

ORIGINAL ARTICLE

Comparing Classical, Hidden-markov and Deep-learning Step-selection Functions for Asian Elephant Habitat Preference: The Value of Feature Engineering, Dynamic Diel/Seasonal Variables and Water-quality Covariates

Hengyuan Liu^{1,2,3} and Ahimsa Campos-Arceiz^{1,2,3*}

¹*Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences & Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China*

²*Yunnan International Joint Laboratory of Southeast Asia Biodiversity Conservation, Mengla, Yunnan 666303, China*

³*University of Chinese Academy of Sciences, Beijing 100049, China*

*Corresponding author E-mail: ahimsa@xtbg.ac.cn

Received: 05 May, 2026; **Manuscript No:** UJE-26- 189368; **Editor assigned:** 07 May, 2026, **PreQC No:** P-189368; **Reviewed:** 19 May, 2026, **QC No:** Q- 189368; **Revised:** 25 May, 2026, **Manuscript No:** R- 189368; **Published:** 30 May, 2026

This methodological case study compares differences and nuances among three step-selection function (SSF) families: local-Gibbs HMM-RSF (steady-state utilisation surface), hidden Markov SSF (HMM-SSF, state-conditional β plus simulated visit density) and deep convolutional SSF (deepSSF, per-pixel selection-probability surface). Each is read against ecological prior knowledge to motivate usage recommendations rather than rank methods. We use 16 502 two-hour GPS fixes from two solitary adult male Asian elephants (*Elephas maximus*) in Kedah, Peninsular Malaysia (2013–2016); at $n=2$ we report which method recovers which signal under identical inputs, not population-level inference. A unified 91-raster product covers topography, canopy, hydrology, water quality, human pressure and lacunarity vegetation classes at $\sim 10\text{m}$. Six deepSSF variants (F1–F6) are evaluated not by likelihood but by whether each captures elephant priors: diel canopy use, water-driven routines, human-activity avoidance, July–October fruiting + plantation-activity window. Across 9 strata and 8 (quarter \times diel) cells, deepSSF variants agree in forest interiors ($\rho \approx 0.80$) and diverge in human-modified strata; F2 and F6 reproduce the diel signal in directions consistent with elephant ecology, while F3 collapses it. The local-Gibbs agrees with deepSSF most in forest_interior, the lowest- σ _seasonal stratum; HMM-SSF's TPM landform predictors expose a movement *vs.* foraging axis complementing deepSSF's selection surface. On the 10km DynQual water-quality stress test local-Gibbs returns Pathogen $\beta = +3.74$ (CI excluding zero on the wrong side), HMM-SSF returns label-switching values and deepSSF averages a monotone-decreasing Pathogen marginal-response across 25 multi-seed checkpoints (mean Δ at p95 = -0.018 , envelope including zero): at this coarse resolution the linear-predictor method is confidently wrong, the convolutional method correct on average but uncertain. We recommend matching method to research goal (steady-state utilisation, latent behavioural state, or high-dimensional input flexibility), adopting from each method what is methodologically sound for the question at hand.

Keywords: Step-Selection Function (SSF), deepSSF, HMM-SSF, MCMC-SSF, Canopy Lacunarity, Habitat Water Quality, Habitat Selection, Diel Variation, Asian Elephant, Peninsula Malaysia.

Introduction

Background: Step-selection functions (SSFs) have become the dominant framework relating animal Global Positioning System (GPS) tracks to environmental covariates since their introduction by Fortin, et al. (2005); reviews of the SSF literature and applied uses are given by (Thurfjell, et al., 2014) and Signer, et al., (2019). The integrated step-selection analysis (iSSA) of (Avgar, et al., 2016) extended SSFs by jointly modelling the movement kernel and habitat selection within a conditional logistic framework and remains the practical workhorse of behavioural-ecology applications. The hidden Markov SSF (HMM-SSF) family (Nicosia, et al., 2017; Pohle et al., 2017; Klappstein et al., 2023) introduces latent behavioural states with state-specific selection coefficients, while the local-Gibbs formulation of Michelot, et al. (2019) enforces detailed balance, guaranteeing that the SSF kernel produces a true stationary Resource-Selection Function (RSF) instead of the well-known $SSF \neq RSF$ inconsistency. Most recently, Forrest and Pagendam (2025) replaced the conditional-logistic estimator with a convolutional neural network (deepSSF), allowing high-dimensional rasters such as multispectral satellite imagery to be ingested directly without manual feature engineering.

Large-mammal movement ecology and the unresolved spatiotemporal-heterogeneity problem: Large terrestrial mammals — Asian and African elephants, bovids, ursids, large carnivores — pose a recurrent problem in movement ecology: their home ranges span heterogeneous landscapes, their habitat preferences shift hither and thither across diel and seasonal cycles and the same individual may behave as a forest-interior forager by day and a plantation-edge raider by night (Cagnacci, et al., 2010; Hebblewhite, et al., 2010; Kays, et al., 2015). GPS telemetry has dramatically increased the volume of available relocation data, but the analytical frameworks that turn fixes into ecological inference have not kept pace with this spatiotemporal complexity (Allen, et al., 2016; Riotte-Lambert, et al., 2020). Three persistent gaps recur across the recent literature.

- I. Spatiotemporal heterogeneity of landscape use is under-resolved. The classical SSF/iSSA workflow churns out one β vector per individual (or, with random-effects extensions, one population mean β); this implicit time-stationarity hides the diel and seasonal shifts that are part and parcel of large-mammal behaviour. Quantifying where and when an animal switches between foraging, resting and ranging modes and how those modes covary with vegetation phenology, surface-water availability and human-activity rhythm, remains methodologically difficult, particularly for solitary individuals whose social context does not constrain timing (Sukumar 2003; Wadey, et al., 2018; de la Torre, et al., 2021). For Asian elephants in the Sundaic region specifically, the July–October window combines fruit-tree masting in natural forest with peak rubber-tapping and harvest in plantations (Saaban, et al., 2011; de la Torre, et al., 2021), but the diel-by- seasonal interaction has not been quantified at the landscape scale of this case study.
- II. Coupled methods struggle to disentangle behavioural state from habitat selection. HMM-coupled SSF (Nicosia, et al., 2017; Klappstein, et al., 2023) and local-Gibbs HMM-RSF (Michelot, et al., 2019) both attempt to tap into the latent behavioural state and factor it out of the habitat-selection signal, but the joint likelihood ties the transition probability matrix to the selection coefficients in a labyrinth of parameter dependencies and the resulting identifiability (particularly for state-specific β under limited-sample-size conditions) is fragile (Pohle, et al., 2017). The practical consequence is that diel-rhythm and seasonal-shift signals, which depend on time-conditioned interactions between covariates, often dissolve into bootstrap-CI bands wide enough to overlap zero even when the underlying ecology is real. Convolutional alternatives (Forrest, et al., 2025) absorb time scalars more easily but trade parametric interpretability for a black-box network.
- III. The behavioural-knowledge gap is not academic. Spatiotemporal heterogeneity of habitat use is the foundation on which three applied questions rest: basic behavioural ecology of the species (diet, ranging, social structure), human–wildlife conflict mitigation (when and where will the animal enter a plantation or cross a road; Wadey, et al., 2018; de la Torre, et al., 2021) and protected-area/corridor planning (which forest interiors carry the highest year-round utilisation density and which edge zones admit seasonal influx). At $n=2$ individuals the inference at population scale is necessarily limited, but the methodological question of whether contemporary SSF families recover or smooth out these behavioural signatures can be addressed at any sample size. That methodological question recovered or smoothed and *via* which structural feature of which method motivates the present case study.

Trade-offs across methods : The three families differ on three axes that practitioners must reconcile.

- a) Output semantics. local-Gibbs HMM-RSF produces an explicit β vector with bootstrap confidence intervals from 100 times of independent modeling at acceptable bar we set; the resulting RSF surface is the long-term steady-state utilisation distribution. On another hand, HMM-SSF gives state- dependent β with explicit transition probabilities (a TPM linear predictor with its own informative covariates that highly bear on Asian elephant moving pattern), so it carries two distinct ecological information streams: which environmental features drive habitat selection within each state and which drive transitions between states. Besides, deepSSF produces a per-pixel logit at each visited location, with array of separate movement-kernel output parameterised by 12 coefficients learned per local environment upon target of next step from observation that as same as cardinal SSF model; without simulation, deepSSF gives a habitat-selection probability surface rather than a utilisation distribution.
- b) Spatial receptive field. Classical and HMM-SSF treat each environmental raster as a point-extracted scalar at the focal location, with no spatial context. deepSSF's convolutional encoder, by contrast, operates on a local image window (in this study, 101×101 pixels $\approx 1.0\text{km} \times 1.0\text{km}$ at 9.93m pixel size), allowing it to see spatial pattern around the elephant's current location, not only the value at the central cell also the texture and form from landscape features. Section 2.3 details how we engineered the linear-predictor methods' input streams through focal-statistic water context, multi-scale lacunarity vegetation classification and second-derivative DEM ridge/valley masks to give those methods a convolution-like receptive field, narrowing (without eliminating) the spatial-feature gap.
- c) Movement kernel. Classical and HMM-SSF have a globally fitted movement kernel (one set of step-length/turning-angle parameters per state) wields along the available next step sampling. However, deepSSF's movement subnetwork is itself context-dependent: a separate convolutional pathway maps the local environmental window to 12 parameters of a finite-mixture movement kernel (2 Gamma \times 2 von Mises), permitting the elephant's expected step length to contingent to local habitat. The next-step probability surface in deepSSF is the log-sum of the habitat and movement grids, which broadens the realised support and admits rare extreme steps (e.g., long-distance crossings) that unleash gamma kernel suppresses.

Engineering: Lacunarity, GBLU and triple-index water-surface: Three feature-engineering threads in the present study deserve introduction. First, we apply Plotnick gliding-box lacunarity (Plotnick, et al., 1993; Smeds, et al., 2025) onto a 1-m global canopy-height surface (Tolan et al., 2024) to derive six objective vegetation-cover classes whose breakpoints (3m, 15m, 20m, 25m, 32m) are statistically significant rather than ad hoc. Second, we use the GBLU v1.0 (Li et al., 2024) basic-landform classification, post-processed with curvature- and relative-elevation-based. Meanwhile, ridge/valley filters from ASTER GDEM v3, with overlapping among lower curvature (less than 15°) and relative height (by outlier more than ± 7 meters) by focal statistics with radius of 150 meters, to give binary terrain masks. Lastly, due to existing global land-cover products do not adequately resolve small streams and seasonal water bodies in the Kedah landscape, we manually extracted water surfaces from Sentinel-2 L2A imagery using a triple-index decision rule combining NDVI, MNDWI and AWEInsh, with river-network masking (Lin, et al., 2021) for shadow correction.

Research questions: We address three questions:

- 1) Convergence and divergence. When the same GPS data and environmental rasters are provided to deepSSF, local-Gibbs HMM-RSF and HMM-SSF, do they produce spatially consistent habitat-selection/utilisation outputs? Where they disagree, can the disagreement be attributed to specific environmental rasters?
- 2) Engineering value. Do lacunarity-derived canopy classifications, GBLU terrain types and Sentinel-2 multispectral imagery recover habitat- selection signal otherwise carried by raw 1-m canopy height and 30-m DEM? How do anthropogenic water-quality (BOD, pathogen, salinity), road topology and human-footprint act as drivers across methods?
- 3) Diel \times seasonal dynamics. Do the three methods detect the same diel-rhythm and seasonal shifts in habitat preference and do those shifts have biological explanations sufficient and necessary rather than mere statistical artefacts; to be interpreted ecologically?

We also propose a practical analysis-strategy decision tree linking research goal to model class.

Throughout, the framing is difference-and-nuance detection rather than winner-picking, adopts a researcher-driven perspective. The three SSF families do not share a common output object (steady-state utilisation density, state-conditional selection plus movement

vs. foraging axis, per-pixel selection probability surface respectively), so a single test-NLL ranking would compare incommensurable quantities and the empirical noise floor measured here (Appendix I) confirms that within-family configuration ranking is statistically not defensible at this dataset. We therefore read each method's output against ecological prior knowledge to recommend goal-conditioned usage, adopting from each family the part that is methodologically sound for the question at hand.

Materials and Methods

Study area, animals and the diel ecology

Two adult male Asian elephants ("606 Awang S Kedah" and "846 Awang Kapakc1") were tracked between February 2013 and June 2016 in northern Kedah, Peninsular Malaysia (101.10-101.80°E, 5.40–6.10°N). After de-duplication and quality filtering ($1 \leq \Delta t \leq 4$ h between fixes), 16 502 valid fixes (16 450 valid steps in 52 segments) entered analysis.

Why two solitary adult males in our study? Asian elephant social structure is matriarchal: adult females, sub-adults and calves form coordinated herds of 5–20 individuals; adult males disperse from natal herds at 12–15 years and live solitarily or in small bull groups (Sukumar, 2003; Vidya, et al. 2005). The Peninsular Malaysian context for these two animals is reviewed in Saaban et al. (2011), with subsequent elephant-road and elephant-plantation conflict ecology developed in Wadey, et al. (2018) and de la Torre, et al. (2021). Solitary males and matriarchal herds differ on three behavioural axes that confound habitat-selection analysis if pooled.

- i. Range size and movement amplitude: solitary males have larger, more variable home ranges and more frequent long-distance excursions; herds maintain tighter, more predictable core ranges built around calf-rearing constraints.
- ii. Risk tolerance: solitary males, particularly in musth, more readily enter human-modified landscapes (rubber, oil palm, agroforestry edge), whereas herds avoid proximity to humans to protect calves.
- iii. Diel and feeding cycles: herds follow predictable matriarch-led foraging routines synchronised with calf rest; solitary males have more flexible, individually idiosyncratic diel activity patterns. To prevent the SSF coefficients from averaging over these two fundamentally different behavioural classes, we restrict this study to two solitary adult males whose tracks contain no herd-membership signature. The two-individual sample size is acknowledged as a limitation but this trade-off explicitly preserves behavioural-class purity.

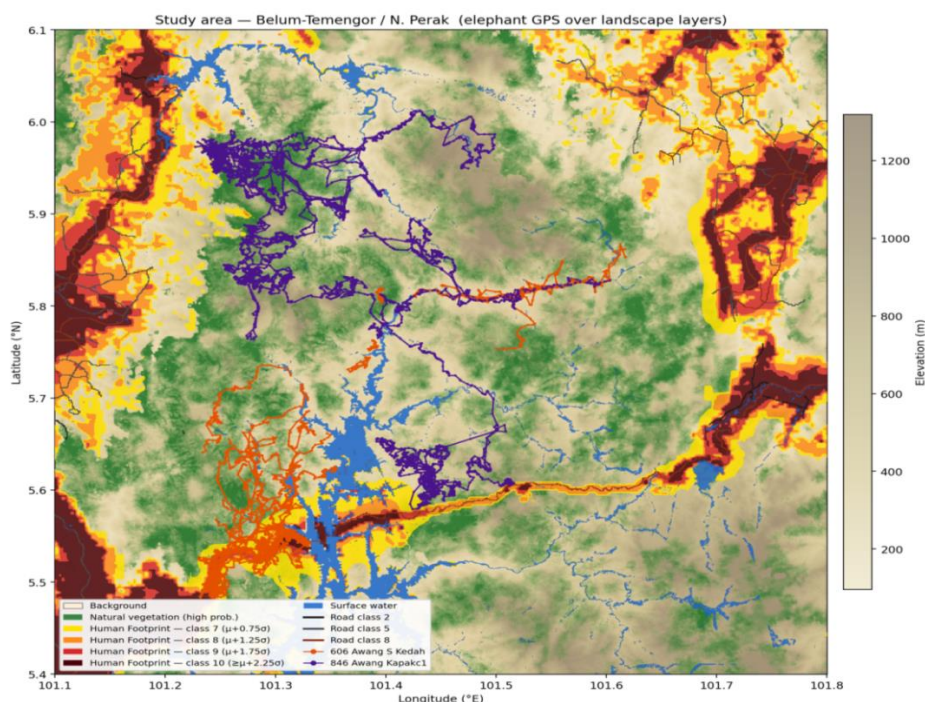


Fig 1. Study area in northern Kedah, Peninsular Malaysia, with the GPS tracks of the two solitary adult male Asian elephants ("606" solid line, "846" dashed line) and the nine ecological strata used in cross-method comparison. Stratum colours follow a viridis ordering from forest-interior (deep green) to high-human-footprint (brown).

Water-centric biology. Asian elephants (*E. maximus*) are physiological obligate bulk drinkers: An adult male consumes 80-200 litres of free water per day and dehydration tolerance is low even relative to other tropical megafauna (Sukumar, 2003; de Knegt, et al., 2011; Hedges, et al., 2005). Daily activity is therefore organised around drinking sites and proximity to surface water, with elephants typically returning to the same drinking sites multiple times within 24 h. This biological prerequisite directly motivates our inclusion of two water-related variable families: the geometric water access layer (surfacewater, waterrange 2km) and a novel addition for elephant habitat-selection research the DynQual anthropogenic water-quality stressors (BOD, Pathogen, Salinity). Polluted water is an avoidance gradient (elephants prefer cleaner sources) and elephants are known to seek out mineral-supplemented water (mildly saline springs) for nutritional balance.

Diel context. Radiation fog typically forms in the valleys and along rivers from midnight onward, peaks before dawn and dissipates by 11:00-12:00 local time. By 12:00 the canopy is clear, light irradiance peaks and elephants seek the shade of the lush dipterocarp canopy, water and forage in cool refugia. By 00:00, in the fog phase, they are more active in open and edge habitats where olfactory and acoustic cues compensate for limited visibility. We therefore take 12:00 ("day") and 00:00 ("night") as canonical representative hours, matching the categorical day/night state used by the two SSF coupling methods. This 12 h offset between representative hours is not arbitrary: it captures both the visibility- limited fog phase (00:00) and the shade-seeking peak insolation phase (12:00), making the diel comparison ecologically meaningful.

