org/10.1111/nyas.12845>.
- [44] Annu Kumari and S. Karthikeyan. “Sentinel-2 Data for Land Use/Land Cover Mapping: A Meta-analysis and Review”. In: *SN Computer Science* 4.6 (Oct. 2023). ISSN: 2661-8907. DOI: 10.1007/s42979-023-02214-0. URL: <http://dx.doi.org/10.1007/s42979-023-02214-0>.

- [45] Scott R. Loarie, Rudi J. van Aarde, and Stuart L. Pimm. “Elephant seasonal vegetation preferences across dry and wet savannas”. In: *Biological Conservation* 142.12 (Dec. 2009), pp. 3099–3107. ISSN: 0006-3207. DOI: 10.1016/j.biocon.2009.08.021. URL: <http://dx.doi.org/10.1016/j.biocon.2009.08.021>.
- [46] Brian S. Maitner, Robert L. Richards, Ben Carlson, John M. Drake, and Cory Merow. “Flexible Methods for Species Distribution Modeling with Small Samples”. In: *bioRxiv* (Apr. 2025). DOI: 10.1101/2025.04.03.647025. eprint: <https://www.biorxiv.org/content/early/2025/04/08/2025.04.03.647025.full.pdf>. URL: <http://dx.doi.org/10.1101/2025.04.03.647025>.
- [47] Jason Matthiopoulos, John Fieberg, Geert Aarts, Hawthorne L. Beyer, Juan M. Morales, and Daniel T. Haydon. “Establishing the link between habitat selection and animal population dynamics”. In: *Ecological Monographs* 85.3 (Aug. 2015), pp. 413–436. ISSN: 1557-7015. DOI: 10.1890/14-2244.1. URL: <http://dx.doi.org/10.1890/14-2244.1>.
- [48] Dipanjan Naha, Stéphanie Périquet, J. Werner Kilian, Caitlin A. Kupferman, Tammy Hoth-Hanssen, and James C. Beasley. “Fencing affects movement patterns of two large carnivores in Southern Africa”. In: *Frontiers in Ecology and Evolution* 11 (June 2023). ISSN: 2296-701X. DOI: 10.3389/fevo.2023.1031321. URL: <http://dx.doi.org/10.3389/fevo.2023.1031321>.
- [49] Robin Naidoo, Cody Aylward, Wendy Elliott, Annika Keeley, Margaret Kinnaird, Michael Knight, Cristian-Remus Papp, Kanchan Thapa, and Rafael Antelo. “From science to impact: Conserving ecological connectivity in large conservation landscapes”. In: *Proceedings of the National Academy of Sciences* 122.31 (July 2025). ISSN: 1091-6490. DOI: 10.1073/pnas.2410937122. URL: <http://dx.doi.org/10.1073/pnas.2410937122>.
- [50] Simon Nampindo and Timothy O. Randhir. “Dynamic modeling of African elephant populations under changing climate and habitat loss across the Greater Virunga Landscape”. In: *PLOS Sustainability and Transformation* 3.1 (Jan. 2024). Ed. by Alka Bharat, e0000094. ISSN: 2767-3197. DOI: 10.1371/journal.pstr.0000094. URL: <http://dx.doi.org/10.1371/journal.pstr.0000094>.

