Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sensitivity of zoonotic malaria suitability models to land-cover representation and spatial support across Borneo

Cesar Ivan Alvarez
cesar.alvarez@uni-a.de
University of Augsburg
Marius Nickel
University of Augsburg
Katharina Waha
University of Augsburg
Inke N. D. Lubis
Universitas Sumatera Utara
Rintis Noviyanti
The National Research and Innovation Agency
Freya M. Shearer
The University of Melbourne
Anisah Jantim
State Public Health Laboratory
Giri Rajahram
Queen Elizabeth Hospital II
Nicholas M. Anstey
Menzies School of Health Research and Charles Darwin University
Matthew J. Grigg
Menzies School of Health Research and Charles Darwin University

Research Article

Keywords: Plasmodium knowlesi, zoonotic malaria, health geography, environmental suitability, land-use and land-cover, remote sensing, spatial epidemiology, machine learning

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11074697/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title

Sensitivity of zoonotic malaria suitability models to land-cover representation and spatial support across Borneo

Author names and affiliations

Cesar Ivan Alvarez[a*], Marius Nickel[a], Katharina Waha[a], Inke N. D. Lubis[b], Rintis Noviyanti[c], Freya M. Shearer[d], Anisah Jantim[e], Giri Rajahram[f,g,h], Nicholas M. Anstey[i,j], and Matthew J. Grigg[i,j,k]

[a] Centre for Climate Resilience, University of Augsburg, Universitäts Strasse 12a, 86159 Augsburg, Germany

[b] Faculty of Medicine, Universitas Sumatera Utara, Medan, Indonesia

[c] Eijkman Research Center for Molecular Biology, The National Research and Innovation Agency, Cibinong, Indonesia

[d] School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia

[e] State Public Health Laboratory, Public Health Laboratory, Kota Kinabalu, Sabah, Malaysia

[f] Department of Medicine, Queen Elizabeth Hospital II, Kota Kinabalu, Sabah, Malaysia

[g] Infectious Diseases Society Sabah-Menzies School of Health Research Clinical Research Unit, Sabah, Malaysia

[h] School of Medicine and Health Sciences, Monash University Malaysia, Kuala Lumpur, Malaysia

[i] Global and Tropical Health Division, Menzies School of Health Research and Charles Darwin University, Darwin, NT, Australia

[j] Infectious Diseases Society Kota Kinabalu Sabah-Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Sabah, Malaysia

[k] Institute for Health Research, University of Notre Dame Australia, Fremantle, Western Australia, Australia

Email address list:

Cesar Ivan Alvarez: cesar.alvarez@uni-a.de

Marius Nickel: marius.nickel@uni-a.de

Katharina Waha: katharina.waha@uni-a.de

Inke N. D. Lubis: inkenadia@gmail.com

Rintis Noviyanti: rnoviyanti2@gmail.com

Freya M. Shearer: freya.shearer@unimelb.edu.au

Anisah Jantime: dranisah@moh.gov.my

Giri Rajahram: gsrajahram@gmail.com

Nicolas M. Anstey: nicholas.anstey@menzies.edu.au

Matthew J. Grigg: Matthew.Grigg@menzies.edu.au

***Corresponding author:**

Cesar Ivan Alvarez

Universität Augsburg, Universitätsstraße 12, 86159 Augsburg, Germany

e-mail: cesar.alvarez@uni-a.de

Tel. : +4915225379102

Abstract:

Background: Geospatial models of zoonotic malaria combine surveillance records with remotely sensed environmental data, but model performance and interpretation may depend on the land-cover product and spatial support used to characterise surrounding landscapes. We assessed how land-cover representation and analytical support influenced model discrimination, predictor rankings, and spatial predictions of *Plasmodium knowlesi* environmental suitability across Borneo.

Methods: Confirmed *P. knowlesi* records and accessible-area background samples were combined with three harmonised land-use and land-cover products at 10 m resolution: ESA WorldCover, ESRI Land Cover, and a regionally refined Hybrid product. Landscape metrics were calculated within analytical windows of 500, 1,000, 2,000, and 5,000 m and combined with topographic, climatic, hydrological, and accessibility variables. Elastic-net logistic regression, Random Forest, Extremely Randomized Trees, and Extreme Gradient Boosting were evaluated using random and spatial five-fold cross-validation. Confirmed negative observations from the Indonesian surveillance study were retained for external evaluation.

Results: Ensemble models achieved the highest discrimination under spatial cross-validation. At the 500 m analytical window, the best product-specific models produced mean AUC values ranging from 0.863 to 0.892. Differences among analytical supports and land-cover products were modest, with overlapping uncertainty ranges. Night-time land-surface temperature, distance to roads, and daytime land-surface temperature were consistently among the highest-ranked predictors. In contrast, the ranking of landscape metrics varied among products, with forest configuration more prominent in the ESA model, forest and wetland configuration in the ESRI model, and built-up configuration in the Hybrid model. The Hybrid model produced lower fold-to-fold spatial variability despite slightly lower mean discrimination. External evaluation showed differences among algorithms in predictions for confirmed negative observations.

Conclusions: Land-cover product choice had a greater influence on landscape interpretation and spatial prediction than on overall discrimination. The higher mean performance at 500 m indicates that locally summarised land-cover configuration around reported surveillance locations added useful predictive information. Interpretation of this spatial support also depends on differences in surveillance-location precision and the native resolution of the environmental predictors. Evaluating land-cover representation, analytical support, spatial uncertainty, and external prediction behaviour together provides a more complete basis for geospatial zoonotic-malaria modelling.

Keywords:
Plasmodium knowlesi; zoonotic malaria; health geography; environmental suitability; land-use and land-cover; remote sensing; spatial epidemiology; machine learning

1. Introduction

Land-use and land-cover (LULC) change is a major form of environmental transformation, with important consequences for ecosystem functioning and human health (Plowright et al., 2021). Deforestation, agricultural expansion, and urbanisation reshape tropical landscapes and create new interfaces among humans, wildlife, and disease vectors (Lambin et al., 2010). These changes can alter habitat suitability, vector ecology, and opportunities for contact with wildlife reservoirs, thereby influencing the spatial distribution of vector-borne and zoonotic diseases (Zaldo-Aubanell et al., 2021).

In Southeast Asia, zoonotic malaria caused by *Plasmodium knowlesi* has become an increasing public-health concern, particularly in areas undergoing rapid environmental change (Tobin et al., 2024). Its transmission involves non-human primate reservoirs and mosquito vectors whose habitats and interactions with people are closely linked to land-use dynamics (Fornace, Alexander, et al., 2019). Deforestation, forest fragmentation, and agricultural expansion have been associated with changes in vector habitat and human exposure, while forest edges and mixed land-use systems have repeatedly been identified as relevant environments for *P. knowlesi* occurrence (Brock et al., 2019; Fornace et al., 2016). In Malaysian Borneo, *P. knowlesi* incidence has continued to increase despite substantial progress in eliminating locally acquired malaria caused by predominantly human-only *Plasmodium* species, highlighting the distinct challenges of zoonotic transmission (Cooper et al., 2020; Grigg et al., 2017).

Remote sensing and geospatial analysis provide valuable tools for characterising the environmental conditions associated with infectious-disease occurrence. Satellite-derived land cover, vegetation, land-surface temperature, and precipitation are increasingly used to develop spatially explicit models of environmental suitability and to support surveillance in areas where epidemiological observations are incomplete or unevenly distributed (Machault et al., 2011; Rakotoarison et al., 2020; Weiss et al., 2019; Harrison et al., 2024). Combined with environmental niche modelling and machine-learning methods, these data allow complex and potentially nonlinear relationships between disease occurrence and environmental conditions to be evaluated across large geographic areas (Elith et al., 2011; Phillips et al., 2009; Tobin et al., 2024).

The representation of environmental conditions within such models remains an important methodological consideration. LULC products differ in their classification frameworks, training data, thematic detail, reference periods, and representation of spatial patterns. These differences can propagate into landscape metrics describing composition, fragmentation, aggregation, connectivity, and patch shape (Turner et al., 2001). Metric values are also sensitive to the spatial unit or radius within which they are calculated (Frazier and Kedron, 2017). As a result, land-cover classification and analytical spatial support may influence model performance, predictor

importance, and the ecological interpretation of landscape–disease associations (Congalton et al., 2014; Prestele et al., 2016; Alvarez et al., 2025).

These issues are particularly relevant when epidemiological records originate from different surveillance systems. Case coordinates may represent households, villages, health-facility catchments, or broader administrative locations, and the recorded location may differ from the place where exposure occurred. Spatial clustering and thinning can reduce pseudo-replication, spatial dependence, and unequal sampling concentration, while multiscale analysis can reveal how model performance and landscape associations change with analytical support. Interpretation also depends on the native resolution of the environmental predictors. Landscape metrics derived from high-resolution land-cover data can be recalculated within alternative spatial windows, whereas land-surface temperature and precipitation products represent environmental conditions over broader and relatively fixed spatial resolutions. Scale-dependent model performance therefore reflects the combined influence of analytical support and predictor resolution.

Beyond land cover, *P. knowlesi* occurrence has been associated with topography, thermal conditions, precipitation, hydrological features, and human accessibility (Paaijmans et al., 2009; Mordecai et al., 2019; Fornace, Brock, et al., 2019). These variables capture complementary dimensions of environmental suitability and potential exposure, but their apparent contribution may vary according to land-cover representation, spatial scale, and modelling approach. Previous studies have demonstrated the value of remotely sensed environmental predictors for mapping *P. knowlesi* suitability, yet the sensitivity of model results to alternative land-cover products and analytical windows has received less systematic evaluation. It remains unclear whether different land-cover representations yield comparable model performance, identify similar landscape associations, and produce equally stable spatial predictions.

