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.

Adaptive Spatial Regularization for Aedes-Borne Disease Risk Prediction: A Case Study of Chikungunya

Shuaishuai Ji
Northwest Minzu University
Jia Mu
284082126@xbmu.edu.cn

Northwest Minzu University
Ling Guo
Yangzhou University
Renqing Suonan
Northwest Minzu University
Zhen Zhao
Northwest Minzu University
Qiwen Yuan
Northwest Minzu University

Research Article

Keywords: CHIKF, ASR-XGBoost, Spatial regularization, Ensemble learning, Risk distribution prediction

Posted Date: October 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11042386/v1

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

Additional Declarations: No competing interests reported.

Adaptive Spatial Regularization for Aedes-Borne Disease Risk Prediction: A Case Study of Chikungunya

Shuaishuai Ji[1], Jia Mu[1*], Ling Guo[2], Renqing Suonan[1], Zhen Zhao[3], Qiwen Yuan[3]

[1*]School of Mathematical Sciences, Northwest Minzu University, Lanzhou, 730000, Gansu, China.
[2]School of Information and Artificial Intelligence, Yangzhou University, Yangzhou, 225009, Jiangsu, China.
[3]China National Institute of Ethnic Information Technology, Northwest Minzu University, Lanzhou, 730000, Gansu, China.

*Corresponding author(s). E-mail(s): 284082126@xbmu.edu.cn; mujia88@163.com;
Contributing authors: 18035630120@163.com; gling@yzu.edu.cn;

Abstract

Chikungunya fever (CHIKF), caused by the chikungunya virus (CHIKV), is a widespread Aedes-borne disease whose range is expanding under climate change. Reliable risk prediction from presence-only surveillance data is essential for vector control but faces two obstacles: underreporting makes reported absence an unreliable label, and standard models treat neighboring grid cells as spatially independent. We propose ASR-XGBoost (Adaptive Spatial Regularization XGBoost), whose centerpiece embeds spatial dependence directly into the training objective: a gated penalty integrates 8-neighborhood mean soft labels with block-adaptive strength, constraining the predicted probability surface rather than adding sample coordinates. Ecologically stratified pseudo-absence sampling reduces label bias, while an anomaly-correction step (ASRAC) that extends ASR repairs spatially inconsistent cells. On a global CHIKF dataset, ASR-XGBoost achieves competitive discrimination (test AUC = 0.9334) with the best spatial generalization and calibration (AUC(CV) = 0.9187, Brier(CV) = 0.1234); ASRAC further improves spatial continuity (Moran's I = 0.8200) while preserving predictive performance. These components reduce label bias and keep spatial dependence explicit; they are also applicable to other Aedes-borne diseases,

supporting reliable risk prediction in data-sparse regions for forward-looking surveillance planning.

Keywords: CHIKF, ASR-XGBoost, Spatial regularization, Ensemble learning, Risk distribution prediction

1 Introduction

Chikungunya fever (CHIKF), transmitted primarily by *Aedes aegypti* and *Aedes albopictus* Côrtes et al. (2023); Burt et al. (2017), imposes a substantial and growing global health burden, with an estimated 35 million annual infections across over 104 countries Ribeiro dos Santos et al. (2025); Cortes-Azuero et al. (2026). Historically confined to tropical Africa and Asia, the causative chikungunya virus (CHIKV) has repeatedly demonstrated its capacity for explosive emergence. Although acute infection is rarely fatal, a substantial proportion of patients develop chronic arthralgia that can persist for months to years, making CHIKF a source of long-term morbidity rather than merely a transient febrile illness Weaver and Lecuit (2015). The two vectors also differ ecologically in ways that shape the spatial risk distribution: *Ae. aegypti* is a highly anthropophilic, container-breeding mosquito concentrated in urban and peri-urban settings, whereas *Ae. albopictus* tolerates cooler climates and has established self-sustaining populations in temperate regions, extending the potential transmission zone beyond the tropics and subtropics Weaver and Lecuit (2015); Kraemer et al. (2019). Ongoing climate warming expands the latitudinal and altitudinal suitable habitats of Aedes vectors Kraemer et al. (2019); Lim et al. (2025), while intensified human mobility accelerates cross-border virus importation, raising the frequency of local outbreaks Ribeiro dos Santos et al. (2025); Farooq et al. (2025). As the suitable zone shifts and expands into regions with no prior transmission history, anticipating where transmission risk is highest can help inform resource allocation, vector surveillance, and epidemic early warning.

Global risk prediction is not simply regional modeling scaled up. Regional studies, such as reconstructions of chikungunya circulation in the Philippines Salje et al. (2016) and in Burkina Faso and Gabon Lim et al. (2023), mathematical modeling of chikungunya in Brazil Cortes-Azuero et al. (2026), and short-term disease forecasting for Brazilian cities Lober et al. (2024), work within a bounded domain where environmental relationships can be treated as approximately stationary, background points can be sampled under a controlled design, and random validation splits draw test samples from the same climatic and epidemiological regime as the training data. A global model faces a different problem. Occurrence records are too sparse once spread across the whole globe to support the time-series models used for regional prediction, so records from every year are pooled into a single static map Nsoesie et al. (2016); Messina et al. (2019), and the target becomes the spatial distribution of risk rather than the number of cases expected over a future period.

Working at global scale also constrains what data can be used. Presence-absence sampling would require systematic testing across the entire globe, which no surveillance program provides, so only reported occurrences are available. A missing record therefore reflects the absence of detection at least as much as the absence of transmission, and most arbovirus distribution modeling works in exactly this setting, relying on presence-only occurrence records aggregated from multi-source surveillance systems Lim et al. (2025); Messina et al. (2019); Yang et al. (2025), which brings two inherent data limitations from a spatial statistical perspective. First, uneven medical diagnostic capacity worldwide leads to severe underreporting: a large proportion of asymptomatic or mild infections fail to be documented; consequently, random pseudo-absence sampling strategies become unreliable, because the absence of reported cases does not imply absence of transmission Meyer et al. (2025); Salje et al. (2016); Lim et al. (2023). Second, geographic data almost universally exhibits spatial autocorrelation consistent with Tobler's first law of geography: adjacent land parcels share similar climate, vector habitat, and human activity conditions, meaning neighboring grid cells should have correlated transmission risk Tobler (1970).

However, shared environmental drivers explain only part of this spatial correlation in infectious diseases. Following introduction into a receptive region, transmission can generate local dependence as secondary cases cluster around primary cases and outbreaks expand from discrete foci, so neighboring locations may be coupled by transmission as well as by shared covariates; this combination of local expansion and repeated introduction is evident in the modern epidemiology of chikungunya Weaver and Lecuit (2015). Spatial statistics supplies tools for representing such dependence, and they have been applied directly to infectious disease data, for example in spatio-temporal models with conditional autoregressive priors used to track COVID-19 cases and deaths jointly across English local authorities Sahu and Böhning (2022), within a Bayesian disease mapping tradition whose scope continues to grow MacNab (2022). More broadly, mature disciplines treat spatial structure as an explicit part of the model, through covariance functions in model-based geostatistics Local Burden of Disease Child Growth Failure Collaborators (2020), spatial weight matrices in spatial econometrics Knappe et al. (2026), or quantified uncertainty in digital mapping Poggio et al. (2021); Meyer and Pebesma (2022). Although spatially structured covariates can allow a conventional machine-learning model to recover some geographic pattern, its usual pointwise loss does not explicitly represent dependence between observations, which motivates the spatially explicit objective developed in this study.

To realize this spatially explicit approach, we need a learner whose training objective can accommodate an extra spatial penalty. XGBoost Chen and Guestrin (2016) is the natural choice, because it supports user-defined objectives and is among the strongest learners for presence-only data of the type considered here, and it therefore serves as the modeling backbone of this study. Although deep tabular models Borisov et al. (2024); Somvanshi et al. (2026) have shown competitive accuracy on large-scale nonlinear benchmarks, tree-based methods generally match or exceed them on moderate-size presence-only datasets Grinsztajn et al. (2022). Tree ensembles have been widely used for global arbovirus risk mapping, yielding global risk layers for dengue Messina et al. (2019), chikungunya Nsoesie et al. (2016); Yang et al. (2025), and

the broader Aedes-borne disease complex Lim et al. (2025); Leta et al. (2018). Several spatial extensions have been proposed to overcome grid-wise independence. Geographically weighted XGBoost Grekousis (2025); Gao et al. (2026) fits local models but requires enough local samples and is sensitive to the chosen bandwidth; Gaussian process boosting Sigrist (2022) adds a spatial random effect but uses sample coordinates as inputs, constraining extrapolation beyond the sampled envelope; and coordinate proxy features can cause over-reliance on location and misleadingly favorable validation Milà et al. (2024). Standard non-spatial validation further overestimates predictive power when spatial autocorrelation is ignored Ploton et al. (2020). None of these strategies repairs the label bias of presence-only surveillance data, and a spatial regularizer acting directly on the predicted probability surface avoids both limitations.

To respond to these two limitations, the proposed framework integrates three components. (i) To repair the label bias of presence-only records, an ecologically stratified pseudo-absence pool samples only from locations with no reported CHIKF, combining three complementary sources—regions endemic for yellow fever or Zika, areas of negligible vector suitability, and populated suitable regions without reported cases—with farthest-point spatial dispersion, so that each 30-ensemble sub-model trains on informative, balanced negatives rather than on random background samples (Section 2.2). (ii) As the core of the framework, ASR-XGBoost embeds spatial dependence directly into the training objective. It integrates pixel-wise 8-neighborhood mean soft labels, computed from an initial ensemble prediction surface, into the XGBoost objective through a gated spatial penalty whose weight is largest where predictions are most uncertain ($p \approx 0.5$) and negligible for confident predictions; the regularization strength is selected per administrative block within the theoretically supported range, adapting to regional spatial heterogeneity rather than applying a fixed global smoothness (Section 2.3). Because the spatial constraint acts on the predicted probability surface rather than on sample coordinates, it avoids coordinate leakage and remains applicable beyond the historical occurrence envelope. (iii) Finally, because even a well-trained classifier can leave isolated, spatially inconsistent cells in the final map, ASRAC, an anomaly-correction post-processing step built on ASR, replaces such cells with spatially aware ASR predictions, improving map coherence while preserving ecological prediction fidelity (Section 3.5).

To assess the framework, we employ a comprehensive set of metrics, including standard predictive metrics (area under the ROC curve (AUC), Brier score, Recall), spatial block cross-validation (CV), and dedicated spatial smoothness indicators (Moran's I, continuous edge density, isolated-cell ratio). Spatial block CV partitions the globe into $5^\circ \times 5^\circ$ blocks and refits the model with the test blocks excluded, avoiding the over-optimistic bias of random splits for spatially autocorrelated data Roberts et al. (2017); Ploton et al. (2020). Multiple tree-based and deep tabular learning baselines are included for comparative analysis, and cross-disease replication on dengue, yellow fever, and Zika—Aedes-borne diseases sharing the same vectors—assesses transferability beyond CHIKF, while synthetic experiments with known noise isolate the effect of the adaptive regularization. The remainder of this paper is organized as follows: Section 2 introduces the data and the full ASR-XGBoost methodology, Section 3

presents experimental results, Section 4 provides discussion, and Section 5 concludes the study.

2 Materials and methods

The following subsections describe the data (Section 2.1), the negative sampling strategy (Section 2.2), the ASR-XGBoost algorithm (Section 2.3), its theoretical properties (Section 2.4), and the two-stage prediction pipeline (Section 2.5). The negative sampling strategy draws pseudo-absences from yellow fever- and Zika-endemic regions, while the ASR objective embeds spatial continuity into the training objective based on the spatial structure of the prediction surface alone.

2.1 Data acquisition and preprocessing

Longitude and latitude data for CHIKF, yellow fever, Zika, and dengue occurrence records from 1927 to 2024 were obtained from the Arbo risk maps public repository published by Lim et al. (2025). Only records from 2000 onward were used. CHIKF occurrence records were supplemented with additional cases collated by Yang et al. (2025). Pixel-level spatial thinning was applied to retain one record per grid cell.

Nineteen bioclimatic variables (1970–2000 averages) were obtained from WorldClim (https://www.worldclim.org/). Long-term (1970–2000) climatological averages are used to characterize the stable climatic niche of the vectors rather than contemporaneous conditions, following standard practice in species distribution modeling. Mosquito suitability data were obtained from Kraemer et al. (2015), provided as GeoTIFF at 2.5 arcminute resolution.

Global population data for 2020 were obtained from WorldPop (https://hub.worldpop.org/). Healthcare accessibility, measured as motorized travel time to healthcare facilities (2020), was sourced from the Malaria Atlas Project (https://malariaatlas.org/). Nighttime lights data for 2020, serving as a proxy for economic development and infrastructure level, were obtained from the harmonized DMSP-VIIRS dataset Li et al. (2020). All datasets were provided at 1×1 km resolution.

All raster datasets were reprojected and resampled to match the 2.5 arcminute resolution of the first bioclimatic variable (Bio1) using bilinear interpolation.

For dengue, chikungunya, Zika, and yellow fever, Lim et al. assembled 58,361 occurrence records spanning 1927 to March 2024 from the HealthMap event platform and cross-referenced them against the peer-reviewed literature and epidemiological bulletins, incorporating additional records only where subnational-scale information was available. Spatial duplicates were removed by collapsing presence records over time, so that repeated reports of the same outbreak would not skew the geographical representation, and the resulting records were shared with the WHO Technical Advisory Group on Arbovirus for expert feedback. The compilation itself documents that detection, diagnosis, and reporting capacity vary across regions, so the observed spatial pattern partly reflects surveillance intensity. The supplementary CHIKF records of Yang et al. were assembled through a systematic literature search of PubMed, Web of Science, Scopus, and Embase, complemented by WHO, PAHO, and ECDC reports and by GenBank submissions with validated spatial metadata. The search followed

a PROSPERO-registered protocol and the PRISMA reporting standard, with two authors independently screening records and resolving disagreements by consensus, and spatial and temporal standardization defined an occurrence as a unique location on a 10 km grid, yielding 3,327 geolocated records spanning 1955 to 2023. Because CHIKF reporting was sporadic before 2000 and accelerated with the emergence of the Indian Ocean lineage, only records from 2000 onward were retained. Residual reporting heterogeneity is inherent to presence-only data and is discussed in Section 4.

The covariate layers are established global products with documented provenance. WorldClim bioclimatic surfaces are interpolated from a large network of station records at 2.5 arcminute resolution, although station density varies across regions. The vector suitability maps of Kraemer et al. are ensemble species distribution predictions trained on occurrence records and constrained by expert opinion. WorldPop population counts are dasymetric disaggregations of national censuses at approximately 1 km resolution, so their accuracy reflects the quality and timeliness of the underlying censuses. The Malaria Atlas Project travel-time layer is computed as least-cost motorized travel time to healthcare facilities over a global friction surface. The harmonized DMSP-VIIRS nighttime lights product calibrates two satellite sensors into a consistent time series, although the signal saturates in bright urban cores and is absent in unlit rural areas.