- [51] Intisar Omar, Muhammad Khan, Andrew Starr, and Khaled Abou Rok Ba. “Automated Prediction of Crack Propagation Using H2O AutoML”. In: *Sensors* 23.20 (Oct. 2023), p. 8419. ISSN: 1424-8220. DOI: 10.3390/s23208419. URL: <http://dx.doi.org/10.3390/s23208419>.
- [52] Liudmila Osipova, Moses M. Okello, Steven J. Njumbi, Shadrack Ngene, David Western, Matt W. Hayward, and Niko Balkenhol. “Fencing solves human-wildlife conflict locally but shifts problems elsewhere: A case study using functional connectivity modelling of the African elephant”. In: *Journal of Applied Ecology* 55.6 (Aug. 2018). Ed. by Matthew Struebig, pp. 2673–2684. ISSN: 1365-2664. DOI: 10.1111/1365-2664.13246. URL: <http://dx.doi.org/10.1111/1365-2664.13246>.
- [53] Michela Pacifici, Piero Visconti, Stuart H. M. Butchart, James E. M. Watson, Francesca M. Cassola, and Carlo Rondinini. “Species’ traits influenced their response to recent climate change”. In: *Nature Climate Change* 7.3 (Feb. 2017), pp. 205–208. ISSN: 1758-6798. DOI: 10.1038/nclimate3223. URL: <http://dx.doi.org/10.1038/nclimate3223>.
- [54] Steven J. Phillips and Miroslav Dudík. “Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation”. In: *Ecography* 31.2 (Mar. 2008), pp. 161–175. ISSN: 1600-0587. DOI: 10.1111/j.0906-7590.2008.5203.x. URL: <http://dx.doi.org/10.1111/J.0906-7590.2008.5203.X>.
- [55] Robert Gilmore Pontius and Marco Millones. “Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment”. In: *International Journal of Remote Sensing* 32.15 (Aug. 2011), pp. 4407–4429. ISSN: 1366-5901. DOI: 10.1080/01431161.2011.552923. URL: <http://dx.doi.org/10.1080/01431161.2011.552923>.
- [56] Zak Ratajczak, Scott L. Collins, John M. Blair, Sally E. Koerner, Allison M. Louthan, Melinda D. Smith, Jeffrey H. Taylor, and Jesse B. Nippert. “Reintroducing bison results in long-running and resilient increases in grassland diversity”. In: *Proceedings of the National Academy of Sciences* 119.36 (Aug. 2022). ISSN: 1091-6490. DOI: 10.1073/pnas.2210433119. URL: <http://dx.doi.org/10.1073/pnas.2210433119>.
- [57] Armand Rausell-Moreno, Núria Galiana, Babak Naimi, and Miguel B. Araújo. “Improving species distribution models by optimising background points: Impacts on current and future climate projections”. In: *Ecological Modelling* 507 (Aug. 2025), p. 111177. ISSN: 0304-3800. DOI:

- 10.1016/j.ecolmodel.2025.111177. URL: <http://dx.doi.org/10.1016/j.ecolmodel.2025.111177>.
- [58] Raimundo Real and Juan M. Vargas. “The Probabilistic Basis of Jaccard’s Index of Similarity”. In: *Systematic Biology* 45.3 (Sept. 1996). Ed. by Richard Olmstead, pp. 380–385. ISSN: 1063-5157. DOI: 10.1093/sysbio/45.3.380. URL: <http://dx.doi.org/10.1093/sysbio/45.3.380>.
- [59] Federico Riva, Caroline Jean Martin, Carmen Galán Acedo, Erwan Nicolas Bellon, Petr Keil, Alejandra Morán-Ordóñez, Lenore Fahrig, and Antoine Guisan. “Incorporating effects of habitat patches into species distribution models”. In: *Journal of Ecology* 112.10 (Aug. 2024), pp. 2162–2182. ISSN: 1365-2745. DOI: 10.1111/1365-2745.14403. URL: <http://dx.doi.org/10.1111/1365-2745.14403>.
- [60] David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure”. In: *Ecography* 40.8 (Mar. 2017), pp. 913–929. ISSN: 1600-0587. DOI: 10.1111/ecog.02881. URL: <http://dx.doi.org/10.1111/ecog.02881>.
- [61] Sylvain Schmitt, Robin Pouteau, Dimitri Justeau, Florian de Boissieu, and Philippe Birnbaum. “<scp>ssdm</scp>: An <scp>r</scp> package to predict distribution of species richness and composition based on stacked species distribution models”. In: *Methods in Ecology and Evolution* 8.12 (Aug. 2017). Ed. by Nick Golding, pp. 1795–1803. ISSN: 2041-210X. DOI: 10.1111/2041-210x.12841. URL: <http://dx.doi.org/10.1111/2041-210X.12841>.
- [62] Imogen A. Schwandner, Thomas A. Morrison, J. Grant C. Hopcraft, Jake Wall, Lacey Hughey, Randall B. Boone, Joseph O. Ogutu, Andrew F. Jakes, Shem C. Kifugo, Campaign Limo, Stephen Ndambuki Mwiu, Vasco Nyaga, Han Olff, Gordon O. Ojwang, Wilson Sairowua, Jackson Sasine, Jully S. Senteu, Daniel Sopia, Jeffrey Worden, and Jared A. Stabach. “Predicting the impact of targeted fence removal on connectivity in a migratory ecosystem”. In: *Ecological Applications*