This study evaluates how land-cover representation, analytical spatial support, and modelling approach influence geospatial models of *P. knowlesi* environmental suitability across Borneo. We compare three harmonised land-cover products, derive landscape metrics within spatial windows ranging from 500 m to 5 km, and evaluate linear and ensemble models using spatial cross-validation and independent confirmed negative observations. We assess whether model discrimination and spatial stability differ among land-cover products, how performance changes across analytical supports, and which environmental and landscape associations remain consistent across products and modelling approaches. The analysis provides a comparative assessment of how methodological choices in environmental representation influence the performance and interpretation of geospatial models for zoonotic-malaria surveillance.

2. Materials and methods

2.1. Study Area

The study was conducted on the island of Borneo, one of the largest tropical islands in the world, shared by Malaysia, Indonesia and Brunei. The region is characterized by high biodiversity, extensive tropical forests, and rapidly changing land use patterns driven by deforestation, agricultural expansion (particularly oil palm plantations), and infrastructure development (Gaveau et al., 2016). These landscape transformations have been closely associated with the emergence and spatial distribution of zoonotic malaria, particularly *P. knowlesi* (Muh et al., 2025). The study area includes both Malaysian Borneo (Sabah and Sarawak), where most epidemiological data are available, and Indonesian Kalimantan (with a focus on North Kalimantan province), where data coverage is more limited (Fig. 1).

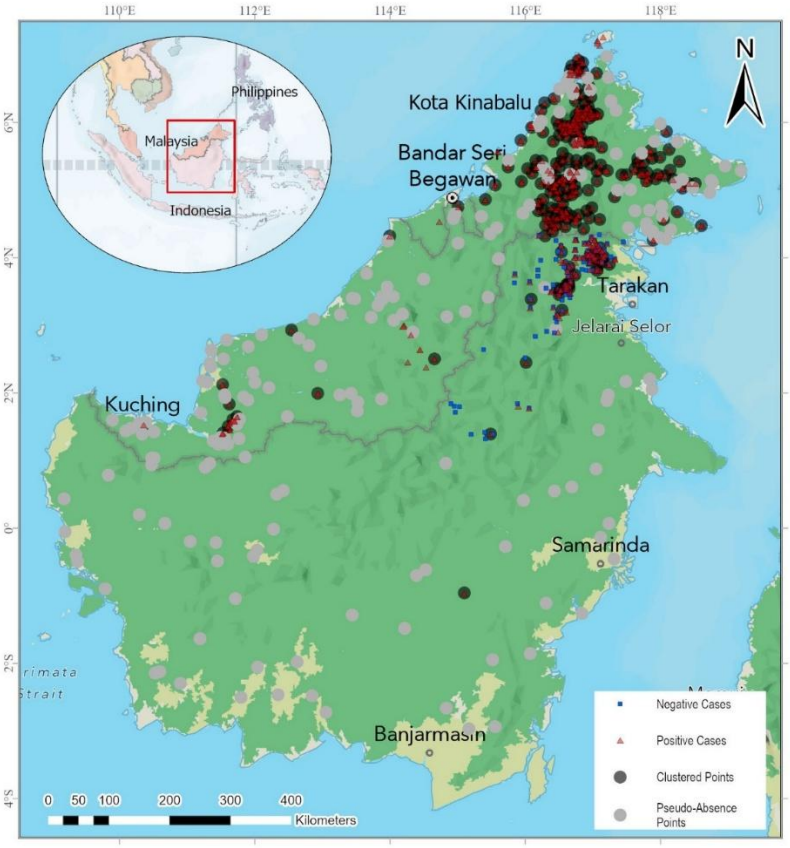

Fig. 1. Study area and spatial distribution of *P. knowlesi* observations across Borneo. The map shows the locations of confirmed positive cases (red triangles), confirmed negative cases (blue squares), spatially clustered observations (black circles), and pseudo-absence points (light grey circles) used for model calibration.

2.2. P. knowlesi (PK) case data sources

2.2.1 PK presence and confirmed negative observations

The epidemiological dataset comprised georeferenced records of PCR-confirmed *P. knowlesi* infection derived from health-facility surveillance and compiled epidemiological studies across Borneo. Coordinates were interpreted as reported locations associated with detected cases. Depending on the source, they represented households, villages, or broader administrative locations and did not necessarily identify the site where infection occurred.

For Malaysia, data were obtained from two main sources: (i) Sabah malaria case notification data (*P. knowlesi* human infections confirmed by PCR (Nuin et al., 2020)) from three districts (Ranau, Nabawan and Kinabatangan) from 2021-2023 (Malaysian MREC: NMRR-15-168-24821) (Cooper et al., 2020), and (ii) collated *P. knowlesi* case data from multiple studies across Sabah and Sarawak as reported in Tobin et al. (2024).

For Indonesia, data were collected as part of the ZOOMAL project through health-facility surveillance conducted in North Kalimantan. P. knowlesi infections were confirmed using molecular diagnostic methods, with confirmed negative observations retained for independent external evaluation (Trianty et al., 2025), confirmed by PCR (Braima et al., 2025). Participants used offline maps and local landmarks to geolocate their households through a participatory mapping procedure (Fornace et al., 2018). These coordinates therefore represent participant household locations rather than presumed locations of infection. Full details of the clinical surveillance, diagnostic procedures and study design are provided by Trianty et al. (2025).

In total, 1,139 *P. knowlesi* case presence records were available for Malaysia and 41 for Indonesia. In addition to presence data, 84 confirmed negative observations derived from the Indonesian surveillance study were included, providing an independent source of absence information. These negative observations were used exclusively for external validation. Malaysian records were predominantly available at village level, with some at district administrative level. Indonesian participants used offline maps and local landmarks to geolocate their households through a participatory mapping procedure (Fornace et al., 2018). The difference in geolocation procedures was considered when interpreting results across analytical supports

2.2.2 Spatial filtering and background sampling

Spatial filtering was adapted to the distribution of observations within each surveillance dataset. In Indonesia, densely clustered positive records within health-facility catchment areas were aggregated within a 1 km neighbourhood using DBSCAN, reducing local pseudo-replication while retaining the spatial distribution of observed cases (Hahsler et al., 2019). Malaysian positive records were more geographically dispersed and were thinned using a minimum separation distance of 7 km to reduce spatial dependence and the overrepresentation of intensively sampled locations (Kramer-Schadt et al., 2013; Descombes et al., 2022).

The filtered records from Malaysia and Indonesia were combined in a pooled Borneo dataset. This design retained the broader geographic and environmental coverage of the Malaysian observations while incorporating the environmental conditions represented by the Indonesian surveillance data. After spatial thinning and clustering, the pooled modelling dataset contained 191 positive records called as clustered positive points and 182 pseudo-absence points.

Background samples were generated within populated areas identified using WorldPop to represent environmental conditions in locations where case detection was plausible (Tatem, 2017; Phillips et al., 2009). A minimum separation distance of 5 km was applied between background samples to limit spatial clustering, avoiding the spatial autocorrelation (Chevalier et al.,2021). The total number of background samples was balanced to the number of filtered positive records and distributed between Malaysia and Indonesia at proportions of 65% and 35%, respectively. This allocation increased the representation of environmental conditions in the less-sampled Indonesian region while maintaining broad coverage across Borneo (Barbet-Massin et al., 2012). The final modelling dataset comprised spatially filtered positive records and accessible-area background samples. The 84 confirmed negative observations from the Indonesian surveillance study were excluded from model training and retained exclusively for independent external evaluation (Fig. 1).

2.3 Land cover datasets and landscape metrics

Three land-use and land-cover products at 10 m spatial resolution were included: ESA WorldCover 2021 (Zanaga et al., 2022), ESRI Land Cover 2022 (Esri, 2022), and the regionally refined Hybrid Land Cover 2020 product (Hulthen and Waha, 2022). ESA WorldCover is derived from Sentinel-1 and Sentinel-2 imagery, while ESRI Land Cover uses a global classification framework based on multispectral satellite imagery. The Hybrid product was developed to represent regional land-use patterns, including forest degradation, oil-palm expansion, plantations, and mixed land-use systems.

All products were reclassified to a common thematic scheme comprising forest, plantation, grass/shrub/agriculture, urban/built-up, bare/open land, and water/wetland. Cloud pixels were excluded from the analysis. The same harmonised classes were used for all products to ensure consistent calculation and comparison of landscape metrics.

Fig. 2 shows the harmonised representation of forest, plantation, and grass/shrub/agriculture classes across the three products for the same geographic area. Differences in class boundaries, patch configuration, and local land-cover composition are visible among the products.

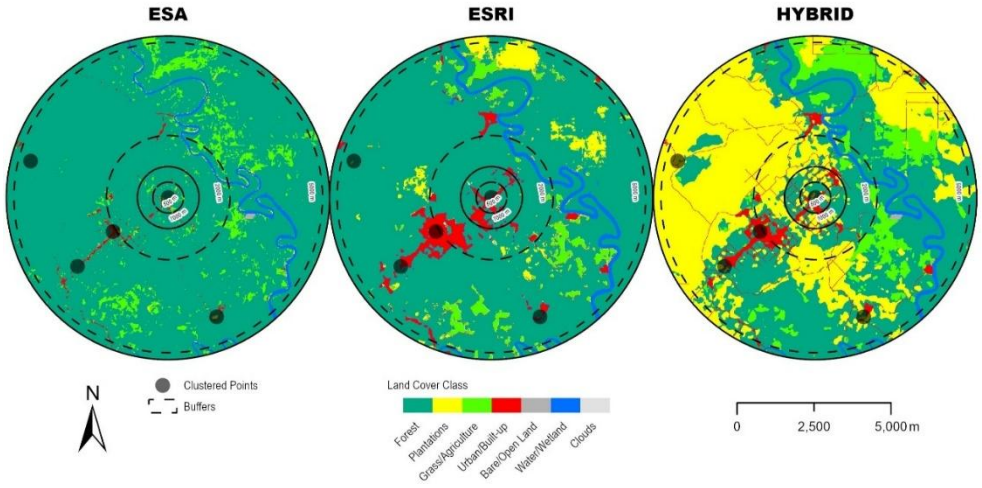

Fig. 2. Comparison of harmonized land cover representation across the ESA WorldCover, ESRI Land Cover, and Hybrid datasets for the same location in Borneo. Circular buffers indicate the four spatial scales used to compute landscape metrics (500 m, 1000 m, 2000 m, and 5000 m).