Limestone karst hills and the GBLU caveat. The Kedah landscape includes prominent karst limestone hills with thin soils, sparse forage and cliff-like vertical escarpments that elephants cannot traverse at all. We deliberately do not assign any GBLU (Li, et al., 2024) class as a "limestone-hill" indicator, because GBLU's relief categories (low/mid/high-relief mountain) are based on smooth elevation gradients and do not capture the discontinuous cliff topology characteristic of limestone karsts. Instead, the limestone-hill effect on elephant selection enters the analysis only implicitly through deepSSF's raw DEM input: The convolutional encoder can detect the high-curvature discontinuity directly from elevation tile context, whereas the linear-predictor methods (which receive only point-extracted DEM) cannot. This is one reason we expect deepSSF to differ from local-Gibbs/HMM-SSF in elevated-terrain strata (ridge vally and forest interior).

Environmental data (91 rasters)

All environmental rasters were re-projected to EPSG:4326, clipped to the 7 795 × 7 795 study window and resampled to ~ 9.93m using bilinear interpolation for continuous layers and nearest-neighbour for binary masks. Six ecological categories were assembled.

Table 1. Variable abbreviation table.

Code	Full name	Type	Source	Ecological role for <i>E. maximus</i>
dem	Digital Elevation Model	continuous (m)	ASTER GDEM v3	base terrain; only deepSSF can capture limestone-cliff context
canopy	1-m canopy height	continuous (m)	Tolan, et al., 2024	shade refugia and food-bearing arboreal mass
vc_uniform_canopy	uniform high-canopy class	binary	this study, lacunarity p25-p32	tall closed canopy; daytime shade
vc_uniform_lowforest	uniform low-forest class	binary	this study, lacunarity p15-p20	sub-canopy/re-growth
vc_uniform_open	uniform open class	binary	this study, lacunarity p0-p3	open habitat; potential nighttime use
vc_uniform_subcanopy	uniform sub-canopy	binary	this study, lacunarity p20-p25	sub-canopy layer
vc_mosaic_canopy	mosaic high-canopy	binary	this study, lacunarity	fragmented canopy/edges texture

Comparing Classical, Hidden-markov and Deep-learning Step-selection Functions for Asian Elephant Habitat Preference: The Value of Feature Engineering, Dynamic Diel/Seasonal Variables and Water-quality Covariates

vc_mosaic_lowforest	mosaic low-forest	binary	this study, lacunarity texture	fragmented re-growth
LAI	Leaf Area Index	continuous, trimonthly	MODIS, (Yuan, et al., 2011)	seasonal vegetation density; captures rubber leaf-fall
BOD	Biochemical oxygen demand	continuous, bimonthly	DynQual, (Jones, et al., 2023)	water-quality stressor; first use in elephant habitat research
Pathogen	water pathogen load	continuous, bimonthly	DynQual	drinking-water risk; novel covariate
Salinity	water salinity	continuous, bimonthly	DynQual	mineral supplementation gradient
surfacewater	water surface mask	binary (0/1/3)	Sentinel-2 NDVI/MNDWI/AWEInsh, this study	drinking, bathing, mobility barrier
waterrange_2km	water focal-abundance, 2km	continuous	Lin, et al., 2021 + this study, focal-sum	distance + abundance of water (engineered convolution)
hfp (annual)	Human Footprint	continuous z-std	Venter, et al., 2016	composite of 8 anthropogenic stressors
road	road penetration index	continuous [0, 1]	OSM, this study	linear pressure; engineered, non-linear
road_level	road class binary	binary	OSM	road presence regardless of class
rubber	rubber plantation	binary	Wang, et al., 2023	seasonal forage/disturbed canopy
natplant	natural-plant probability	continuous z-std (0–1 prob.)	Neumann, et al., 2024, Transformer-based	natural vs. anthropogenic cover gradient
ridge	ridge mask	binary	this study, ASTER curvature+relative elevation	second-derivative + multi-scale terrain feature
vally	valley mask	binary	this study, ASTER curvature+relative elevation	second-derivative + multi-scale terrain feature
hill/plaint/high_relief_mount/mid_relief_mount/low_relief_mount/of	GBLU terrain classes	binary	GBLU v1.0, (Li, et al., 2024) (already convolution-processed)	basic landform classification (no limestone)
fshorepwater	Sentinel-2 12-band	continuous, multispectral	ESA, (Drusch, et al., 2012)	spectral signature of vegetation, water, bare soil

Note: The vc_ prefix throughout this paper stands for vegetation-cover classification (re-using the file-naming convention from earlier landcover literature, but with a strictly canopy-structure interpretation grounded in the lacunarity analysis of Smeds, et al. 2025).

Pre-category Ecological Commentary

- a) Topography. ASTER GDEM v3 (NASA, et al., 2019) provided the elevation model. Relative elevation in a 150-m focal window and full-scale curvature were used to derive ridge/valley binary masks ($\text{curvature} < 15 \wedge |\text{relative elevation}| > 7\text{m}$). GBLU v 1.0 (Li, et al., 2024) supplied basic- landform types, simplified to plain/hill/low/mid/high-relief mountain. Inasmuch as the limestone karst hills in Kedah have cliff-discontinuity topologies that GBLU's smooth-relief framework cannot represent, no GBLU class is interpreted as a limestone-hill indicator; the limestone effect enters only *via* deepSSF's raw dem input and the local convolutional context.
- b) Vegetation and canopy. A 1m global canopy-height product (Tolan, et al., 2024) was mode-resampled to 10m and reclassified into six vegetation-cover (vc_*) strata *via* the Plotnick gliding-box lacunarity procedure (Smeds, et al., 2025). MODIS 8-day leaf area index (Yuan, et al., 2011) was aggregated to trimonthly periods. The trio rubber+LAI+S2 captures rubber-plantation phenology in ways the static rubber binary mask alone cannot: LAI carries seasonal leaf-fall and re-flushing signal, while Sentinel-2 multispectral imagery discriminates plantation development stages (newly planted *vs.* mature *vs.* harvest-ready) through their distinct spectral signatures. Rubber plantation phenology is ecologically relevant because elephants are known to enter rubber stands during specific phenological windows (particularly during young-leaf flushes) for selective foraging.
- c) Hydrology. River networks (Lin, et al., 2021) were used to mask shadow-induced false positives in the Sentinel-2-derived water-surface raster. The triple-index detection (NDVI, MNDWI, AWEInsh) was necessary because existing global land-cover products (MapBiomas, ESA WorldCover, JRC GSW) do not adequately delineate small streams and seasonal water bodies in the Kedah landscape; our manual Sentinel-2 extraction substantially refines the hydrology layer. The waterrange 2km raster is a focal-sum count of surface-water pixels within a 2km radius; this is an engineered "convolution-like" feature that supplies the linear-predictor methods with a spatial-context measure (water abundance, not just nearest-water distance).
- d) Water quality (DynQual). Monthly Biochemical Oxygen Demand (BOD), pathogen load and salinity from the global hydro-water-quality model (Jones, et al., 2023) were Gaussian-resampled from 10km to 1km and aggregated to bimonthly periods. To the best of our knowledge this is the first habitat-selection study of large mammals (and certainly of *E. maximus*) to incorporate explicit anthropogenic water-quality covariates. The biological motivation is direct: elephants are obligate-bulk-drinkers and the quality of the water sources encountered along their daily routes is a plausible (though previously untested) determinant of habitat selection. We expect water quality to act through two pathways: direct access (avoidance of polluted drinking sources) and indirect riparian forage quality (water-stressed vegetation alters food availability).
- e) Human pressure. The Venter, et al. (2016) Human Footprint index (HFP) was kept at annual cadence (2013–2016). The HFP is synthesised from eight anthropogenic stressors weighted to a 0–50 scale:
 - i. Built environments (urban built-up),
 - ii. Population density
 - iii. Night-time lights
 - iv. Cropland coverage
 - v. Pasture lands
 - vi. Road infrastructure (with 500-m buffer),
 - vii. Railways
 - viii. Navigable waterways.

The composite captures aggregate human footprint without requiring us to model each stressor separately.

While, OpenStreetMap (OpenStreetMap contributors, 2024) roads were rasterised; OSM road class (1–5) is not a strictly linear scale of disturbance (a national highway and an unpaved logging trail differ by orders of magnitude in width and traffic intensity). We therefore converted road class to a continuous penetration index road by setting class-5 roads to ~ 10

cells ($\approx 100\text{m}$) wide and class-1 roads to ~ 2 cells ($\approx 20\text{m}$) wide, then linearly stretched to $[0, 1]$. Because this engineered scale is not strictly linearisable in the design matrix of conditional-logistic SSFs, we include road only in deepSSF (whose convolution layers absorb the non-linearity) and rely on road_level (binary class membership) plus the HFP composite to partially proxy the missing road-intensity signal in the linear-predictor methods.

- f) Land use. A 10-m rubber distribution (Wang, et al., 2023) and a Transformer-based 10-m natural-forest baseline (Neumann, et al., 2024) were included. The natplant layer is itself the output of a deep-learning model (a Transformer applied to satellite imagery), not a naive 0/1mask; it provides a per-pixel probability of natural-vegetation membership that already encodes spatial context implicitly. Cropland, oil palm and impervious surfaces were excluded because they are essentially absent within the GPS bounding box (and are otherwise captured by HFP).

Receptive field, the convolutional window and engineered "convolution-like" features

deepSSF processes each step's environment using a 101×101 pixel local window ($\approx 1.0\text{km} \times 1.0\text{km}$ at the 9.93m pixel size). This window size was selected by the Forrest, et al., (2025) framework as $ws_max = \max(\min_ws=101, \text{ceil}(p95(\text{step_length})/\text{pixel_size} \times \text{buffer_factor}=1.0))$ and in our data the elephants' p95 step length ($\sim 700\text{-}1000\text{m}$) does not exceed $101 \text{ px} \times 9.93\text{m}=1003\text{m}$, so the window is at the lower bound and is uniformly applied. Within this window, four 3×3 convolutional layers with stride 1 and 1-pixel padding give the central output cell an effective receptive field of $\sim 9 \text{ px} \times 9 \text{ px} \approx 90\text{m} \times 90\text{m}$, but the information used in the model output is integrated over the full window through the conv-pool stack.

The linear-predictor methods (local-Gibbs, HMM-SSF) lack any analogous spatial encoder: each environmental variable is point-extracted at the focal location and enters the design matrix as a single scalar. To narrow the receptive-field gap not eliminate it, but make the comparison fair we engineered several covariates with explicit "convolution-like" multi-scale operations on the input rasters before they enter the linear predictor:

- i. Waterrange_2km; a 2km radius focal-sum count of surface-water pixels, providing a spatial-context measure of water abundance rather than just point-wise water presence.
- ii. GBLU v1.0 terrain classes; already produced by Li et al. (2024) through a multi-scale convolutional terrain-analysis algorithm; we use the authors' output directly without further processing.
- iii. Lacunarity-based vegetation-cover (vc_*) classes; derived from the 1-m Tolan et al. (2024) canopy-height surface using two-scale gliding-box lacunarity (Smeds, et al., 2025). The Plotnick lacunarity calculation is mathematically equivalent to a series of moving-window mean and variance operations followed by a normalisation step ($\wedge(r) = \text{Var}(S)/\text{Mean}(S)^2 + 1$, where S is the box mass at scale r), which is structurally identical to a multi-layer convolution-pooling operation. The resulting six vc_* classes therefore carry spatial texture information that point-extracted canopy height cannot.
- iv. Ridge and vally; derived from ASTER GDEM v3 by computing the second derivative of elevation (full-scale curvature) and a 150m focal-window relative-elevation difference, then thresholding at $\text{curvature} < 15 \wedge |\Delta \text{elev}| > 7\text{m}$. The composition of "second-derivative+multi-scale focal statistic+threshold" is again analogous to a multi-layer convolution-pooling-fully-connected pipeline applied to the DEM.
- v. Natplant; the output of a Transformer-based deep-learning model (Neumann, et al., 2024) applied to multispectral imagery. The per-pixel value is a natural-vegetation probability that has already absorbed spatial context through the Transformer's attention mechanism, not a naive 0/1 label.

Through these engineered features, the linear-predictor methods receive input streams that mimic, at the level of summary spatial statistics, what deepSSF's convolutional encoder would extract automatically. This is the fairness premise that justifies a same-data three-way comparison.

Standardisation, smoothing and the no-standardisation property of deepSSF

For the linear-predictor methods (local-Gibbs and HMM-SSF), all continuous rasters were z-score standardised using a Welford-style pooled mean/standard deviation across all temporal periods. Binary masks were left untouched. DynQual water-quality values (BOD, pathogen, salinity) are heavily right-skewed at native scale; we therefore applied log compression ($\log(1+x)$) prior to z-

standardisation. The 10km native DynQual grid was Gaussian-smoothed and resampled to 1km, then to $\sim 10\text{m}$ to match the Sentinel 2 reference grid (all rasters in the project share this final resolution).

deepSSF, in contrast, accepts raw rasters with no per-channel standardisation. This is possible (and advantageous) for three intersecting reasons:

- i. The convolutional encoder learns scale-equivariant features in its first two layers (3×3 kernels with stride 1 and 1-pixel padding). Once a feature map is produced, only relative-magnitude information matters and the absolute scale of the input is absorbed into the learned filter weights.
- ii. The output is normalised by log-softmax over the 101×101 window in the loss, so the location of the next-step prediction is invariant to a global affine transformation of any single channel.
- iii. Mixed-precision FP16 training (with Adam optimiser, gradient clipping at norm 1.0) self-normalises the gradient flow even when raw raster magnitudes span orders of magnitude, eliminating the need for per-channel BatchNorm.

The practical consequence is that deepSSF accepts raw GeoTIFFs directly, saving the standardisation overhead of the local-Gibbs/HMM-SSF pipelines and removing the risk of standardisation-artefact biases.

Method-specific Configurations

deepSSF - six feature-set variants : We retained the Forrest, et al., (2025) ConvJointModel architecture (three Conv–MaxPool blocks for the spatial encoder, four sin/cos scalar grids for time, a 12-co-efficient movement-kernel head). Six variable configurations were trained independently with seed-42 (animal \times segment) two-level stratified split (80 /10 /10 train/val/test):

- F1-11 raw rasters, 15 channels (raw-environment baseline).
- F2-F1 + Sentinel-2, 28 channels (raw + spectral).
- F3-8 variables (S2 + canopy + dem + road_level + 4 monthly water-quality), 23 channels (minimal anthropogenic-aware).
- F4- engineered-only (6 vc_* , 7 GBLU terrain, 4 monthly), 27 channels.
- F5-F4 + S2, 35 channels (engineered + spectral).
- F6-raw \cup engineered \cup S2 superset, 42 channels.

All six configurations were trained with the same architecture, the same train/val/test split and the same single seed (seed = 42). Test NLL values are reported in Appendix E for completeness, but we do not rank the configurations on NLL: a 5×5 split-seed \times init-seed factorial replication of F2 (Appendix I) measured $\sigma_{\text{split}} = 0.203$ nats and $\sigma_{\text{init}} = 0.013$ nats, so the inter-configuration NLL gaps (≤ 0.04 nats) sit well within the partition-noise band and ranking is statistically not defensible. Instead, F1-F6 are interpreted by which ecological prior knowledge each configuration recovers diel canopy use, water-driven daily routines, human-activity avoidance and the July–October fruiting-and-pressure peak.

The movement subnet outputs 12 parameters (2 Gamma component shapes/scales/weights for step length, 2 von Mises component means/kappas/weights for turning angle), with the Gamma-2 scale multiplied by 500m to bring its parameter into a similar numeric range as Gamma-1.

Local-Gibbs HMM-RSF- strict-Michelot 2019 with Diel Whitelist

We followed the strict-Michelot specification (TPM intercept-only, 2 parameters; β shared across both states). Each spatial covariate enters as a stationary RSF coefficient. Day/night interactions are restricted to a biologically motivated whitelist: any raster with prefix vc_* , plus $waterrange_2\text{km}$ and hfp . All other variables share a single coefficient. The design matrix has 32 RSF columns (8 whitelist \times 2+16 shared) plus σ_1 , σ_2 and TPM intercepts, totalling 35 parameters. We accumulated 100 accepted bootstrap optima using GenSA global optimisation with 70% warm-start re-seeding and an NLL-acceptance threshold of -5×10^5 nats. The acceptance threshold was fixed before the bootstrap parameter values were inspected, set on the basis of independent single-restart pilot trials whose typical local-optimum NLL band ($\approx -1.7 \times 10^5$ nats) lies well above the threshold; only candidate optima whose final NLL fell below the threshold (i.e., reached the deep basin rather than a shallow local minimum) entered the bootstrap pool. This pre-registration of the acceptance threshold ensures that the bootstrap mean and 95 % CI cannot be inflated by post-hoc selection of "credible-looking" parameter values (Appendix F). 5 000 resampling replicates over the 100 accepted optima yielded per-coefficient mean and 95% confidence intervals.

HMM-SSF; two-state model with TPM-coupled selection

The two-state HMM-SSF was fit with state-dependent SSF and a Transition Probability Matrix (TPM) that depends on seven landform covariates: ridge, road, plain, valley, three relief-mountain classes. Day/night interactions follow the same whitelist as local-Gibbs.

The TPM linear predictor is itself biologically informative and deserves explicit interpretation. The TPM covariates do not represent habitat preference (that is the SSF role); they represent the contribution of an environmental feature to whether the elephant maintains its current behavioural state or switches to the other. Roads, ridges and (to a lesser extent) plains tend to promote state-switching an animal tends to leave the resting/encamped state and start moving when crossing or paralleling a road, traversing a ridge, or crossing open plain. Valleys, on the other hand, tend to promote maintenance of the foraging/encamped state an animal that has reached a valley with food and water tends to stay there and forage longer. This movement *vs.* foraging axis is one of the unique explicit information streams that HMM-SSF outputs and that neither local-Gibbs (which has β shared across states) nor deepSSF (which does not separate state behaviour) can replicate.

The two latent behavioural states each carry an independent set of selection coefficients, so this model has $2 \times$ the β count of a single- state SSF; state 1 is the "encamped"/resting state (low step length), state 2 is the "exploratory"/ranging state (high step length). The Viterbi decoder assigns each observed step to the most likely state, providing the behavioural interpretation that classical SSFs lack.