2.2 Negative sample generation and data partitioning

To address underreporting caused by regional differences in medical capacity, we designed an ecologically stratified pseudo-absence sampling strategy. Pseudo-absence selection strategies have been systematically evaluated in the species distribution modeling (SDM) literature Barbet-Massin et al. (2012); Renner et al. (2015) and have recently been extended to account for sampling biases in zoonotic disease contexts Simoncini et al. (2025). Our stratified approach incorporates cross-disease occurrence information into the pseudo-absence selection process. Negative samples were drawn from three ecologically distinct sources, all restricted to populated locations without reported CHIKF cases: (i) yellow fever (YF)/Zika virus (ZIKV)-endemic regions Lim et al. (2025), drawing inspiration from the target-group background sampling framework Phillips et al. (2009), which uses occurrence records of related taxa as background points; here, records of co-circulating Aedes-borne diseases with distinct clinical profiles serve as reliable pseudo-absence labels; yellow fever presents clinically distinct symptoms (jaundice, hemorrhagic features), and ZIKV can be reliably distinguished from Chikungunya virus (CHIKV) by serological and molecular assays Cardona-Trujillo et al. (2022), so the presence of these diseases signals established arbovirus surveillance capacity; co-circulating CHIKV would likely be detected if present; (ii) mosquito suitability < 0.1, populated (30% target); and (iii) mosquito suitability ≥ 0.1, no reported cases, where human population exists but no cases have been reported. The remaining quota after allocating (ii) was filled from (iii). A threshold of 0.1 was chosen as a conservative criterion for defining negligible mosquito suitability; a stricter threshold would restrict the available candidate pool. The final pool comprised 31.1% YF/ZIKV-endemic, 30.0% mosquito suitability < 0.1, and 38.9% mosquito suitability ≥ 0.1 with no reported cases (Table 1). All samples were restricted to populated areas.

Table 1 Pseudo-absence sample proportions

Source	Count	Proportion
YF/ZIKV-endemic	4,559	31.1%
Mosquito suitability < 0.1	4,391	30.0%
Mosquito suitability ≥ 0.1	5,689	38.9%
Total	14,639	100%

To ensure that negative samples from disease-free suitable areas are spatially dispersed and well-separated from CHIKF occurrence records, farthest point sampling (FPS) Eldar et al. (1997) was applied to this category. FPS iteratively selects the candidate grid cell whose distance to the nearest already-selected point is maximal, ensuring maximum spatial dispersion from the existing set of CHIKF presence locations and already-chosen negative samples:

$$\mathbf{x}_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{C}_{\text{bg}}} \min_{b\in\mathcal{B}_t} \|\mathbf{x} - b\|,$$

where $\mathcal{C}_{\text{bg}}$ is the candidate pool of disease-free suitable areas, $\mathcal{B}_t$ contains the t samples already selected, and $b \in \mathcal{B}_t$ is an individual selected point. The FPS criterion reduces to maximizing spatial dispersion among the remaining background candidates only, with a minimum spacing of 5 km enforced between selected points.

The candidate pool for FPS ($\mathcal{C}_{\text{bg}}$) therefore consists of grid cells that satisfy all of the following: (i) mosquito suitability ≥ 0.1; (ii) population ¿ 0; (iii) no reported cases of CHIKF, YF, ZIKV, or dengue; and (iv) distance ≥ 30 km from every CHIKF occurrence record, a conservative buffer that prevents pseudo-absences from being drawn too close to reported cases, where pseudo-absence labels would be least reliable. The full dataset was divided into training (70%), validation (20%), and test (10%) sets using a double-stratification approach to ensure consistent proportions of positive and negative samples in each fold.

This stratification separates negative samples by data reliability, reducing the risk that unreliable pseudo-absence labels dominate the training signal. Random background sampling from presence-only data tends to mix easy negatives with hard negatives, where pseudo-absence labels are less certain. Down-weighting samples from the less reliable stratum reduces the impact of potentially noisy pseudo-absence labels on the decision boundary. The 1:1 positive-to-negative ratio per bootstrap round maintains class balance and ensures that the limited number of positive occurrence records receive sufficient attention during training, preventing the model from being overwhelmed by negative samples. Rather than drawing a fixed multiple of negatives per round, we paired each round of bootstrap with an equal number of negatives (n_{pos}) drawn proportionally from the three strata. Stratified proportional sampling ensures that hard negatives are well-represented in each round, and the 30-ensemble provides broad coverage of the negative pool through resampling with replacement.

One negative stratum is defined by a mosquito suitability threshold, which raises the question of whether the model reproduces this threshold rather than learning genuine risk signals. We address this question empirically in the Results, where the model is re-evaluated on test subsets that exclude low-suitability cells (Table 5).

2.3 Adaptive spatial regularization XGBoost

ASR-XGBoost integrates spatial continuity into XGBoost training through a regularization objective. Throughout this paper, Adaptive Spatial Regularization (ASR) denotes the adaptive variant in which the regularization weight λ is selected per spatial block (Eq. (4)); the fixed-strength variant is denoted Spatial Regularization (SR). ASR-XGBoost denotes this regularized model, and the anomaly-correction step described in Section 3.5 is a post-processing extension of ASR. ASR regularizes the final model toward a fixed, spatially aggregated teacher surface derived from the initial ensemble; it encourages, but does not impose, spatial continuity at inference.

The training proceeds in two stages.

Initial ensemble. The base model uses XGBoost's default binary cross-entropy (CE) objective:

$$Obj^{(t)} = \sum_{i=1}^{n} \ell_{\mathrm{CE}} \left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i) \right) + \Omega(f_t), \tag{1}$$

where $Obj^{(t)}$ is the objective at the t-th boosting iteration, n is the number of training samples, $y_i \in \{0, 1\}$ is the observed label at location $\mathbf{x}_i$, $\hat{y}_i^{(t-1)}$ is the cumulative prediction from the previous $t-1$ trees, $f_t(\mathbf{x}_i)$ is the output of the t-th tree, $\ell_{\mathrm{CE}}(\cdot, \cdot)$ is the per-sample binary cross-entropy loss, and $\Omega(f_t)$ is the tree complexity penalty. An ensemble of 30 XGBoost models is trained via this standard cross-entropy on bootstrap-balanced subsets. Each base model used the following hyperparameters: tree depth 5, learning rate 0.03, subsample ratio 0.7, column subsample ratio 0.7, minimum child weight 5, gamma 0.1, L1 and L2 regularization both 0.1, and 1,000 trees per model.

To capture spatial correlation of the prediction surface, for each training sample at geographic location $\mathbf{x}_i$, a soft label $\tilde{p}_i$ is computed as the mean predicted probability of its eight nearest neighbor grid cells (cardinal and diagonal) in the grid, using the initial ensemble predictions:

$$\tilde{p}_i = \frac{1}{8} \sum_{\mathbf{x}_j \in \mathcal{N}_8(\mathbf{x}_i)} \hat{p}^{(0)}(\mathbf{x}_j),$$

where $\mathcal{N}_8(\mathbf{x}_i)$ denotes the 8-connected neighbors and $\hat{p}^{(0)}$ is the initial ensemble mean prediction. The 8-connected neighborhood spans a spatial range of approximately 7 km near the equator, comparable to the resolution of the global covariates used in this study. At higher latitudes the east–west extent of the 8-neighborhood contracts as longitude spacing shrinks, so the physical scale of the neighborhood is latitude-dependent; the $\sim$7 km figure refers to the equatorial resolution. Rather than enforcing a fixed spatial smoothness constraint, the soft label shifts the gradient direction during optimization, nudging the model toward spatially coherent predictions, while retaining the flexibility to deviate from the local neighborhood when warranted by the feature space.

Regularized training. The final ensemble is trained with a modified objective that augments the standard cross-entropy with an adaptive spatial regularization term. The combined objective ($Obj_{\mathrm{ASR}}^{(t)}$) replaces ℓ_{CE} in Eq. (1) with the ASR loss:

$$Obj_{\mathrm{ASR}}^{(t)} = \sum_{i=1}^{n} \ell_{\mathrm{ASR}} \left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i), \tilde{p}_i \right) + \Omega(f_t), \tag{2}$$

where the per-sample ASR loss is:

$$\ell_{\mathrm{ASR}}(y_i, \hat{y}_i, \tilde{p}_i) = \ell_{\mathrm{CE}}(y_i, \hat{y}_i) + \lambda_{b(i)} \cdot r(p_i, \tilde{p}_i),$$

where $p_i = \sigma(\hat{y}_i)$ is the predicted probability from the binary cross-entropy loss ℓ_{CE} ($\hat{y}_i$ is the raw margin before the sigmoid), λ is the block-specific regularization weight that controls the strength of the spatial penalty relative to the cross-entropy, selected from the theoretically supported range $[0, 3.5]$, and r is the gated spatial penalty:

$$r(p_i, \tilde{p}_i) = \delta(p_i) \cdot (p_i - \tilde{p}_i)^2, \tag{3}$$

where $\delta(p_i) = 1 - 2|p_i - 0.5|$ is the adaptive spatial gating function. When p_i is near 0 or 1 (confident), $\delta \to 0$ and the regularization is weakest; when p_i is near 0.5, $\delta \to 1$ and the regularization is fully active. When the model is unsure about its judgment, spatial regularization assists the model by pulling its prediction toward the average predicted probability of its neighboring grid cells. Consequently, predictions where the model is already confident ($p_i \approx 0$ or $p_i \approx 1$) remain nearly unchanged by the spatial penalty, effectively preserving the correct classification decisions. Instead of a single global constant, the regularization weight is selected per administrative block (country or first-level province; small countries with sparse data are merged with their nearest-neighbor larger block): for each block b, the strength λ_b minimizes the block's held-out Brier score over the theoretically supported range:

$$\lambda_b = \arg \min_{\lambda \in [0,\, \lambda_{\max}]} \mathrm{Brier}_b(\lambda), \qquad \lambda_{\max} = 3.5, \tag{4}$$

where $\mathrm{Brier}_b(\lambda)$ is the Brier score of the model trained with strength λ evaluated on block b's held-out samples. In the main experiment, λ_b is selected on the training partition (train–validation) per administrative block, and the test partition is strictly held out and used only for final evaluation. In the spatial CV, the anchor models are retrained on the training folds only, and λ_b is selected per held-out $5^\circ \times 5^\circ$ block on that block's own held-out labels. In both settings the evaluated regions never contribute to model training; in the main experiment they also do not contribute to the selection of λ_b. Blocks with fewer than five held-out samples or a single class (blocks with only positive samples or only negative samples) retain the unregularized CE prediction, so the regularization is applied only where the data support it. Since every selected strength satisfies $\lambda_b \in [0, 3.5]$, the theoretical guarantees below apply unchanged with λ replaced by the block-selected λ_b. The ASR spatial regularization term is thus obtained from Eqs. (2)–(3).

$$Obj_{\mathrm{ASR}}^{(t)} = \sum_{i=1}^{n} \left[\ell_{\mathrm{CE}}(y_i, \hat{y}_i) + \lambda_{b(i)} \cdot \delta(p_i) \cdot (p_i - \tilde{p}_i)^2 \right] + \Omega(f_t),$$

where $b(i)$ denotes the spatial block containing sample i.

2.4 Theoretical properties

We establish two theoretical properties of the ASR loss: gradient boundedness and Hessian positivity, to ensure its compatibility with the XGBoost optimization framework. These properties are engineering-feasibility guarantees for the custom objective within XGBoost's Newton-step machinery; we state them formally because they justify the admissible range $[0, 3.5]$ used throughout the experiments.

Gradient boundedness of the ASR loss

Proposition 1 (Gradient boundedness) *For any sample $(y_i, \hat{y}_i, \tilde{p}_i)$ with $y_i \in \{0, 1\}$, $p_i = \sigma(\hat{y}_i) \in [0, 1]$, $\tilde{p}_i \in [0, 1]$, and block-selected weight $\lambda_{b(i)} \in [0, \lambda_{\max}]$, the per-sample gradient of the ASR loss with respect to the raw margin $\hat{y}_i$ satisfies*

$$\left| \frac{\partial \ell_{ASR,i}}{\partial \hat{y}_i} \right| \leq 1 + \frac{\lambda_{\max}}{2}.$$

Proof. At the t-th boosting iteration, the predicted probability p_i and the raw margin $\hat{y}_i$ satisfy:

$$\begin{cases} p_i = \sigma(\hat{y}_i) = \dfrac{1}{1 + e^{-\hat{y}_i}}, \\ \hat{y}_i = \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i). \end{cases}$$

The binary cross-entropy loss ℓ_{CE} is defined as:

$$\ell_{\text{CE}}(y_i, p_i) = -[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)].$$

Differentiating the above definitions gives the per-sample derivatives needed for the chain rule:

$$\begin{cases} \dfrac{\partial \ell_{\text{CE}}}{\partial p_i} = -\dfrac{y_i}{p_i} + \dfrac{1 - y_i}{1 - p_i} = \dfrac{p_i - y_i}{p_i(1 - p_i)}, \\ \dfrac{\partial r}{\partial p_i} = 2\delta(p_i)(p_i - \tilde{p}_i), \\ \dfrac{dp_i}{d\hat{y}_i} = \dfrac{e^{-\hat{y}_i}}{(1 + e^{-\hat{y}_i})^2} = \dfrac{1}{1 + e^{-\hat{y}_i}} \cdot \dfrac{e^{-\hat{y}_i}}{1 + e^{-\hat{y}_i}} = p_i(1 - p_i), \\ \dfrac{\partial \hat{y}_i}{\partial f_t} = 1. \end{cases} \tag{5}$$

Substituting these into the gradient with respect to the margin yields the per-sample gradient of the cross-entropy loss:

$$\nabla \ell_{\text{CE},i} = \frac{\partial \ell_{\text{CE}}}{\partial f_t} = \frac{\partial \ell_{\text{CE}}}{\partial p_i} \cdot \frac{dp_i}{d\hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial f_t} = \frac{p_i - y_i}{p_i(1 - p_i)} \cdot p_i(1 - p_i) = p_i - y_i. \tag{6}$$

For the per-sample spatial penalty r, the adaptive spatial gating function δ from Eq. (3) is treated as a constant during differentiation:

$$\nabla r_i = \frac{\partial r}{\partial f_t} = \frac{\partial r}{\partial p_i} \cdot \frac{dp_i}{d\hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial f_t} = 2\delta(p_i) \cdot (p_i - \tilde{p}_i) \cdot p_i(1 - p_i), \tag{7}$$

The per-sample gradients in Eqs. (6) and (7) give $\nabla \ell_{\mathrm{CE},i}$ and ∇r_i. The gating function $\delta(p_i)$ depends only on p_i and is treated as a local constant during differentiation. Because $\delta(p_i)$ changes as the predictions change during training, it is re-computed only at the start of each boosting round and then treated as a fixed weight while that round's tree is built. The block-selected weight $\lambda_{b(i)}$ is a constant determined before training (Eq. (4)) and is likewise frozen within each round, so the surrogate objective remains a fixed quadratic in f_t with constant per-sample weights. XGBoost therefore optimizes a surrogate objective in which δ is frozen within each round. This frozen-gating treatment ensures that the objective supplied to XGBoost remains continuous and differentiable within each boosting iteration, so that the per-sample gradient and Hessian analyzed in Propositions 1 and 2 are well-defined at every step. Note that δ itself is not differentiable at $p_i = 0.5$; under this convention it is assigned the value $\delta = 1$ there, and no derivative of δ enters the update. Retaining the full derivative of δ would add a term of magnitude at most λ to the gradient bound and would require a case-by-case Hessian analysis; this is not the implemented scheme. Under this surrogate formulation, each boosting round minimizes a fixed convex quadratic approximation whose per-sample Hessian is positive for $\lambda \leq 3.5$ (Proposition 2).