- 35.1 (Jan. 2025). ISSN: 1939-5582. DOI: 10.1002/eap.3094. URL: <http://dx.doi.org/10.1002/eap.3094>.
- [63] Austin M Smith, Cesar Capinha, and Andrew M Kramer. “Incorporating environmental time series into species distribution models”. In: (Oct. 2022). DOI: 10.1101/2022.10.26.513922. URL: <http://dx.doi.org/10.1101/2022.10.26.513922>.
 - [64] Robert H. Somers. “A New Asymmetric Measure of Association for Ordinal Variables”. In: *American Sociological Review* 27.6 (1962). JSTOR Stable URL <https://www.jstor.org/stable/2090408>, pp. 799–811. DOI: 10.2307/2090406. URL: <https://doi.org/10.2307/2090408>.
 - [65] Stylianos Stasinou, Martino Mensio, Elena Lazovik, and Athanasios Trantaf. *BioCube: A Multimodal Dataset for Biodiversity Research*. 2025. arXiv: 2505.11568 [q-bio.QM]. URL: <https://arxiv.org/abs/2505.11568>.
 - [66] Lauren L. Sullivan, Adam T. Clark, David Tilman, and Allison K. Shaw. “Mechanistically derived dispersal kernels explain species-level patterns of recruitment and succession”. In: *Ecology* 99.11 (Oct. 2018), pp. 2415–2420. ISSN: 1939-9170. DOI: 10.1002/ecy.2498. URL: <http://dx.doi.org/10.1002/ecy.2498>.
 - [67] Alexandra D. Syphard and Janet Franklin. “Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors”. In: *Ecography* 32.6 (Dec. 2009), pp. 907–918. ISSN: 1600-0587. DOI: 10.1111/j.1600-0587.2009.05883.x. URL: <http://dx.doi.org/10.1111/j.1600-0587.2009.05883.x>.
 - [68] Junchi Tian and Chang Che. “Automated Machine Learning: A Survey of Tools and Techniques”. In: *Journal of Industrial Engineering and Applied Science* 2.6 (Dec. 2024), pp. 71–76. ISSN: 3005-608X. DOI: 10.70393/6a69656173.323336. URL: <http://dx.doi.org/10.70393/6a69656173.323336>.
 - [69] Sérgio Timóteo, Jörg Albrecht, Beatriz Rumeu, Ana C. Norte, Anna Traveset, Carol M. Frost, Elizabete Marchante, Francisco A. López-Núñez, Guadalupe Peralta, Jane Memmott, Jens M. Olesen, José M. Costa, Luís P. da Silva, Luísa G. Carvalheiro, Marta Correia, Michael Staab, Nico Blüthgen, Nina Farwig, Sandra Hervías-Parejo, Sergei Mironov, Susana Rodríguez-Echeverría, and Ruben Heleno. “Tripartite

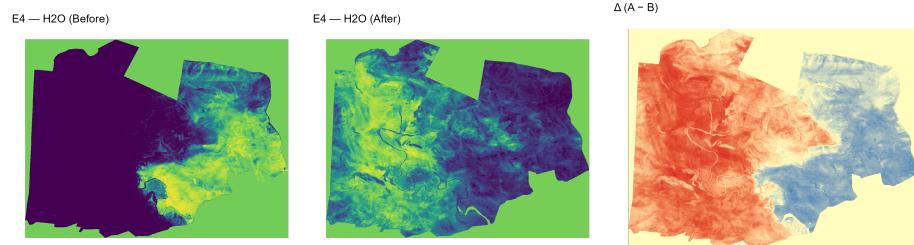
- networks show that keystone species can multitask”. In: *Functional Ecology* 37.2 (Oct. 2022), pp. 274–286. ISSN: 1365-2435. DOI: 10.1111/1365-2435.14206. URL: <http://dx.doi.org/10.1111/1365-2435.14206>.
- [70] Joseph A. Tobias, James M. Bullock, Lynn V. Dicks, Brenna R. Forester, and Orly Razgour. “Biodiversity conservation requires integration of species-centric and process-based strategies”. In: *Proceedings of the National Academy of Sciences* 122.31 (July 2025). ISSN: 1091-6490. DOI: 10.1073/pnas.2410936122. URL: <http://dx.doi.org/10.1073/pnas.2410936122>.
- [71] Devis Tuia, Benjamin Kellenberger, Sara Beery, Blair R. Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W. Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D. Couzin, Grant van Horn, Margaret C. Crofoot, Charles V. Stewart, and Tanya Berger-Wolf. “Perspectives in machine learning for wildlife conservation”. In: *Nature Communications* 13.1 (Feb. 2022). ISSN: 2041-1723. DOI: 10.1038/s41467-022-27980-y. URL: <http://dx.doi.org/10.1038/s41467-022-27980-y>.
- [72] Roozbeh Valavi, Jane Elith, José J. Lahoz-Monfort, and Gurutzeta Guillera-Arroita. “<scp>block</scp><scp>CV</scp>: An <scp>r</scp> package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models”. In: *Methods in Ecology and Evolution* 10.2 (Nov. 2018). Ed. by David Warton, pp. 225–232. ISSN: 2041-210X. DOI: 10.1111/2041-210x.13107. URL: <http://dx.doi.org/10.1111/2041-210X.13107>.
- [73] Antoinette Van de Water, Marion Garaï, Matthew Burnett, Michelle Henley, Enrico Di Minin, Jarryd Streicher, Lucy Bates, and Rob Slotow. “Integrating a “One Well-being” approach in elephant conservation: evaluating consequences of management interventions”. In: *Ecology and Society* 29.3 (2024). ISSN: 1708-3087. DOI: 10.5751/es-15193-290315. URL: <http://dx.doi.org/10.5751/ES-15193-290315>.
- [74] Vulcan Inc. *EarthRanger software*. Software, data monitored via LoRa system. 2025. URL: <https://earthranger.com/>.
- [75] Jake Wall, George Wittemyer, Brian Klinkenberg, and Iain Douglas-Hamilton. “Novel opportunities for wildlife conservation and research with real-time monitoring”. In: *Ecological Applications* 24.4 (June 2014),