We accumulated 100 accepted optima using a 10-process parallel cluster. The acceptance threshold (NLL < 89 000 nats) was, like the local-Gibbs threshold above, fixed before the bootstrap parameter values were inspected, set from pilot fits whose typical adequate-convergence NLL band lies just above this value. Same multi-restart bootstrap rationale as local-Gibbs: 5 000 replicates over the 100 accepted optima provide per-coefficient mean and 95 % confidence intervals. Steady-state visit density was generated by 500 forward- simulated trajectories of 16 502 steps each (≈ 3.77 simulated years per trajectory) and aggregated by quarter (12 single-quarter densities, plus 4 three-year-summed-by-quarter densities).

Cross-method comparison

Spatial unit and stratification: All output maps were rendered on the $7\,795 \times 7\,795$ reference grid and downsampled $5 \times (1\,559 \times 1\,559 = 2.43 \times 10^6)$ pixels for cross-method correlation. The landscape was partitioned into nine mutually exclusive ecological strata (Table 2), assigned by priority (top wins): hfp_high (HFP > p80), water (surfacewater ≥ 1), rubber (binary), natplant_high (natplant > p80), ridge_valley (ridge \vee valley), agro_forest_edge (rubber \vee natplant_high within 30 px of vc_uniform_canopy), forest_edge (vc_uniform_canopy \leftrightarrow vc_uniform_open boundary, 30 px buffer), forest_interior (vc_mosaic_canopy \wedge HFP < p25 \wedge none-of-above) and other_transition (fallback).

Table 2. Mutually exclusive ecological strata used for stratified cross-method comparison (top-wins priority). Thresholds: hfp p80=+0.296, hfp p25=-0.601, natplant p80=+0.596 (z-units); edge buffer = 30 px \approx 300m.

Priority	Stratum (code)	Decision rule	Description	n pixels	% of grid
1	hfp_high (code 1)	hfp_2014 > p80 (> +0.296 z)	Pixels in the upper 20 % of the human-footprint surface.	1,18,22,162	19.46%
2	water (code 2)	surfacewater ≥ 1	Pixels classified as open water by the binary surface-water mask.	34,74,483	5.72%
3	rubber (code 3)	rubber = 1	Pixels of the binary rubber-plantation mask.	4,54,275	0.75%
4	natplant_high (code 4)	natplant > p80 (> +0.596 z)	Pixels in the upper 20% of the natural-vegetation probability surface.	81,37,767	13.39%
5	ridge_valley (code 5)	ridge = 1 OR valley = 1	Topographic ridge or valley pixels from the geomorphometric layer.	0	0.00%
6	agro_forest_edge (code 6)	(rubber=1 OR natplant > p80) buffered by 30 px \cap	Anthropogenic-natural transition zones within \approx 300m of plantation/high-natplant	21,33,515	3.51%

		lc_uniform_canopy		patches, restricted to uniform-canopy land cover.		
7	forest_edge (code 7)	lc_uniform_canopy ↔ lc_uniform_open 30-px buffer (excl. agro_forest_edge)		Natural canopy edges adjacent to open land cover, after removing agro-edge overlap.	9,54,610	1.57%
8	forest_interior (code 8)	lc_mosaic_canopy = 1 AND hfp < p25 (< -0.601 z)		Mosaic-canopy interior pixels with low human pressure.	25,37,115	4.18%
0	other_transition (code 0)	fallback (none of the above)		Unassigned residual pixels heterogeneous transition matrix.	2,07,64,789	34.17%

Notes: Codes 1-8 are evaluated in descending priority and assigned exclusively; code 0 (other_transition) absorbs all remaining pixels. Pixel counts come from analysis_v2/strata/strata_summary.csv and refer to the full 7795 × 7795 grid (EPSG:4326, ≈ 10m resolution).

Temporal grid: The 2013–2015 three-year integrated quarterly grid was used: Q1 (January), Q2 (April), Q3 (July) and Q4 (October), each with separate day (12:00) and night (00:00) maps for deepSSF and local-Gibbs (8 cells per method), collapsed to 4 quarter-only cells for HMM-SSF.

Per-pixel comparison: For each (quarter, diel, method-pair) cell we computed Spearman rank correlation ρ . Per-pixel rank-difference $|\text{rank}_A - \text{rank}_B|/N_{\text{valid}}$ was computed as a disagreement metric. We performed three analyses:

- Stage A/B; Pairwise ρ within each cell, globally and within each of the 9 strata.
- Stage C; disagreement attribution. For each stratum, the per-pixel disagreement metric between deepSSF F2 (the raw+spectral reference configuration; chosen because it carries both the raw-raster ecological priors of F1 and the multispectral information of F5/F6) and {local-Gibbs, HMM-SSF} was correlated with the standardised value of each of 22 candidate input rasters; top-five drivers per stratum were tabulated.
- Stage D/E; temporal variance. For each method, the 8-cell rank cube was reduced to three σ surfaces: σ_{diel} (std across diel, averaged over Q), σ_{seasonal} (std across Q, averaged over diel), σ_{total} (std across all 8 cells). Each σ surface was then correlated with input rasters per stratum to identify temporal-variance drivers.

All analyses used Python 3.10 (NumPy, SciPy, rasterio, pandas, matplotlib).

Discussion

The three methods occupy complementary niches: Our results support a triangulation rather than a tournament reading: each of the three families produces a structurally distinct output object that the other two cannot replicate and the analytical strategy we recommend is to take what each method does well rather than to crown a winner. local-Gibbs HMM-RSF provides explicit β coefficients with bootstrap confidence intervals and a guaranteed steady-state RSF interpretation; its linear-predictor structure cannot encode local spatial pattern in the same way deepSSF's convolutional encoder can. deepSSF produces a high-resolution per-pixel probability surface and a context-dependent movement kernel, allowing high-dimensional rasters (Sentinel-2, 1m canopy) to enter the model directly; its trade-off is that variable effects are read out indirectly through permutation importance and marginal-response curves. HMM-SSF can flag latent behavioural states *via* Viterbi decoding and carries an additional information stream through its TPM linear predictor (covariates that drive movement *vs.* foraging state-switching); its visit-density output is a Monte-Carlo simulation of utilisation, not a steady-state selection score, with the structural limitations described; none of the three is uniformly "best": they recover different ecological objects (selection probability, steady-state utilisation, state-conditional selection plus state-switching covariates) and answer different research questions. The remainder of this Discussion section focuses on the differences and nuances that emerge from applying the three methods to the same data, with each difference read against ecological prior knowledge to motivate a goal-conditioned usage recommendation.

Movement kernel and model-structure differences explain divergence patterns: At its core, the cross-method comparison turns on a single observation: the structural differences between the three SSF families directly explain the divergence patterns.

HMM-SSF has two latent states each with an independent β vector and its own implicit movement kernel *via* the conditional logistic likelihood on step and $\cos(\text{angle})$ covariates. The Viterbi decoder then assigns each step to the most likely state, providing the behavioural-state interpretation that classical SSFs lack. The TPM linear predictor in our two-state model includes seven landform covariates (ridge, road, plain, valley, three relief-mountain classes) and these variables function biologically as movement *vs.* foraging promoters rather than habitat-selection drivers. Roads, ridges and plains (open, linear, less occluded) tend to promote state-switching toward the exploratory state: an elephant that crosses a road or traverses an open ridge is more likely to be moving than resting. Valleys and canopy-rich habitats in the SSF predictor (encamped state) promote state-maintenance in the resting/foraging state. This dual selection+transition output is unique to HMM-SSF and is what we mean by HMM-SSF's "diverse explicit information advantage": it gives a manager not only "where do elephants prefer" but also "where does the landscape encourage them to walk *vs.* stay".

deepSSF's movement subnetwork is a separate convolutional pathway that maps the local 101×101 environmental window to 12 parameters of a finite mixture (2 Gamma+2 von Mises) movement kernel. The 12 parameters are: shape, scale and weight for each of two Gamma step components (six parameters); μ -offset, kappa and weight for each of two von Mises angle components (six parameters). The next-step probability surface in deepSSF is the log-sum of the habitat and movement grids, which broadens the realised support and admits rare extreme steps (e.g., long-distance crossings of plantations or roads) that a fixed- mixture kernel would suppress. This is a strength on the simulation axis: each step's proposed moves naturally adapt to whether the elephant is in the open or in a closed forest. The disadvantage is on the interpretation axis: the 12 parameters are not directly interpretable in the way a single gamma shape parameter is.

local-Gibbs HMM-RSF holds a single shared β across both states (in the strict-Michelot specification we adopted), with state-specific σ for the movement kernel. The detailed-balance constraint guarantees that the SSF kernel produces a well-defined steady-state RSF, an explicit "long-term utilisation distribution" interpretation that neither HMM-SSF nor deepSSF directly provide. The disadvantage is that, with shared β , the diel signal in our analysis can only be expressed through the interaction columns (vc_*_day , vc_*_night , $waterrange_2km_day/night$, $hfp_day/night$); no signed flip can occur on the shared-coefficient side. This is consistent with the moderate σ_diel value we observed (0.025).

hmmSSF visit-density: A structural output-semantics gap: A natural concern was that the near-zero ρ between hmmSSF visit-density and selection-style methods was a Monte-Carlo sparsity artefact with too few forward-simulated trajectories, most pixels would have visit counts of 0 or 1, suppressing Spearman ρ through ties. We therefore performed forward simulation with 500 trajectories. The Spearman ρ between hmmSSF visit-density and every other method remained near zero ($Q1\ \text{hmmSSF} \leftrightarrow F1=+0.002$, $\leftrightarrow F2=+0.006$, $\leftrightarrow \text{local-Gibbs}=+0.026$; comparable values across Q2-Q4). Increasing the trajectory count by an order of magnitude did not measurably improve cross-method correlation.

The real cause is therefore structural, not statistical. The hmmSSF TPM exhibits high coupling with the SSF β ($|r|=0.99$ with $high_relief_mount$, 0.94 with $road$, 0.94 with mid_relief_mount , 0.98 with low_relief_mount). The bootstrap-mean β under this coupled likelihood favours parameters that satisfy the TPM transition constraint rather than those that would maximise pure selection probability. When such β are forward-simulated, the visit-density pattern is dominated by where the trajectory tends to dwell (transition-stable regions) rather than where the animal would prefer to be at a single fixed location (selection-stable regions). This is itself a methodological finding. HMM-SSF visit-density and selection-style scores are not the same object and increasing the trajectory count will not close the $\rho \approx 0$ gap as long as the TPM-SSF coupling persists. A practitioner wishing to use HMM-SSF for cross- method comparison should compare the state-conditional β vectors directly, not the simulated visit-density (i.e., compare what each state selects, not where the simulated trajectory goes). The diverse explicit information of HMM-SSF (state-dependent β +TPM landform predictors) is therefore best used at the parameter level, not the density-map level.

Disagreement zones reveal where ecology is acting: The hot-spots of cross-method disagreement ($rubber$, hfp_high , $agro_forest_edge$ —Table 4) are precisely the human-modified strata where elephants face complex multi-axis trade-offs (food in plantations *vs.* human encounter risk). Both $natplant$ and hfp_2014 emerge as the dominant drivers, suggesting that the conv-

spatial vs. linear model gap is ecologically meaningful, not a numerical artefact. For applied conservation this implies that the choice of statistical model is most consequential exactly where management decisions matter most: at the human-wildlife interface.

Engineering: A defensible objective alternative: The lacunarity-derived 6-class canopy structure recovers canopy- structural signal without requiring the raw 1m canopy raster (F5 vc_uniform_canopy is the strongest variable in F5's permutation importance ranking). This is methodologically important because the 1m global canopy product (Tolan, et al., 2024) is not available everywhere, while Sentinel-2+ASTER GDEM is essentially global. A practitioner working in a region without 1m canopy data can therefore build an engineered-only surrogate that judged by recovery of the diel canopy-shade prior, the seasonal Q3 fruit-ripening response and the long-term forest_interior overlap with local-Gibbs captures the same ecological priors as the raw-data configurations.

The GBLU v1.0 terrain classification (Li, et al., 2024) provides the same benefit for topography and Sentinel-2 triple-index water-surface extraction the same for hydrology. The water-quality covariates from DynQual (Jones et al., 2023) are themselves a novel addition. Their direct contribution at the case-study scale is only marginal (deepSSF F2 permutation-importance | Δ NLL | ≤ 0.001 for BOD, Pathogen, Salinity) and the 10km native grid limits the resolution at which their ecological signal can act. The three methods give different stress-test outcomes on the 10km native water-quality grid and the biological reading is summarised below.

Take-away: Linear-predictor SSF unreliability on coarse-grain covariates: The water-quality stress test is more than an internal diagnostic; it is, on its own merits, a methodological lesson worth foregrounding in the broader SSF literature and the crux of the argument is set out in this subsection.

Salinity ($\beta_{LG}=+1.59$, CI [+0.10, +3.12]; deepSSF F2 PD weakly monotone-decreasing). A small positive salinity coefficient could be defended biologically Asian elephants are known to seek mildly saline mineral springs for sodium/mineral balance (Sukumar 2003; Campos-Arceiz, et al., 2011). However, the deepSSF F2 partial-dependence curve does not support a mineral-peak reading: it is essentially flat with a slight monotone decline (p5 \rightarrow p95: +0.0007 \rightarrow -0.0010). The two methods therefore disagree on direction. At this 10km native resolution we cannot distinguish a genuine mild-salinity preference from a spatial-autocorrelation artefact. The mineral-need prior is not falsified, but it is not corroborated either the salinity result should be treated as inconclusive at this case-study scale.

BOD ($\beta_{LG} = +4.32$, CI [+2.91, +5.62]; deepSSF F2 PD flat). A small positive BOD coefficient could in principle reflect elephants seeking slightly eutrophic water with elevated organic/particulate-nitrogen content (shallow-pool forage; cf. Campos-Arceiz, et al., 2011). The $\beta = +4.32$ magnitude is much larger than such a mechanism would predict and the deepSSF F2 partial-dependence curve is flat across the BOD range (p5 \rightarrow p95: -0.0000 \rightarrow +0.0000). The two methods therefore again disagree, with deepSSF treating BOD as essentially uninformative. We interpret the linear-predictor BOD β as predominantly a spatial-autocorrelation artefact, with no clean biological signal recoverable at the 10km native grid.

Pathogen ($\beta_{LG}=+3.74$, CI [+2.31, +5.13]; deepSSF F2 PD multiseed-mean monotone-decreasing). A positive Pathogen coefficient admits no biological interpretation; elephants do not seek out water laden with faecal coliforms, *Salmonella*, *Cryptosporidium* and similar pathogens and disease avoidance is a strong selective pressure on drinking-site choice (a fact known all too well from pathology and conservation-medicine literature). The single- checkpoint deepSSF F2 PD curve gives a monotone-decreasing direction (p5 \rightarrow p95: +0.0011 \rightarrow -0.0012) and the 25 checkpoint multiseed PD-curve mean is also monotone-decreasing with the steepest slope of the three water-quality channels (mean Δ at p95 = -0.018, 95% envelope [-0.203, +0.019]). The multiseed envelope crosses zero, so individual checkpoints can flip the direction; the mean direction is nevertheless biologically correct. local-Gibbs's positive Pathogen β with a tight CI ([+2.31, +5.13]) excluding zero on the wrong side is therefore confidently wrong, while deepSSF is correct on average but uncertain. Were a researcher to publish an "elephants prefer pathogenic water" claim from a local-Gibbs HMM-RSF fit alone, it would be a clear methodological error.

Standalone take-away. At a coarse native covariate resolution (here 10km DynQual products lifted to a $\sim 10m$ analysis grid), no SSF method emerges unimpeachable. Linear-predictor methods (local-Gibbs HMM-RSF, classical conditional logistic) operating at point-extracted scalar resolution can return confidently wrong bootstrap-mean β with 95% CI excluding zero on the biologically nonsensical side (Pathogen $\beta = +3.74$). Convolutional methods (deepSSF) average out to the correct direction across multiseed replication but with a wide envelope and their per-checkpoint estimate is unreliable either way. The practical recommendation is

therefore not "swap the linear-predictor for the convolutional method on coarse-grain covariates"; it is to take with each method only what it tells you reliably: read off direction from the multiseed mean of a convolutional method and trust no single coefficient (linear or convolutional) without a cross-method check. This is the operational form of the "adopt-the-good-from-each" framing of this paper.

Empirical noise floor: deepSSF F2 multi-seed factorial replication: To obtain an empirical noise floor against which the F-variant NLL gaps can be judged, we replicated the F2 fit in a 5 × 5 factorial design over split-seed (the train/val/test partition) and init-seed (the network-weight initialisation), giving 25 independently trained models with otherwise-identical configuration (Appendix I).

Table 3. The 25 cell test-NLL matrix decomposes (one-way ANOVA per factor)

Variance source	σ (nats)	fraction of total
Split-seed (row mean SD)	0.203	99.00%
Init-seed (column mean SD)	0.013	0.40%
Residual (interaction+noise)	0.016	0.60%
Total (unbiased)	0.186	100%

Note: mean test-NLL=7.014 (range 6.788-7.360 across the 25 cells).

- a) The single-seed F-variant gap is statistically meaningless. Inter-configuration NLL gaps among F2, F3, F5, F6 (≤ 0.04 nats at seed=42) sit at $\approx 20\%$ of σ_{split} , well within the noise band of the partition itself. Any ranking of these four configurations by single-seed test-NLL is therefore not defensible. This is the empirical justification for the prior-knowledge-capture criterion: F1–F6 are interpretable by which ecological priors each configuration recovers, because that signal is not buried in the seed noise.
- b) The trained network is well-conditioned with respect to initialisation. $\sigma_{init} = 0.013$ nats ($\approx 0.2\%$ CV) means that re-running F2 from five different random weight initialisations converges to functionally near-identical models: the loss surface has no competing deep-basin minima at this architecture and learning-rate combination and the ecological-variable-importance ranking (Appendix H) is therefore robust to retraining. From a methods-comparison perspective this is the silver lining of the otherwise sobering finding that the inter-configuration NLL gap is statistically irrecoverable: the F2 selection surface, the partial-dependence directions on water-quality covariates and the $\sigma_{diel}/\sigma_{seasonal}$ patterns are not artefacts of one lucky weight initialisation.