Landscape metrics were calculated independently for each land-cover product. Circular analytical windows with radii of 500, 1,000, 2,000, and 5,000 m were centred on each spatially filtered positive record, background sample, and confirmed negative observation. Calculations were performed using the *landscapemetrics* package in R, which implements landscape metrics derived from FRAGSTATS (Hesselbarth et al., 2019; McGarigal et al., 2021).

The selected metrics were percentage of landscape (PLAND), edge density (ED), clumpiness index (CLUMPY), aggregation index (AI), mean Euclidean nearest-neighbour distance (ENN_M), patch density (PD), and mean perimeter–area ratio (PARA_MN). These metrics characterize complementary dimensions of landscape structure, including composition, fragmentation, aggregation, connectivity, and patch shape. Metrics were calculated by harmonised land-cover class within each analytical window and subsequently combined with the environmental and accessibility variables used in the modelling framework.

2.4 Environmental and landscape variables

To complement landscape metrics, a set of topographic, climatic, and accessibility variables was included to capture key environmental drivers of *P. knowlesi* transmission.

Topographic variables included elevation and slope, derived from the Shuttle Radar Topography Mission (SRTM) at 30 m resolution (NASA, 2013). Accessibility and hydrological factors were represented by distance to roads and distance to rivers, computed from OpenStreetMap (OSM) data and hydrographic layers (Geofabrik GmbH, 2025). These variables capture human mobility, landscape disturbance, and proximity to potential mosquito breeding habitats.

Climatic variables included annual precipitation from ERA5-Land reanalysis (Muñoz Sabater, 2019) and land surface temperature (LST) derived from MODIS (MOD11A1.061), including both daytime and nighttime conditions (Wan et al., 2021). These variables capture key thermal and moisture conditions influencing vector ecology and parasite development.

Environmental variables were extracted for each clustered presence point, pseudo-absence sample, and independent negative observation, and subsequently combined with landscape metrics to construct the final modelling datasets across all buffer radii. A summary of all mean annual composition variables used in the analysis is provided in Table 1. Given the coarser spatial resolution of MODIS LST (1 km), extracted values may be relatively homogeneous within smaller buffer radii (e.g. 500 m), and should therefore be interpreted as representing broader thermal gradients rather than fine-scale variability.

Table 1. Environmental and landscape variables used in the analysis. We used the annual median values for topography and climate variables.

Category	Variable	Unit	Year	Source with Spatial Resolution
Topography	Elevation	m	2013	SRTM 30 m
	Slope	degrees	2013	
Climate	Precipitation	mm	2023	ERA5-Land - ~9 km
	LST Day	°C	2023	MODIS MOD11A1.061 – 1 km
	LST Night	°C	2023	
Accessibility	Distance to roads	m	2023	OSM – 10 m
	Distance to rivers	m	2023	
Landscape Metrics derived from LULC	PLAND	%	2020–2022	ESA WorldCover (2021); ESRI Land Cover (2022); Hybrid Indonesia (2020) – 10 m
	ED	m/ha	2020–2022	
	CLUMPY	index	2020–2022	
	AI	index	2020–2022	
	ENN_M	m	2020–2022	
	PD	patches/km²	2020–2022	
	PARA_MN	ratio	2020–2022	

2.5 Machine learning modelling and evaluation

Four modelling approaches were implemented to evaluate linear and nonlinear associations with *P. knowlesi* occurrence: logistic regression with elastic-net regularisation (LOGIT_EN), Random Forest (RF), Extremely Randomized Trees (ET), and Extreme Gradient Boosting (XGB) (Breiman, 2001; Chen and Guestrin, 2016).

Models were trained using the spatially filtered positive–background dataset. The confirmed negative observations from the Indonesian surveillance study were excluded from model training and retained for independent external evaluation.

Models were fitted separately for each combination of land-cover product and analytical window. For each experiment, predictors included the landscape metrics derived from the corresponding land-cover product at 500, 1,000, 2,000, or 5,000 m, together with the environmental, topographic, hydrological, and accessibility variables described in Section 2.4. The non-LULC predictors were held constant across land-cover products for each observation.

Model performance was evaluated using random stratified five-fold cross-validation and spatial five-fold cross-validation. Spatial folds were defined using existing block identifiers where available or generated from observation coordinates using a 25 km grid (Miller et al., 2026). Spatial cross-validation was used as the principal evaluation framework, while random cross-validation was included for comparison.

Performance was assessed using the area under the receiver operating characteristic curve (AUC), the area under the precision–recall curve (PR-AUC), and the Brier score (Allouche et al., 2006). The confirmed Indonesian negative observations were used to calculate the false-positive rate of model predictions during external evaluation.

Predictor contributions were evaluated using permutation importance for RF, ET, and XGB under spatial cross-validation. Coefficients from the elastic-net logistic regression were used to interpret the LOGIT_EN model, and SHAP values were calculated for the best model for each LULC to examine the relative contribution of individual predictors (Lundberg and Lee, 2017). Predictor importance was compared across land-cover products, analytical windows, and modelling approaches. Fig. 3 summarises the complete analytical workflow.

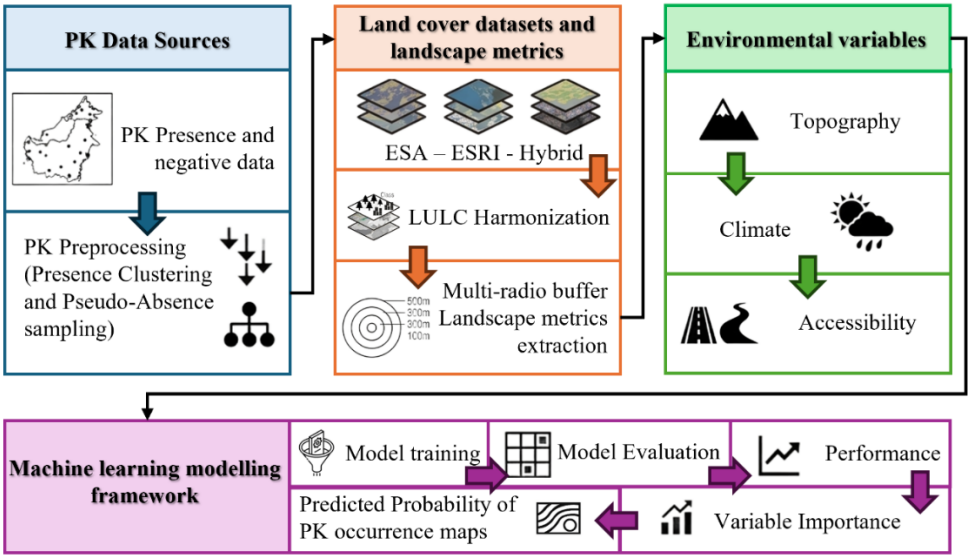

Fig. 3. Workflow diagram of the study

3. Results

3.1. Model performance under random and spatial validation

Model performance differed by algorithm and validation strategy (Supplementary Table S1). RF, ET, and XGB generally produced higher mean discrimination than LOGIT_EN, consistent with nonlinear relationships and spatial heterogeneity in the predictor set.

Random cross-validation produced mean AUC values approximately 0.02-0.04 higher than spatial cross-validation across models (Fig. 4). This difference indicates that random partitions benefited from spatial dependence between training and test observations. Under spatial validation, XGB and RF reached mean AUC values close to 0.89, whereas LOGIT_EN had lower discrimination and greater sensitivity to spatial structure. PR-AUC showed a similar ranking (Supplementary Fig. S1). Spatial cross-validation was therefore used for the principal comparisons.

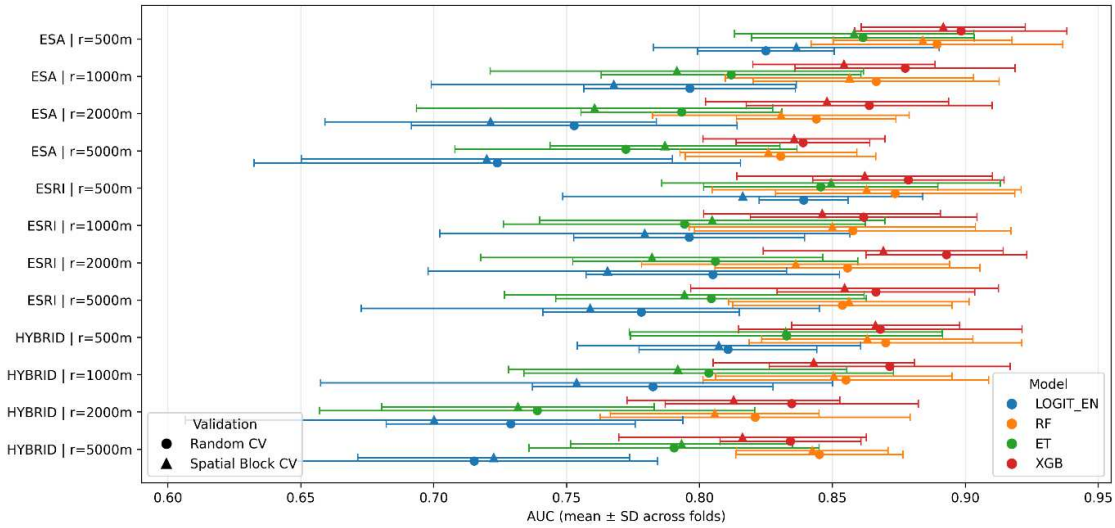

Fig. 4. Mean AUC under random and spatial cross-validation across modelling approaches.