- pp. 593–601. ISSN: 1939-5582. DOI: 10.1890/13-1971.1. URL: <http://dx.doi.org/10.1890/13-1971.1>.
- [76] Le Wang, Chunyuan Diao, and Ying Lu. “The role of remote sensing in species distribution models: a review”. In: *International Journal of Remote Sensing* 46.2 (Nov. 2024), pp. 661–685. ISSN: 1366-5901. DOI: 10.1080/01431161.2024.2421949. URL: <http://dx.doi.org/10.1080/01431161.2024.2421949>.
- [77] Caitlin Wilkinson, Ulrich Brose, Alexander Dyer, Myriam R. Hirt, and Remo Ryser. “A Mechanistic Approach to Animal Dispersal—Quantifying Energetics and Maximum Distances”. In: *Ecology Letters* 28.2 (Feb. 2025). ISSN: 1461-0248. DOI: 10.1111/ele.70085. URL: <http://dx.doi.org/10.1111/ele.70085>.
- [78] Damaris Zurell, Janet Franklin, Christian König, Phil J. Bouchet, Carsten F. Dormann, Jane Elith, Guillermo Fandos, Xiao Feng, Gurutzeta Guillera-Arroita, Antoine Guisan, José J. Lahoz-Monfort, Pedro J. Leitão, Daniel S. Park, A. Townsend Peterson, Giovanni Rapacciulo, Dirk R. Schmaltz, Boris Schröder, Josep M. Serra-Díaz, Wilfried Thuiller, Katherine L. Yates, Niklaus E. Zimmermann, and Cory Merow. “A standard protocol for reporting species distribution models”. In: *Ecography* 43.9 (June 2020), pp. 1261–1277. ISSN: 1600-0587. DOI: 10.1111/ecog.04960. URL: <http://dx.doi.org/10.1111/ecog.04960>.

Appendix A. Appendix

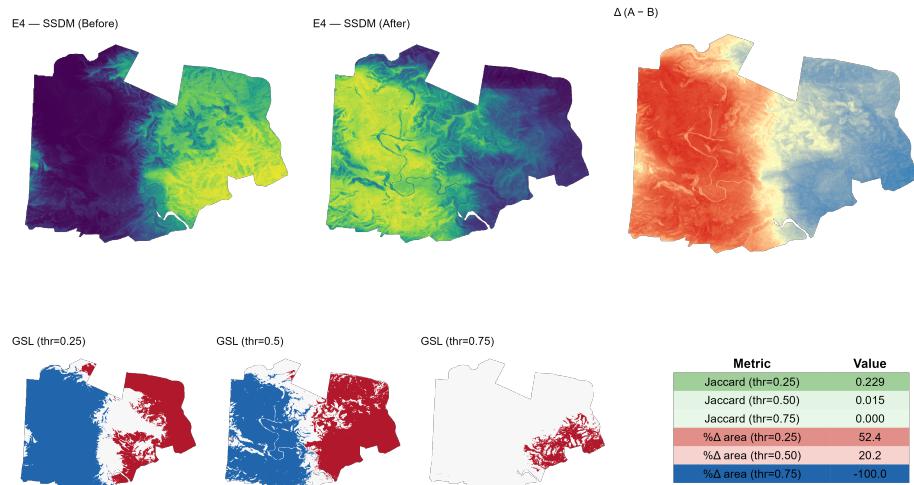
Appendix A.1. Per-elephant panel alignment of H_2O AutoML (top) and SSDM (bottom) results for comparison of habitat change.