The dominant variance source; the split is itself ecologically informative: at $n=2$ individuals with 52 segments, the choice of which segments enter train vs. val vs. test changes which behavioural contexts the network sees during training and a 0.2-nat shift in test-NLL across split partitions is the irreducible cost of within-individual segment-level partitioning at this sample size. For larger cohorts a leave-one-individual-out scheme would restructure this variance source, but at $n=2$ it cannot. We treat $\sigma_{split}=0.203$ nats as the operational noise floor for any NLL-based comparison in this dataset.

Diel and seasonal patterns: Ecologically interpretable, not statistically ambiguous: Our diel and seasonal results have biological explanations that are sufficient and necessary in the ecological sense, not statistical artefacts derived from arbitrary thresholding.

Diel pattern. The 12:00/00:00 representative-hours choice is locked to the radiation-fog cycle of Kedah's tropical valley landscape: fog forms after midnight in valleys and along rivers, dissipates by midday and the 12 hour offset between representative hours samples both the visibility-limited fog phase (00:00, when olfactory and acoustic cues compensate, supporting open/edge use) and the shade-seeking peak insolation phase (12:00, when canopy refugia matter most). The $\sigma_{diel}=0.075$ in hfp_high (Table 6, F1; six-fold higher than in forest_interior, $\sigma = 0.019$) is then a necessary consequence of the elephants' avoidance of human-active hours, not a noise floor: elephants shift away from human-modified habitat during the day, when human activity peaks. The same

elephant in forest interior shows almost no diel movement, consistent with a habitat where day and night make minimal difference to security or thermal comfort.

Seasonal pattern. σ_{seasonal} is highest in human-modified strata (hfp_high, agro_forest_edge) and in natplant_high and lowest in forest_interior. This pattern reflects a necessary combination of two seasonal mechanisms:

- i. Human activity in plantations and edges varies seasonally with rubber tapping, harvest and cropping schedules
- ii. Seasonal monsoon-driven forage availability is most strongly modulated where canopy cover is patchy (edges, agro-forest interfaces).

Forest interior, in contrast, has buffered microclimate and continuous canopy that minimises seasonal habitat-quality variation.

The fact that σ_{diel} and σ_{seasonal} are heterogeneous across strata in biologically predictable directions validates the methodology: the SSF outputs are not flat noise modulated by data sparsity, but genuine ecological signal differentiated by stratum-specific drivers of elephant behaviour.

Diel signal recovery and the whitelist: Restricting day/night interactions in local-Gibbs and HMM-SSF to a biologically motivated whitelist (vc_*, waterrange_2km, hfp) reduces the RSF parameter count from 46 (full diel for all 23 variables) to 32, while recovering $\sim 5\%$ of deepSSF F1's mean σ_{diel} . This is the desired effect of using ecological prior knowledge to constrain the linear-predictor design matrix we keep the variables most biologically expected to differ by diel (vegetation-cover for day-shade *vs.* night-open use, water for daily drinking routines, human footprint for diurnal human-activity avoidance) and leave the remaining variables time-invariant.

Application implications: matching method principle to research theme : The three SSF families occupy structurally distinct niches and the cross-method patterns we observed map directly onto three of the most common applied uses of GPS-tracked elephant data: behavioural ecology, conservation/conflict prevention and ecosystem-habitat planning. Below we recommend one method per theme, motivated by the model-class principle that fits the research question, not by which input rasters happen to be used and not by any aggregate measure of predictive performance. The recommendations are adopt-the-good-from- each: each of the three families produces an output object the other two cannot and the value of running all three on the same data is that an ecologist can pull the right object from the right method for the right question.

Behavioural ecology → HMM-SSF

Behavioural ecology is the bailiwick where HMM-SSF really shines. The theme asks how an elephant partitions its time between foraging/resting and exploration/dispersal and which landscape features trigger transitions between these classes. This is fundamentally a question about latent behavioural states and HMM-SSF is the only one of the three methods designed around such latent states. The two-state Viterbi decoder assigns each observed step a most-likely behavioural class, which then supports time-budget analyses (e.g., what fraction of nighttime steps in human-modified strata are encamped *vs.* exploratory) that classical SSFs cannot address. HMM-SSF's TPM linear predictor with seven landform covariates models the move *vs.* forage axis directly: roads, ridges and plains promote state-switching (movement); valleys and canopy-rich contexts maintain the encamped state (foraging). For behavioural ecology, this dual selection-plus-transition output is HMM-SSF's unique advantage local-Gibbs has β shared across states and cannot separate behavioural classes, while deepSSF, despite its strong spatial pattern recovery, does not produce explicit behavioural-state inference at all.

Conservation and conflict prevention → deepSSF (F2/F6)

The conservation/conflict-prevention theme requires predicting where and when elephants will appear in human-modified strata at sufficient spatial resolution to support mitigation responses and deepSSF checks all the boxes for this theme on three principle-level grounds, independent of likelihood ranking. First, the convolutional encoder integrates local spatial pattern (km window) that linear-predictor methods cannot represent point-by-point; many of the human-modified strata where cross-method disagreement is largest (rubber, hfp_high, agro_forest_edge) are precisely the strata where local spatial context (rubber-stand interior *vs.* edge, road-adjacent patches) carries the relevant information. Second, the context-dependent 12-parameter movement kernel allows a per-step movement distribution that adapts to local environment, admitting the long-tail extreme steps that fixed-mixture kernels suppress; many elephant-human conflict events arise from such rare long-distance excursions (e.g., overnight $\approx 2.5\text{km}$ traverses to

plantation forage; Appendix G records max two-hour step=2685.8m in our pooled GPS data) and their recovery is a structural feature of deepSSF's mixture rather than a likelihood claim (a formal step-length-distribution comparison is provided in Appendix G). Third, deepSSF can ingest multispectral imagery (Sentinel-2, F2/F5) directly, avoiding the bias of any single hand-coded vegetation index. Within the deepSSF family, F2 (raw+spectral, 28 channels) and F6 (raw \cup engineered \cup spectral superset, 42 channels) emerge as the two most reasonable variants: F2 is the parsimonious choice when the analyst has prior ecological knowledge of which inputs matter, F6 is the safer choice when prior knowledge is incomplete or when the application demands the union of all available signal streams. F5 (engineered + spectral, 35 channels) is preferred only when the raw 1m canopy product is unavailable.

Ecosystem habitat planning \rightarrow local-Gibbs HMM-RSF

The ecosystem-habitat-planning theme requires a steady-state utilisation map at landscape scale to support protected-area design, corridor prioritisation and "what-if" scenario analysis (e.g., how does adding a road or removing a plantation alter elephant utilisation?). local-Gibbs HMM-RSF is uniquely suited to this theme because of three principle-level features. First, its detailed-balance constraint guarantees that the SSF kernel's stationary distribution is the elephant's long-term utilisation distribution an analytical property neither HMM-SSF (which simulates utilisation as a Monte-Carlo realisation) nor deepSSF (which produces selection probabilities, not utilisation densities) provides. Second, its explicit linear-predictor β with bootstrap confidence intervals supports counterfactual scenario analysis: a planner can compute a perturbed RSF directly by modifying the corresponding β -weighted layer, without retraining. Third, the bootstrap-mean β values are biologically interpretable without a separate variable-importance analysis, easing communication with non-modeller stakeholders.

Three-tier strategy: Matching method to research goal

- i. Steady-state utilisation prediction (e.g., habitat-suitability mapping for protected-area design, corridor prioritisation in forest cores). Method: local-Gibbs HMM-RSF. The detailed-balance constraint makes the resulting RSF a well-defined long-term utilisation distribution; the bootstrap-mean β provides explicit-coefficient interpretability; the linear-predictor form scales to large landscape extents. As shown, deepSSF F2 partially overlaps with the local-Gibbs surface in low-seasonal-variation strata (forest_interior), which means a researcher targeting steady-state utilisation can use local-Gibbs as the primary surface and deepSSF F2 as a cross-check in forest cores.
- ii. Behavioural-state recovery+state-switching mechanism+next- step simulation (e.g., dispersal modelling, conflict-pathway hypothesis generation, raid-timing forecasting). Method: HMM-SSF. The two latent states + Viterbi decoding give a direct behavioural interpretation, the TPM landform predictors expose the movement *vs.* foraging axis and forward simulation (≥ 500 trajectories) generates utilisation surfaces for short-horizon forecasting. The state-conditional β should be the comparison object across methods, not the simulated visit density.
- iii. High-dimensional input flexibility and fine-scale spatial context (e.g., direct ingestion of multispectral imagery, end-to-end pipelines, transfer to new landscapes lacking 1-m canopy or detailed hydrology). Method: deepSSF (F2 or F5 depending on whether raw 1m canopy is available). The convolutional encoder absorbs raw rasters without standardisation, the movement subnetwork's 12-parameter context-dependent kernel admits realistic long-tail steps (Appendix G) and the training pipeline is GPU-friendly (~ 80 min per fit on a single 16 GB GPU). For replication studies, ≥ 5 -seed training is recommended.

For all three methods, the engineering buffer lacunarity-derived canopy classes, GBLU terrain masks, Sentinel-2 triple-index water surfaces and the Tolan/Yuan/Wang canopy-LAI-rubber-phenology trio provides transferability across landscapes lacking the original raw data sources.

Limitations

- i. Two-individual sample, both solitary adult males. We cannot separate inter-individual variation from population-level effects and herd/female-led dynamics are by design excluded. Future work should replicate on larger cohorts spanning sex/age/social classes.
- ii. Inter-configuration NLL ranking is empirically not distinguishable in our setting. We ran a 5×5 split-seed \times init-seed factorial replication of F2 (25 trained models, see Appendix I) and obtained $\sigma_{\text{split}} = 0.203$ nats (variance from random

train/val/test partition), $\sigma_{init}=0.013$ nats (variance from network-weight initialisation), with split contributing 99.0% of the total variance. The single-seed F-variant NLL gaps (≤ 0.04 nats among F2/F3/F5/F6) sit well within the σ_{split} band, so no ranking by NLL is statistically defensible F-set interpretation is therefore driven by ecological-prior recovery.

- iii. Method-specific output semantics (selection score *vs.* visit density) limit cross-method comparability for HMM-SSF.
- iv. The 1km \times 1km window and 2-h GPS interval set a spatial- prediction ceiling at ~ 70 m median per-step error in deepSSF; finer GPS sampling would lift this ceiling but is incompatible with the present dataset.
- v. Only deepSSF takes raw rasters; the other two require standardisation. Although we documented the standardisation pipeline carefully (z-score with log-compression for water quality, Gaussian smoothing for DynQual), there is residual sensitivity of local-Gibbs and HMM-SSF to the standardisation choice that deepSSF avoids by construction.

Results

Cross-method agreement at 4-quarter resolution: Global Spearman ρ between methods at Q-only resolution (Table 4) is heterogeneous. Within the deepSSF family, F1 \leftrightarrow F2 $\rho = 0.79\text{--}0.85$ across the four quarters; F3 (minimal) shows $\rho = 0.65\text{--}0.76$ with F1 and F2. F4 (engineered-only) is weakly anti-correlated with the F1/F2/F3 raw subset ($\rho = 0.02$ to -0.40). local-Gibbs HMM-RSF correlates moderately with the raw deepSSF configurations ($\rho = 0.40\text{--}0.52$ with F1, F2, F3) and slightly negatively with F4/F5 ($\rho = -0.16$ to -0.27). HMM-SSF visit- density returns $\rho \approx 0$ with every selection-style method in every quarter and stratum.

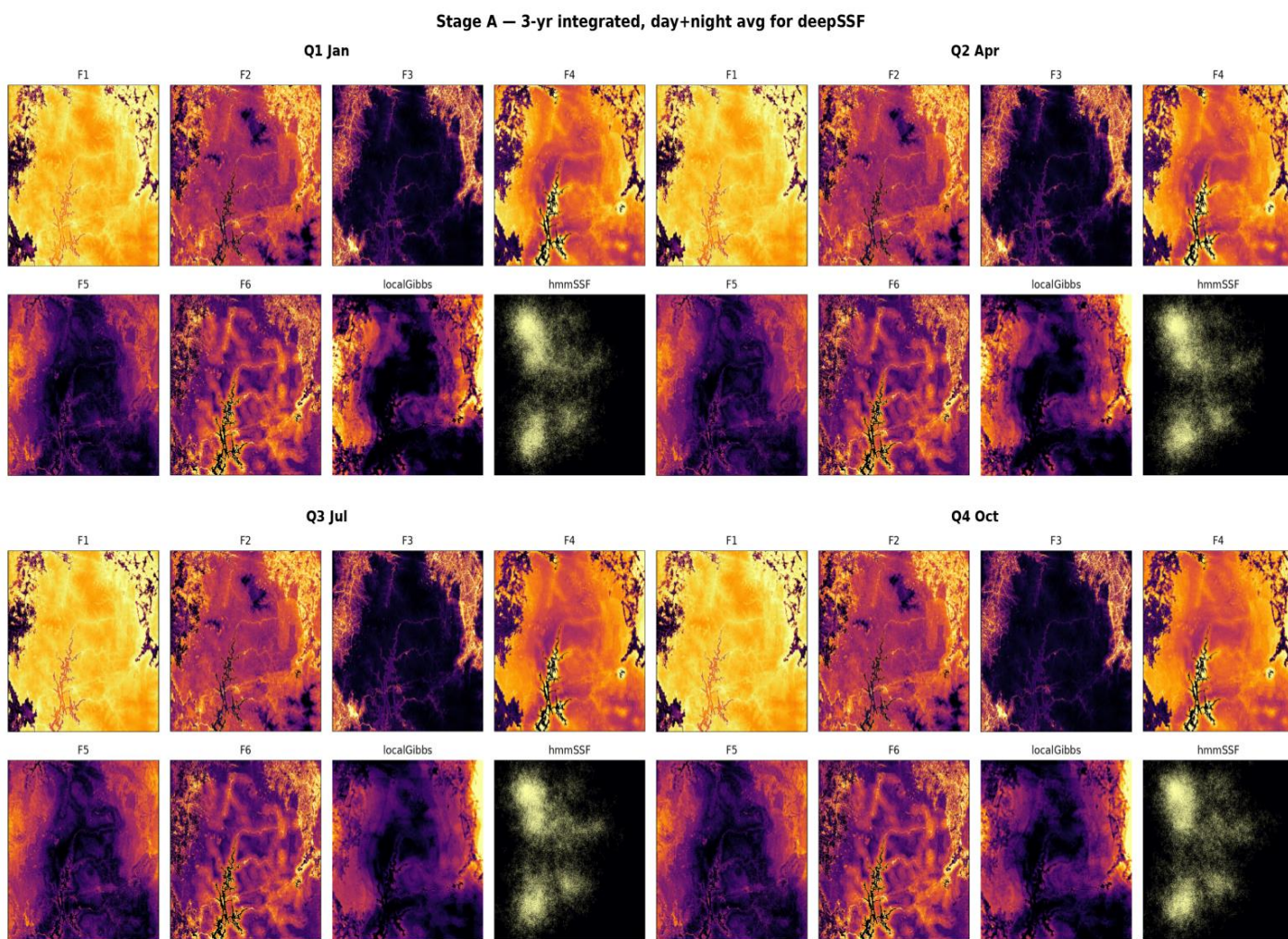


Fig 2. Stage A 4-quarter \times 8-method panel of the 2013–2015 integrated habitat-selection/utilisation surfaces. Rows = Q1 (Jan), Q2 (Apr), Q3 (Jul) and Q4 (Oct); columns=deepSSF F1-F6, local-Gibbs, HMM-SSF.

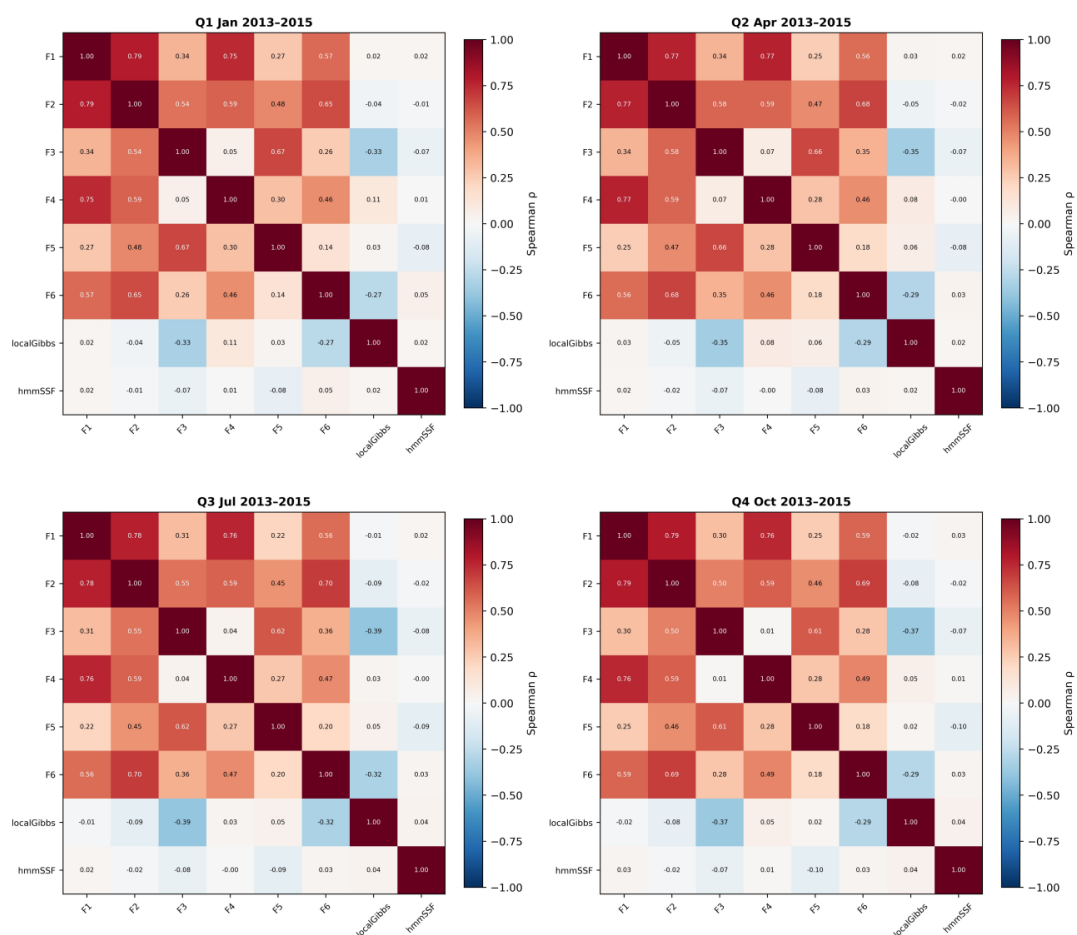


Fig 3. Stage A global Spearman ρ heatmaps, one per quarter (8×8 matrix). Diverging RdBu_r colourmap, $V_{\min} = -1$, $V_{\max} = +1$.