3.2. Sensitivity to analytical support and land-cover representation

Model discrimination varied moderately across analytical supports. Models using 500 m landscape summaries produced the highest mean AUC, with an average advantage of approximately 0.05 relative to 5000 m. However, standard-deviation ranges overlapped across supports. The result therefore demonstrates sensitivity to the spatial support used to summarize LULC configuration, but it does not identify a precise ecological, movement, or exposure radius. Larger windows smoothed local landscape heterogeneity and generally produced slightly lower discrimination.

Differences among harmonized LULC products were smaller than differences among algorithms or validation strategies. ESA produced slightly higher mean discrimination than ESRI, while the

Hybrid product had slightly lower mean AUC after harmonization. The Hybrid product nevertheless represented more complex local patterns and produced more stable spatial predictions. This apparent contrast indicates that harmonization can reduce thematic distinctions while local spatial configuration still affects predictor interpretation and prediction stability.

Table 2 summarizes the best-performing model for each LULC product at the 500 m analytical support. XGB with ESA produced the highest mean AUC (0.892, SD 0.031), followed by XGB with ESRI (0.879, SD 0.036) and RF with the Hybrid product (0.863, SD 0.040). The overlapping uncertainty ranges indicate that product-level differences in discrimination were modest. Brier scores ranged from 0.145 to 0.158 and were similarly close across products.

Table 2. Best models under spatial cross-validation at 500 m analytical support

LULC	Model	AUC	AUC SD	PR-AUC	PR-AUC SD	Brier Score	Brier Score SD
ESA	XGB	0.892	0.031	0.887	0.068	0.145	0.025
ESRI	XGB	0.879	0.036	0.892	0.034	0.155	0.036
HYBRID	RF	0.863	0.040	0.876	0.047	0.158	0.016

Evaluation against the independent confirmed Indonesian negatives revealed a trade-off between discrimination and conservative prediction (Fig. 5). Ensemble models with the highest internal AUC tended to produce higher false-positive-rate proxies, whereas LOGIT_EN was generally more conservative. Because the confirmed negatives came only from the Indonesian surveillance study, these results were used to compare model behaviour rather than to estimate overall specificity across Borneo.

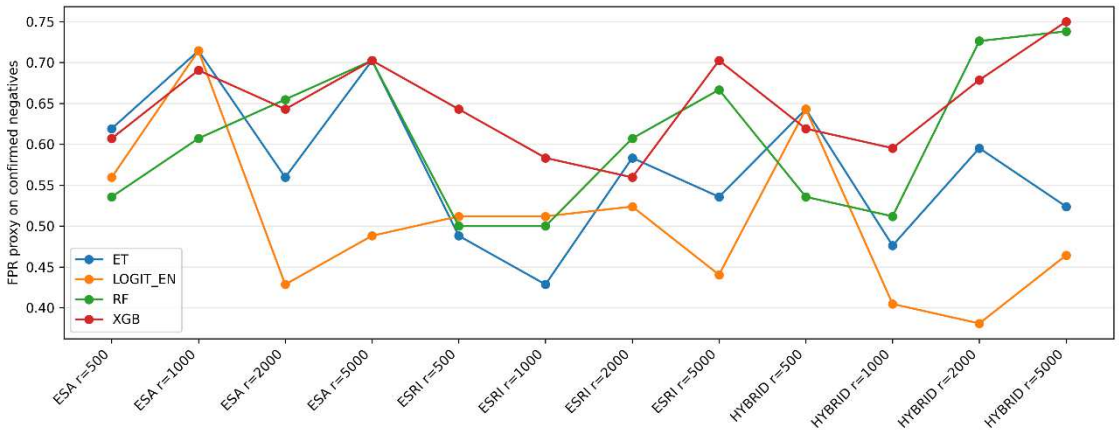

Fig. 5. External evaluation using confirmed Indonesian negative observations.

3.3. Predictor importance across land-cover products

Predictor importance was evaluated for the best-performing model associated with each land-cover product at the 500 m analytical window (Table 2). Complete SHAP-based importance results across predictors and analytical windows are provided in Supplementary Table S2.

Night-time land-surface temperature, distance to roads, and daytime land-surface temperature were among the highest-ranked predictors in the ESA, ESRI, and Hybrid models, with relative importance values ranging from approximately 9% to 13%. Elevation also ranked among the principal environmental predictors, although with lower relative importance than temperature and road proximity.

The ranking of landscape metrics differed among land-cover products. Urban/built-up metrics occupied five of the nine highest-ranked landscape-metric positions across the three product-specific models. In the ESA model, forest patch density was the highest-ranked landscape metric, with a relative importance of 5.3%, followed by urban/built-up patch density at 4.5% and urban/built-up edge density at 3.6%.

In the ESRI model, forest patch density had a relative importance of 7.8%. Metrics describing water/wetland configuration were also prominent, including mean perimeter–area ratio at 6.3% and edge density at 4.1%.

In the Hybrid model, the highest-ranked landscape metrics were associated with urban/built-up land cover. Percentage of landscape had a relative importance of 9.6%, followed by aggregation index at 6.0% and clumpiness at 5.7%. Urban/built-up edge density and patch density contributed an additional 5.1% and 2.2%, respectively. Together, these five metrics accounted for 28.6% of the normalized predictor importance in the Hybrid model.

Overall, temperature and road proximity remained highly ranked across the three product-specific models, whereas the relative importance of individual landscape classes and configuration metrics varied among the ESA, ESRI, and Hybrid products (Fig. 6).

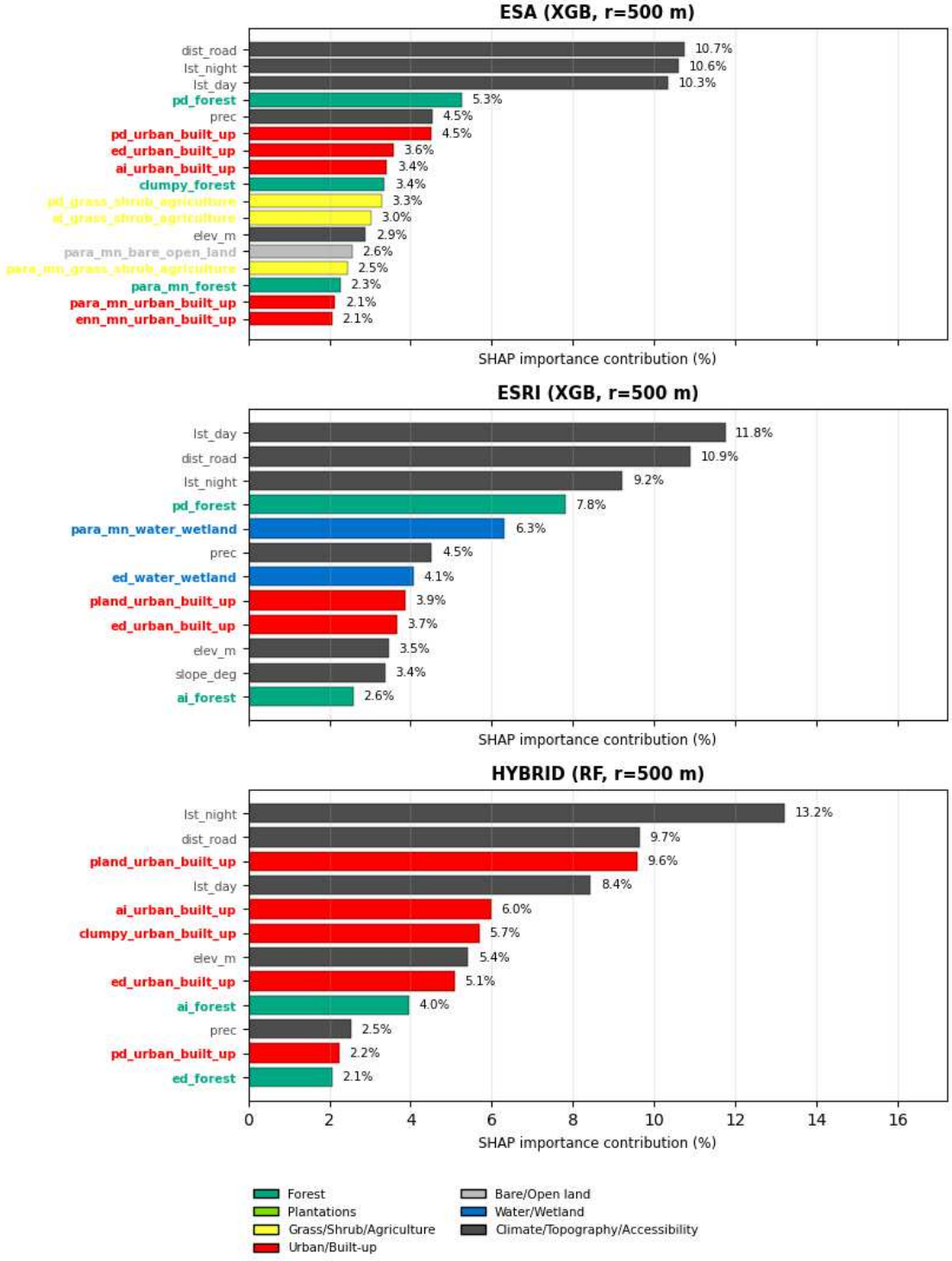

Fig. 6. Normalised global SHAP importance for the best-performing models at the 500 m analytical window. Mean absolute SHAP values were normalised to sum to 100% within each model. The selected models were XGB for ESA and ESRI and Random Forest for the Hybrid product. Colors indicate land-cover classes; percentages represent relative contributions.

3.4 Spatial distribution and uncertainty of predicted environmental suitability

Spatial predictions showed broadly consistent large-scale patterns across Borneo, with higher predicted environmental suitability concentrated mainly in northern and central parts of the island (Fig. 7). Many of these areas occurred within heterogeneous landscapes comprising fragmented forest and mixed land-use mosaics and extended across upland regions in the island interior. Elevation contributed 2.9%, 3.5%, and 5.4% of total predictor importance in the ESA, ESRI, and Hybrid models, respectively.