Elephant E4 — H₂O | Before vs After: base maps, Δ (A–B), GSL strip, metrics



(a) E4 — H₂O AutoML

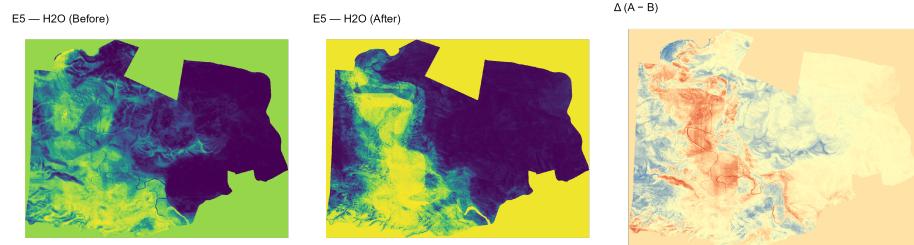
Elephant E4 — SSDM | Before vs After: base maps, Δ (A–B), GSL strip, metrics



(b) E4 — SSDM

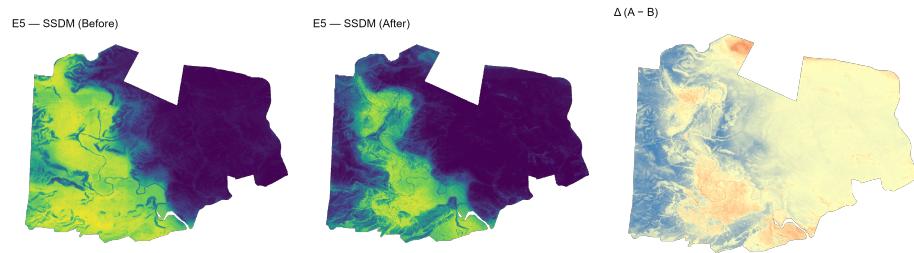
Figure A.10: Per-elephant comparison for Half Moon (E4). The H₂O AutoML and SSDM frameworks are shown vertically, highlighting contrasts in post-fence suitability patterns and change maps.

Elephant E5 — H₂O | Before vs After: base maps, Δ (A–B), GSL strip, metrics



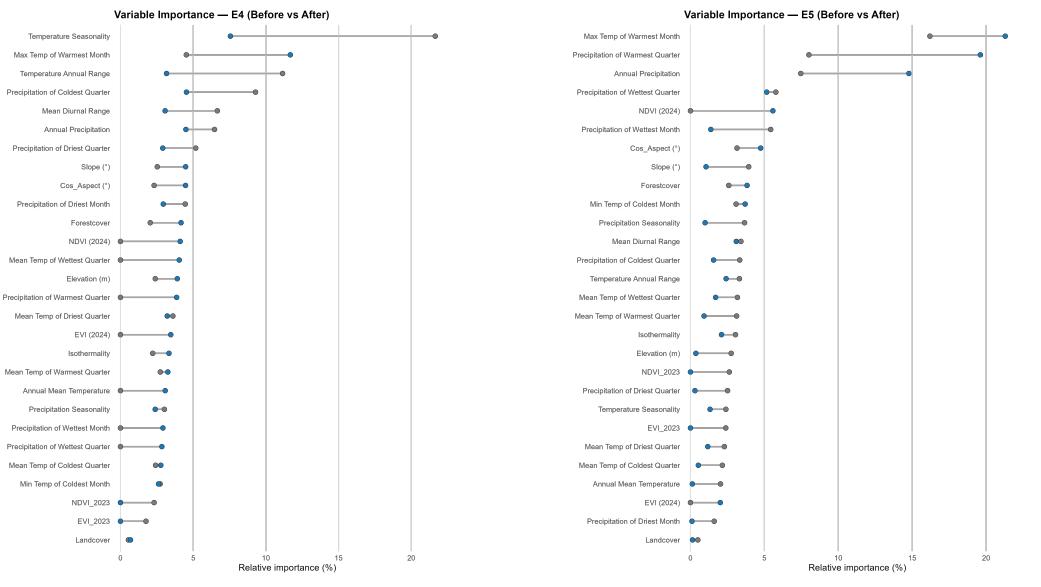
(a) E5 — H₂O AutoML

Elephant E5 — SSDM | Before vs After: base maps, Δ (A–B), GSL strip, metrics



(b) E5 — SSDM

Figure A.11: Per-elephant comparison for Beauty (E5). Stacked panels show vertical alignment of H₂O AutoML (top) and SSDM (bottom) results for clearer visual comparison of habitat change.



(a) Variable importance for Elephant E4 ("Half-Moon").

(b) Variable importance for Elephant E5A ("Beauty").

Figure A.12: Variable importance comparisons for the female elephants E4 ("Half-Moon") and E5A ("Beauty") under the AutoML and SSDM frameworks.