The full ASR objective then expands to:

$$\nabla \ell_{\mathrm{ASR},i} = \underbrace{(p_i - y_i)}_{\nabla \ell_{\mathrm{CE}}} + \lambda_{b(i)} \cdot \underbrace{2\,\delta(p_i)\,(p_i - \tilde{p}_i)\,p_i(1 - p_i)}_{\nabla r_i}. \tag{8}$$

We show that this gradient remains bounded throughout training.

Since $p_i \in [0, 1]$ and $y_i \in \{0, 1\}$, we have $|\nabla \ell_{\mathrm{CE}}| = |p_i - y_i| \leq 1$.

Regularization term. The adaptive spatial gating function satisfies $0 \leq \delta(p_i) \leq 1$, with extremes at $p_i = 0.5$ ($\delta = 1$) and $p_i = 0, 1$ ($\delta = 0$). The soft-label residual satisfies $|p_i - \tilde{p}_i| \leq 1$ because both probabilities lie in $[0, 1]$. The variance term $p_i(1 - p_i) \leq \frac{1}{4}$, with equality at $p_i = 0.5$. Combining these bounds:

$$|\nabla r_i| = 2\,\delta(p_i)\,|p_i - \tilde{p}_i|\,p_i(1 - p_i) \leq 2 \times 1 \times 1 \times \frac{1}{4} = \frac{1}{2}.$$

By the triangle inequality:

$$|\nabla \ell_{\mathrm{ASR},i}| \leq |\nabla \ell_{\mathrm{CE}}| + \lambda_{b(i)}\,|\nabla r_i| \leq 1 + \frac{\lambda_{b(i)}}{2} \leq 1 + \frac{\lambda_{\max}}{2}.$$

Since $\lambda_{b(i)} \leq \lambda_{\max}$ holds for every block, the bound is uniform across samples and a constant multiple of the cross-entropy gradient bound, confirming that ASR does not introduce gradient explosion. The modest increase in gradient magnitude is concentrated precisely in those uncertain samples (p_i near 0.5) that benefit most from spatial regularization. □

Curvature and Newton-step validity

The preceding analysis establishes that the ASR gradient remains bounded. We now examine the second-order curvature, which governs the Newton-step update and the tree-splitting criterion in XGBoost. At each boosting iteration, XGBoost approximates the loss by a second-order Taylor expansion around the current prediction $\hat{y}_i^{(t-1)}$:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \Big[\ell(\hat{y}_i^{(t-1)}, y_i) + g_i\, f_t(\mathbf{x}_i) + \frac{1}{2} h_i\, f_t(\mathbf{x}_i)^2 \Big],$$

where $g_i = \partial \ell / \partial \hat{y}$ and $h_i = \partial^2 \ell / \partial \hat{y}^2$ are the per-sample gradient and Hessian of the loss with respect to the prediction. The quadratic approximation decomposes into n independent one-dimensional quadratics $\frac{1}{2} h_i\, t^2 + g_i\, t$ in $f_t(\mathbf{x}_i)$. To guarantee a valid Newton step throughout tree construction, every training sample must satisfy positive curvature $h_i > 0$. Even if the sum H_j of second derivatives in a current leaf is positive, any individual $h_i \leq 0$ creates a critical risk: splitting this leaf can yield child nodes with $H_L \leq 0$ or $H_R \leq 0$, where the quadratic surrogate loses a finite minimum and the analytical update rule collapses.

XGBoost uses a diagonal approximation of the Hessian that treats each sample's contribution independently, giving a scalar h_i per sample. We therefore examine the per-sample Hessian of the ASR loss with respect to the raw margin $\hat{y}_i$, which follows from differentiating the gradient expression (Eq. (8)). For the cross-entropy component, $h_{\mathrm{CE},i} = p_i(1-p_i) \geq 0$.

By Eq. (5) and Eq. (7):

$$\begin{aligned}
h_{r,i} &= \frac{\partial}{\partial \hat{y}} [2\delta(p_i) \cdot (p_i - \tilde{p}_i) \cdot p_i(1-p_i)] \\
&= 2\delta(p_i) \cdot \Big[\frac{\partial}{\partial \hat{y}} (p_i - \tilde{p}_i) \cdot p_i(1-p_i) + (p_i - \tilde{p}_i) \cdot \frac{\partial}{\partial \hat{y}} (p_i(1-p_i)) \Big] \\
&= 2\delta(p_i) \cdot [p_i(1-p_i) \cdot p_i(1-p_i) + (p_i - \tilde{p}_i) \cdot (1-2p_i) \cdot p_i(1-p_i)] \\
&= 2\delta(p_i) \cdot p_i(1-p_i) \cdot [p_i(1-p_i) + (p_i - \tilde{p}_i)(1-2p_i)].
\end{aligned}$$

The total per-sample Hessian is therefore:

$$\begin{aligned}
h_{\mathrm{ASR},i} &= p_i(1-p_i) + \lambda_{b(i)} \cdot 2\delta(p_i) \cdot p_i(1-p_i) \cdot [p_i(1-p_i) + (p_i - \tilde{p}_i)(1-2p_i)] \\
&= p_i(1-p_i) \big\{ 1 + 2\lambda_{b(i)} \delta(p_i) [p_i(1-p_i) + (p_i - \tilde{p}_i)(1-2p_i)] \big\}. \qquad (9)
\end{aligned}$$

The cross-entropy component $h_{\mathrm{CE},i} = p_i(1-p_i) \geq 0$ is always non-negative. The regularization component $h_{r,i}$ can be positive or negative. To keep the per-sample Hessian of the surrogate objective strictly positive, we restrict the block-selected weight to the theoretically admissible range $\lambda_{b(i)} \in [0, 3.5]$, whose upper bound $\lambda_{\max} = 3.5$ is the largest strength for which $h_{\mathrm{ASR},i} > 0$ holds for all $p_i \in (0,1)$ (Proposition 2).

Proposition 2 (Hessian positivity) *For any $p_i, \tilde{p}_i \in [0,1]$ and any block-selected regularization weights $\lambda_{b(i)} \in [0, 3.5]$, the per-sample Hessian $h_{ASR,i} = \partial^2 \ell_{ASR,i} / \partial \hat{y}_i^2$ of the ASR surrogate loss (with frozen gating) satisfies $h_{ASR,i} > 0$ for all $p_i \in (0,1)$.*

Proof. Factor the Hessian from Eq. (9):

$$h_{\mathrm{ASR},i} = p_i(1-p_i)[1 + 2\lambda\,\delta(p_i)\,F(p_i,\tilde{p}_i)], \qquad F(p,\tilde{p}) = p(1-p) + (p-\tilde{p})(1-2p),$$

where $\delta(p_i) = 1 - 2|p_i - 0.5|$ is the adaptive spatial gating function from Eq. (3). Since $p_i(1-p_i) > 0$ for $p_i \in (0,1)$, it suffices to show $1 + 2\lambda\delta F > 0$.

The function F is linear in $\tilde{p}$ with coefficient $-(1-2p)$. For each fixed p, the minimum over $\tilde{p} \in [0,1]$ is therefore attained at an endpoint:

$$F_{\min}(p) = \min_{\tilde{p}\in[0,1]} F(p,\tilde{p}) = \begin{cases} (1-p)(3p-1), & p < \frac{1}{2}\ (\tilde{p}=1), \\ \frac{1}{4}, & p = \frac{1}{2}\ (\tilde{p}\in[0,1]), \\ p(2-3p), & p > \frac{1}{2}\ (\tilde{p}=0). \end{cases}$$

Now consider $p < \frac{1}{2}$, where $\delta = 2p$ and $F_{\min} = (1-p)(3p-1)$. The product

$$\delta F = 2p(1-p)(3p-1) =: g(p), \qquad p \in [0, \tfrac{1}{2}],$$

is a cubic polynomial. Differentiating,

$$g'(p) = -18p^2 + 16p - 2 = 0 \quad \Longrightarrow \quad p^* = \frac{4-\sqrt{7}}{9},$$

which lies in $(0, \frac{1}{2})$. The minimum value

$$g(p^*) = 2p^*(1-p^*)(3p^*-1) = \frac{40 - 28\sqrt{7}}{243}.$$

The case $p > \frac{1}{2}$ is symmetric ($p \mapsto 1-p$, $\delta = 2(1-p)$, $F_{\min} = p(2-3p)$) and yields the identical bound. Hence $\delta F \geq \frac{4(10-7\sqrt{7})}{243}$ for all admissible $(p, \tilde{p})$.

Consequently,

$$1 + 2\lambda\delta F \geq 1 + 2\lambda \cdot \frac{4(10-7\sqrt{7})}{243} = 1 - \frac{8(7\sqrt{7}-10)}{243}\lambda \geq 0, \qquad \lambda \leq \frac{243}{8(7\sqrt{7}-10)} < 3.57,$$

and therefore $h_{\mathrm{ASR},i} = p_i(1-p_i)\,(1 + 2\lambda\delta F) > 0$ for all $p_i \in (0,1)$, $\tilde{p}_i \in [0,1]$, and $\lambda \in [0, 3.5]$. The Hessian is strictly positive without case splitting or additional constraints. □

XGBoost's Newton step operates on *leaf-level* sums rather than individual samples. The per-leaf Hessian sum is

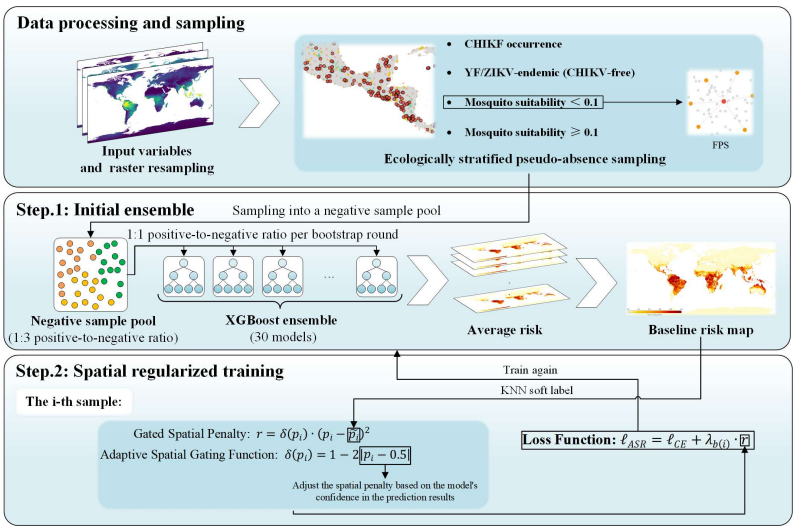


Fig. 1 Construction of the global CHIKF risk distribution prediction model

$$H_j + \lambda_{\text{reg}} = \sum_{i \in I_j} h_{\text{ASR},i} + \lambda_{\text{reg}},$$

where $\lambda_{\text{reg}} > 0$ is XGBoost's native regularization hyperparameter. Since every individual per-sample Hessian is positive ($h_{\text{ASR},i} > 0$), the leaf sum $H_j + \lambda_{\text{reg}}$ is strictly positive at every leaf. This ensures:

1. The leaf weight $w_j^* = -G_j/(H_j + \lambda_{\text{reg}})$ is always well-defined and corresponds to a minimizer of the quadratic approximation.
2. The split gain is guaranteed non-negative in the absence of the pruning term γ, because the Cauchy–Schwarz argument (Section 2.3) holds for $H_L, H_R > 0$.
3. The Newton direction remains a descent direction.

In summary, the ASR loss produces strictly positive per-sample curvature under our parameterisation. Consequently, the standard XGBoost Newton-step machinery applies without modification or numerical safeguards.

2.5 Global CHIKF risk distribution prediction model

The overall framework is structured into two stages, preceded by data preprocessing and negative sample generation (Fig. 1). In the data preparation stage, all raster covariates were resampled to a unified 2.5 arcminute resolution. Among the 19 WorldClim bioclimatic variables, five (isothermality, temperature seasonality, minimum temperature of the coldest month, temperature annual range, and annual precipitation) were retained according to Gain-based variable importance from a pilot XGBoost model to reduce overfitting risk. The remaining five covariates comprised healthcare access, population density, nighttime lights, and the habitat suitability indices

of Aedes aegypti and Aedes albopictus. Ecologically stratified pseudo-absence samples were drawn from the three-source negative pool (Section 2.2) at a global 1:3 positive-to-negative ratio, rebalanced to 1:1 within each bootstrap round.

In Stage 1 (initial ensemble), an XGBoost ensemble of 30 models is trained using the cross-entropy objective. For each bootstrap round, the negative pool is subsampled to a 1:1 positive-to-negative balance before training each ensemble member. The averaged predictions across the 30 models form the Baseline Risk Map, which supplies the soft label priors for the next stage.

In Stage 2 (spatial regularized training), pixel-wise 8-neighborhood mean soft labels $\tilde{p}_i$ are computed from the Baseline Risk Map using the 8-neighborhood mean of each grid cell. An adaptive spatial gating function $\delta(p_i) = 1 - 2|p_i - 0.5|$ dynamically scales the spatial penalty $r = \delta(p_i) \cdot (p_i - \tilde{p}_i)^2$, concentrating regularization on ambiguous predictions where the model is uncertain. A new XGBoost ensemble is retrained with the combined loss $\ell_{\text{ASR}} = \ell_{\text{CE}} + \lambda \cdot r$, which embeds spatial continuity directly into the training objective. The final risk prediction at each grid cell $\mathbf{x}$ is obtained by averaging across all 30 ASR-trained models.

We additionally implemented three neural network baselines to benchmark against tree-based models under identical data conditions, namely a multilayer perceptron (MLP), a TabResNet He et al. (2016), and an FT-Transformer Gorishniy et al. (2021). The MLP consisted of three hidden layers (64, 32, 16 neurons) with ReLU activation, dropout (0.2) after each hidden layer except the last, weight decay 10^{-5}, and was trained using AdamW with binary cross-entropy loss and early stopping (patience 30). The TabResNet followed the rtdl implementation with $d = 256$ and three residual blocks. The FT-Transformer used the rtdl implementation with $d_{\text{token}} = 192$, three transformer blocks, eight attention heads, and a single output head. Both TabResNet and FT-Transformer were trained with AdamW (learning rate 10^{-3}, weight decay 10^{-5}) and identical early stopping. All models applied z-score standardization to the input features. For direct comparability with the tree-based ensemble evaluation, each NN was trained on the same 7:2:1 data splits with 1:1 class balancing and Platt calibration.

3 Results

3.1 Spatial smoothness metrics

To quantify the spatial smoothness of a predicted risk surface p, we use three indices.