Differences were evident in the spatial configuration of predictions among land-cover products. The ESA-based model produced relatively continuous areas of higher suitability across central Borneo (Fig. 7a). The ESRI-based model generated a more heterogeneous pattern, with higher values distributed across smaller and more fragmented areas (Fig. 7b). The Hybrid-based model produced a broader and more spatially dispersed distribution of elevated suitability values (Fig. 7c).

Higher predicted suitability was generally located outside major urban centres. At the 500 m analytical window, urban/built-up configuration metrics were particularly prominent in the Hybrid model. Urban percentage of landscape contributed 9.6% of total importance, followed by aggregation index at 6.0%, clumpiness at 5.7%, edge density at 5.1%, and patch density at 2.2%. Together, these urban/built-up metrics accounted for 28.6% of the normalized predictor importance in the Hybrid model.

Predicted scores ranged from 0 to 1, with higher values indicating stronger modelled environmental suitability within the positive–background modelling framework. Fold-based uncertainty maps showed spatial variation in prediction agreement among the three products (Fig. 8). The ESA- and ESRI-based models displayed higher standard-deviation values in several northern and central areas, whereas the Hybrid-based model generally produced lower standard-deviation values across the study region.

Overall, the three land-cover products identified similar broad areas of higher environmental suitability while producing distinct patterns in spatial continuity, local configuration, and prediction uncertainty.

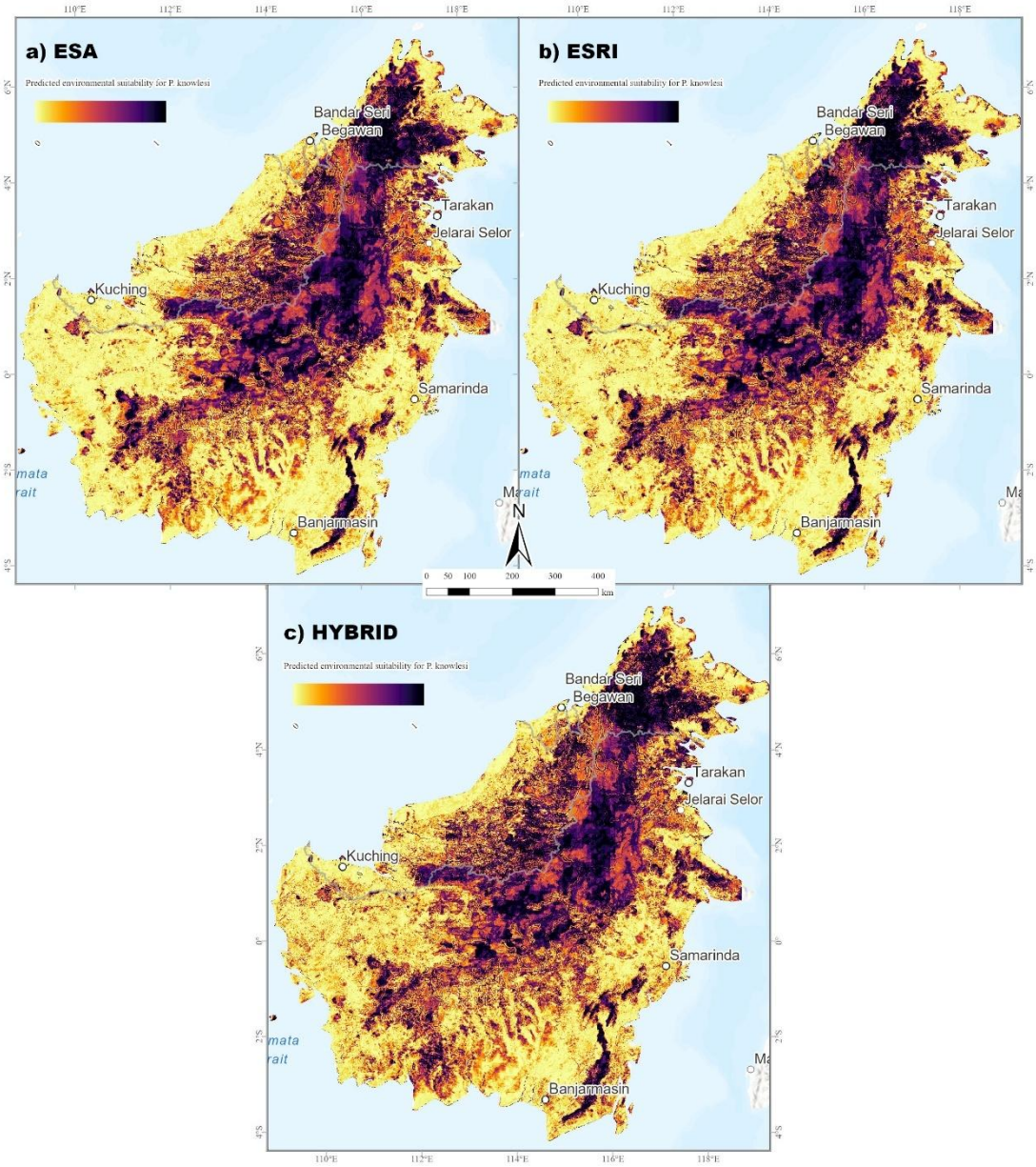

Fig. 7. Predicted environmental suitability for Plasmodium knowlesi across Borneo. Predictions are shown for the best-performing model at the 500 m analytical window for the ESA, ESRI, and Hybrid land-cover products. (a) ESA (XGB), (b) ESRI (XGB), and (c) Hybrid (RF). Values range from 0 to 1 and represent relative environmental suitability rather than absolute probability of occurrence.

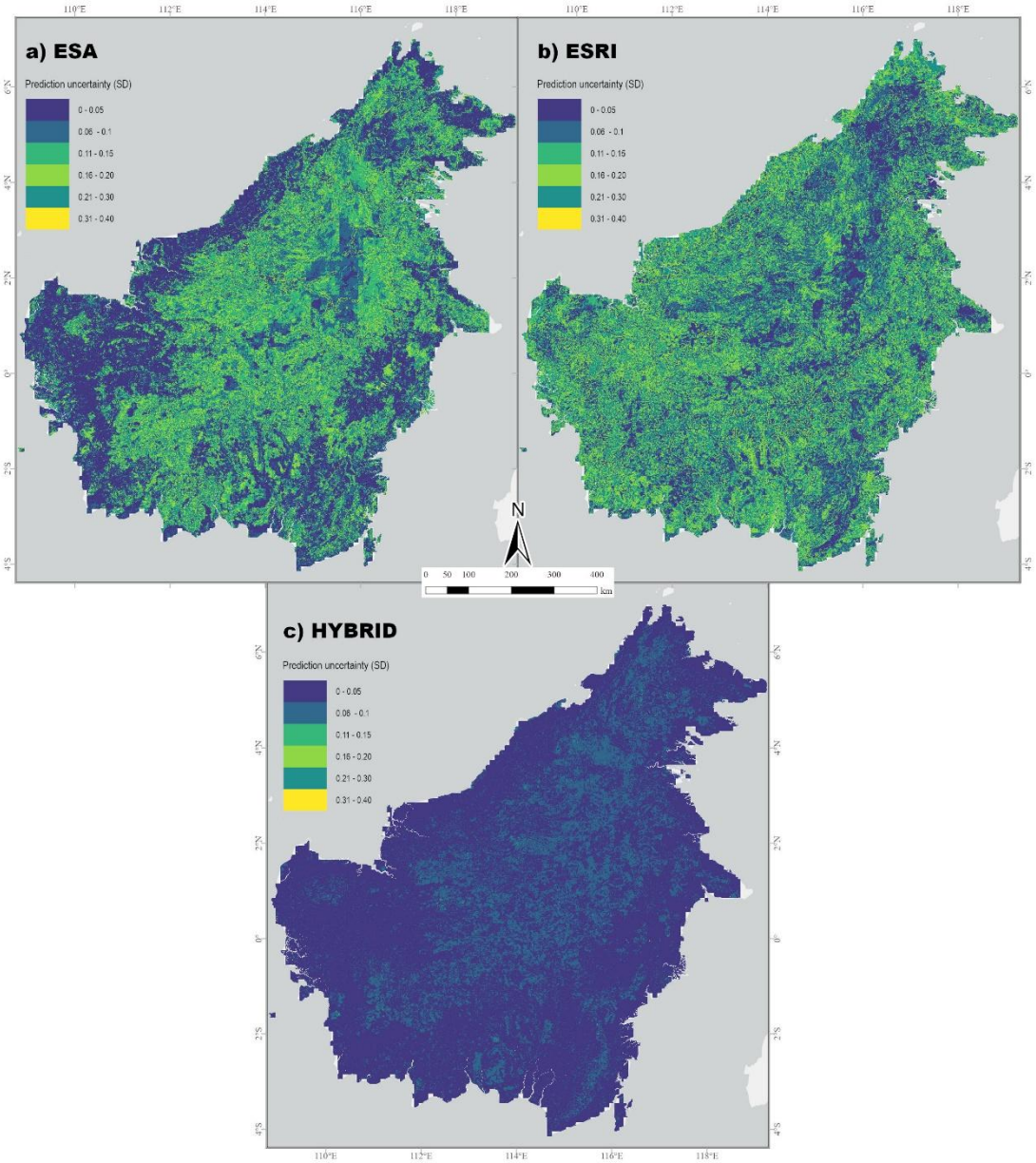

Fig. 8. Spatial uncertainty (SD) of predicted environmental suitability. Maps show the standard deviation of predictions across spatial cross-validation folds for the best-performing model associated with each land-cover product (a) ESA (XGB), (b) ESRI (XGB), and (c) Hybrid (RF). Lower values indicate greater agreement among model predictions.