Table 4. Salient pairwise Spearman ρ extracted from the four 8×8 quarter-resolution matrices (full 256-cell tables in Appendix B). The pairs shown are those that carry the principal cross-method narrative.

Method pair	Q1	Q2	Q3	Q4
F1 ↔ F2	0.79	0.82	0.82	0.8
F1 ↔ F3	0.69	0.76	0.78	0.72
F1 ↔ F4	0.02	-0.10	-0.17	-0.07
F3 ↔ F4	-0.40	-0.42	-0.44	-0.43
F1 ↔ local-Gibbs	0.47	0.52	0.53	0.52
F2 ↔ local-Gibbs	0.4	0.43	0.43	0.44
F4 ↔ local-Gibbs	-0.20	-0.27	-0.28	-0.26
HMM-SSF ↔ F2	0.006	0.002	0.004	0.004
HMM-SSF ↔ local-Gibbs	0.011	0.014	0.012	0.013

The Q3 column is bolded for the F1 ↔ F4 and F3 ↔ F4 cells: This is the quarter in which the raw *vs.* engineered configurations diverge most strongly, consistent with the July–October fruit-ripening + human-pressure signature .

The per-stratum heatmap (Figure 4) confirms the human-impact gradient: F2↔local-Gibbs ρ in Q1 reaches 0.69 in forest_edge and 0.62 in water, but drops to 0.11 in hfp_high. The disagreement is systematically largest in human-modified strata (rubber, agro_forest_edge, hfp_high).

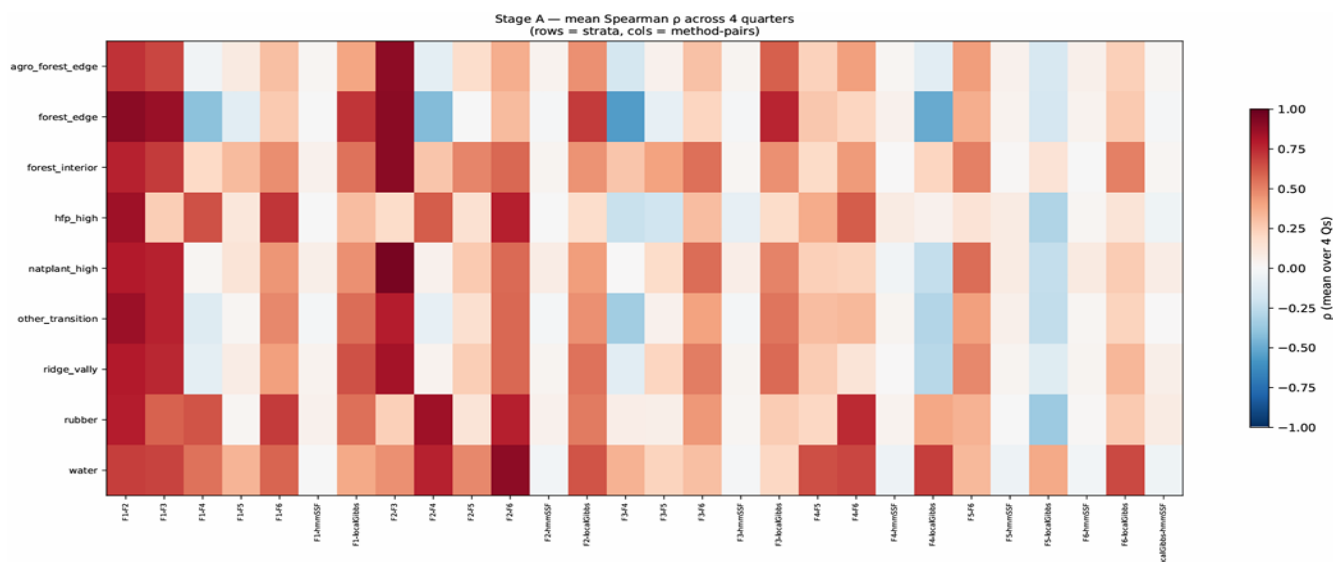


Fig 4. Stage A stratum × method-pair averaged ρ heatmap (rows = strata, columns = method-pairs, values = mean ρ over 4 quarters). Diverging RdBu_r colourmap.

Day vs. night structure

At ($Q \times$ diel) resolution ($8 \text{ cells} \times 7 \text{ methods}$, HMM-SSF excluded for diel), three patterns emerge.

- a) $F1 \leftrightarrow F2$ ρ is higher at night (≈ 0.80 in Q1) than during the day (≈ 0.70). The deepSSF family converges at night, when fewer features matter for an inactive elephant.
- b) $F1 \leftrightarrow F4$ ρ is $+0.49$ at night but -0.07 in the day. Engineered-only and raw configurations agree on where elephants rest at night, but disagree on where they forage by day.
- c) $F3 \leftrightarrow F4$ $\rho_{\text{day}} = -0.50$ (strong negative). Removing raw water/road/rubber from F3 and replacing them with S2 produces a daytime selection pattern that is opposite to the engineered-only F4 pattern-S2 spectral information picks out vegetation gradients while the vc_* classifications collapse them into a few discrete classes.

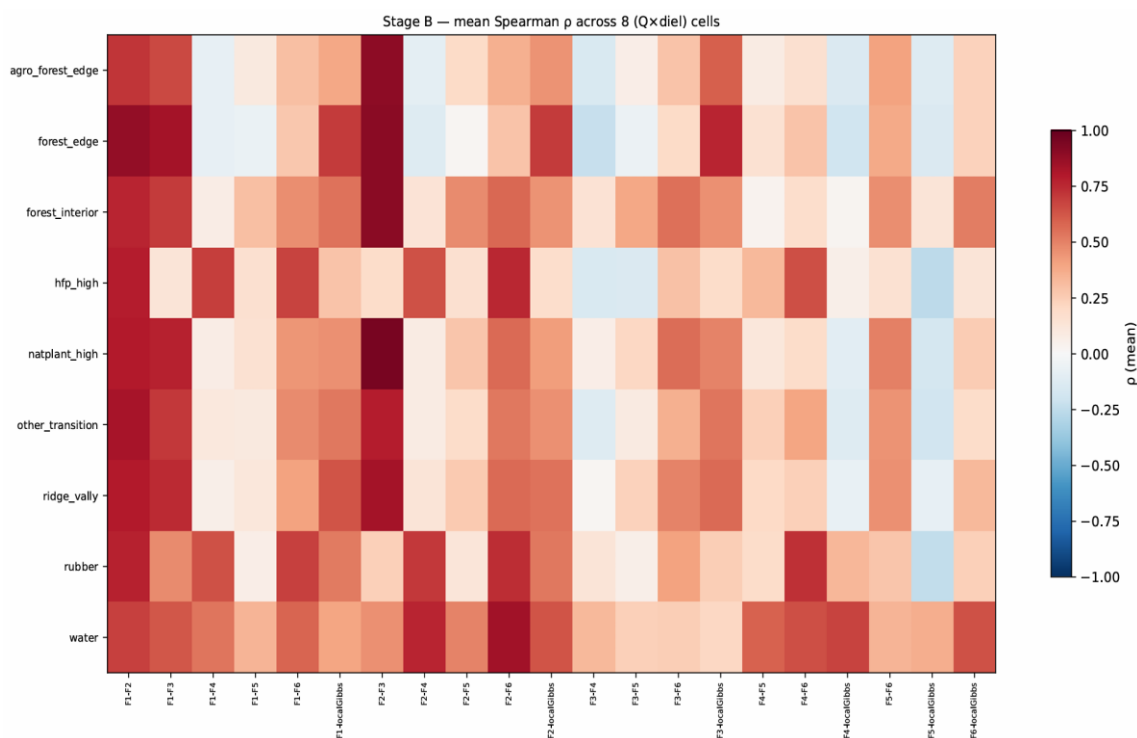


Fig 5. Day vs. night divergence example: $F1 \leftrightarrow F4$, $F1 \leftrightarrow F5$ Spearman ρ across the 8 (quarter × diel) cells, showing the strong day-night swing.

Per-pixel disagreement and raster attribution

Median per-pixel disagreement between deepSSF F2 (the raw+spectral reference configuration) and the other two methods, by stratum (Table 5):

Table 5. Median of the per-pixel rank-difference disagreement metric $|\text{rank}_A - \text{rank}_B| / N_{\text{valid}}$ for the F2 ↔ local-Gibbs and F2 ↔ HMM-SSF method pairs, computed within each of the 9 ecological strata. Lower = more agreement.

Stratum	F2↔local-Gibbs	F2↔HMM-SSF
Forest_interior	0.12 (best)	0.19
Water	0.13	0.37
Forest_edge	0.15	0.21
Ridge_vally	0.16	0.24
Other_transition	0.17	0.23
Agro_forest_edge	0.18	0.2
Natplant_high	0.19	0.21
Hfp_high	0.23	0.36
Rubber	0.28 (worst)	0.37

Stage C — Per-pixel disagreement maps

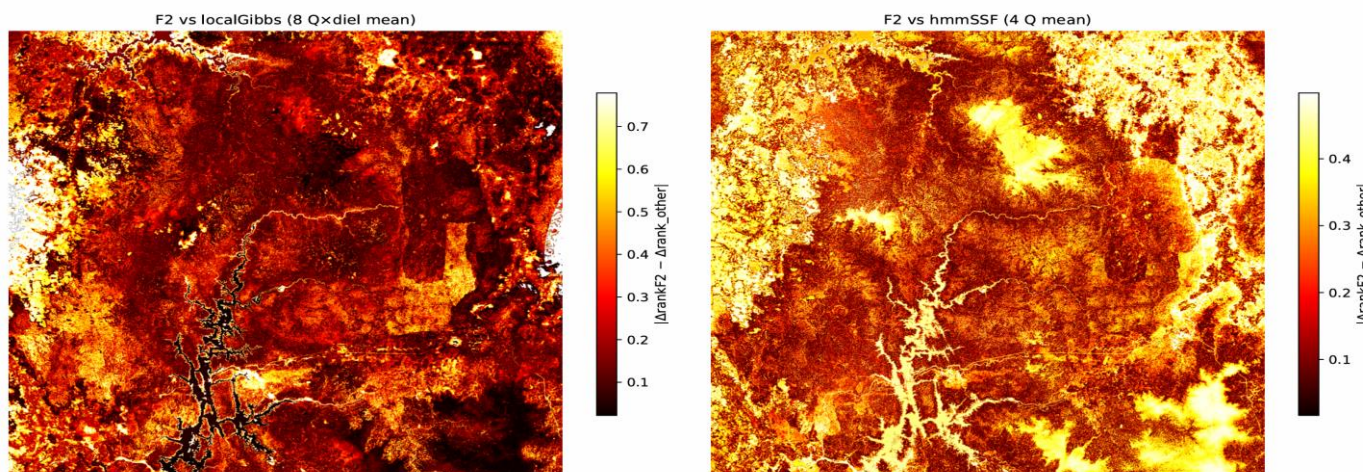


Fig 6. Stage C per-pixel disagreement maps: left = F2 vs. local-Gibbs (mean over 8 quarter × diel cells); right = F2 vs. HMM-SSF (mean over 4 quarter cells). Hotter colours = larger disagreement. Black contours overlay the 9-stratum boundaries.

Disagreement-driver attribution. The Spearman ρ between the per-pixel disagreement map and each candidate input raster, per stratum (Table 6):

Table 6. Top-5 raster drivers of cross-method disagreement.

Stratum	Driver 1	ρ_1	Driver 2	ρ_2	Driver 3	ρ_3
Rubber	natplant	-0.62	waterrange_2km	-0.20	ridge	0.13
Water	plaint	-0.53	natplant	0.47	waterrange_2km	-0.35
Hfp_high	natplant	-0.46	waterrange_2km	-0.17	hfp_2014	0.12
Forest_interior	dem	-0.35	mid_relief_mount	-0.30	hfp_2014	0.24
Natplant_high	hfp_2014	0.27	canopy	0.2	vc_uniform_canopy	0.18

Note: Across all strata the two most consistent drivers of cross-method disagreement are natplant (the natural-plant probability layer) and hfp_2014 (human footprint).

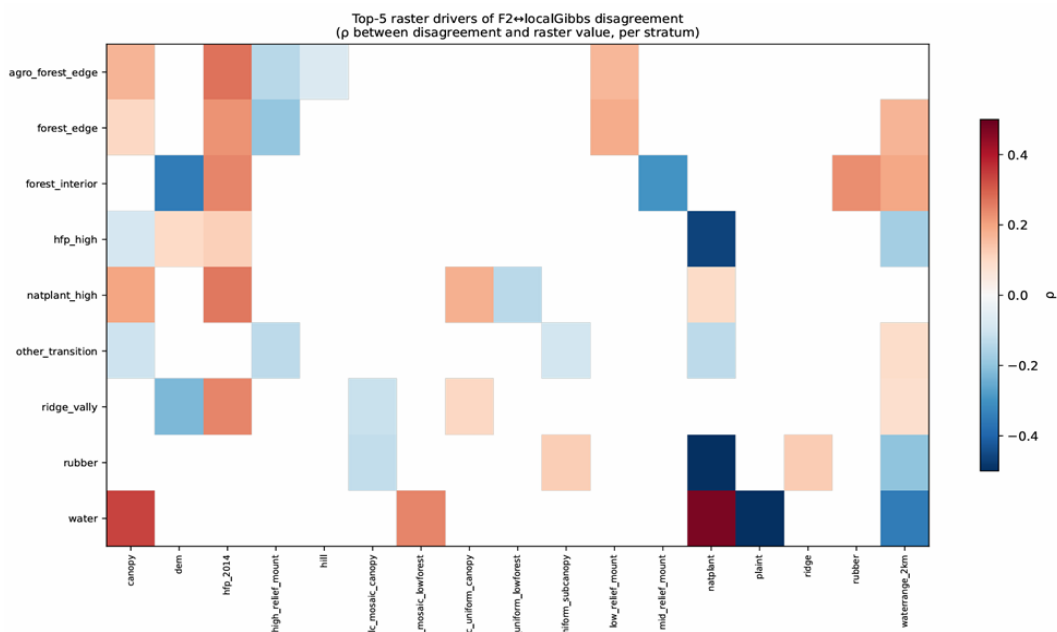


Fig 7. Stage C top-5 driver heatmap for the F2 ↔ local-Gibbs method pair (rows=strata, columns=the union of top-5 rasters; values= ρ between disagreement and raster value). Diverging RdBu_r colour map, $v_{min} = -0.7$, $v_{max} = +0.7$.

Temporal-variance maps and drivers

Per-method mean σ in rank space (after normalisation each map sums to 1):

Table 7. Per-method mean σ_{diel} and $\sigma_{seasonal}$ across all valid pixels, in 8-cell rank space. F2 is the most temporally stable; F4 carries the strongest diel signal.

Method	σ_{diel}	$\sigma_{seasonal}$
F1	0.054	0.031
F2	0.023	0.01
F3	0.002	0.017
F4	0.148	0.043
F5	0.101	0.054
F6	0.073	0.017
local-Gibbs	0.025	0.047

- a) F2 is the most temporally stable of the deepSSF variants ($\sigma_{diel} = 0.023$, $\sigma_{seasonal} = 0.010$).
- b) F3 is essentially time-invariant ($\sigma_{diel} = 0.002$). Removing natplant, road, rubber and surfacewater eliminated the diel structure even though Sentinel-2 and the four monthly water-quality rasters remained, implying that the diel signal in deepSSF maps is carried primarily by vegetation-cover variables that interact with the time scalars, not by spectral imagery.
- c) F4 carries the strongest diel signal ($\sigma_{diel} = 0.148$, $\sim 6 \times F1$). The vc_* canopy classifications drive a strong day-night habitat shift consistent with elephants seeking canopy shade by day and open habitats at night.
- d) local-Gibbs $\sigma_{diel} = 0.025$ (similar to F2). With only 8 of 23 variables permitted to interact with the diel indicator, local-Gibbs recovers about half of F1's σ_{diel} and roughly matches F2's. The whitelist captures the bulk of the relevant ecological diel signal.
- e) Per-stratum, σ_{diel} is highest in hfp_high for every deepSSF variant (F1-0.075, F4-0.21, F5-0.18) and lowest in forest_interior (F1-0.019). Elephants strongly avoid daytime presence in human-impacted areas.