(i) **Moran's I** measures whether similar risk values cluster in space. A value near 1 means that high-risk cells tend to be surrounded by high-risk cells and low-risk cells by low-risk cells; a value near 0 means the risk pattern is spatially random. This index follows the spatial autocorrelation coefficient introduced by Moran Moran (1950).

The computation compares each cell with its four directly adjacent neighbors (up, down, left, right). For any cell k, let $z_k = p_k - \bar{p}$ denote how far its risk value deviates from the global mean, so z_i and z_j in the sums below are the deviations of the two cells in each neighbor pair; let $w_{ij} = 1$ if cells i and j are horizontally or vertically

adjacent (and 0 for diagonal neighbors). Moran's I is then

$$I = \frac{n}{\sum_{i,j} w_{ij}} \cdot \frac{\sum_{i,j} w_{ij} z_i z_j}{\sum_i z_i^2},$$

which is the sum of the products $z_i z_j$ over all directly adjacent pairs, divided by the total squared deviation and rescaled by the average number of neighbors per cell.

(ii) **Continuous edge density** (ContED) measures how abruptly the risk surface changes between neighboring cells. It is the mean absolute difference between horizontally or vertically adjacent cells,

$$\text{ContED} = \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} |p_i - p_j|,$$

where $\mathcal{E}$ is the set of valid adjacent pairs; lower values indicate a smoother surface. It can be viewed as the global average of the local edge magnitudes (gradient magnitudes between adjacent samples) underlying edge-detection methods in landscape ecology Fortin (1994).

(iii) **Isolated-cell ratio** (Iso ratio) measures surface fragmentation. It is the proportion of cells whose prediction differs from the mean of their eight neighbors by more than 0.05,

$$\text{Iso} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{|p_i - \bar{p}_{\mathcal{N}_8(i)}| > 0.05\},$$

where $\bar{p}_{\mathcal{N}_8(i)}$ is the mean risk of the eight cells surrounding cell i (horizontally, vertically, and diagonally adjacent). Isolated cells are the discrete analogue of the spurious oscillations that total-variation denoising removes while preserving sharp signals Rudin et al. (1992); the ratio therefore quantifies residual fragmentation of the risk surface.

3.2 ASR-XGBoost performance

Model performance was evaluated by a comprehensive comparison of all ensemble configurations (Table 4). All models used a 30-ensemble with 1,000 trees per model on identical 70/20/10 stratified data splits. Standard test-set discrimination and calibration metrics (AUC, Brier score, Recall) are reported alongside spatial smoothness indices (Moran's I, continuous edge density, isolated-cell ratio). To assess spatial generalization, we additionally report spatial block 10-fold CV metrics (AUC(CV), Brier(CV)): following the spatial-blocking CV recommendations of Roberts et al. (2017) and Ploton et al. (2020), under which random splits would underestimate predictive error for spatially autocorrelated data, the globe is partitioned into $5° \times 5°$ blocks, blocks are assigned to ten folds without splitting (block-stratified, $\geq$10 samples and both classes per block), and models are retrained per fold with the test blocks excluded from training. The adaptive variant (ASR) in this CV selects λ per held-out $5° \times 5°$ block on that block's own held-out labels. The outer evaluation used $5° \times 5°$ spatial blocks, whereas the adaptive weights for the final risk map were selected per administrative block on the training partition only (Section 2.3).

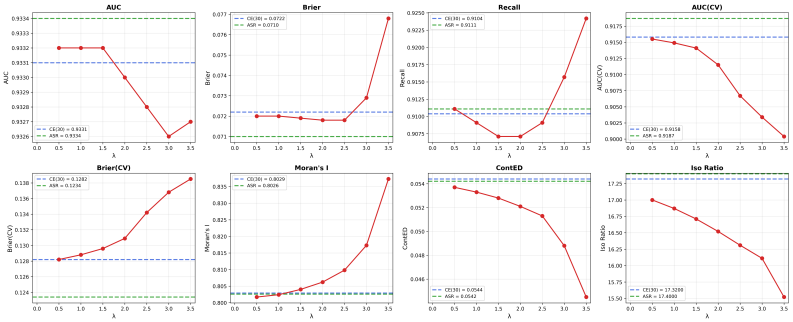


Fig. 2 Stability of SR-XGBoost spatial regularization across the theoretically supported range. Each panel tracks one metric as the regularization weight λ increases from 0.5 to 3.5 (red curve); the dashed blue and green horizontal lines mark the CE baseline and the adaptive ASR reference, respectively

The ASR model preserves and slightly improves the spatial generalization of the CE baseline (AUC(CV) 0.9187 versus 0.9158, with ASR exceeding CE in 10 of 10 folds; Table 2), indicating that locally adaptive regularization strength does not incur an extrapolation cost. Calibration improvement is established on the strictly held-out test set (Brier 0.0710 versus 0.0722).

Table 2 Per-fold spatial block CV metrics ($5° \times 5°$ blocks) for CE and ASR. ASR exceeded CE on AUC in all ten folds

Fold	AUC		Brier	
	CE(CV)	ASR(CV)	CE(CV)	ASR(CV)
1	0.9027	**0.9066**	0.1220	**0.1142**
2	0.8565	**0.8605**	0.2230	**0.2201**
3	0.9245	**0.9304**	0.1076	**0.1021**
4	0.9365	**0.9404**	0.1414	**0.1325**
5	0.9115	**0.9122**	0.1395	**0.1361**
6	0.9394	**0.9449**	0.1031	**0.0964**
7	0.9286	**0.9296**	0.0753	**0.0728**
8	0.8967	**0.8971**	0.1422	**0.1402**
9	0.9216	**0.9237**	0.1269	**0.1241**
10	0.9394	**0.9413**	0.1010	**0.0956**
Mean	0.9158	**0.9187**	0.1282	**0.1234**

Table 3 examines the stability of SR-XGBoost (the fixed-strength variant) across the theoretically supported regularization range. As the regularization weight λ increases from 0.5 to 3.5, test-set AUC remains essentially unchanged (0.9326–0.9332) and the Brier score is only mildly elevated at $\lambda = 3.5$ (0.0768 versus 0.0718–0.0729 for $\lambda \leq 3.0$). Spatial smoothness improves monotonically (Moran's I from 0.8017 to 0.8373, isolated-cell ratio from 17.00% to 15.52%). Under the spatial block CV, AUC(CV) declines modestly with stronger regularization (0.9155 at $\lambda = 0.5$ to 0.9004 at $\lambda = 3.5$), while Brier(CV) rises from 0.1282 to 0.1385, reflecting the expected smoothness–fidelity trade-off under strict spatial extrapolation. Proposition 2 guarantees positive per-sample curvature of the frozen-gating surrogate within the specified

Table 3 Stability of SR-XGBoost across the theoretically supported regularization range ($\lambda \in [0, 3.5]$)

λ	AUC	Brier	Recall	AUC(CV)	Brier(CV)	Moran's I	ContED	Iso. ratio
0(CE)	0.9331	0.0722	0.9104	**0.9158**	**0.1282**	0.8029	0.0544	17.32%
0.5	**0.9332**	0.0720	0.9111	0.9155	**0.1282**	0.8017	0.0537	17.00%
1.0	**0.9332**	0.0720	0.9091	0.9149	0.1288	0.8024	0.0533	16.87%
1.5	**0.9332**	0.0719	0.9071	0.9141	0.1296	0.8040	0.0528	16.71%
2.0	0.9330	**0.0718**	0.9071	0.9115	0.1309	0.8062	0.0521	16.52%
2.5	0.9328	**0.0718**	0.9091	0.9067	0.1342	0.8098	0.0513	16.31%
3.0	0.9326	0.0729	0.9157	0.9034	0.1368	0.8173	0.0488	16.11%
3.5	0.9327	0.0768	**0.9242**	0.9004	0.1385	**0.8373**	**0.0445**	**15.52%**

Table 4 Comparison of baseline model configurations. MLP, TabResNet, and FT-Transformer are neural network baselines; RF and BRT are tree-based 30-ensembles; XGBoost is the standard cross-entropy 30-ensemble; ASR-XGBoost is the proposed model with per-block adaptive regularization. AUC(CV) and Brier(CV) are the spatial block 10-fold CV AUC and Brier score, respectively. All models share identical training data and feature set

Metric	MLP	TabResNet	FT-Transformer	RF	BRT	XGBoost	ASR-XGBoost
AUC	0.8006	0.7974	0.8416	0.9041	0.9287	0.9331	**0.9334**
Brier	0.1438	0.1446	0.1403	0.1072	0.0801	0.0722	**0.0710**
Recall	0.7731	0.7675	0.7804	0.8847	0.8933	0.9104	**0.9111**
AUC(CV)	0.7092	0.7042	0.6969	0.8852	0.9144	0.9158	**0.9187**
Brier(CV)	0.1785	0.1806	0.1809	0.1424	**0.1234**	0.1282	**0.1234**

range, supporting the numerical validity of the XGBoost second-order update; predictive stability is evaluated empirically: regardless of the strength chosen locally, predictive accuracy remains within a narrow, reliable band (Table 3), while fixed-strength SR delivers its intended spatial smoothing with increasing λ. The adaptive ASR reference in Figure 2 matches or exceeds every fixed-strength curve on the discrimination metrics (test AUC 0.9334 and Brier 0.0710 both better than the best fixed-strength values of 0.9332 and 0.0718, and spatial block CV AUC 0.9187 and Brier 0.1234 likewise optimal), while its spatial smoothness lies between the CE baseline and strongly regularized SR, reflecting the per-block trade-off between discrimination and smoothing.

Table 4 compares all model configurations, including neural network baselines (MLP, TabResNet, FT-Transformer) and tree-based models (random forest (RF), boosted regression trees (BRT), XGBoost, ASR-XGBoost). The three neural network baselines underperform their tree-based counterparts on predictive discrimination, consistent with prior work reporting that tree-based models remain state-of-the-art on tabular datasets of medium size (∼10K samples) Grinsztajn et al. (2022). Among the proposed models, SR-XGBoost with fixed strength (λ=3.0) combined improved spatial smoothing (ContED 0.0488, Iso ratio 16.11%) with competitive discrimination (AUC 0.9326, Brier 0.0729) and Recall (0.9157). The ASR variant attains the best CV discrimination among all models (AUC(CV) 0.9187 versus 0.9158 for XGBoost) and the lowest CV Brier (0.1234, tied with BRT), while the ASRAC post-processing further improves spatial coherence (Iso ratio 6.12%) at comparable discrimination (AUC 0.9331).

To confirm that this discrimination does not rely on the suitability threshold used to define one negative stratum, we re-evaluated the model on nested subsets

of the test set that progressively remove cells defined by low mosquito suitability (Table 5). Removing the low-suitability negatives (488 cells) left the AUC high at 0.9238. Restricting the evaluation entirely to cells with mosquito suitability ≥ 0.1 gave an AUC of 0.9356, and the strictest comparison, between suitable cells with reported cases and suitable cells without reports, in which no suitability threshold can separate the two groups, yielded an AUC of 0.9474. The CE 30-ensemble behaved similarly across these subsets (0.9234, 0.9354 and 0.9473). Discrimination is therefore not carried by the suitability-defined negatives. Within the suitable range the model separates case-free from case-positive cells, where the pseudo-absence labels come from the absence of reports rather than from vector unsuitability.

Table 5 Discrimination on nested test subsets that progressively remove the low-suitability stratum. Row 1, full test set; row 2, low-suitability negatives (suitability < 0.1) removed; row 3, cells with suitability ≥ 0.1 only, also excluding 19 positives below the threshold; row 4, reported versus unreported cells within the suitable range

Subset	n	**Positive**	**AUC(CE)**	**AUC(ASR)**
Full test set	2,169	542	0.9331	0.9334
Excluding low-suitability negatives	1,681	542	0.9234	0.9238
Suitability ≥ 0.1 only	1,662	523	0.9354	0.9356
Reported vs. unreported (suitability ≥ 0.1)	1,155	523	0.9473	0.9474

We further examine the spatial effect of the direct regularization by comparing the ASR-XGBoost (block-wise) risk surface directly against the CE baseline. Figure 3 maps the probability difference $p_{\mathrm{ASR}} - p_{\mathrm{CE}}$ across all land grid cells (land without predicted risk in medium gray, ocean in white), with deep red indicating regions where ASR increases risk relative to CE and deep blue where it decreases risk. Near-zero differences ($|\Delta p| \leq 0.003$) are rendered in light gray to suppress visual noise. Red regions (ASR $>$ CE) are predominantly concentrated near areas with known CHIKV transmission, while blue regions (ASR $<$ CE) are concentrated in currently disease-free areas. ASR regularization applies the strongest adjustments in uncertain regions (near $p = 0.5$) where the adaptive gating function is most active: grid cells with intermediate predictions surrounded by higher-risk neighbors are adjusted upward, while those in lower-risk neighborhoods are adjusted downward. The adjustments are modest in magnitude for most cells, with near-zero differences ($|\Delta p| \leq 0.003$) dominating the map; the spatial pattern shows that ASR predominantly boosts endemic fringe areas where the CE baseline may have underestimated risk due to sparse training records. Figure 4 shows the distribution of differences centered near zero, with annotated percentiles (P5, P50, P95) and the sample mean.