Appendix A.2. H_2O AutoML Results (After Fence Removal)

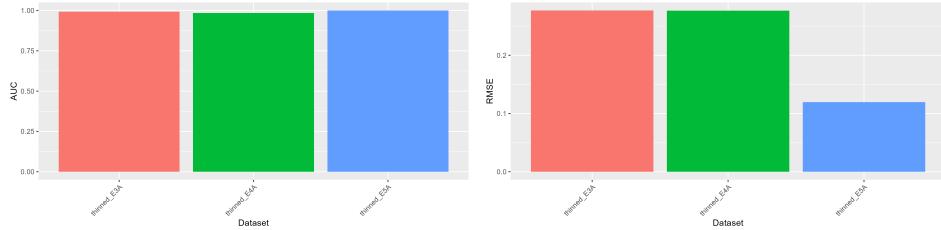


Figure A.13: Model performance summary (After): AUC and RMSE across elephants.

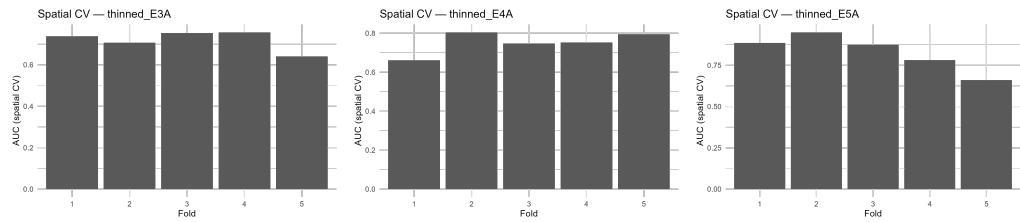


Figure A.14: Spatial block cross-validation AUC for individual elephants (After).

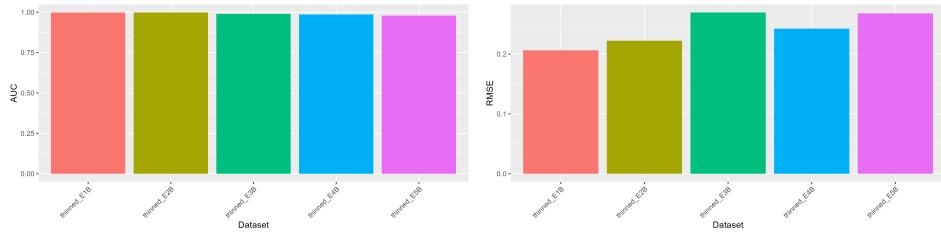


Figure A.15: Model performance summary (Before): AUC and RMSE across elephants.

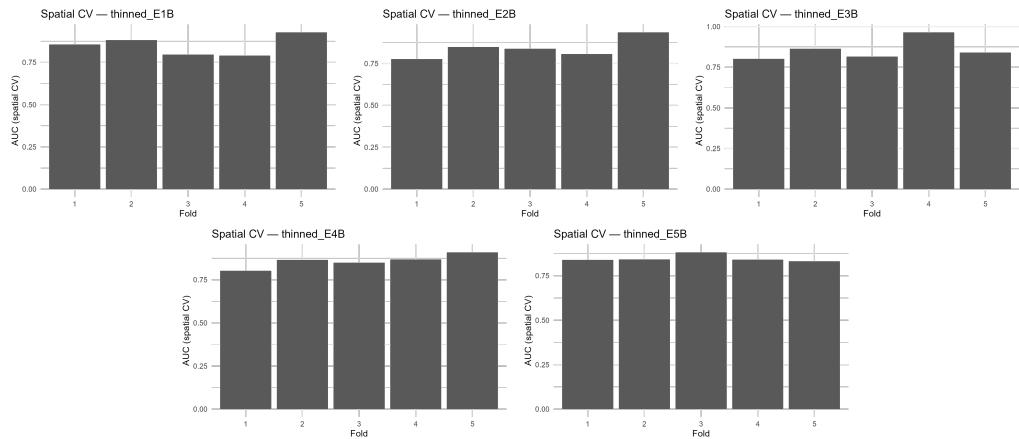
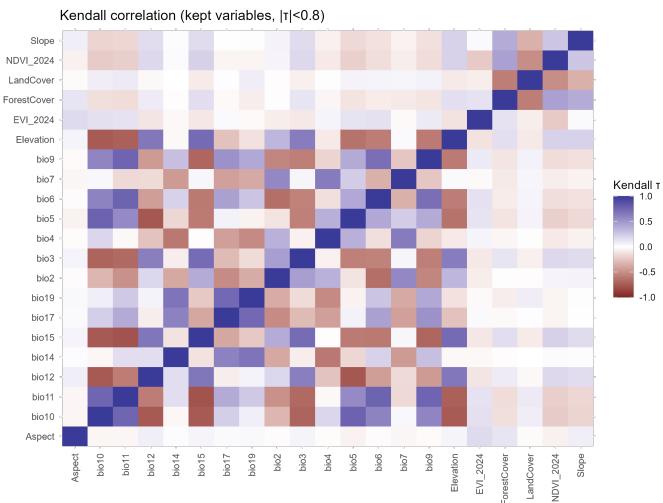
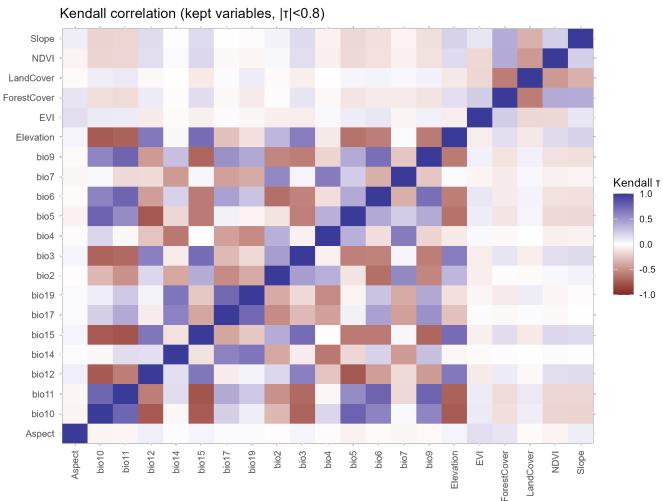