4. Discussion

4.1 Environmental associations across modelling choices

Across the three land-cover products, night-time land-surface temperature, daytime land-surface temperature, and distance to roads remained among the highest-ranked predictors. By comparison, the ranking of individual landscape metrics varied more clearly among ESA, ESRI, and Hybrid. Ensemble models reached spatial cross-validation AUC values close to 0.89, and

differences in overall discrimination among the harmonised products were modest. However, the spatial distribution and fold-based uncertainty of the predictions differed among products. ESA achieved the highest mean discrimination, whereas the Hybrid model produced lower variability among cross-validation folds.

Temperature is relevant to mosquito survival, biting activity, and parasite development, and daily thermal conditions can influence malaria transmission potential (Shapiro et al., 2017; Shocket et al., 2025). Remotely sensed temperature has been incorporated into malaria suitability models for *P. falciparum* and *P. knowlesi* across different regions (Gething et al., 2011; Weiss et al., 2014; Tobin et al., 2024). The consistent ranking of daytime and night-time land-surface temperature in the present analysis is therefore in line with the established role of temperature in malaria ecology (Mordecai et al., 2019).

Daytime and night-time temperature describe related components of the thermal environment. Their joint spatial variation may be more informative than either variable considered separately. Modified landscapes, including forest edges and plantation systems, can experience different thermal conditions from intact forest, creating links between land-cover configuration and temperature-dependent suitability. Elevation also remained moderately important across products and may capture additional thermal, topographic, and ecological gradients across Borneo (Brown et al., 2024).

4.2 Analytical spatial support and surveillance locations

Models based on landscape metrics calculated within the 500 m analytical window achieved the highest mean discrimination, although differences among analytical supports were moderate and their uncertainty ranges overlapped. Smaller windows retained local variation in forest edges, plantations, built-up patches, and mixed land-use mosaics that became increasingly averaged at broader supports (Herdiana et al., 2018). This local variation may contribute useful information for distinguishing occurrence and background observations in heterogeneous tropical landscapes (LaDeau et al., 2015; Naserrudin et al., 2022).

The spatial meaning of this result depends on the characteristics of the surveillance locations. Malaysian observations were predominantly reported at village level, with some available at broader administrative resolution, whereas Indonesian locations were mapped at household level through participatory methods (Fornace et al., 2018; Trianty et al., 2025). These coordinates describe the locations associated with reported cases, while the place of infection may differ because exposure can occur during agricultural, plantation, occupational, or forest-related activities. The higher discrimination obtained at 500 m therefore identifies the analytical support at which the selected landscape summaries were most informative around the recorded

epidemiological locations. It does not define a common infection or exposure radius across the surveillance datasets.

Spatial clustering and thinning reduced local pseudo-replication, spatial dependence, and unequal sampling concentration within the Malaysian and Indonesian datasets. These procedures supported the pooled analysis across Borneo while retaining the geographical contribution of the smaller Indonesian dataset. Differences in the original spatial precision of the records nevertheless remain relevant when interpreting fine-scale landscape associations.

The scale comparison also reflects differences in the spatial resolution of the predictors. Land-cover metrics were recalculated at 500, 1,000, 2,000, and 5,000 m, whereas MODIS land-surface temperature represented conditions at approximately 1 km and ERA5-Land precipitation described considerably broader spatial gradients. The performance difference at 500 m therefore mainly reflects variation in the spatial summarisation of land-cover composition and configuration, combined with climatic predictors operating at coarser spatial supports.

4.3 Accessibility and landscape configuration

Distance to roads remained highly ranked across products and modelling approaches. Roads structure access to settlements, agricultural land, plantations, and forest edges and are closely connected to human movement and land-use change (Fornace, Alexander, et al., 2019). Agricultural expansion, logging, and infrastructure development can alter mosquito and macaque habitats while increasing human activity within forest and peri-forest environments (Grigg et al., 2017; Johnson et al., 2024). This interpretation is consistent with epidemiological evidence linking farming, logging, plantation work, and other forest-related occupations with *P. knowlesi* infection (Grigg et al., 2017; Fornace, Brock, et al., 2019). Within the present models, road proximity likely captures a combination of accessibility, landscape disturbance, and the spatial distribution of human activities.

Landscape metrics added information on the composition and organisation of these environments. Patch density, edge density, aggregation, clumpiness, and land-cover proportion describe different aspects of fragmentation and spatial heterogeneity that may influence contact among people, mosquito vectors, and non-human primate hosts (Talukdar et al., 2021; Rulli et al., 2025; Muh et al., 2025). Their individual contributions were generally lower than those of temperature and road proximity, suggesting that landscape configuration refined suitability patterns within broader climatic and accessibility gradients (Rodewald et al., 2016).

The leading landscape metrics differed among land-cover products. Forest patch density was prominent in the ESA model, forest and water/wetland configuration were prominent in the ESRI model, and built-up configuration was prominent in the Hybrid model. These differences

are consistent with variation in class boundaries, small-patch representation, and the depiction of mixed land-use systems among products (Zhang et al., 2023). Previous studies have similarly associated *P. knowlesi* occurrence with combinations of forest disturbance, human presence, accessibility, and fragmented land-use mosaics (Fornace et al., 2016; Davidson et al., 2019; Tobin et al., 2024).

In the Hybrid model, percentage of built-up land, aggregation, clumpiness, edge density, and patch density collectively accounted for 28.6% of normalised predictor importance. The mapped higher-suitability areas were generally outside major urban centres. In this context, the built-up metrics are more consistent with dispersed settlements and infrastructure embedded within rural, agricultural, and forested landscapes than with dense urban development. Their importance therefore points to the spatial configuration of settlement–landscape interfaces represented by the Hybrid product. The importance values describe the contribution of these variables to model predictions, while their geographical meaning depends on the mapped landscape context.

4.4 Implications of land-cover product choice

The comparison among ESA, ESRI, and Hybrid indicates that similar overall discrimination can coexist with different predictor rankings and spatial predictions. The three harmonised products produced comparable broad-scale performance, but they differed in the continuity, extent, and uncertainty of predicted suitability. Land-cover representation therefore influenced the geographical interpretation of the models even when differences in AUC were small. Differences in classification frameworks and spatial patterns propagated into the derived landscape metrics and affected which land-cover characteristics ranked most highly (Liu et al., 2013). Harmonisation allowed equivalent classes to be compared across products but also reduced some of the native thematic detail of the regional Hybrid classification. This may partly account for the slightly lower mean discrimination of the Hybrid model. At the same time, its lower fold-based uncertainty suggests that the regional representation of plantations, degraded forest, dispersed built-up areas, and mixed land-use mosaics supported more consistent spatial predictions.

These findings are relevant to land-cover selection in health-geographic modelling. Global products may be suitable when the main objective is broad spatial prediction across large areas. Regionally refined products may offer additional value when local landscape configuration, plantation structure, forest edges, or dispersed settlements are central to interpretation. Product choice should therefore consider both predictive performance and how well the land-cover classes represent the environmental processes relevant to the disease system. Comparing more than one product is particularly useful when conclusions depend on individual landscape metrics

or on the location of predicted high-suitability areas (Rendenieks et al., 2017; Alvarez et al., 2025).

The comparison of algorithms also provided complementary information. XGB achieved the highest discrimination under spatial cross-validation but produced higher false-positive rates when evaluated against the independent Indonesian negative observations. Random Forest achieved slightly lower average discrimination and produced more conservative predictions, consistent with differences in calibration and generalisation among ensemble methods (Chollet Ramampiandra et al., 2023). Spatial cross-validation evaluated geographical transfer among folds, while the confirmed negative observations provided an additional assessment in the Indonesian surveillance setting.

4.5 Strengths and limitations

The pooled dataset combined surveillance records that differed in sample size, sampling design, spatial concentration, and coordinate precision. Most positive observations originated from Malaysia, while the Indonesian records increased environmental coverage and provided independent confirmed negatives. The resulting associations describe the combined Borneo study domain and may vary between countries or surveillance settings.

Background samples represented populated and accessible environmental conditions where case detection was plausible. Remote forest environments may be less represented, particularly where exposure occurs but surveillance access is limited. The confirmed negative observations came from Indonesia and therefore assess model behaviour within that geographical and surveillance context.

The land-cover products represented different reference years, during which deforestation, plantation expansion, and infrastructure development may have changed local landscapes. Harmonisation supported direct comparison but reduced native thematic detail, particularly within the regional Hybrid product. The comparison therefore evaluates how alternative land-cover representations affect derived metrics and model results rather than providing a formal assessment of product accuracy.

Environmental predictors also differed in spatial and temporal resolution. MODIS land-surface temperature had a native resolution of approximately 1 km, while ERA5-Land precipitation represented broader spatial conditions. These climatic layers described environmental context across the study area and were coarser than the smallest LULC analytical window. This difference is important when interpreting the apparent advantage of the 500 m landscape summaries.

The use of three land-cover products, four analytical windows, multiple modelling approaches, spatial cross-validation, and independent confirmed negatives allowed methodological sensitivity to be assessed from several perspectives. Temperature and road proximity remained consistently important, while landscape metrics provided product-dependent information about forest, wetland, agricultural, and built-up configuration. Together, the findings indicate that thermal conditions, accessibility, and heterogeneous rural landscapes contribute to the spatial pattern of *P. knowlesi* environmental suitability. These areas may provide useful priorities for surveillance where environmental conditions and changing land-use configurations coincide (Lynggaard et al., 2025).

5. Conclusion

This study compared three harmonised land-cover products and four analytical spatial supports for modelling Plasmodium knowlesi environmental suitability across Borneo. Temperature and distance to roads remained among the most important predictors across products and modelling approaches, while the contribution of landscape metrics varied according to how land cover was represented.

Models using landscape metrics calculated within the 500 m window achieved the highest mean discrimination, showing that local landscape configuration around reported surveillance locations contained useful information for the models. The differences from larger analytical windows were modest, however, indicating that broader environmental patterns also contributed to predicted suitability.