The resulting global CHIKF risk distribution under current climate is shown in Figure 5. The risk surface produced by ASR-XGBoost (the block-wise variant) identifies spatially contiguous high-risk zones concentrated in several major regions: (i) sub-Saharan Africa, covering Nigeria, central Democratic Republic of the Congo, Cameroon, and coastal Ghana; (ii) northern South America, including the Amazon Basin of Brazil, Colombia, and Venezuela; (iii) south and southeast Asia, spanning India, Bangladesh, Thailand, southern Vietnam, Indonesia, and the Philippines; (iv)

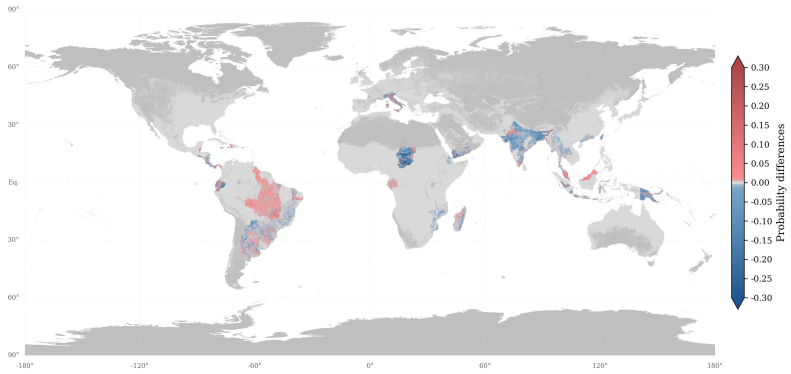


Fig. 3 Spatial distribution of the probability difference between ASR-XGBoost (block-wise) and CE-XGBoost 30-ensemble. Deep red (positive) indicates regions where ASR predicts higher risk than CE; deep blue (negative) where ASR predicts lower risk. Near-zero differences are shown in neutral gray to reduce noise. Base map distinguishes land (gray) from ocean (white); grid lines at 60°longitude and 30°latitude intervals

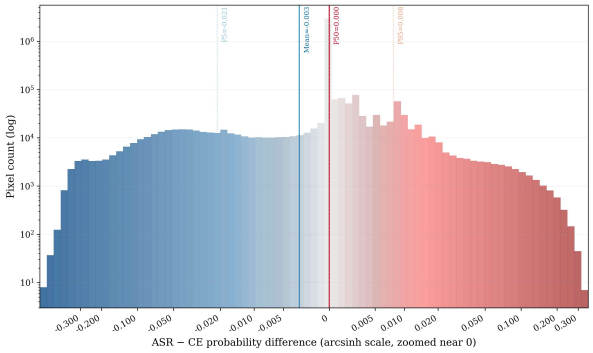


Fig. 4 Distribution of probability differences $p_{\mathrm{ASR}} - p_{\mathrm{CE}}$ across all land grid cells. Dashed vertical line marks zero; coloured bars follow the diverging map colour scheme. Annotations show the sample mean, median (P50), and selected percentiles (P5, P95)

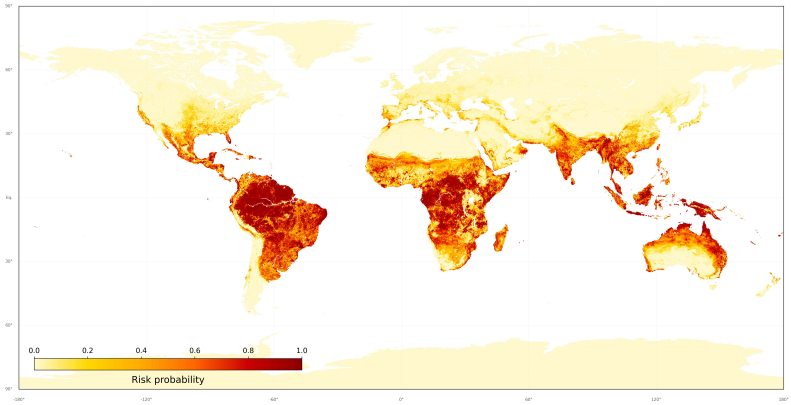


Fig. 5 Global CHIKF risk distribution under current climate, with the regularization strength selected within each administrative block

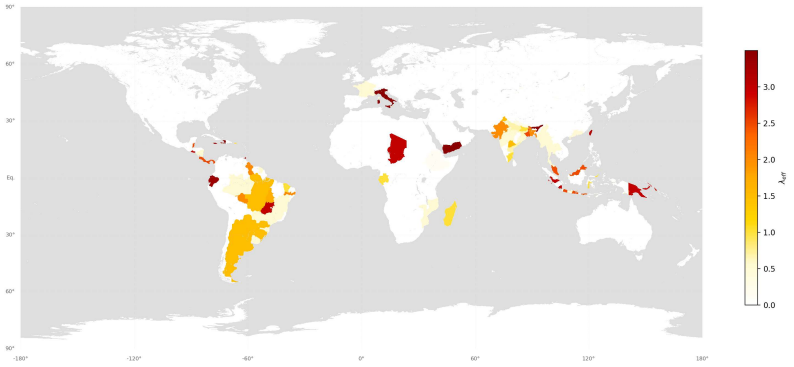


Fig. 6 Regularization strength selected independently within each administrative block. Blocks are coloured by the selected λ^*, blocks where no regularization was selected remain pale, and ocean is shown in grey

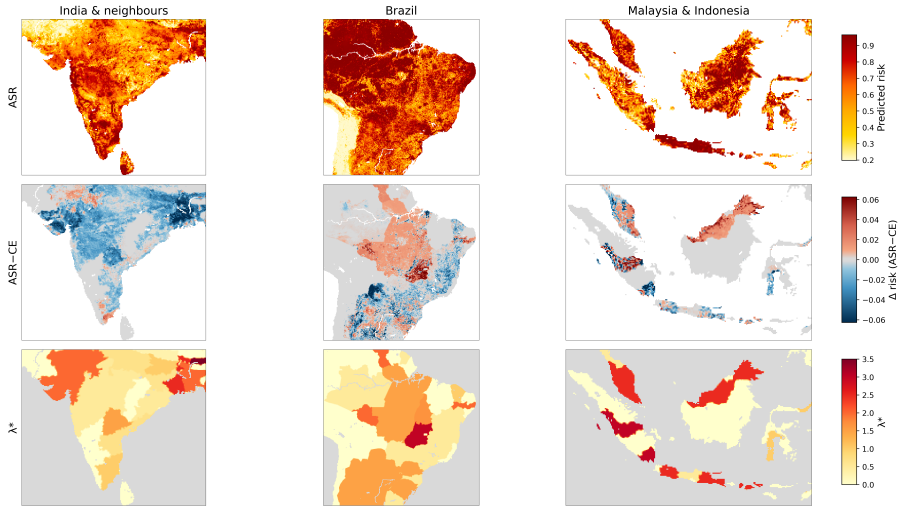


Fig. 7 Regional detail for India and neighbours, Brazil, and Malaysia and Indonesia. Rows show the ASR risk surface, its difference from the CE baseline ($p_{\mathrm{ASR}} - p_{\mathrm{CE}}$), and the regularization strength selected for each administrative block

southern Mexico and Central America; and (v) northern tropical coastal areas of Australia, among others. These regions are characterized by climatic conditions optimal for the survival and reproduction of Aedes mosquito vectors.

Figure 6 maps the regularization strength selected independently within each administrative block. Of the 104 blocks with sufficient training records, the selected values span the admissible range up to 3.5 with a median of 0.50, and 34% of the blocks select $\lambda^* \geq 1.5$, whereas the best single global value is 0.24. The selection is spatially structured rather than uniform, with the strongest regularization in South-east Asian blocks (median 2.50 for Indonesia) and weaker values in Brazilian blocks (median 0.50).

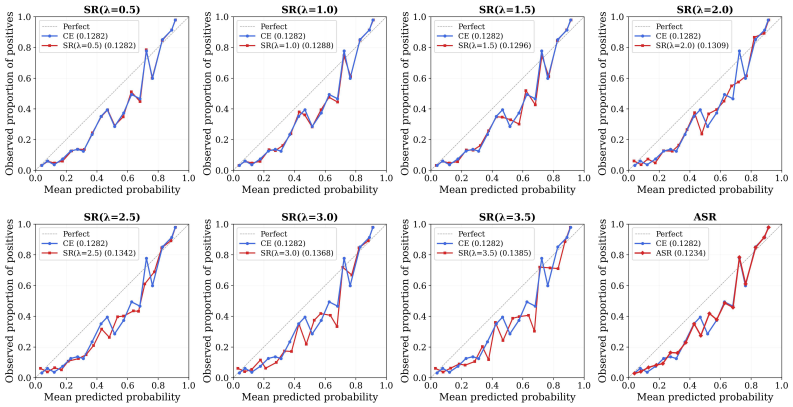


Fig. 8 Calibration curves for CE (blue baseline) and SR-XGBoost at individual regularization strengths (red), with the adaptive ASR model shown in the final (eighth) panel; the dashed diagonal represents perfect calibration

Figure 7 examines three representative regions in detail. The ASR surface retains the spatial pattern of the CE baseline, the difference map shows where the regularizer acts, and the bottom row shows the block-level λ^* that produced it.

3.3 Calibration analysis

To evaluate how the ASR spatial regularization affects probability calibration, we construct calibration curves for CE and SR-XGBoost at all λ values (Figure 8) using the standard reliability-diagram construction: test predictions are sorted into 20 equal-width probability bins, and within each bin the mean predicted probability is compared with the observed positive frequency, with the CE baseline shown in blue and the model curves in red. A well-calibrated model should lie along the diagonal where predicted probabilities match observed frequencies.

Across the regularization spectrum, the calibration curves remain close to the diagonal and to the CE baseline over the full probability range, indicating that the ASR spatial regularization preserves probability calibration across the entire range of regularization strengths.

3.4 Variable importance analysis

To identify the key environmental and socioeconomic drivers of CHIKF risk in our model, we computed Gain-based variable importance from the CE 30-ensemble (Figure 9). The ten features used in this study were drawn from an initial candidate pool of 24 variables (19 WorldClim bioclimatic variables plus five socioeconomic and mosquito suitability covariates). To mitigate overfitting given the sparse CHIKF occurrence record set, the five retained bioclimatic variables were those ranked highest by Gain-based importance in a pilot XGBoost model; the remaining five covariates (medical access, population, nightlights, and the two Aedes suitability indices) were retained directly due to their established ecological relevance.

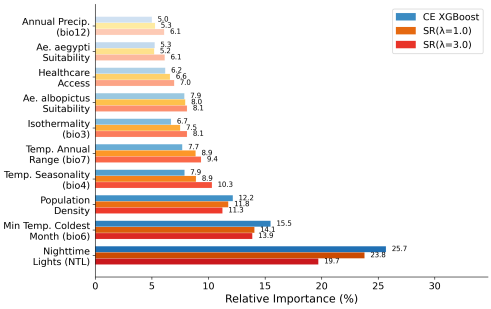


Fig. 9 Gain-based variable importance for the CE 30-ensemble. Blue bars represent the CE XGBoost model and orange/red bars represent the SR-XGBoost models at $\lambda = 1.0, 3.0$, illustrating how the regularization strength rebalances feature contributions

Nighttime lights consistently emerged as the dominant predictor in both the CE and SR-XGBoost models. The strength of nighttime lights reflects its role not only as a proxy for infrastructure and surveillance capacity, but also as a factor that can influence mosquito vector behavior to some extent Coetzee et al. (2023); Fiorta et al. (2024). Therefore, the high importance of nighttime lights is reasonable. However, areas with weak nighttime lights tend to have limited surveillance capacity, which may lead to systematic underreporting of CHIKF cases and cause the model to over-rely on nighttime lights as a proxy for detection bias. Across the SR models, the importance of nighttime lights is lower at $\lambda = 3.0$ than at $\lambda = 1.0$ (Figure 9), suggesting that the spatial regularization mitigates some of the bias introduced by nighttime lights.

The relatively moderate importance of the two mosquito suitability variables can be attributed to two factors. First, the suitable ranges of Ae. aegypti and Ae. albopictus overlap considerably across tropical and subtropical regions, so their shared predictive signal is distributed across both variables rather than concentrated in a single feature. Nevertheless, both vectors are competent for CHIKV transmission and their spatial distributions are not identical: Ae. albopictus extends into cooler temperate zones, so retaining both is epidemiologically necessary. Second, mosquito habitat suitability is strongly mediated by temperature and precipitation, resulting in partial correlation with the retained bioclimatic variables. Under Gain-based importance, this shared information is distributed across related covariates rather than captured solely by the suitability layers. From a disease transmission perspective, climate influences not only vector population dynamics but also viral replication and extrinsic incubation periods Caldwell et al. (2021), making the inclusion of both bioclimatic and suitability variables ecologically necessary.

3.5 Anomaly correction post-processing

An anomaly-correction post-processing is applied to the baseline CE ensemble: the fixed-strength variant is corrected by SR-based anomaly correction (SRAC), and the adaptive variant by ASRAC. The ASRAC step reuses the trained ASR model as a spatially-aware corrector for the baseline CE ensemble. Unlike purely spatial smoothing approaches that impose smoothness without ecological constraints, the

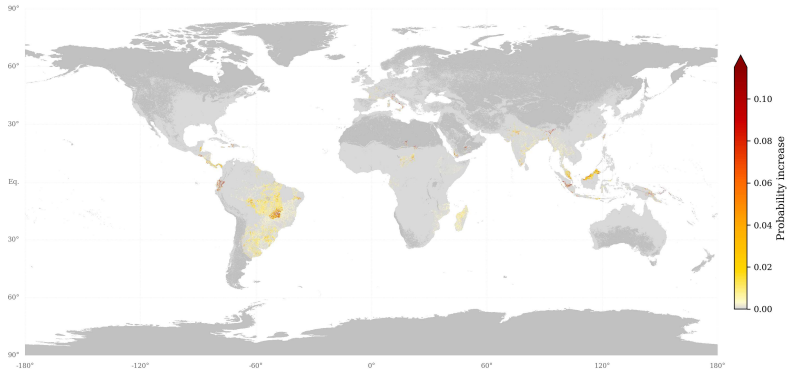


Fig. 10 Spatial distribution of probability increases from CE to ASRAC: warmer colours (yellow to dark red) indicate larger increases, while grid cells with predicted risk but negligible adjustment ($\Delta p \leq 0.003$) are shown in light gray and cells without predicted risk in medium gray; ocean is shown in white

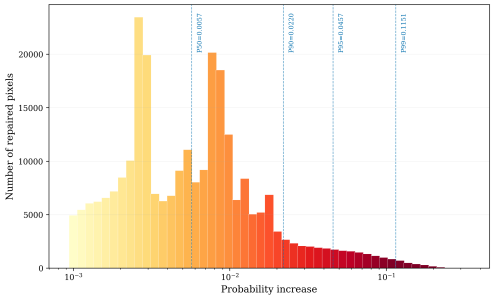


Fig. 11 Distribution of repair magnitudes (CE $\rightarrow$ ASRAC probability increases) for all repaired grid cells. Dashed vertical lines indicate the 50th, 90th, 95th, and 99th percentiles

ASR-based correction exploits the ecologically meaningful spatial relationships learned during training, thereby achieving spatial smoothing with less fluctuation in prediction accuracy.

The ASRAC is designed as a unidirectional post-processing step. Anomalous grid cells, identified as those where the CE prediction deviates from its 3×3 neighborhood mean by more than 0.1, are replaced by the ASR prediction only when the ASR model predicts a higher probability than the CE baseline, and never lowered. This design choice reflects the underlying epidemiology. In infectious disease risk mapping, high-risk areas drive spatial propagation, making it ecologically appropriate to allow risk to spread upwards while avoiding the suppression of potentially legitimate risk signals in data-sparse regions. Despite the slight increase in Brier score, this upward-only strategy has the potential to more effectively flag regions where underreporting is likely. This upward-only asymmetry implicitly assumes that the cost of missing a true high-risk area exceeds the cost of over-flagging, an appropriate default for screening-oriented risk mapping; the resulting positive probability shift manifests as a slightly elevated Brier score for ASRAC. To quantify the effect of this unidirectional design, we

Table 6 Impact of anomaly correction method on spatial and predictive accuracy

Metric	BRT base					XGBoost base				
	CE	KNN	SRAC λ=1.0	SRAC λ=3.0	ASRAC	CE	KNN	SRAC λ=1.0	SRAC λ=3.0	ASRAC
Recall	0.8933	0.8300	0.9012	**0.9058**	0.9012	0.9104	0.8485	0.9117	**0.9150**	0.9104
AUC(CV)	0.9144	0.7793	0.9159	0.9103	**0.9170**	0.9158	0.7963	0.9158	0.9087	**0.9161**
Brier(CV)	**0.1234**	0.2189	0.1270	0.1342	0.1260	**0.1282**	0.2166	0.1289	0.1384	0.1283
Moran's I	0.8052	0.9761†	0.8213	**0.8263**	0.8213	0.8029	0.9784†	0.8192	**0.8279**	0.8200
ContED	0.0606	0.0078†	0.0211	**0.0207**	0.0211	0.0544	0.0071†	0.0185	**0.0182**	0.0184
Iso. ratio	20.22%	4.29%†	**7.12%**	**7.12%**	**7.12%**	17.32%	3.74%†	6.13%	6.14%	**6.12%**

†: Invalid spatial indices. KNN replaces each cell by the arithmetic mean of its 8 spatial neighbors; its near-unity Moran's I and near-zero edge density reflect uniform smoothing rather than ecologically meaningful spatial structure, and the accompanying loss of predictive accuracy precludes reliable performance evaluation via these metrics.

compared it against a bidirectional variant that replaces every anomalous cell with the ASR prediction regardless of direction, reusing the identical base map and correction source. Despite revising roughly three times as many cells (1,588,920 versus 579,508), the bidirectional variant yields nearly identical performance (test AUC 0.9331 versus 0.9331; Brier 0.0715 versus 0.0720; Moran's I 0.8202 versus 0.8200; isolated-cell ratio 6.11% versus 6.12%). The near-identical metrics confirm that the aggregate effect of the upward-only design is small; because ASRAC only replaces anomalous cells when the correction increases risk, it is intended for sensitivity-oriented screening and may introduce a positive probability shift, and should therefore not be interpreted as a calibrated probability map.