Figure A.16: Spatial block cross-validation AUC for individual elephants (Before).



(a) After fence removal (Set A).



(b) Before fence removal (Set B).

Figure A.17: Kendall's τ correlation heatmaps of retained environmental predictors. The upper panel shows variable relationships for the post-fence (After) period, while the lower panel represents pre-fence (Before) conditions. Predictors exceeding $\tau > 0.8$ were excluded to minimize multicollinearity.

Appendix A.3. Base Learner Contributions

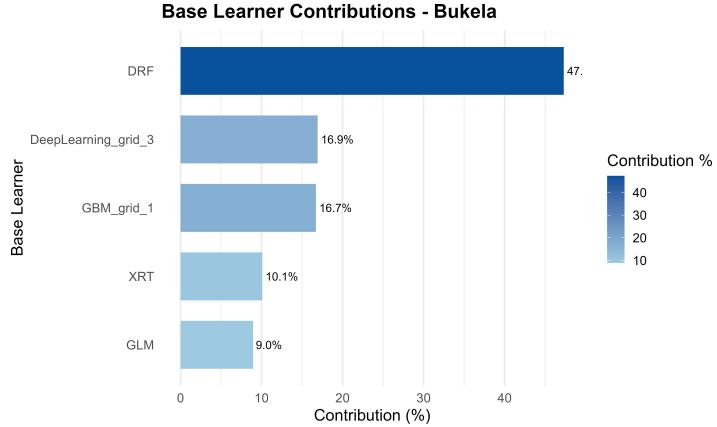


Figure A.18: E3A — Bukela

For H₂O AutoML, the contributions of base learners in the ensemble demonstrate how stacked models combine multiple algorithms to achieve improved accuracy and calibration. Figure A.20 shows the contribution of each base learner to the final ensemble model. In the E3A model, XRT (Extremely Randomized Trees) contributed 35% of the total, followed by GBM (23.1%), DeepLearning (16.2%), and DRF (15.6%), showing a strong reliance on tree-based methods, with GLM contributing a smaller 9.7%. In contrast, the E4A ensemble model was led by DeepLearning_grid_2 (35%), DRF (23.1%), and DeepLearning_grid_1 (18.2%), while GBM and GLM contributed less. This pattern suggests that E4A required more complex nonlinear modeling structures to capture elephant habitat use dynamics effectively.