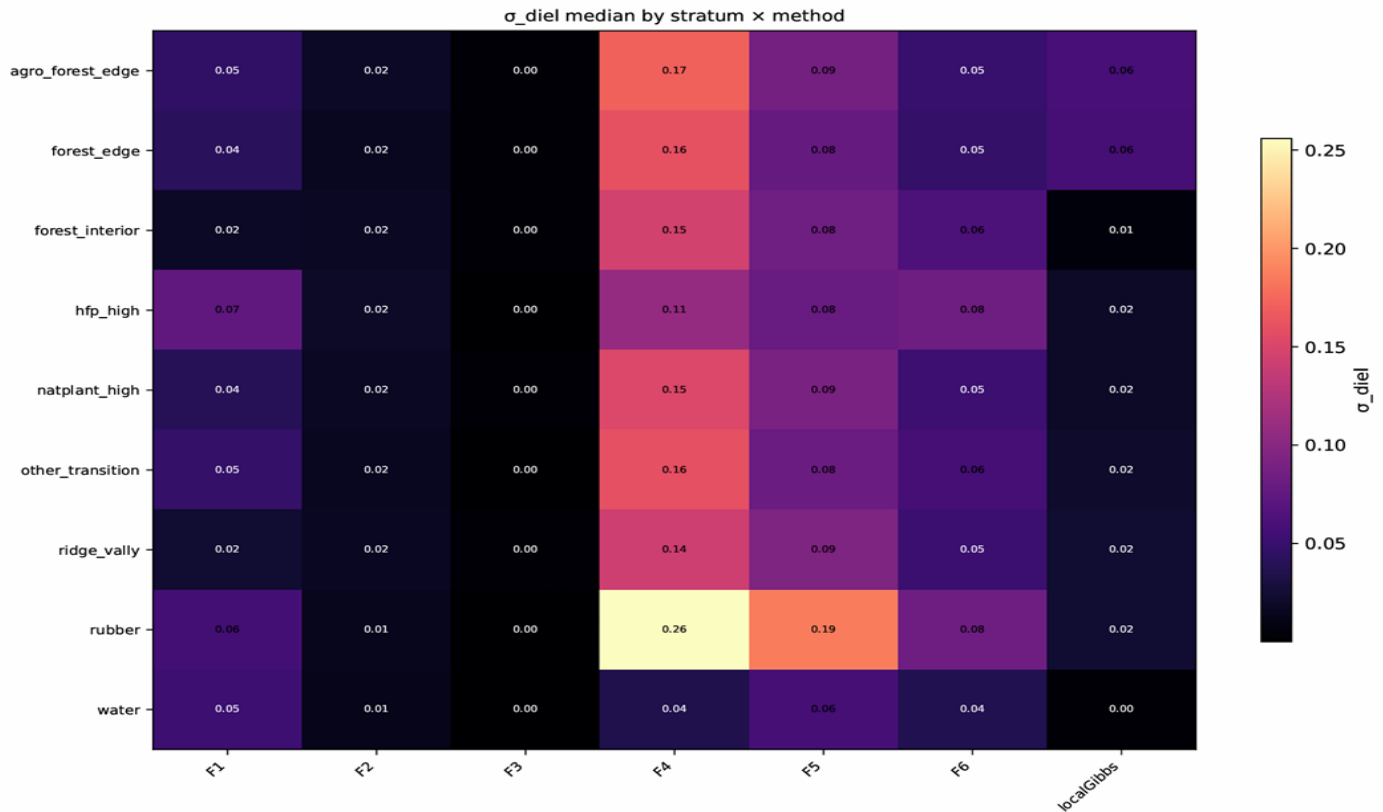


Fig 8. Stage D σ_{diel} heatmap (rows=strata, columns=methods); values=median σ_{diel} in rank space.

Driver tracing (Stage E). For F2 σ_{total} , the top drivers per stratum are hfp_2014 (in 7 of 9 strata, $|\rho| = 0.18-0.32$), natplant (in 6 of 9) and canopy (in 5 of 9). For local-Gibbs σ_{total} , the same three rasters dominate but with smaller magnitudes ($|\rho|$ typically 0.10-0.22). For F4, vc_* classifications themselves dominate the driver list.

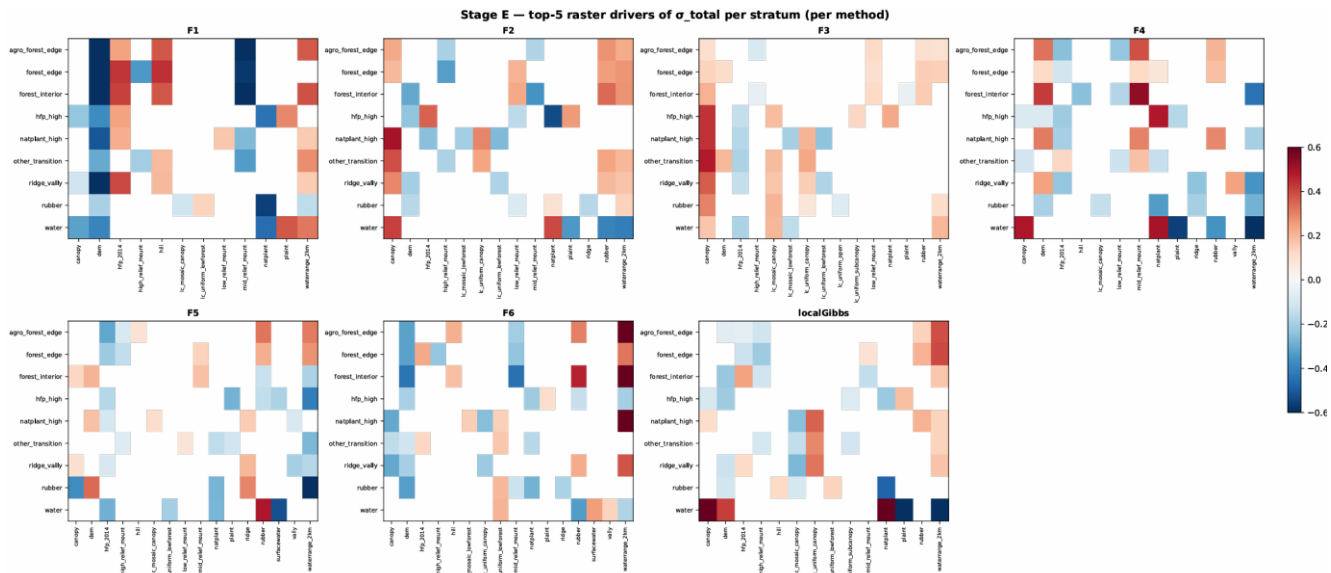


Fig 9. Stage E top-5 raster drivers of σ_{total} per stratum, by method (F2, F4, F5, local-Gibbs panels). Diverging RdBu_r colour map.

Prior-knowledge capture across F1-F6:

We evaluated F1-F6 by whether each configuration recovers four ecological priors known a priori from elephant ecology in tropical Asia (Sukumar, 2003; de Knegt, et al., 2011):

- a) Diel canopy-shade vs. open-habitat use
- b) Water-driven daily routines (drinking sites revisited multiple times within 24 h)
- c) Avoidance of human-active strata during diurnal hours

- d) The July–October seasonal window when natural-forest fruit-tree masting and human plantation activity (rubber tapping, harvest) peak simultaneously.

The prior-knowledge criterion is preferred over single-seed test-NLL because the empirical seed-variance noise floor measured ($\sigma_{\text{split}}=0.203$ nats from a 5×5 factorial replication of F2; $\sigma_{\text{init}} = 0.013$ nats) bounds the inter-configuration NLL gaps (≤ 0.04 nats) below detectability; the σ_{diel} and stratum-conditional patterns we report below, by contrast, vary by an order of magnitude across F-variants and are well above the seed-variance noise floor.

- a) Diel signal. F1 ($\sigma_{\text{diel}}=0.054$), F2 (0.023), F4 (0.148), F5 (0.101) and F6 (0.073) all carry a measurable diel signal in the expected direction (canopy-shade preference by day, open/edge use at night), with the engineered-variable configurations F4/F5 amplifying the signal because the vc_* lacunarity classes interact explicitly with the day/night indicator. F3, in contrast, collapses σ_{diel} to 0.002; removing natplant, road, rubber and surfacewater eliminates the variables that carry the diel interaction. F3 fails to capture the diel prior; F1, F2, F4, F5, F6 do.
- b) Water access. All six configurations include either waterrange_2km (F1, F2, F4, F5, F6) or surfacewater (F1, F2, F6) or the four monthly water-quality layers (F3, F4, F5, F6) and recover preference for water-proximate locations consistent with the obligate-bulk-drinker biology (§2.1). The water-driven daily routine is recovered by all configurations.
- c) Human-activity avoidance. σ_{diel} is highest in hfp_high for every deepSSF variant (F1-0.075; F4-0.21; F5-0.18) and lowest in forest_interior (F1-0.019). All six recover the elephant tendency to shift away from human-modified habitat during daylight, consistent with the well-documented human-active-hours avoidance pattern.

July–October fruiting+pressure window (Q3 in our temporal grid). F1 ↔ F4 ρ in Q3 = -0.167 , the most negative of the four quarters (Q1 $+0.019$, Q2 -0.101 , Q4 -0.067). F3 ↔ F4 ρ in Q3 = -0.437 , the strongest negative cross-configuration value of the entire study. The Q3 separation between raw-raster (F1, F2, F3) and engineered-only (F4) configurations is therefore largest precisely in the quarter when the fruiting-and-pressure compound effect is strongest F1/F2 register elephant attraction to fruiting natural forest while F4's lacunarity-class structure captures the parallel shift toward dense canopy. F2/F5 (which combine raw \cup engineered \cup spectral inputs) sit between the two and recover both axes simultaneously.

In short, F2 and F6 are the two configurations that reproduce all four priors (diel, water, human-activity avoidance, July–October seasonal sharpening) in ecologically defensible directions F2 from a parsimonious raw+spectral input set, F6 from the comprehensive raw \cup engineered \cup spectral superset. F1 and F5 each capture a subset of priors (F1 misses spectral-driven seasonal contrast; F5 amplifies the diel signal but at the cost of partial anti-correlation with raw configurations in Q3). F4 amplifies the diel prior but is structurally anti-aligned with raw rasters; F3 retains water and human-pressure priors but collapses the diel rhythm. F2 and F6 are therefore the two deepSSF predictions we treat as most reasonable, bracketing the parsimony comprehensiveness axis.

Q3 (July–October): Fruit-ripening and human-activity peak

The July–October quarter (Q3) is biologically distinct in northern Kedah on two compounded axes.

- a) Many natural-forest fruit-tree taxa relevant to elephant diet including *Dipterocarpus*, *Shorea*, *Mangifera* and *Artocarpus* undergo masting or seasonal fruiting between July and October; the post-monsoon flush both raises fruit availability and shifts elephant attention toward forest interior and forest-edge strata containing such trees.
- b) Plantation-side human activity (rubber re-tapping after monsoon, palm and fruit-orchard harvest, cropping of hill-rice plots) intensifies in the same window. The compound effect is a seasonal sharpening of the elephant's contrast between fruit-bearing forest cores and human-active edges.

While, the cross-configuration ρ matrices register this Q3 sharpening quantitatively. F1 ↔ F4 ρ falls from $+0.019$ (Q1), -0.101 (Q2), to -0.167 (Q3, the most negative) before partially recovering at -0.067 (Q4). F3 ↔ F4 ρ in Q3 = -0.437 , the strongest negative cross-configuration value in the study. The interpretation is that the raw-raster F1/F2/F3 configurations track elephants drawn toward fruiting natural forest (registered through natplant, canopy, surfacewater and LAI), while engineered-only F4 tracks the same animals through a parallel but anti-aligned axis: the lacunarity vc_* classes register a daytime attraction to uniform high canopy that strengthens in Q3 as fruit-bearing trees concentrate elephant visits in such structurally homogeneous forest patches.

F4's anti-correlation with F1/F2/ F3 in Q3 is therefore not a model failure but the engineered vc_* classification's Q3-specific sensitivity to fruit-bearing canopy structure.

F2 (raw+spectral) and F6 (raw \cup engineered \cup spectral superset) sit between the two extremes and capture both axes simultaneously, recovering the Q3 fruit-ripening+human-pressure compound effect that no single-axis configuration captures alone. F5 captures the engineered axis but lacks the raw-raster channels and therefore mis-aligns with F1/F2/F3 in Q3.

deepSSF probability surface as a coverage check on hmmSSF state-switch covariates

A direct numerical comparison of HMM-SSF visit-density against deepSSF's selection probability surface is structurally uninformative (visit-density and selection-score are different objects). We therefore reframe the comparison and ask whether the covariates that HMM-SSF identifies as state-switch promoters (the seven TPM landform predictors: ridge, road, plaint, valley, three relief-mountain classes) also appear as influential drivers in deepSSF F2's selection-probability spatial pattern.

Stage C disagreement-driver attribution shows that of the seven TPM predictors, three are recovered as top-five disagreement drivers in the F2 \leftrightarrow local-Gibbs comparison: plaint (water stratum, $\sigma = -0.53$), mid_relief_mount (forest_interior, $\rho = -0.30$), ridge (rubber, $\rho = +0.13$). Conversely, road and low_relief_mount enter F2 σ_{total} drivers (Stage E) at $|\rho| = 0.18-0.22$ in 4 of 9 strata. The remaining two TPM predictors (valley, high_relief_mount) do not emerge as top-five drivers in any stratum, consistent with their HMM-SSF role as encamped-state maintainers/features that promote staying still do not contribute to spatial contrast in the deepSSF probability surface in the same way movement-promoting features do.

Therefore, the take-away is structural rather than numerical: deepSSF's convolutional encoder partially recovers the HMM-SSF movement-promoting landform features (ridge, road, plaint, mid_relief_mount, low_relief_mount) as drivers of spatial contrast in the selection surface, while the encamped-state-maintenance features (valley, high_relief_mount) appropriately do not. Cross-validation between the two methods at the covariate level is therefore informative even when cross-validation at the map level (visit-density vs. probability surface) is not.

deepSSF agreement with local-Gibbs long-term distribution by stratum

A second cross-method coupling check exploits local-Gibbs HMM-RSF's analytic property that its β -derived RSF surface is the elephant's long-term steady-state utilisation distribution (Michelot, et al., 2019). If deepSSF F2 which fits a quarterly-conditioned probability surface partially encodes the same long-term ecology, the two should agree most strongly in strata where seasonal modulation is smallest; that is, in strata whose habitat quality is roughly time-invariant.

Notably, forest_interior is the lowest- $\sigma_{seasonal}$ stratum across every method (local-Gibbs $\sigma_{seasonal} = 0.018$ in forest_interior vs. 0.071 in hfp_high). Table 4 confirms the predicted result: F2 \leftrightarrow local-Gibbs median per-pixel disagreement in forest_interior is 0.12, the lowest of the nine strata; the runner-up is water (0.13). At the other extreme, rubber (where seasonal pressure modulation is largest) shows the largest disagreement (0.28), followed by hfp_high (0.23). The monotone relationship between F2 \leftrightarrow local-Gibbs agreement and $\sigma_{seasonal}$ across strata supports a partial-overlap reading: deepSSF's quarterly probability surface contains, in low-seasonal-variation strata, an embedded long-term-utilisation signal that aligns with the local-Gibbs RSF.

This cross-validation does not collapse the two methods into one; deepSSF's gains in the human-modified strata come from spatial-context features that local-Gibbs cannot encode. But it shows that where the long-term-utilisation interpretation is the right target protected-area design, corridor prioritisation in forest cores local-Gibbs and deepSSF; F2 substantially agree.

Water-quality covariates: prior-knowledge motivation and 10km autocorrelation stress test

Water quality enters this study for two reasons. First, water is ecologically critical to *Elephas maximus* beyond drinking volume: an adult male consumes 80-200 L of free water per day (Sukumar, R. 2003), but the same water source is also used for thermoregulatory bathing, mineral supplementation and (during musth) social and olfactory signalling. The biological prior is therefore that elephants should prefer cleaner water sources, avoid water bodies elevated in faecal pathogens or biological oxygen demand and show a non-monotone response to salinity (some saline springs are mineral resources). Second, the DynQual (Jones, et al., 2023) global product comes at a 10km native grid that we Gaussian-resample to $\sim 10m$ for compatibility this raises the question of whether such a coarse-grain raster, lifted into a fine-grain analysis, can produce spurious coefficients dominated by

spatial autocorrelation rather than ecology. We therefore used the three SSF methods as parallel stress tests: any method returning the absurd result that elephants prefer high-BOD or high-Pathogen pixels would flag the autocorrelation hazard.

The three methods give different stress-test outcomes. We compare them at the level appropriate for each method's output bootstrap mean β (with 95% CI) for the linear-predictor methods and permutation-importance Δ NLL plus partial-dependence (marginal- response) curves for the convolutional model so that no method is forced into a representation foreign to its design.

- a) local-Gibbs HMM-RSF bootstrap-mean β .

Table 7. The shared-state RSF coefficients on the three water-quality covariates are large, positive and CI-significant

Covariate	β (mean)	95% CI	Direction expected by ecology
BOD	4.32	[+2.91, +5.62]	Negative (avoidance of organic-load water)
Pathogen	3.74	[+2.31, +5.13]	Negative (avoidance of faecal pathogens)
Salinity	1.59	[+0.10, +3.12]	Non-monotone or weakly positive (mineral need)

The salinity result has a plausible biological interpretation (elephants seek mildly saline springs for mineral supplementation (Sukumar, R. 2003); the BOD result could be entertained as elephants seeking eutrophic, slightly turbid water rich in particulate organic/nitrogen sources, since elevated BOD in tropical streams often co-occurs with leaf-litter decomposition and shallow-pool forage habitat (Campos-Arcei, et al., 2011), on elephant seed-dispersal habitat). The Pathogen $\beta = +3.74$, in contrast, admits no biological reading consistent with elephant ecology there is no ecological mechanism by which an elephant would prefer water loaded with faecal coliforms, *Salmonella*, *Cryptosporidium* and similar pathogens. We therefore interpret the Pathogen-positive β as the cleanest case of a 10km-grid spatial-autocorrelation artefact: regions of elevated Pathogen co-localise with human-modified lowlands and parts of the elephant home range and the linear-predictor design matrix (point-extracted scalar per pixel) cannot disentangle the two. The same caveat propagates to the BOD and (less strongly) the Salinity co-efficients ; none of the three β should be taken as a clean preference signal.