KNN, the purely spatial 8-neighborhood grid cell replacement without environmental or ecological constraint, achieves the spatial smoothness metrics. This mechanical smoothing substantially reduces predictive performance (AUC(CV) 0.7963) and cannot be used as a valid risk map. Compared to the CE model, SRAC provides better spatial coherence (ContED 0.0182, Iso ratio 6.14%) while maintaining predictive accuracy. CE and SR are both XGBoost-based 30-ensembles: CE is trained with the standard cross-entropy objective; SR is the spatially regularized ensemble with fixed strength ($\lambda = 3.0$); KNN applies an 8-neighborhood mean replacement to anomalous CE grid cells; SRAC applies SR-model-based replacement to anomalous CE grid cells. The same pattern holds when the base model is a BRT 30-ensemble (Table 6): SRAC(λ=3.0) improves spatial coherence while maintaining or slightly improving accuracy, whereas the spatial KNN replacement under the BRT base again collapses predictive accuracy (AUC(CV) 0.7793), confirming that the SR-based correction, rather than spatial smoothing, drives the spatial coherence improvement.

To further examine the effect of ASRAC, we summarise the global repair statistics in Table 7. The correction magnitude is defined as $\Delta p = p(\text{ASRAC}) - p(\text{CE})$. Among the 3,912,486 grid cells with positive baseline risk, 268,257 cells (6.86%) received corrections with $\Delta p > 0.001$ (579,508 cells showed any positive change, 311,251 of them with $0 < \Delta p \leq 0.001$); the mean Δp over the corrected cells is 0.0117 and the median 0.0057. Newly classified as high-risk ($p > 0.5$) numbered 944 cells. Figure 10 shows the spatial distribution of the corrections, with the largest increases concentrated in

Table 7 Summary statistics of ASRAC on the global risk map. Pixels are defined at 2.5 arcminute resolution

Statistic	Value
Total Pixels	37,324,800
$p(\text{CE}) > 0$	3,912,486
Materially corrected ($\Delta p > 0.001$)	268,257 (6.86%)
Any positive replacement ($\Delta p > 0$)	579,508 (14.81%)
Newly classified as high-risk	944
Mean Δp	0.0117
Median Δp	0.0057

fringe areas of known transmission zones. Figure 11 shows the distribution of correction magnitudes. Because the correction is upward-only, the final surface carries a systematic positive shift.

3.6 Cross-disease replication: transferability to other Aedes-borne diseases

To evaluate generalization to ecologically similar Aedes-borne diseases, we trained CE and ASR-XGBoost models on occurrence data for each disease, using the same pipeline and identical model architecture as the CHIKF experiment. Table 8 reports model performance across all three diseases. In addition to the standard SR configurations (λ=1.0 and λ=3.0), we evaluate the block-wise variant (ASR), which selects the regularization strength per administrative block (country or first-level province, with sparse small countries merged via nearest-neighbor grouping) by minimizing the block's held-out Brier score, exactly as in the CHIKF main experiment. For all models we report spatial 10-fold cross-validation metrics (AUC(CV), Brier(CV)), following the same block-stratified protocol as the CHIKF experiment, so that extrapolation performance is directly comparable across diseases. The λ=1.0 setting serves as a conservative baseline to avoid over-regularizing data-sparse diseases. The λ=3.0 setting matches the CHIKF main experiment and tests whether the stronger regularizer transfers to other diseases. Spatial smoothness indices (Moran's I, ContED and Iso. ratio) are computed on the full predicted risk surface for all three diseases (Table 8); block-wise test-set evaluation at the $5^\circ \times 5^\circ$ scale was additionally examined for dengue, whose larger sample size allows meaningful spatial partitioning of the test set, whereas the limited sample size for yellow fever and Zika prevents such partitioning, although the block-stratified CV protocol remains applicable to all three diseases.

Cross-disease replication was restricted to Aedes-borne diseases because dengue, yellow fever, and Zika share the primary vectors (Aedes aegypti and Aedes albopictus) and environmental drivers with CHIKF, including the Aedes suitability features, so the identical feature pipeline transfers without modification. Their established arbovirus surveillance also makes their occurrence records reliable external labels. These diseases thus provide a transfer test ecologically close enough to be meaningful, yet independent in pathogen and reporting system.

Under the spatial block CV protocol, ASR consistently improves discrimination and calibration over both the CE baseline and the fixed-strength SR variants across all three diseases. For dengue, ASR attains the best AUC(CV) (0.9082 versus 0.9038

Table 8 Cross-disease replication across three Aedes-borne diseases. AUC(CV) and Brier(CV) are spatial 10-fold cross-validation metrics; ASR selects the regularization strength per administrative block on the training data

Disease	Metric	CE	SR(λ=1.0)	SR(λ=3.0)	ASR	SRAC(λ=1.0)	SRAC(λ=3.0)	ASRAC
	AUC(CV)	0.9038	0.9033	0.8936	**0.9082**	0.9036	0.8970	0.9042
	Brier(CV)	0.1349	0.1349	0.1398	**0.1280**	0.1359	0.1444	0.1355
Dengue	Recall	**0.8711**	0.8706	0.8585	0.8673	0.8740	0.8779	0.8753
(N=20,858)	Moran's I	0.8115	0.8085	0.8172	0.8108	0.8279	**0.8330**	0.8294
	ContED	0.0571	0.0570	0.0555	0.0573	0.0196	0.0197	**0.0195**
	Iso. ratio	18.50%	18.37%	18.00%	18.57%	6.57%	6.59%	**6.55%**
	AUC(CV)	0.8169	0.8166	0.8075	**0.8238**	0.8147	0.8053	0.8189
	Brier(CV)	0.1637	0.1656	0.1721	**0.1596**	0.1653	0.1713	0.1620
YF	Recall	0.8887	0.8925	0.9020	0.8970	0.8924	**0.9065**	0.8962
(N=4,179)	Moran's I	0.9403	0.9327	**0.9446**	0.9399	0.9350	0.9434	0.9386
	ContED	0.0232	0.0248	0.0233	0.0233	0.0084	**0.0081**	**0.0081**
	Iso. ratio	8.74%	9.21%	9.59%	8.81%	3.17%	**3.15%**	3.18%
	AUC(CV)	0.8075	0.8093	0.8077	**0.8176**	0.8084	0.8083	0.8105
	Brier(CV)	0.1900	0.1908	0.1934	**0.1847**	0.1919	0.1952	0.1906
Zika	Recall	0.7527	0.7636	0.7734	0.7650	0.7645	**0.7764**	0.7636
(N=4,316)	Moran's I	0.9018	0.9112	0.9056	0.9021	**0.9131**	0.9035	0.9091
	ContED	0.0415	0.0393	0.0402	0.0414	**0.0141**	0.0151	0.0143
	Iso. ratio	17.99%	16.69%	15.91%	17.98%	**6.30%**	**6.30%**	6.39%

for CE) and the lowest Brier(CV) (0.1280 versus 0.1349), while the fixed λ=3.0 setting shows a modest extrapolation cost (AUC 0.8936). For yellow fever, ASR again leads (AUC 0.8238 versus 0.8169 for CE; Brier 0.1596 versus 0.1637), and for Zika the improvement is largest in relative terms (AUC 0.8176 versus 0.8075; Brier 0.1847 versus 0.1900). The smaller CV gains for Zika can be attributed to its highly dispersed occurrence distribution: the 1,079 positive training records are scattered across 82 distinct 10-degree blocks worldwide (compared with 39 blocks for yellow fever and a more aggregated dengue distribution), leaving fewer spatially contiguous high-risk zones for the block-wise adaptive regularization to exploit. Under such sparsely distributed data, the adaptive selection finds limited local structure to exploit, yet it still improves CV calibration relative to a fixed strength (Brier 0.1847 versus 0.1900). Both SRAC(λ=1.0) and SRAC(λ=3.0) provided substantial spatial smoothing across all diseases (Table 8) while maintaining predictive accuracy, confirming that the ASR framework transfers across Aedes-borne diseases. The adaptive ASRAC post-processing also maintains competitive CV discrimination across diseases (dengue AUC(CV) 0.9042, YF 0.8189, Zika 0.8105), consistent with the main CHIKF experiment.

3.7 Simulation evidence for adaptive regularization

The simulation study evaluates spatial regularization under controlled spatial statistical conditions and examines the sensitivity of its results to several land–sea configurations. The data-generating mechanisms do not contain the neighbourhood-averaging term or the ASR penalty, thereby avoiding direct incorporation of the proposed objective into the simulated truth.

Each dataset was generated on a 60×60 grid. Three standardized covariates were drawn as nested two-scale Gaussian random fields with Matérn covariance ($\nu = 1.5$), and the true suitability was $p_{\text{true}} = \sigma(\alpha\eta)$ with $\eta = X_1 + 0.8X_2 + 0.5X_3 - 0.6X_1X_2 + 0.4X_2X_3 - 0.2$, so that suitability follows nonlinear covariate responses. For the

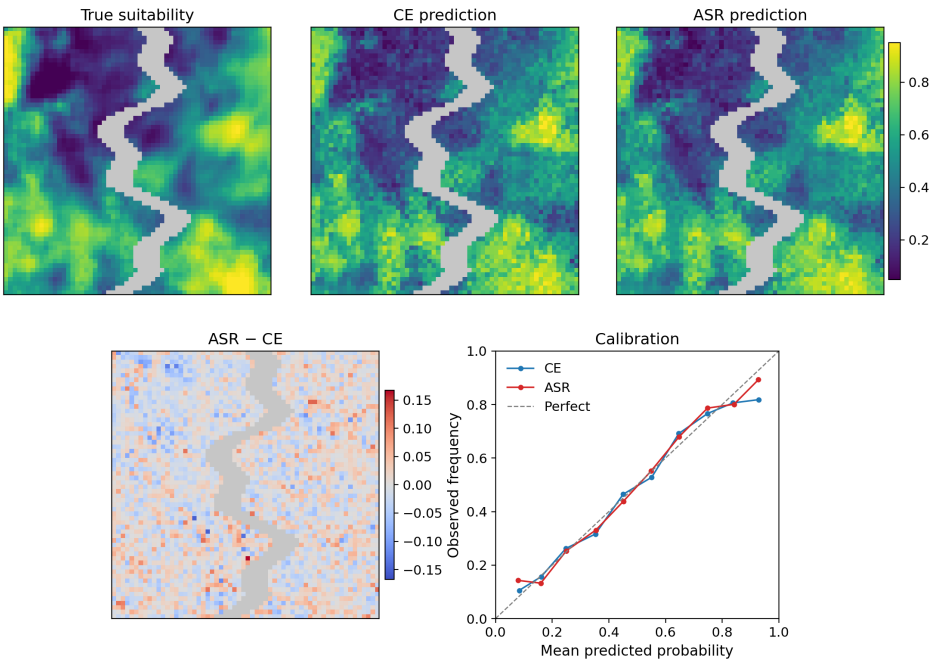


Fig. 12 A single repetition of the medium-noise scenario, drawn with all land pixels sampled so that the predicted surface is complete. The panels show the true suitability field, the risk predicted by the cross-entropy baseline (CE) and by block-adaptive regularization (ASR), their difference, and a calibration curve for both methods, at the outer held-out pixels

symmetric-noise scenarios, the observable occurrence probability was

$$p_{\text{obs}}(s) = E_{\varepsilon}[\text{clip}\{p_{\text{true}}(s) + \varepsilon(s), 0.001, 0.999\}], \qquad \varepsilon(s) \sim \mathcal{N}(0, \sigma_b^2),$$

and $y(s) \sim \text{Bernoulli}\{p_{\text{obs}}(s)\}$. The region-specific σ_b provides a symmetric proxy for heterogeneous observation reliability rather than a mechanistic model of underreporting. The detection scenario instead used $p_{\text{obs}}(s) = \text{clip}\{p_{\text{true}}(s)q(s), 0.001, 0.999\}$ with a smoothly varying reporting probability $q(s)$. Five scenarios were considered: low noise, medium noise, weak signal, heterogeneous detection, and a non-stationary covariate process formed by a spatially varying mixture of long- and short-range Gaussian fields. The last construction changes the effective smoothness across the grid without placing an ASR term in the data-generating process.

Land pixels were subsampled at 40% and evaluated with three outer folds, so that every sampled pixel was predicted once by a model that had not seen it. Regularization strengths were chosen by inner three-fold cross-validation, with a separate 15% subset of each training fold used for early stopping. No outer-fold held-out label entered model fitting, early stopping or strength selection. Two methods were evaluated under both random-pixel and spatial-block outer folds: CE and block-wise ASR. ASR used the candidate set $\{0, 0.5, 1, 2, 3.5\}$ and selected strengths by inner-fold Brier score. Results

Table 9 Simulation study: predictive performance of the cross-entropy baseline (CE) and block-wise adaptive spatial regularization (ASR), by scenario and evaluation protocol, over 20 independent repetitions per scenario. The column labels rand. and spat. denote random-pixel and spatial-block outer folds, respectively; under spatial-block folds, complete spatial blocks are held out together. $\mathrm{MSE}_{p_{\mathrm{true}}}$ is measured against the known latent suitability field. The better of the two methods is shown in bold for each scenario and metric; small differences should be interpreted together with the paired Monte Carlo uncertainty described in the text

		Low noise		**Medium noise**		**Weak signal**		**Non-stationary**		**Detection**	
		rand.	spat.	rand.	spat.	rand.	spat.	rand.	spat.	rand.	spat.
AUC(CV)	CE	0.7925	0.7750	0.7436	0.7288	0.5845	0.5687	0.7747	0.7517	0.6901	0.6567
	ASR	**0.7966**	**0.7793**	**0.7485**	**0.7340**	**0.5874**	**0.5692**	**0.7787**	**0.7574**	**0.6974**	**0.6588**
Brier(CV)	CE	0.1835	0.1909	0.2043	0.2100	0.2450	**0.2466**	0.1913	0.2007	0.1928	0.1998
	ASR	**0.1816**	**0.1889**	**0.2026**	**0.2083**	**0.2442**	0.2469	**0.1898**	**0.1989**	**0.1908**	**0.1989**
$\mathrm{MSE}_{p_{\mathrm{true}}}$	CE	0.0188	0.0263	0.0167	0.0243	0.0138	0.0172	0.0181	0.0291	**0.0505**	**0.0554**
	ASR	**0.0163**	**0.0239**	**0.0145**	**0.0225**	**0.0127**	**0.0167**	**0.0161**	**0.0271**	0.0511	0.0568

were summarized over 20 independent repetitions per scenario; Monte Carlo standard errors describe uncertainty arising from the finite number of repetitions.