ESA and ESRI produced slightly higher discrimination, whereas the Hybrid product generated more consistent predictions across spatial cross-validation folds. Forest configuration was more prominent in the ESA model, forest and wetland patterns in the ESRI model, and built-up configuration in the Hybrid model. These differences show that similar model performance can lead to different interpretations of the landscapes associated with P. knowlesi occurrence.

Land-cover selection therefore matters not only for predictive accuracy but also for the spatial patterns and environmental relationships identified by the model. Across products, areas combining suitable thermal conditions, road accessibility, and heterogeneous rural landscapes emerged as the main settings of higher predicted suitability and provide a useful basis for spatial surveillance across Borneo.

Acknowledgements

We thank all the co-authors for their efforts in this study.

Availability of data and material

The individual-level and georeferenced epidemiological data are not publicly available because they contain sensitive health and location information and are subject to data-governance and ethical restrictions. Access requests may be directed to the corresponding author and will require approval from the relevant data custodians and ethics bodies.

Statements and Declarations

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by the ZOOMAL project ('Evaluating zoonotic malaria and agricultural land use in Indonesia'; #LS-2019-116), Australian Centre for International Agricultural Research and the Indo-Pacific Centre for Health Security, Department of Foreign Affairs and Trade, Australian Government. BFS was supported by a James Cook University Postgraduate Research Scholarship.

Author's Contributions

C.I.A. conceptualized the study, developed the methodology, supervised the research, and led the writing and revision of the manuscript. M.N. developed the code and processed the data. K.W., M.J.G., R.N., F.M.S., G.R., and N.M.A. contributed to the interpretation of the results and provided critical revisions to the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, language editing tools were used to improve clarity and readability. All content was reviewed and approved by the authors.

References

- Alvarez, C. I., Surendra, H., Perlambang, Y. A., & Waha, K. (2025). The impact of land use classification accuracy on computing landscape metrics for health risk and agricultural models. En Remote Sensing for Agriculture, Ecosystems, and Hydrology XXVII (Vol. 13666, p. 1366617). SPIE. https://doi.org/10.1117/12.3066275
- Barbet-Massin, M., Jiguet, F., Albert, C.H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? Methods in

Ecology and Evolution, 3(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

- Braima, K. A., Piera, K. A., Lubis, I. N. D., Noviyanti, R., Rajahram, G. S., Kariodimedjo, P., Nainggolan, I. R. A., Permatasari, R., Trianty, L., Amalia, R., Sakam, S. S., Tan, A. F., William, T., Westaway, J. A. F., Lee, P., Daim, S., Surendra, H., Christy, N., Letizia, A. G., . . . Grigg, M. J. (2025). Improved limit of detection for zoonotic *Plasmodium knowlesi* and *P. cynomolgi* surveillance using reverse transcription for total nucleic acid preserved samples or dried blood spots. *PLOS Neglected Tropical Diseases*, *19*(2), Article e0012129. https://doi.org/10.1371/journal.pntd.0012129
- Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
- Brock, P. M., Fornace, K. M., Grigg, M. J., Anstey, N. M., William, T., Cox, J., Drakeley, C. J., & Ferguson, H. M. (2019). Predictive analysis across spatial scales links zoonotic malaria to deforestation. Proceedings of the Royal Society B: Biological Sciences, 286(1894), 20182351. https://doi.org/10.1098/rspb.2018.2351
- Brown, O., Flegg, J. A., Weiss, D. J., & Golding, N. (2024). A global mathematical model of climatic suitability for Plasmodium falciparum malaria. Malaria Journal, 23(1), Article 306. https://doi.org/10.1186/s12936-024-05122-7
- Chevalier, M., Mod, H., Broennimann, O., Di Cola, V., Schmid, S., Niculita-Hirzel, H., Pradervand, J.-N., Schmidt, B. R., Ursenbacher, S., Pellissier, L., & Guisan, A. (2021). Low spatial autocorrelation in mountain biodiversity data and model residuals. Ecosphere, 12(3), Article e03403. https://doi.org/10.1002/ecs2.3403
- Chollet Ramampiandra, E., Scheidegger, A., Wydler, J., & Schuwirth, N. (2023). A comparison of machine learning and statistical species distribution models: Quantifying overfitting supports model interpretation. Ecological Modelling, 481, Article 110353. https://doi.org/10.1016/j.ecolmodel.2023.110353
- Congalton, R. G., Gu, J., Yadav, K., Thenkabail, P., & Ozdogan, M. (2014). Global Land Cover Mapping: A Review and Uncertainty Analysis. Remote Sensing, 6(12), 12070-12093. https://doi.org/10.3390/rs61212070
- Cooper, D. J., Rajahram, G. S., William, T., Jelip, J., Mohammad, R., Benedict, J., Alaza, D. A., Malacova, E., Yeo, T. W., Grigg, M. J., Anstey, N. M., & Barber, B. E. (2020). Plasmodium knowlesi malaria in Sabah, Malaysia, 2015–2017: Ongoing increase in incidence despite near-elimination of the human-only Plasmodium species. Clinical Infectious Diseases, 70(3), 361–367. https://doi.org/10.1093/cid/ciz237

- Davidson, G., Chua, T.H., Cook, A. et al. The Role of Ecological Linkage Mechanisms in Plasmodium knowlesi Transmission and Spread. EcoHealth 16, 594–610 (2019). https://doi.org/10.1007/s10393-019-01395-6
- Descombes, P., Chauvier, Y., Brun, P., Righetti, D., Wüest, R.O., Karger, D.N., Zurell, D., & Zimmermann, N.E. (2022). Strategies for sampling pseudo-absences for species distribution models in complex terrain. bioRxiv. https://doi.org/10.1101/2022.03.24.485693
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x
- Esri. "Land Cover Map of the World" [map]. 2022. Scale not given. "ArcGIS Living Atlas of the World". https://livingatlas.arcgis.com/landcover/ (Accessed: March 20, 2026).
- Fornace, K. M., Abidin, T., Alexander, N., Brock, P., Grigg, M. J., Murphy, A., William, T., Menon, J., Drakeley, C. J., & Cox, J. (2016). Association between landscape factors and spatial patterns of Plasmodium knowlesi infections in Sabah, Malaysia. Emerging Infectious Diseases, 22(2), 201–208. https://doi.org/10.3201/eid2202.150656
- Fornace, K. M., Alexander, N., Abidin, T. R., Brock, P. M., Chua, T. H., Vythilingam, I., Ferguson, H. M., Manin, B. O., Wong, M. L., Ng, S. H., Cox, J., & Drakeley, C. (2019). Local human movement patterns and land use impact exposure to zoonotic malaria in Malaysian Borneo. eLife, 8, Article e47602. https://doi.org/10.7554/eLife.47602
- Fornace, K. M., Brock, P. M., Abidin, T. R., Grignard, L., Herman, L. S., Chua, T. H., Daim, S., William, T., Patterson, C. L. E. B., Hall, T., Grigg, M. J., Anstey, N. M., Tetteh, K. K. A., Cox, J., & Drakeley, C. J. (2019). Environmental risk factors and exposure to the zoonotic malaria parasite Plasmodium knowlesi across northern Sabah, Malaysia: A population-based cross-sectional survey. The Lancet Planetary Health, 3(4), e179–e186. https://doi.org/10.1016/S2542-5196(19)30045-2
- Fornace, K.M., Surendra, H., Abidin, T.R. et al. Use of mobile technology-based participatory mapping approaches to geolocate health facility attendees for disease surveillance in low resource settings. Int J Health Geogr 17, 21 (2018). https://doi.org/10.1186/s12942-018-0141-0
- Frazier, A. E., & Kedron, P. (2017). Landscape metrics: Past progress and future directions. Current Landscape Ecology Reports, 2, 63–72.

https://doi.org/10.1007/s40823-017-0026-0

- Gaveau, D., Sheil, D., Husnayaen et al. Rapid conversions and avoided deforestation: examining four decades of industrial plantation expansion in Borneo. Sci Rep 6, 32017 (2016). https://doi.org/10.1038/srep32017
- Geofabrik GmbH. (2025). Asia OpenStreetMap Data Extracts. Geofabrik Download Server. https://download.geofabrik.de/asia.html
- Gething, P. W., Van Boeckel, T. P., Smith, D. L., Guerra, C. A., Patil, A. P., Snow, R. W., & Hay, S. I. (2011). Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax. Parasites & Vectors, 4(92). https://doi.org/10.1186/1756-3305-4-92
- Grigg, M. J., Cox, J., William, T., Jelip, J., Fornace, K. M., Brock, P. M., von Seidlein, L., Barber, B. E., Anstey, N. M., Yeo, T. W., & Drakeley, C. J. (2017). Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: A case-control study. The Lancet Planetary Health, 1(3), e97–e104. https://doi.org/10.1016/S2542-5196(17)30031-1
- Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast density-based clustering with R. Journal of Statistical Software, 91(1), 1–30. https://doi.org/10.18637/jss.v091.i01
- Hesselbarth, M. H. K., Sciaini, M., With, K. A., Wiegand, K., & Nowosad, J. (2019). landscapemetrics: An open-source R tool to calculate landscape metrics. Ecography, 42(10), 1648–1657. https://doi.org/10.1111/ecog.04617
- Herdiana, H., Irnawati, I., Coutrier, F.N. et al. Two clusters of Plasmodium knowlesi cases in a malaria elimination area, Sabang Municipality, Aceh, Indonesia. Malar J 17, 186 (2018). https://doi.org/10.1186/s12936-018-2334-1
- Hulthen, Andy; & Waha, Katharina (2022): Indonesian Hybrid Landcover 2020 - GeoTIFF. v4. CSIRO. Data Collection. https://doi.org/10.25919/8624-rg56
- Johnson, E., Sharma, R. S. K., Ruiz Cuenca, P., Byrne, I., Salgado-Lynn, M., Shahar, Z. S., Lin, L. C., Zulkifli, N., Saidi, N. D. M., Drakeley, C., Matthiopoulos, J., Nelli, L., & Fornace, K. (2024). Landscape drives zoonotic malaria prevalence in non-human primates. eLife, 12, Article RP88616. https://doi.org/10.7554/eLife.88616
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Würsten, C., Belant, J.L., Hofer, H., & Wilting, A. (2013). The importance of correcting for

sampling bias in MaxEnt species distribution models. Diversity and Distributions, 19(11), 1366–1379. https://doi.org/10.1111/ddi.12096