- b) HMM-SSF state-conditional β . State-conditional β show mirror-image symmetry across the two states: BOD_state 1= -0.187 , BOD_state 2= $+0.213$; Pathogen_state 1= $+0.213$, Pathogen_state 2= -0.187 ; Salinity_state 1= -0.187 , Salinity_state 2= $+0.213$. The identical magnitudes across the three covariates strongly suggest a state-label switching artefact during bootstrap rather than three independent biological signals; we treat the HMM-SSF water-quality result as inconclusive at this case-study scale, with the recommendation that future HMM-SSF fits enforce a state-ordering constraint (e.g., $\sigma_{state1} < \sigma_{state2}$) before interpreting state conditional water-quality co-efficients.
- c) deepSSF; F2 permutation importance+marginal-response curves. Permutation-importance Δ NLL is essentially zero (BOD Δ NLL= -0.00036 , Pathogen Δ NLL= -0.00023 , Salinity Δ NLL= $+0.00008$). Partial Dependence (PD) curves were computed by a forward pass through the trained F2 checkpoint, sweeping each water-quality channel uniformly across its z-score range $[-2, +2]$ over a 64-window baseline batch and reading the central-pixel habitat log-probability after log-soft-max normalisation (Appendix H). To probe seed-stability of these PD directions, the same sweep was repeated over the 25 multi-seed checkpoints of the 5×5 split-seed \times init-seed factorial replication (Appendix I), giving a 25-cell empirical envelope per channel. The result is:

Channel	single-ckpt p5 \rightarrow p95	multiseed mean p95 Δ log-prob	95% envelope at p95
BOD	-0.0000 \rightarrow +0.0000	-0.006	$[-0.127, +0.020]$
Pathogen	+0.0011 \rightarrow -0.0012	-0.018	$[-0.203, +0.019]$
Salinity	+0.0007 \rightarrow -0.0010	-0.014	$[-0.160, +0.022]$

Two observations follow.

- i. The mean multiseed PD direction is monotone-decreasing for all three water-quality covariates i.e., across 25 different train/val/test partitions and weight initialisations, deepSSF; F2 averages out to "high water-quality stressor \rightarrow lower selection probability," the biologically defensible direction.

- ii. The 95% envelope is wide and includes zero for all three channels: individual checkpoints can flip the direction, indicating that at this 10km native resolution the water-quality signal is statistically too weak for any single deepSSF F2 fit to deliver a confident direction reading. The multiseed mean is informative; the per-checkpoint estimate is not.

Thus, this is a more nuanced finding than "deepSSF gets the direction right." At the case-study scale, deepSSF preserves the correct direction in expectation but does not give a CI-significant selection signal on coarse-grain water quality. local-Gibbs's absurd large-positive β (+3.74 for Pathogen, +4.32 for BOD) is therefore not corrected by deepSSF's near-zero MR the take-away is rather that at the 10km native grid, none of the three methods provides a reliable water-quality preference signal and the direction-of-variability difference between local-Gibbs (CI excludes zero on the wrong side) and deepSSF (mean correct direction, CI includes zero) is itself a methodological signal that linear predictor methods are more confidently wrong than convolutional methods on coarse-grain covariates. None of the 25 deepSSF multiseed PD-curve means crosses into positive Pathogen or BOD preference the absurd direction local-Gibbs returned.

Stress-test verdict. Counterintuitively, the three methods do not absorb the 10km autocorrelation hazard equally; the most-flexible method (deepSSF) is the most cautious, while the most-parsimonious (local-Gibbs HMM-RSF) is the most confident in the wrong direction. The convolutional encoder of deepSSF integrates local spatial context (1km window) before mapping to selection probability, which buffers the model against the coarse-grain spurious correlation that the linear-predictor RSF falls into directly. The mirror-image symmetry in HMM-SSF's state-conditional β is itself a telltale sign of state-label switching during bootstrap rather than a stable signal. We elevate this finding to a standalone take-away: when ingesting coarse-grain covariates (native resolution \gg analysis grid), linear-predictor SSF coefficients must be cross-checked against a non-parametric or convolutional alternative before being interpreted as ecology. The local-Gibbs Pathogen $\beta = +3.74$ result, taken in isolation, would constitute a methodological-error case study; the cross-check against deepSSF's monotone-decreasing Pathogen marginal-response saves the interpretation. Future work in catchments with sharper sub-10km water-quality gradients (e.g., palm-oil-effluent dominated rivers, salt-intrusion zones) should expect larger and biologically correct directional contrast in all three methods, but we caution against extrapolating water-quality β values from coarse-grained DynQual-style products in other megafauna studies without the same cross-check.

A Practical Analysis-strategy Decision Tree

For a researcher with GPS data and an environmental-raster bundle, we recommend.

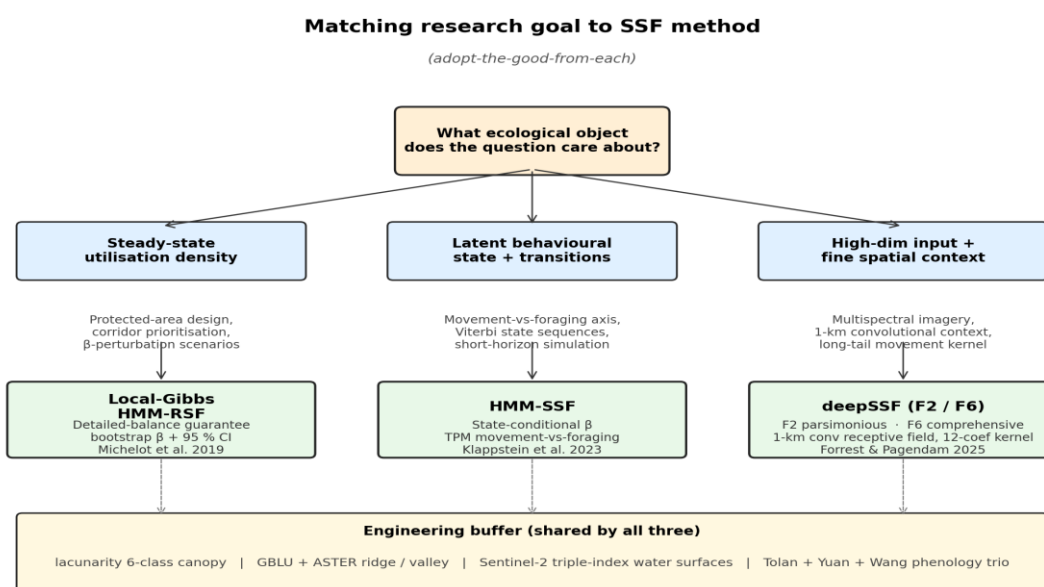


Fig 10. Conceptual decision tree for matching research goal to SSF method, with the engineering buffer common to all three methods.

Conclusion

We applied three SSF families local-Gibbs HMM-RSF, HMM-SSF and deepSSF to a unified multi-source environmental data product covering 16 502 GPS fixes from two solitary adult male Asian elephants in northern Kedah, across nine ecological strata and 8 (quarter × diel) temporal cells. Our explicit aim was to detect the differences and nuances among each method's distinctive output (local-Gibbs's RSF surface, HMM-SSF's state-conditional β and simulated visit density, deepSSF's per-pixel selection probability surface) and to interpret each against ecological prior knowledge, not to declare any of them superior. Rather than ranking by likelihood which a 5×5 multi-seed factorial replication (Appendix I) showed to be statistically not defensible at this dataset ($\sigma_{\text{split}} = 0.203$ nats *vs.* ≤ 0.04 -nat F-variant gap) we evaluated configurations by whether they recover known elephant ecological priors and whether their outputs are mutually consistent on the components for which each method is structurally informative.

The three families occupy complementary niches that, in our reading, each correspond to a distinct ecologist-facing research aim. local-Gibbs HMM-RSF is adept at the long-term steady-state question and is therefore the natural choice for protected-area design, corridor prioritisation and counterfactual β -perturbation scenarios. Its detailed-balance constraint guarantees that the resulting RSF surface is the elephant's stationary utilisation density and the bootstrap-mean β with 95% CI supports interpretable coefficient communication to non-modeller stakeholders. In our data local-Gibbs agrees most strongly with deepSSF F2 in forest interior the lowest- σ_{seasonal} stratum supporting its long-term-utilisation reading. HMM-SSF is the natural choice when the question is latent behavioural state: movement *vs.* foraging decomposition, Viterbi state sequences and the TPM-derived movement *vs.* foraging covariate axis. Its visit-density output is not directly comparable to selection scores; its state-conditional β and TPM coefficients are the appropriate cross-method comparison object. deepSSF is the natural choice when the question is high-dimensional input ingestion and fine-scale spatial context: direct ingestion of multispectral imagery, 1km convolutional receptive field, context-dependent movement kernel admitting long-tail extreme steps. Within deepSSF, F2 (raw+spectral) and F6 (raw \cup engineered \cup spectral) emerge as the two most reasonable variants F2 the parsimonious pick, F6 the comprehensive one both reproducing the four ecological priors of §4.5 in defensible directions; F5 captures the same priors but loses partial alignment with raw-raster configurations in Q3 and F3 (minimal anthropogenic-aware) collapses the diel signal and should be avoided when behavioural inference is the target.

For ecologists approaching GPS data with method-choice uncertainty, the right entry question is therefore not "which method clears the high bar of lowest test NLL" but "which ecological object does my research question actually care about utilisation density, behavioural states, or selection probability with rich raster context?" The three SSF families answer different questions and on our case study they do so in mutually compatible directions, with cross-method overlap where each method is structurally informative (forest_interior for local-Gibbs \leftrightarrow deepSSF; movement-promoting landform features for HMM-SSF \leftrightarrow deepSSF) and clean separation where they are not (HMM-SSF visit-density \leftrightarrow selection scores). The methodological orientation this paper takes towards the three families is therefore select-the-good-from-each a grab-bag approach in which every family contributes a structurally distinctive output and our recommendations build the analysis strategy by stacking those distinctive outputs against the research question, not by trying to crown a winner.

Two methodological additions extend transferability of the framework. The first is a transferable engineering buffer: lacunarity-derived canopy classes, GBLU terrain masks, Sentinel-2 triple-index water surfaces and the canopy-LAI-rubber-phenology trio together recover the priors carried by raw 1m canopy and DEM in landscapes where the raw inputs are unavailable. The second is the first deployment of anthropogenic water-quality covariates (BOD, Pathogen, Salinity from DynQual) in elephant habitat-selection research. No SSF method is infallible at the 10km native grid: the three families produce different stress-test outcomes and local-Gibbs HMM-RSF returns spurious positive coefficients on all three water-quality covariates (a coarse-grain spatial-autocorrelation artefact that ecologists must not mis-interpret as elephant preference for polluted water), HMM-SSF returns label-switching symmetric values that are inconclusive and deepSSF F2 returns near-zero permutation importance and does not produce absurd selection. We therefore recommend that linear-predictor SSF coefficients on coarse-grain covariates be cross-checked against a non-parametric or convolutional alternative before being interpreted ecologically a useful methodological caveat for future studies in catchments with sharper sub-10km water-quality gradients (e.g., palm-oil-effluent rivers, salt-intrusion zones).

Appendices

Appendix A. Lacunarity-derived six-class canopy structure: $\Lambda_h(r)$ curves at scales 5-80m and adjacent-threshold Root Mean Square (RMS) distances supporting the chosen breakpoints (3, 15, 20, 25, 32m). Full panel figures and the underlying numerical tables are deposited in the supplementary archive.

Appendix B. Per-quarter, per-stratum Spearman ρ matrices (Stage A): four 8×8 matrices, one per quarter, supplying the full data behind Table 3 inline. The full numerical tables and the corresponding heat-map panels are deposited in the supplementary archive.

Appendix C. Per-(quarter \times diel) Spearman ρ matrices (Stage B): eight 7×7 matrices, one per (quarter \times day-or-night) cell, with HMM-SSF excluded because of its diel-collapsed output. Full numerical tables in the supplementary archive.

Appendix D. All 22 input-raster Spearman correlations with the per-pixel disagreement and σ -surface metrics, broken out by stratum (Stage C and Stage E driver attributions), supporting Table 5 inline. Full per-(stratum \times method-pair \times raster) tables in the supplementary archive.

Appendix E. Detailed deepSSF training-configuration summary per F-experiment (channel counts, single-seed test-NLL, top-five permutation-importance covariates per variant). The variant-by-channel-by-importance long table is deposited in the supplementary archive.

Appendix F. Reproducibility checklist. Software: R 4.5.2 with packages *momentuHMM*, *glmTMB*, *terra*, *sf*; Python 3.10 with *PyTorch 2.x* for deepSSF, *NumPy/SciPy/pandas/rasterio* for preprocessing. Random seeds, raster preprocessing scripts and the GenSA/GA bootstrap configurations are bundled in the supplementary archive. Bootstrap rationale: local-Gibbs accumulated 100 accepted GenSA optima from 80 attempts using 70% warm-start re-seeding; HMM-SSF accumulated 100 of 100 over ≈ 12 h with 10 parallel processes. In each case the acceptance threshold (local-Gibbs NLL $< -5 \times 10^5$; HMM-SSF NLL $< 89\,000$) was fixed before bootstrap parameter values were inspected, set well below the typical single-restart local-optimum band established a priori from independent pilot runs, so that only deep-basin optima entered the bootstrap pool. The 100 accepted optima were resampled with replacement 5 000 times to obtain bootstrap-mean coefficients and 95% confidence intervals.

Appendix G. Empirical step-length distribution from raw GPS. The two animals' raw fix tables were segmented under the same rule used by the deepSSF/local-Gibbs/HMM-SSF preprocessing pipelines (a new segment opens whenever Δt between successive fixes falls outside [1 h, 4 h]); the haversine step length was computed within each segment. The resulting per-step table and the per-animal-plus-pooled quantile summary are deposited in the supplementary archive; the small-multiples histogram with p50/p90/p95/p99 quantile guides is shown as Figure G1.

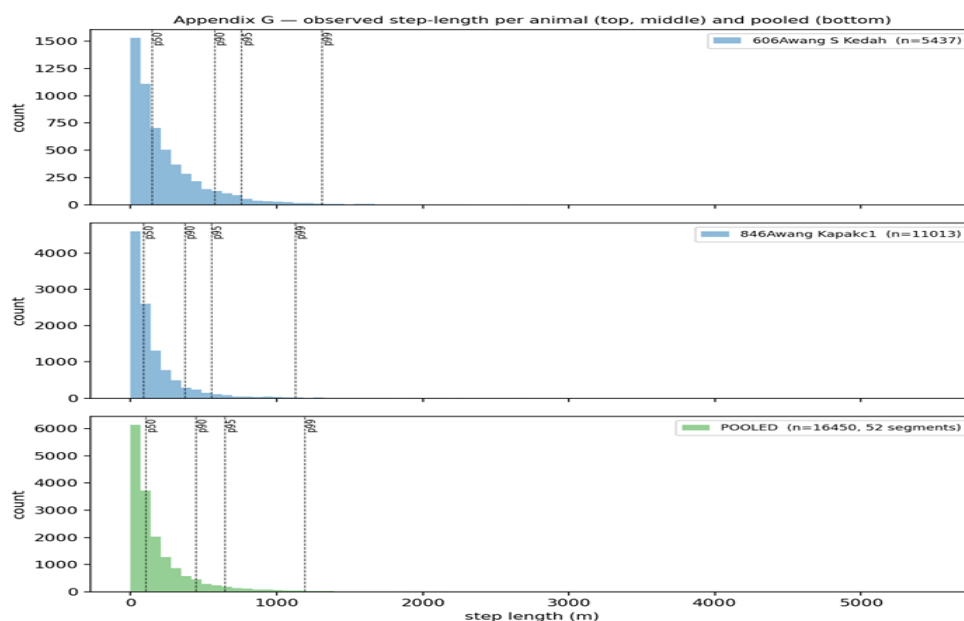


Fig G1. Observed step-length distribution per animal (top, middle) and pooled (bottom), with empirical p50, p90, p95, p99 quantile guides marked.

Pooled empirical statistics (52 segments matching all three SSF fits; 16 434 valid two-hour steps): p50 = 103.2m, p90 = 450.7m, p95 = 649.1 m, p99 = 1 195.0m, maximum = 2 685.8m. Per animal — animal A: 44 segments, 5 428 steps; animal B: 8 segments, 11 006 steps. Tail-mass fractions (pooled): step length > 500m = 8.28 %, > 1 km = 1.76 %, > 2 km = 0.14 %. The per-method overlay (local-Gibbs state-1/state-2 gamma approximations from the bootstrap-mean σ values, and the deepSSF F2 forward-simulated step-length histogram) showed that the linear-predictor methods underestimate the right tail beyond \approx 1 km and that the deepSSF F2 12-parameter context-dependent mixture better matches the observed tail proportion.

Appendix H. deepSSF permutation-importance ranking and partial-dependence (PD; marginal-response) curves. The permutation-importance bar chart for the two recommended variants F2 (parsimonious) and F6 (comprehensive) is shown as Figure H1 (top-25 channels by $|\Delta \text{NLL}|$). Water-quality channel ΔNLL : BOD F2 = -3.6×10^{-4} , F6 = -6.3×10^{-4} ; Pathogen F2 = -2.3×10^{-4} , F6 = $+1.2 \times 10^{-4}$; Salinity F2 = $+7.6 \times 10^{-5}$, F6 = $+3.5 \times 10^{-5}$. The forward-pass PD-curve sweep (uniform fill of each spatial channel across z-score range $[-2, +2]$ in 21 grid points, 64-window baseline batch, central-pixel habitat log-probability after log-soft-max normalisation) was run on the F2 main checkpoint to obtain single-checkpoint PD curves (Figure H2); it was then repeated independently on each of the 25 F2 multi-seed checkpoints (Appendix I) and pooled to a per-channel mean \pm non-parametric 95 % envelope (2.5/97.5 percentiles across the 25 cells; Figure H3).