Figure 12 shows one repetition of the medium-noise scenario and is used only as a qualitative illustration. Table 9 reports the aggregated predictive results. Relative to CE, ASR increased mean AUC by 0.0029–0.0073 and reduced mean Brier score by 0.0008–0.0020 under random-pixel folds. The 95% Monte Carlo interval for the paired mean difference excluded zero for AUC in four scenarios and for Brier score in all five. Under spatial-block folds, corresponding improvements were supported for both AUC and Brier score in the low-noise, medium-noise and non-stationary scenarios. In the weak-signal spatial setting the differences were small and uncertain, while in the detection setting ASR improved prediction of the observed labels but moved farther from the latent suitability field and reduced recall.

An exploratory analysis localized changes to regions where the baseline model was uncertain. The gating weight $\delta(\hat{p}) = 1 - 2|\hat{p} - 0.5|$ is largest at $\hat{p} = 0.5$ and vanishes at 0 and 1. Within bands defined by p_{CE} (Table 10), ASR showed higher rank correlation with the known suitability field and higher band-local AUC. These contrasts should not be interpreted as independent confirmation of the gating mechanism: selecting pixels by the CE prediction mechanically restricts the range of CE within the band, while ASR is free to reorder the same pixels. The results are therefore reported as a descriptive localization analysis rather than as an estimate of the overall treatment effect of gating; the overall predictive comparisons are those in Table 9.

A supplementary robustness check repeated the low-noise, medium-noise and non-stationary scenarios under four land–sea configurations using the regular grid partition. The ASR–CE differences were qualitatively similar across the tested configurations. This is not an isolated causal test of coastline geometry: the masks change the number of available land pixels and sampled observations, and they can also alter the generated regions and observation process. Fragmented masks additionally increased the number of blocks that inherited the globally selected strength, limiting how fully the block-wise adaptation was exercised.

Table 10 Exploratory localization analysis within bands defined by the CE baseline prediction (20 repetitions per scenario and protocol; rand. and spat. denote random-pixel and spatial-block outer folds, as in Table 9). Rank correlation is the Spearman correlation between predicted risk and the known suitability field, and AUC uses observed labels. Because selection by p_{CE} restricts the CE prediction range but not the ASR range, CE–ASR contrasts in this table are descriptive and should not be interpreted as independent tests of the gating mechanism

			Low noise		**Medium noise**		**Weak signal**		**Non-stationary**		**Detection**	
			rand.	spat.	rand.	spat.	rand.	spat.	rand.	spat.	rand.	spat.
$p_{\mathrm{CE}} \in [0.45, 0.55]$	Rank corr.	CE	0.1518	0.1326	0.2333	0.2062	0.3887	0.3018	0.1988	0.1829	0.1793	0.1136
		ASR	**0.4267**	**0.3870**	**0.4738**	**0.3924**	**0.4701**	**0.3444**	**0.4200**	**0.3822**	**0.3348**	**0.2343**
	AUC(CV)	CE	0.5124	0.5241	0.5516	0.5499	0.5379	0.5404	0.5352	0.5438	0.5356	0.5108
		ASR	**0.5813**	**0.5693**	**0.5891**	**0.5843**	**0.5481**	**0.5433**	**0.5684**	**0.6014**	**0.5666**	**0.5477**
$p_{\mathrm{CE}} \in [0.4, 0.6]$	Rank corr.	CE	0.3606	0.3353	0.4335	0.3976	0.5093	0.4007	0.3786	0.3165	0.3391	0.2386
		ASR	**0.5200**	**0.4792**	**0.5782**	**0.5111**	**0.5650**	**0.4227**	**0.5101**	**0.4466**	**0.4380**	**0.3112**
	AUC(CV)	CE	0.5643	0.5800	0.5848	0.5821	0.5605	0.5538	0.5718	0.5643	0.5439	0.5225
		ASR	**0.5928**	**0.6049**	**0.6093**	**0.6080**	**0.5648**	**0.5542**	**0.5957**	**0.6028**	**0.5678**	**0.5396**
$p_{\mathrm{CE}} \in [0.35, 0.65]$	Rank corr.	CE	0.4966	0.4655	0.5986	0.5288	0.5727	0.4466	0.5359	0.4551	0.4669	0.3619
		ASR	**0.6033**	**0.5664**	**0.6877**	**0.6063**	**0.6176**	**0.4647**	**0.6171**	**0.5382**	**0.5362**	**0.4091**
	AUC(CV)	CE	0.5970	0.6108	0.6153	0.6193	0.5704	0.5630	0.6039	0.6099	0.5706	0.5527
		ASR	**0.6164**	**0.6285**	**0.6328**	**0.6374**	**0.5741**	**0.5631**	**0.6204**	**0.6311**	**0.5858**	**0.5588**

4 Discussion

4.1 Contributions

Existing disease risk mapping studies often incorporate prior knowledge to guide model predictions toward epidemiologically plausible patterns. Messina et al. (2019) mapped current and future global dengue suitability using boosted regression trees with bioclimatic predictors, projecting changes in population at risk under climate and socioeconomic scenarios. Lim et al. (2025) corrected surveillance biases through joint niche models across multiple arboviruses, and Yang et al. (2025) mapped the global risk of CHIKV endemicity and of autochthonous transmission following importation. Across these approaches, spatial information enters primarily at the level of input data and variable specification. The proposed ASR framework takes a different route: rather than modifying the input structure or the internal architecture of the prediction algorithm, it embeds spatial continuity directly into the training objective function through 8-neighborhood mean soft labels. This allows the model to learn spatial patterns while maintaining computational efficiency. The proposed framework has three components, whose core mechanism is further validated on synthetic data with known noise.

1. Ecologically stratified pseudo-absence sampling. Each of the 30-ensemble submodels independently draws a balanced random subset of negatives from the three sampling sources, ensuring that every sub-model benefits from informative negatives and class balance while reducing label bias inherent in random background sampling. Removing the suitability-defined negative source lowers test AUC only from 0.9334 to 0.9238, so discrimination does not rest on that source alone (Section 2.2).

2. ASR. Pixel-wise 8-neighborhood mean soft labels derived from an initial ensemble are combined with a gated spatial penalty $r = \delta(p_i) \cdot (p_i - \tilde{p}_i)^2$ to embed spatial continuity directly into the XGBoost objective function, incorporating spatial correlation of infectious disease risk into the prediction. The regularization strength is selected per block within a theoretically admissible range whose upper bound is fixed

by the Hessian-positivity condition of Proposition 2; the gradient-boundedness and Hessian-positivity guarantees (Propositions 1 and 2) hold for any base loss that retains these properties, and because the constraint acts on the predicted probability surface rather than on sample coordinates, it avoids coordinate leakage and remains applicable beyond the historical occurrence envelope. Because each soft label is the mean of neighboring initial-ensemble probabilities generated from the neighbors' full covariate vectors, the spatial reference is feature-mediated rather than a smoothing of probability values alone.

3. ASRAC, a post-processing extension of ASR that replaces spatially inconsistent grid cells in the baseline risk map with ASR predictions, substantially improving spatial coherence while maintaining predictive accuracy (Section 3.5).

4. Controlled synthetic validation. Because comparisons on empirical data bundle the regularizer with the sampling design and the post-processing step, we isolate the regularization mechanism in a controlled simulation of risk fields with known ground truth and controlled block-wise label noise (Section 3.7). Across five stress scenarios and two evaluation protocols, spatial regularization generally improved prediction relative to CE, with weaker or uncertain effects under spatial validation in the weak-signal and heterogeneous-detection settings. A separate three-scenario robustness check produced qualitatively similar ASR–CE contrasts across the tested land–sea configurations.

4.2 Comparison with spatially explicit learning methods

Spatial structure can enter risk models through different modelling logics, such as geographically weighted fitting (G-XGBoost Grekousis (2025), M-GWXGB Gao et al. (2026)), a Gaussian-process random effect (GPBoost Sigrist (2022)), or regularization of the predicted probability surface (ASR). Geographically weighted approaches follow a local-modelling logic, assuming that the covariate–response relationship itself varies in space: G-XGBoost fits spatially local models with spatial weights, at the cost of requiring enough positive records in each locality. M-GWXGB relaxes this requirement by assigning each variable to either a global or a local estimator within an additive structure, but the method is inference-oriented: it is designed to reveal where and how a variable's effect varies across space, whereas risk mapping calls for a probability surface that remains reliable beyond the sampled area. Its local components also require place-specific calibration that a few thousand spatially thinned presence-only records can hardly support.

GPBoost follows a random-effects logic: it augments a global tree model with a Gaussian-process spatial random effect in the residual covariance, assuming that residual autocorrelation carries the spatial signal. We evaluated GPBoost under the identical data protocol used for the proposed models: under the same spatial block CV it attains the highest discrimination among all models (AUC(CV) = 0.9238 versus 0.9158 for CE and 0.9187 for ASR) with competitive calibration (Brier(CV) = 0.1254). We deliberately keep GPBoost out of the direct comparison tables, however, because the comparison would be information-asymmetric: the GP random effect uses sample coordinates (longitude/latitude) as inputs, granting the model access to location information that the covariate-only CE and ASR models do not use.

ASR follows a regularization logic: it regularizes predictions toward their neighborhood mean through a gated penalty, encouraging spatial continuity without imposing it, making no assumption about non-stationarity or residual covariance structure and requiring no coordinate inputs. For presence-only records that are globally sparse and unevenly distributed, ASR avoids both the local-sample requirement of geographically weighted fitting and the coordinate dependence of the Gaussian-process random effect.

The ASR objective is not specific to XGBoost. The soft labels are computed from the initial model output, and the gated penalty $r = \delta(p_i) \cdot (p_i - \tilde{p}_i)^2$ depends only on the predicted probability p_i and the soft label $\tilde{p}_i$, rather than on the internal structure of the classifier. The theoretical guarantees ensure that adding the ASR penalty preserves gradient boundedness and Hessian positivity for any base model whose original loss has these properties. With a BRT 30-ensemble as the base model, SRAC(λ=3.0) improves spatial coherence while maintaining or slightly improving accuracy, whereas spatial KNN replacement collapses predictive accuracy (Table 6). Unlike spatial proxy approaches Milà et al. (2024), which add coordinate-based features to the predictor set, such approaches risk excessive reliance on coordinates, and generalization can suffer where coordinate ranges fall outside the training domain. Because the ASR regularizer operates on the predicted probability surface itself rather than on coordinates, it does not inherit this risk.

4.3 Pseudo-absence sampling and its evaluation

The ecologically stratified pseudo-absence sampling strategy (Section 2.2) combines three complementary sources to mitigate shared biases: YF/ZIKV-endemic regions provide informative negatives under established surveillance capacity, low-suitability areas provide reliable negatives, and farthest-point sampling covers suitable case-free areas. The sampling procedure draws negatives away from reported cases and maximizes inter-point distances, keeping the pseudo-absence labels reliable and well separated. The ensemble aggregation across rounds further mitigates underreporting-driven noise.

The nested evaluation in Table 5 speaks to the validity of the stratified design. One of the three negative sources is defined by a threshold on mosquito suitability, a covariate that the model also receives as input; if the model had simply learned to reproduce this threshold, discrimination would collapse once the stratum is removed. It does not: excluding the low-suitability negatives lowers the AUC only modestly (0.9238 versus 0.9334 on the full test set). Discrimination rises on the later subsets only because progressively harder samples are removed, namely positive records that the model under-rates and YF/ZIKV-endemic negatives whose environmental profiles resemble CHIKV-positive cells; the AUC values across rows are therefore not directly comparable, and the high value in the last row does not indicate better discrimination. The informative signal comes from the contrast between suitable cells with and without reported cases, which is exactly the contrast that the case-free suitable source is designed to provide; the three-source strategy therefore supplies reliable pseudo-absence labels without teaching the model an artefactual threshold rule.

4.4 Design choices of the ASR regularizer

The 8-neighborhood mean soft labels in the ASR penalty use a 3×3 (8-neighborhood) window to compute the local mean predicted probability $\tilde{p}_i$. This window is the smallest non-trivial neighborhood, chosen to keep the spatial prior strictly local. Because the soft label and gating penalty are computed per grid cell, a larger window would make the neighborhoods of adjacent cells overlap extensively: the same region would be repeatedly referenced by the regularization terms of many cells, turning the local penalty into a broad, entangled smoothing whose spatial effect is difficult to interpret. The minimal 8-neighborhood (approximately 7 km near the equator, Section 2.3) regularizes each cell only toward its immediate neighbours, preserving a clear locality for the prior.

The adaptive spatial gating function $\delta(p_i) = 1 - 2|p_i - 0.5|$ in Eq. (3) is used for its simplicity. The non-differentiability of δ at $p_i = 0.5$ has no effect on the optimization under the frozen-gating convention described in Section 2.4, and it yields the explicit Hessian-positivity range $\lambda \in [0, 3.5]$ guaranteed by Proposition 2, which is exactly the admissible range from which each block selects its strength (Eq. (4)). Smoother alternatives with the same endpoints and peak, such as the parabola $\delta(p_i) = 4p_i(1 - p_i)$, would still depend on the prediction and therefore still require the frozen-gating treatment, while narrowing the guaranteed range of λ below the admissible upper bound $\lambda_{\max} = 3.5$.

The soft label is the arithmetic mean of neighboring initial-ensemble probabilities. Its ecological content is indirect: each probability is generated from the neighboring cell's full covariate vector. The spatial prior is therefore feature-mediated, in contrast to purely spatial smoothing (e.g., Gaussian filtering or coordinate-based interpolation), which operates on probability values alone.

4.5 Anomaly correction and cross-disease transfer

The anomaly correction stage is useful when the real-world phenomenon exhibits strong spatial autocorrelation but the baseline model produces fragmented predictions. By applying a trained SR model to repair isolated grid cells in a baseline risk map, the correction procedure achieves spatial smoothing while maintaining relatively stable predictive accuracy. The SR-based correction generates replacement values from the full covariate space, so the corrected surface remains ecologically consistent, in contrast to purely spatial smoothing, which degrades accuracy (Section 3.5). This two-stage design separates the roles of the base predictor and the spatial regularizer, allowing practitioners to improve spatial coherence of existing risk maps without retraining from scratch.