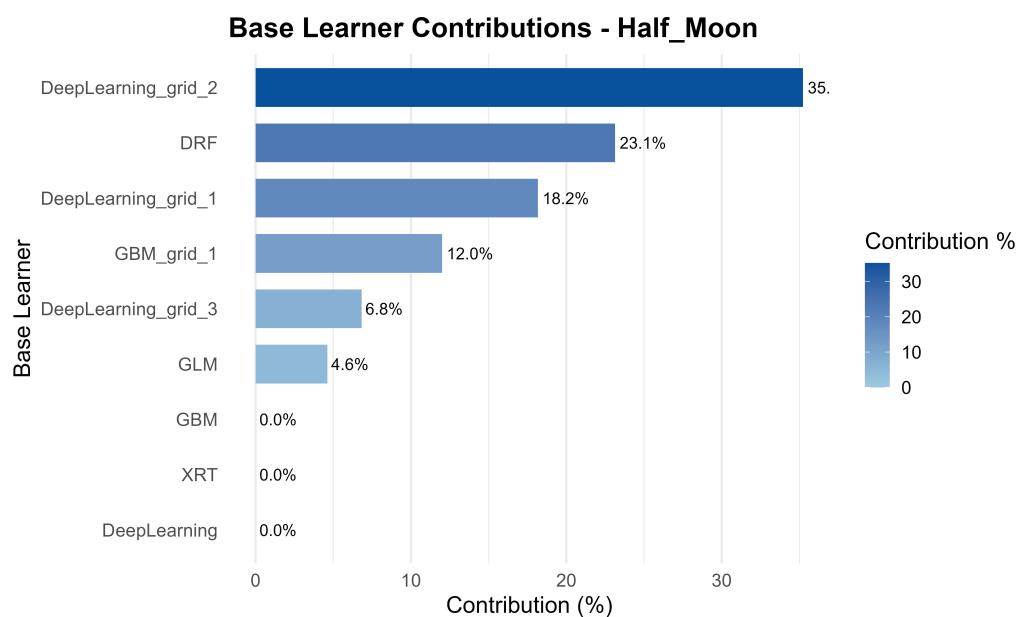


Figure A.19: E4A — Half Moon

Figure A.20: Base learner contributions for H₂O AutoML ensemble models. Each bar plot shows the relative weight of individual algorithms (e.g., GBM, DRF, XGBoost, Deep Learning) contributing to the final stacked ensemble for two elephants (E3A and E4A).

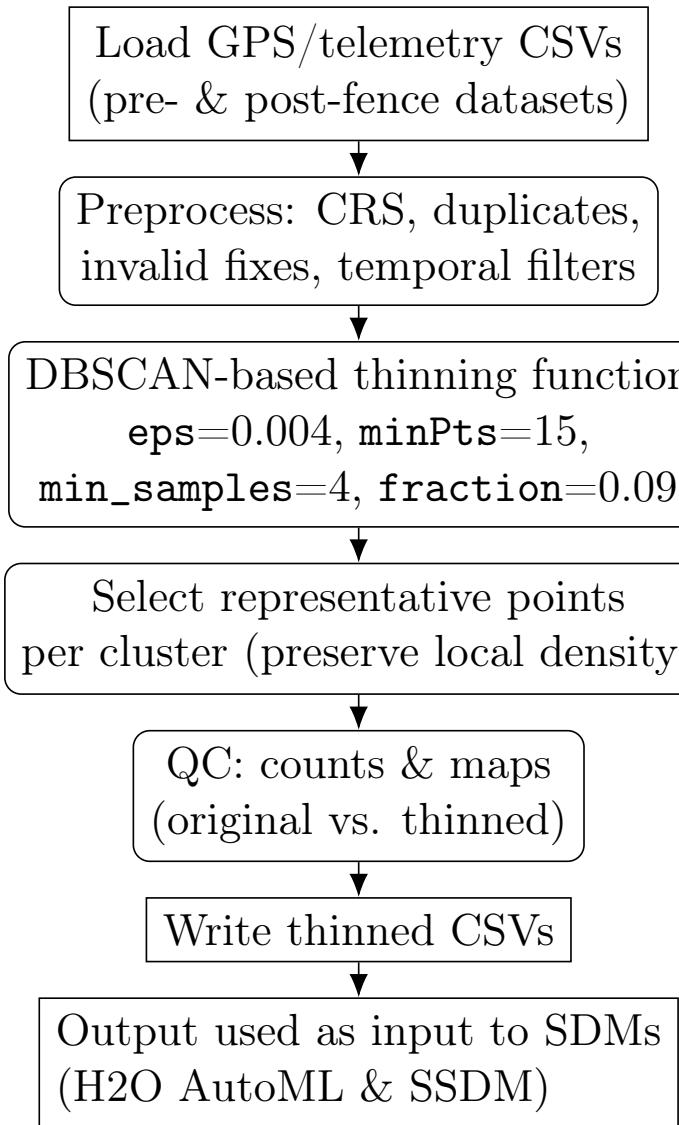


Figure A.21: DBSCAN-based thinning workflow used to reduce spatial autocorrelation and computational load while preserving local density patterns. Parameters were tuned empirically (`eps=0.004, minPts=15, min_samples=4, fraction=0.09`).