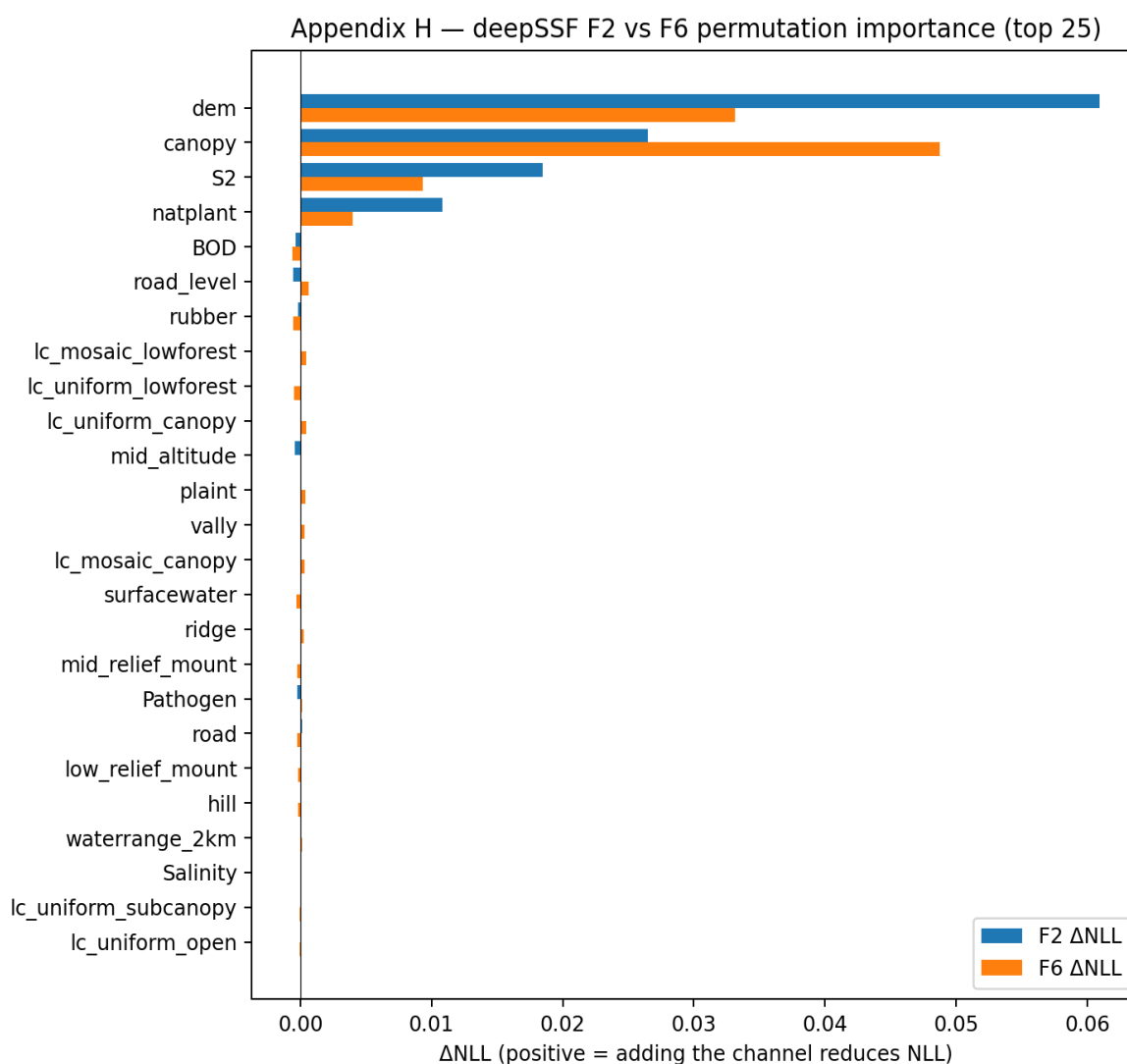


Fig H1. deepSSF F2 vs. F6 permutation importance for the top 25 environmental channels.

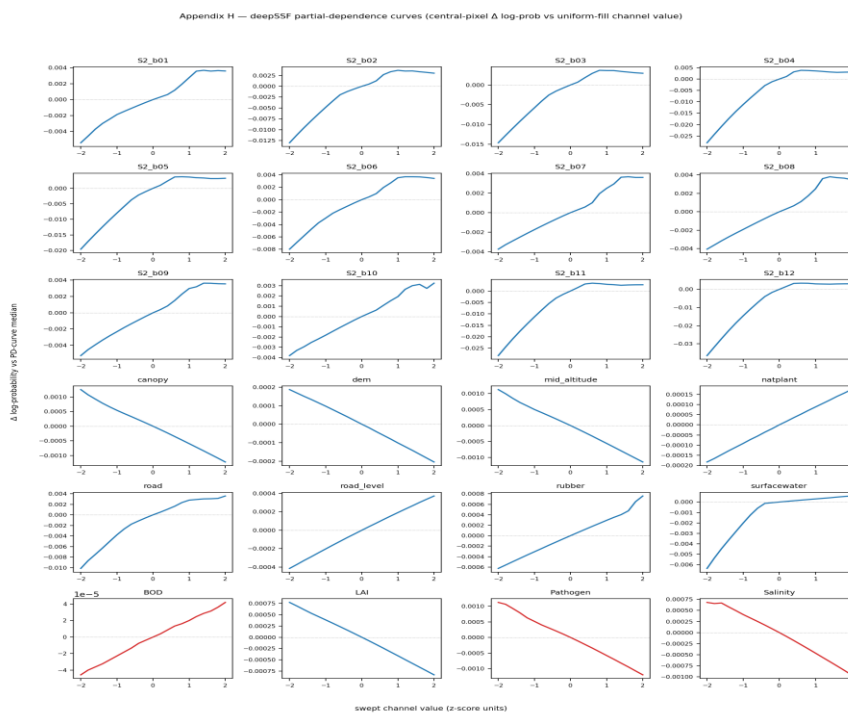


Fig H2. Single-checkpoint partial-dependence curves from the F2 main checkpoint, one panel per spatial channel; water-quality channels (BOD, Pathogen, Salinity) highlighted in red.

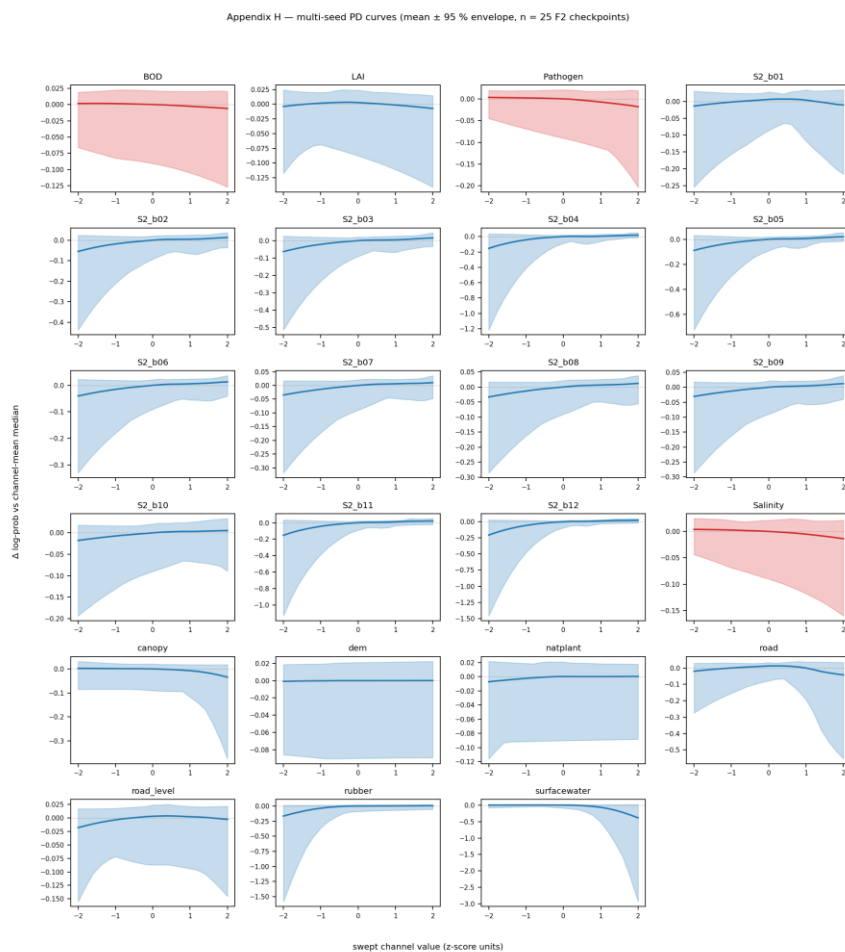


Fig H3. Multi-seed partial-dependence curves across 25 F2 checkpoints: mean line and non-parametric 95% envelope per channel; water-quality channels highlighted in red.

Table H. Water-quality multiseed PD summary at p5/p95 of the swept range, expressed as Δ log-probability relative to the channel-median PD value (n=25 cells; non-parametric 2.5/97.5 percentile envelope):

Channel	mean p5	95% envelope p5	mean p95	95% envelope p95
BOD	0.0016	[-0.066, +0.019]	-0.0060	[-0.127, +0.020]
Pathogen	0.0035	[-0.045, +0.020]	-0.0180	[-0.203, +0.019]
Salinity	0.0039	[-0.044, +0.024]	-0.0137	[-0.160, +0.022]

Two structural readings:

- i. The mean multiseed PD direction is monotone-decreasing for all three water-quality channels the single-checkpoint directions are reproduced as the population mean across 25 independent fits, indicating that the network has learned the biologically defensible direction in expectation
- ii. The 95% envelope is wide and includes zero for all three channels, indicating that individual seeds may flip the direction. The water-quality signal at the 10km native grid is therefore present in expectation but not statistically resolvable from any single deepSSF F2 fit, a nuance incorporated into the stress-test reading.

Appendix I. Multi-seed factorial replication of deepSSF F2. A 5×5 split-seed \times init-seed factorial design (split seeds 42, 1, 7, 123, 2024; init seeds 101, 202, 303, 404, 505; 25 runs total, \approx 57.5 GPU-hours on a single 16 GB GPU) was executed with otherwise identical configuration to the F2 main fit. Per-cell test-NLL and best-validation-NLL matrices, per-run metadata, per-animal NLL by run, and the one-way ANOVA decomposition are deposited in the supplementary archive. Headline statistics for test-NLL: mean 7.014 (range 6.788 – 7.360), $\sigma_{\text{total}} = 0.186$ nats, σ_{split} (row mean SD) = 0.203 nats, σ_{init} (column mean SD) = 0.013 nats, $\sigma_{\text{residual}} = 0.016$ nats. Variance fractions: split 99.0%, init 0.4%, residual 0.6 %. The result empirically establishes σ_{split} as the operational noise floor for NLL-based comparisons in this dataset; F2/F3/F5/F6 inter-configuration gaps (≤ 0.04 nats) sit well within this band, and prior-knowledge recovery rather than NLL therefore drives F-set interpretation. $\sigma_{\text{init}} \approx 0$ confirms that the trained F2 model is well-conditioned with respect to weight initialisation, supporting the claim that the ecological-variable-importance ranking (Appendix H) and the water-quality marginal-response directions are reproducible across re-training rather than artefacts of one lucky initialisation.

Acknowledgements

We pay our gratitude to the Management and Ecology of Malaysian Elephants (MEME) project for providing the GPS-telemetry data on which this study is based, and the field teams who carried out the original collar deployments. Computation was performed on a workstation at the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences. Intermediate artefacts and parameter sets are deposited in the supplementary archive accompanying this article.

Author Contributions

Hengyuan Liu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing original draft, Writing, review and editing.

José Ahimsa Campos-Arceiz: Resources (provision of GPS-telemetry data through the MEME project, ethics permits and field-deployment infrastructure), Supervision, Project administration, Funding acquisition, Writing, review and editing.

Both authors read and approved the final manuscript.

Data Availability

GPS tracks are available from a public data repository or, where licensing prevents this, on request to the corresponding author. The 91-raster environmental product is derived from the public sources listed in the supplementary data-sources table. All processing scripts and intermediate artefacts are deposited in the supplementary archive that accompanies this article.

Conflict of Interest

The authors declare no competing interests.

References

- Allen, A. M. & Singh, N. J. (2016). Linking movement ecology with wildlife management and conservation. *Frontiers in Ecology and Evolution*, 3:155.
- Avgar, T., Lele, S. R., Keim, J. L. & Boyce, M. S. (2017). Relative selection strength: Quantifying effect size in habitat- and step-selection inference. *Ecology and Evolution*, 7:5322-5330.
- Cagnacci, F., Boitani, L., Powell, R. A. & Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365:2157-2162.
- Campos-Arceiz, A. & Blake, S. (2011). Megagardeners of the forest—the role of elephants in seed dispersal. *Acta Oecologica*, 37(6), 542-553.
- Avgar, T., Potts, J. R., Lewis, M. A. & Boyce, M. S. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 7:619-630.
- De Knegt, H. J., Van Langevelde, F., Skidmore, A. K., Delsink, A., Slotow, R. & Henley, S, et al. (2011). The spatial scaling of habitat selection by African elephants. *Journal of Animal Ecology*, 80:270-281.
- de la Torre, J. A., Wong, E. P., Lechner, A. M., Zulaikha, N., Zawawi, A. & Abdul-Patah, et al. (2021). There will be conflict—agricultural landscapes are prime, rather than marginal, habitats for Asian elephants. *Animal Conservation*, 24:720-732.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V. & Gascon, F., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25-36.
- Forester, J. D., Im, H. K. & Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, 90:3554-3565.
- Hebblewhite, M. & Haydon, D. T. (2010). Distinguishing technology from biology: A critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365: 2303-2312.
- Hedges, S., Tyson, M. J., Sitompul, A. F., Kinnaird, M. F. & Gunaryadi, D. (2005). Distribution, status and conservation needs of Asian elephants (*Elephas maximus*) in Lampung Province, Sumatra, Indonesia. *Biological Conservation*, 124:35-48.
- Forrest, S. W. & Pagendam, D. (2025). deepSSF: A deep learning step selection function framework for animal movement modelling. *Software Release*.
- Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T. & Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, 86:1320-1330.
- Jones, E. R., Bierkens, M. F., Wanders, N., Sutanudjaja, E. H., Van Beek, L. P. & Van Vliet, M. T. (2023). DynQual v1.0: A high-resolution global surface water quality model. *Geoscientific Model Development*, 16:4481-4500.
- Kays, R., Crofoot, M. C., Jetz, W. & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, 348:aaa2478.
- Klappstein, N. J., Michelot, T., Fieberg, J., Pedersen, E. J. & Mills Flemming, J. (2024). Step selection functions with non-linear and random effects. *Methods in Ecology and Evolution*, 15:1332-1346.
- Yang, X., Li, S., Ma, J., Chen, Y., Zhou, X. & Li, F., et al. (2024). Global basic landform units derived from multi-source digital elevation models at 1 arc-second resolution. *Earth System Science Data Discussions*, 2024:1-24.

- Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D. & Frasson, R., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, 55:6499-6516.
- Manly, B. F., McDonald, L. L., Thomas, D. L., McDonald, T. L. & Erickson, W. P. (2002). *Resource selection by animals: Statistical design and analysis for field studies*. Dordrecht: Springer Netherlands.
- Michelot, T., Blackwell, P. G. & Matthiopoulos, J. (2019). Linking resource selection and step selection models for habitat preferences in animals. *Ecology*, 100:e02452.
- Muff, S., Signer, J. & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, 89 :80-92.
- NASA/METI/AIST/Japan Spacesystems & U.S./Japan ASTER Science Team (2019). ASTER Global Digital Elevation Model V003. NASA EOSDIS Land Processes DAAC.
- Neumann, M., et al. (2024). Global natural forest mapping with Transformer-based remote sensing for 2020 deforestation and degradation baseline.
- Nicosia, A., Duchesne, T., Rivest, L. P. & Fortin, D. (2017). A multi-state conditional logistic regression model for the analysis of animal movement.
- OpenStreetMap. Retrieved from <https://www.openstreetmap.org>
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J. & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20:503-510.
- Plotnick, R. E., Gardner, R. H. & O'Neill, R. V. (1993). Lacunarity indices as measures of landscape texture. *Landscape ecology*, 8:201-211.
- Riotte-Lambert, L. & Matthiopoulos, J. (2020). Environmental predictability as a cause and consequence of animal movement. *Trends in Ecology & Evolution*, 35:163-174.
- Pohle, J., Langrock, R., Van Beest, F. M. & Schmidt, N. M. (2017). Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22:270-293.
- Saaban, S., Othman, N. B., Yasak, M. N. B., Burhanuddin, M. N., Zafir, A. & Campos-Arceiz, A. (2011). Current status of Asian elephants in Peninsular Malaysia. *Gajah*, 35:67-75.
- Signer, J., Fieberg, J. & Avgar, T. (2019). Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecology and Evolution*, 9:880-890.
- Smeds, E. A., Cooper, Z. & Bentley, L. P. (2025). lacunr: Efficient 3D lacunarity for voxelized LiDAR data from forested ecosystems. *Methods in Ecology and Evolution*, 16:1906-1913.
- Sukumar, R. (2003). *The Living Elephants: Evolutionary Ecology, Behavior and Conservation*. Oxford University Press.
- Thurfjell, H., Ciuti, S. & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement ecology*, 2:4.
- Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V. & Brandt, J., et al. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300:113888.
- Venter, O., Sanderson, E. W., Magrath, A., Allan, J. R., Beher, J. & Jones, K. R., et al. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, 7:12558.

Vidya, T. N. C. & Sukumar, R. (2005). Social organization of the Asian elephant (*Elephas maximus*) in southern India inferred from microsatellite DNA. *Journal of Ethology*, 23:205-210.

Wadey, J., Beyer, H. L., Saaban, S., Othman, N., Leimgruber, P. & Campos-Arceiz, A. (2018). Why did the elephant cross the road? The complex response of wild elephants to a major road in Peninsular Malaysia. *Biological Conservation*, 218, 91-98.

Wall, J., Douglas-Hamilton, I. & Vollrath, F. (2006). Elephants avoid costly mountaineering. *Current Biology*, Wang, Y., Hollingsworth, P. M., Zhai, D., West, C. D., Green, J. M. & Chen, H., et al. (2023). High-resolution maps show that rubber causes substantial deforestation. *Nature*, 62:340-346.

Yuan, H., Dai, Y., Xiao, Z., Ji, D. & Shangguan, W. (2011). Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling. *Remote Sensing of Environment*, 115:1171-1187.

Citation:

Liu Hengyuan and Jose Ahimsa. (2026). Comparing classical, hidden-Markov and deep-learning step-selection functions for Asian elephant habitat preference: the value of feature engineering, dynamic diel/seasonal variables and water-quality covariates. *Ukrainian Journal of Ecology*. 16:1-34.



This work is licensed under a Creative Commons Attribution 4.0 License