- LaDeau, S. L., Allan, B. F., Leisnham, P. T., & Levy, M. Z. (2015). The ecological foundations of transmission potential and vector-borne disease in urban landscapes. Functional Ecology, 29(7), 889–901. https://doi.org/10.1111/1365-2435.12487
- Lambin EF, Tran A, Vanwambeke SO, Linard C, Soti V. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. Int J Health Geogr. 2010 Oct 27;9:54. doi: 10.1186/1476-072X-9-54. PMID: 20979609; PMCID: PMC2984574.
- Liu, D., Hao, S., Liu, X. et al. Effects of land use classification on landscape metrics based on remote sensing and GIS. Environ Earth Sci 68, 2229–2237 (2013). https://doi.org/10.1007/s12665-012-1905-7
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
- Lynggaard, C., Gogarten, J. F., & Jokelainen, P. (2025). Integrating environment into One Health approaches: Embracing the complexity. One Health, 20, Article 101021. https://doi.org/10.1016/j.onehlt.2025.101021
- Machault V, Vignolles C, Borchi F, Vounatsou P, Pages F, Briolant S, Lacaux JP, Rogier C. The use of remotely sensed environmental data in the study of malaria. Geospat Health. 2011 May;5(2):151-68. doi: 10.4081/gh.2011.167. PMID: 21590665.
- McGarigal, K., Cushman, S.A., and Ene, E. (2021). FRAGSTATS V4: Spatial Pattern Analysis Program for Categorical Maps (University of Massachusetts Amherst).
- Mordecai, E. A., Cohen, J. M., Evans, M. V., Gudapati, P., Johnson, L. R., Lippi, C. A., Miazgowicz, K., Murdock, C. C., Rohr, J. R., Ryan, S. J., Savage, V., Shocket, M. S., Ibarra, A. S., Thomas, M. B., & Weikel, D. P. (2019). Thermal biology of mosquito-borne disease. Ecology Letters, 22(10), 1690–1708. https://doi.org/10.1111/ele.13335
- Muh, F., Qanita, N. G., Yuwanita, M. R., Wusono, A. D., Rahayu, A. A., Amirudin, F. A., Firdaus, M. E. R., Cahyani, N. K. D., & Han, J. H. (2025). Plasmodium knowlesi and global changes: Socio-ecological drivers and challenges. Malaria Journal, 25(1), Article 5. https://doi.org/10.1186/s12936-025-05687-x
- Muñoz Sabater, J. (2019): ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.e2161bac
- NASA Shuttle Radar Topography Mission (SRTM)(2013). Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF. Accessed 2026-03-20

- Naserrudin, N. A., Hod, R., Jeffree, M. S., Ahmed, K., Culleton, R., & Hassan, M. R. (2022). The Role of Human Behavior in Plasmodium knowlesi Malaria Infection: A Systematic Review. International Journal of Environmental Research and Public Health, 19(6), 3675. https://doi.org/10.3390/ijerph19063675
- Nuin, N. A., Tan, A. F., Lew, Y. L., Piera, K. A., William, T., Rajahram, G. S., Jelip, J., Dony, J. F., Mohammad, R., Cooper, D. J., Barber, B. E., Anstey, N. M., Chua, T. H., & Grigg, M. J. (2020). Comparative evaluation of two commercial real-time PCR kits (QuantiFast™ and abTES™) for the detection of Plasmodium knowlesi and other Plasmodium species in Sabah, Malaysia. Malaria Journal, 19(1), Article 306. https://doi.org/10.1186/s12936-020-03379-2
- Paaijmans, K. P., Read, A. F., & Thomas, M. B. (2009). Understanding the link between malaria risk and climate. Proceedings of the National Academy of Sciences, 106(33), 13844–13849. https://doi.org/10.1073/pnas.0903423106
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications, 19(1), 181–197. https://doi.org/10.1890/07-2153.1
- Plowright, R. K., Reaser, J. K., Locke, H., Woodley, S. J., Patz, J. A., Becker, D. J., Oppler, G., Hudson, P. J., & Tabor, G. M. (2021). Land use-induced spillover: A call to action to safeguard environmental, animal, and human health. The Lancet Planetary Health, 5(4), e237–e245. https://doi.org/10.1016/S2542-5196(21)00031-0
- Prestele, R., Alexander, P., Rounsevell, M. D., Arneth, A., Calvin, K., Doelman, J., Eitelberg, D. A., Engström, K., Fujimori, S., Hasegawa, T., Havlik, P., Humpenöder, F., Jain, A. K., Krisztin, T., Kyle, P., Meiyappan, P., Popp, A., Sands, R. D., Schaldach, R., . . . Verburg, P. H. (2016). Hotspots of uncertainty in land-use and land-cover change projections: A global-scale model comparison. Global Change Biology, 22(12), 3967–3983. https://doi.org/10.1111/gcb.13337
- Rakotoarison, H. A., Rasamimalala, M., Rakotondramanga, J. M., Ramiranirina, B., Franchard, T., Kapesa, L., Razafindrakoto, J., Guis, H., Tantely, L. M., Girod, R., Rakotoniaina, S., Baril, L., Piola, P., & Rakotomanana, F. (2020). Remote Sensing and Multi-Criteria Evaluation for Malaria Risk Mapping to Support Indoor Residual Spraying Prioritization in the Central Highlands of Madagascar. Remote Sensing, 12(10), 1585. https://doi.org/10.3390/rs12101585
- Rendenieks, Z., Tērauds, A., Nikodemus, O., & Brūmelis, G. (2017). Comparison of input data with different spatial resolution in landscape pattern analysis – A case study

from northern Latvia. Applied Geography, 83, 100–106. https://doi.org/10.1016/j.apgeog.2017.03.019

- Rodewald, A.D., Arcese, P. Direct and Indirect Interactions between Landscape Structure and Invasive or Overabundant Species. Curr Landscape Ecol Rep 1, 30–39 (2016). https://doi.org/10.1007/s40823-016-0004-y
- Rulli, M. C., D'Odorico, P., Galli, N., John, R. S., Muylaert, R. L., Santini, M., & Hayman, D. T. S. (2025). Land use change and infectious disease emergence. Reviews of Geophysics, 63, e2022RG000785. https://doi.org/10.1029/2022RG000785
- Shapiro LLM, Whitehead SA, Thomas MB (2017) Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria. PLoS Biol 15(10): e2003489. https://doi.org/10.1371/journal.pbio.2003489
- Shocket, M.S., Bernhardt, J.R., Miazgowicz, K.L. et al. Mean daily temperatures predict the thermal limits of malaria transmission better than hourly rate summation. Nat Commun 16, 3441 (2025). https://doi.org/10.1038/s41467-025-58612-w
- Talukdar, S., Eibek, K. U., Akhter, S., Ziaul, S., Towfiqul Islam, A. R. M., & Mallick, J. (2021). Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh. Ecological Indicators, 126, Article 107612. https://doi.org/10.1016/j.ecolind.2021.107612
- Tatem, A. WorldPop, open data for spatial demography. Sci Data 4, 170004 (2017). https://doi.org/10.1038/sdata.2017.4
- Tobin, R. J., Harrison, L. E., Tully, M. K., Lubis, I. N. D., Noviyanti, R., Anstey, N. M., Rajahram, G. S., Grigg, M. J., Flegg, J. A., Price, D. J., & Shearer, F. M. (2024). Updating estimates of Plasmodium knowlesi malaria risk in response to changing land use patterns across Southeast Asia. PLOS Neglected Tropical Diseases, 18(1), Article e0011570. https://doi.org/10.1371/journal.pntd.0011570
- Trianty, L., Amalia, R., Noviyanti, R., Kariodimedjo, P., Fadila, N., Nisa, F. A., Coutrier, F. N., Yuniarti, T., Irnawati, I., Hidar, Burkot, T. R., Piera, K. A., Barber, B. E., Anstey, N. M., Garjito, T. A., & Grigg, M. J. (2025). Severe and uncomplicated *Plasmodium knowlesi* malaria in North Kalimantan and Sabang, Aceh, and suspected *Anopheles* vectors across Kalimantan, Indonesia. *medRxiv*, 2025.04.29.25326532. https://doi.org/10.1101/2025.04.29.25326532
- Turner, M. G., Gardner, R. H., & O'Neill, R. V. (2001). Landscape ecology in theory and practice: Pattern and process. Springer.

- Wan, Z., Hook, S., & Hulley, G. (2021). MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061 [Data set]. NASA Land Processes Distributed Active Archive Center (LP DAAC). https://doi.org/10.5067/MODIS/MOD11A1.061
- Weiss, D. J., Lucas, T. C. D., Nguyen, M., Nandi, A. K., Bisanzio, D., Battle, K. E., Cameron, E., Twohig, K. A., Pfeffer, D. A., Rozier, J. A., Gibson, H. S., Rao, P. C., Casey, D., Bertozzi-Villa, A., Collins, E. L., Dalrymple, U., Gray, N., Harris, J. R., Howes, R. E., … Gething, P. W. (2019). Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–2017. The Lancet, 394(10195), 322–331. https://doi.org/10.1016/S0140-6736(19)31097-9
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., & Arino, O. (2022). ESA WorldCover 10 m 2021 v200 (Version v200) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7254221
- Zhang, T., Cheng, C. & Wu, X. Mapping the spatial heterogeneity of global land use and land cover from 2020 to 2100 at a 1 km resolution. Sci Data 10, 748 (2023). https://doi.org/10.1038/s41597-023-02637-7

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarymaterialZoomalLULC1.docx