Cross-disease replication across three Aedes-borne diseases (Table 8) confirms that the ASR transfers across diseases: under the same spatial block CV protocol, ASR attains the best AUC(CV) and the lowest Brier(CV) for all three diseases, each improving on the CE baseline and on both fixed-strength SR variants, while both SRAC variants provide substantial spatial smoothing across all diseases at comparable accuracy.

Throughout this study, the ASR regularization weight is not a globally fixed constant but is selected per spatial block: each block chooses the strength that minimizes its held-out Brier score within the theoretically supported range, and the resulting per-block strengths form a block-wise (piecewise-constant) spatial field whose upper bound is $\lambda_{\max} = 3.5$ (Section 2.3). The theoretical guarantee (Proposition 2) constrains this upper bound: the per-sample Hessian of the ASR loss is proven to be positive for $\lambda \leq 3.5$, ensuring that the Newton-type update steps of XGBoost rely on a convex quadratic approximation of the loss function. Beyond this threshold, the Hessian can become non-positive, potentially destabilizing the tree-splitting optimization. Because every block-selected weight satisfies $\lambda_{b(i)} \in [0, 3.5]$, the positivity guarantee holds per sample without further tuning.

4.6 Limitations and future work

This study has several limitations. First, the three-source negative sampling strategy relies on the quality of existing surveillance data; areas with limited diagnostic capacity may still contain undetected CHIKF cases, introducing noise into the negative pool. Second, while the ASR framework embeds spatial continuity into training, the XGBoost backbone itself remains a black-box model, and the interpretability of the learned spatial patterns is limited Ribeiro et al. (2016), which the spatial coupling that the regularizer introduces into neighboring predictions further complicates. Third, the ASR regularizer operates on local spatial continuity, so landscape-scale ecological barriers and connectivity patterns are not represented.

Future work could reduce the noise in the negative pool with independent evidence on case detection, whose regional coverage is constrained by the same reporting heterogeneity documented above. Extending the regularizer to landscape-scale structure would require connectivity information outside the covariate set assembled here and a rederivation of the penalty's theoretical guarantees.

5 Conclusion

Risk at neighboring locations is often spatially correlated because environmental conditions and vector suitability vary smoothly and transmission can generate local dependence after introduction. Standard grid-cell classifiers may recover part of this pattern from spatially structured predictors, but their usual pointwise losses do not explicitly represent dependence between observations and can yield fragmented prediction surfaces. ASR-XGBoost embeds local spatial context in the training objective, using pixel-wise 8-neighborhood mean soft labels as a spatial reference and a gated penalty whose strength increases near uncertain predictions. ASRAC, a post-processing extension, applies a specified spatial-consistency rule to replace selected isolated cells with ASR predictions while leaving cells that satisfy the rule unchanged.

The three components operate at complementary levels. Ecologically stratified sampling reduces dependence on arbitrary background selection, the ASR objective constrains the predicted probability surface during training, and ASRAC reduces fragmentation according to an explicit spatial-consistency rule. The resulting CHIKF maps retain strong predictive performance while producing more spatially coherent

risk surfaces that may support arbovirus surveillance and prioritization. In controlled simulations with known ground truth, spatial regularization generally improved prediction relative to CE across the tested scenarios, while block-wise ASR showed no consistent additional accuracy advantage over a globally tuned regularizer. The simulations therefore support the use of spatial regularization and define the present empirical limits of the adaptive component rather than attributing all gains specifically to regional adaptation. Cross-disease replication on dengue, yellow fever, and Zika suggests that the workflow can be applied to related Aedes-borne mapping tasks.

References

Barbet-Massin M, Jiguet F, Albert C, et al (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3(2):327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x, URL https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00172.x

Borisov V, Leemann T, Seßler K, et al (2024) Deep neural networks and tabular data: A survey. IEEE Trans Neural Netw Learn Syst 35(6):7499–7519. https://doi.org/10.1109/TNNLS.2022.3229161, URL https://ieeexplore.ieee.org/document/9998482/

Burt FJ, Chen W, Miner JJ, et al (2017) Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. Lancet Infect Dis 17(4):e107–e117. https://doi.org/10.1016/S1473-3099(16)30385-1, URL https://linkinghub.elsevier.com/retrieve/pii/S1473309916303851

Caldwell JM, LaBeaud AD, Lambin EF, et al (2021) Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents. Nat Commun 12(1):1233. https://doi.org/10.1038/s41467-021-21496-7, URL https://www.nature.com/articles/s41467-021-21496-7

Cardona-Trujillo MC, Ocampo-Cárdenas T, Tabares-Villa FA, et al (2022) Recent molecular techniques for the diagnosis of Zika and Chikungunya infections: A systematic review. Heliyon 8(8):e10225. https://doi.org/10.1016/j.heliyon.2022.e10225, URL https://linkinghub.elsevier.com/retrieve/pii/S2405844022015134

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 785–794, https://doi.org/10.1145/2939672.2939785, URL https://dl.acm.org/doi/10.1145/2939672.2939785

Coetzee BWT, Burke AM, Koekemoer LL, et al (2023) Scaling artificial light at night and disease vector interactions into socio-ecological systems: a conceptual appraisal. Philos Trans R Soc B Biol Sci 378(1892):20220371. https://doi.org/10.1098/rstb.2022.0371, URL http://royalsocietypublishing.org/rstb/article/109395

Cortes-Azuero O, O'Driscoll M, Ribeiro dos Santos G, et al (2026) The epidemiology of chikungunya virus in Brazil and the potential impact of vaccines: a mathematical

modelling study. Lancet Infect Dis 26(4):406–416. https://doi.org/10.1016/S1473-3099(25)00605-X, URL https://linkinghub.elsevier.com/retrieve/pii/S147330992500605X

Côrtes N, Lira A, Prates-Syed W, et al (2023) Integrated control strategies for dengue, Zika, and Chikungunya virus infections. Front Immunol 14:1281667. https://doi.org/10.3389/fimmu.2023.1281667, URL https://www.frontiersin.org/articles/10.3389/fimmu.2023.1281667/full

Eldar Y, Lindenbaum M, Porat M, et al (1997) The farthest point strategy for progressive image sampling. IEEE Trans Image Process 6(9):1305–1315. https://doi.org/10.1109/83.623193

Farooq Z, Segelmark L, Rocklöv J, et al (2025) Impact of climate and Aedes albopictus establishment on dengue and chikungunya outbreaks in Europe: a time-to-event analysis. Lancet Planet Health 9(5):e374–e383. https://doi.org/10.1016/S2542-5196(25)00059-2, URL https://linkinghub.elsevier.com/retrieve/pii/S2542519625000592

Fiorta MA, Fyie LR, Meuti ME (2024) Light pollution disrupts seasonal reproductive phenotypes and reduces lifespan in the West Nile vector, Culex pipiens. J Insect Physiol 159:104725. https://doi.org/10.1016/j.jinsphys.2024.104725, URL https://linkinghub.elsevier.com/retrieve/pii/S0022191024001136

Fortin MJ (1994) Edge detection algorithms for two-dimensional ecological data. Ecology 75(4):956–965. https://doi.org/10.2307/1939419, URL https://doi.org/10.2307/1939419

Gao F, He SY, Kwan MP (2026) Mixed geographically weighted XGBoost (M-GWXGB) model: A new spatially explicit machine learning model. Ann Am Assoc Geogr 116(4):824–855. https://doi.org/10.1080/24694452.2025.2586039, URL https://www.tandfonline.com/doi/full/10.1080/24694452.2025.2586039

Gorishniy Y, Rubachev I, Khrulkov V, et al (2021) Revisiting deep learning models for tabular data. In: Advances in Neural Information Processing Systems 34 (NeurIPS 2021). Curran Associates, Inc., pp 18932–18943, URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/9d86d83f925f2149e9edb0ac3b49229c-Abstract.html

Grekousis G (2025) Geographical-XGBoost: a new ensemble model for spatially local regression based on gradient-boosted trees. J Geogr Syst 27(2):169–195. https://doi.org/10.1007/s10109-025-00465-4, URL https://link.springer.com/10.1007/s10109-025-00465-4

Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on typical tabular data? In: Koyejo S, Mohamed S, Agarwal

A, et al (eds) Advances in Neural Information Processing Systems, vol 35. Curran Associates, Inc., pp 507–520, https://doi.org/10.52202/068431-0037, URL https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf

He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 770–778, https://doi.org/10.1109/CVPR.2016.90, URL https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Knappe S, Wied D, Jansen T (2026) Unveiling Spatial Dependencies — Investigating the Determinants of Firm Exit in Germany. J Reg Sci 66(2):500–515. https://doi.org/10.1111/jors.70042, URL https://onlinelibrary.wiley.com/doi/10.1111/jors.70042

Kraemer MU, Sinka ME, Duda KA, et al (2015) The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. eLife 4:e08347. https://doi.org/10.7554/eLife.08347, URL https://elifesciences.org/articles/08347

Kraemer MUG, Reiner RC, Brady OJ, et al (2019) Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. Nat Microbiol 4(5):854–863. https://doi.org/10.1038/s41564-019-0376-y, URL https://www.nature.com/articles/s41564-019-0376-y

Leta S, Beyene TJ, De Clercq EM, et al (2018) Global risk mapping for major diseases transmitted by Aedes aegypti and Aedes albopictus. Int J Infect Dis 67:25–35. https://doi.org/10.1016/j.ijid.2017.11.026, URL https://linkinghub.elsevier.com/retrieve/pii/S1201971217303089

Li X, Zhou Y, Zhao M, et al (2020) Harmonization of DMSP and VIIRS nighttime light data from 1992–2024 at the global scale. https://doi.org/10.6084/m9.figshare.9828827.v10, https://doi.org/10.6084/m9.figshare.9828827.v10

Lim A, Shearer FM, Sewalk K, et al (2025) The overlapping global distribution of dengue, chikungunya, Zika and yellow fever. Nat Commun 16(1):3418. https://doi.org/10.1038/s41467-025-58609-5, URL https://doi.org/10.1038/s41467-025-58609-5

Lim JK, Ridde V, Agnandji ST, et al (2023) Seroepidemiological reconstruction of long-term chikungunya virus circulation in Burkina Faso and Gabon. J Infect Dis 227(2):261–267. https://doi.org/10.1093/infdis/jiac246, URL https://academic.oup.com/jid/article/227/2/261/6609553

Lober L, Roster K, Rodrigues F (2024) Forecasting infectious diseases in brazilian cities: Integrating socio-economic and geographic data from related cities through a machine learning approach. Chaos Solitons Fractals 187:115417. https://doi.org/10.1016/j.chaos.2024.115417, URL https://linkinghub.elsevier.com/retrieve/pii/S096007792400969X

Local Burden of Disease Child Growth Failure Collaborators (2020) Mapping child growth failure across low- and middle-income countries. Nature 577(7789):231–234. https://doi.org/10.1038/s41586-019-1878-8, URL https://www.nature.com/articles/s41586-019-1878-8

MacNab YC (2022) Bayesian disease mapping: Past, present, and future. Spat Stat 50:100593. https://doi.org/10.1016/j.spasta.2022.100593, URL https://linkinghub.elsevier.com/retrieve/pii/S2211675322000033

Messina JP, Brady OJ, Golding N, et al (2019) The current and future global distribution and population at risk of dengue. Nat Microbiol 4(9):1508–1515. https://doi.org/10.1038/s41564-019-0476-8, URL https://www.nature.com/articles/s41564-019-0476-8

Meyer AD, Guerrero SM, Dean NE, et al (2025) Predictability of infectious disease outbreak severity: Chikungunya as a case study. Sci Adv 11(40):eadt5419. https://doi.org/10.1126/sciadv.adt5419, URL https://doi.org/10.1126/sciadv.adt5419

Meyer H, Pebesma E (2022) Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat Commun 13(1):2208. https://doi.org/10.1038/s41467-022-29838-9, URL https://www.nature.com/articles/s41467-022-29838-9

Milà C, Ludwig M, Pebesma E, et al (2024) Random forests with spatial proxies for environmental modelling: opportunities and pitfalls. Geosci Model Dev 17(15):6007–6033. https://doi.org/10.5194/gmd-17-6007-2024, URL https://gmd.copernicus.org/articles/17/6007/2024/

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37(1-2):17–23. https://doi.org/10.1093/biomet/37.1-2.17, URL https://doi.org/10.1093/biomet/37.1-2.17

Nsoesie EO, Kraemer MU, Golding N, et al (2016) Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. Euro Surveill 21(20):30234. https://doi.org/10.2807/1560-7917.ES.2016.21.20.30234, URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2016.21.20.30234

Phillips SJ, Dudík M, Elith J, et al (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19(1):181–197. https://doi.org/10.1890/07-2153.1, URL https://esajournals.onlinelibrary.wiley.com/doi/10.1890/07-2153.1

Ploton P, Mortier F, Réjou-Méchain M, et al (2020) Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat Commun 11(1):4540. https://doi.org/10.1038/s41467-020-18321-y, URL https://www.nature.com/articles/s41467-020-18321-y

Poggio L, De Sousa LM, Batjes NH, et al (2021) SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7(1):217–240. https://doi.org/10.5194/soil-7-217-2021, URL https://soil.copernicus.org/articles/7/217/2021/

Renner IW, Elith J, Baddeley A, et al (2015) Point process models for presence-only analysis. Methods Ecol Evol 6(4):366–379. https://doi.org/10.1111/2041-210X.12352, URL https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12352

Ribeiro M, Singh S, Guestrin C (2016) Why should I trust you?: Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 1135–1144, https://doi.org/10.1145/2939672.2939778, URL https://dl.acm.org/doi/10.1145/2939672.2939778

Roberts DR, Bahn V, Ciuti S, et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40(8):913–929. https://doi.org/10.1111/ecog.02881, URL https://nsojournals.onlinelibrary.wiley.com/doi/10.1111/ecog.02881

Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60(1-4):259–268. https://doi.org/10.1016/0167-2789(92)90242-F, URL https://doi.org/10.1016/0167-2789(92)90242-F

Sahu SK, Böhning D (2022) Bayesian spatio-temporal joint disease mapping of Covid-19 cases and deaths in local authorities of England. Spat Stat 49:100519. https://doi.org/10.1016/j.spasta.2021.100519, URL https://linkinghub.elsevier.com/retrieve/pii/S2211675321000294

Salje H, Cauchemez S, Alera MT, et al (2016) Reconstruction of 60 years of chikungunya epidemiology in the Philippines demonstrates episodic and focal transmission. J Infect Dis 213(4):604–610. https://doi.org/10.1093/infdis/jiv470, URL https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiv470

Ribeiro dos Santos G, Jawed F, Mukandavire C, et al (2025) Global burden of chikungunya virus infections and the potential benefit of vaccination campaigns. Nat Med 31(7):2342–2349. https://doi.org/10.1038/s41591-025-03703-w, URL https://www.nature.com/articles/s41591-025-03703-w

Sigrist F (2022) Gaussian process boosting. J Mach Learn Res 23(232):1–46. URL http://jmlr.org/papers/v23/20-322.html

Simoncini A, Giunchi D, Marcucci M, et al (2025) A framework for predicting zoonotic hosts using pseudo-absences: the case of echinococcus multilocularis. Ecol Inform 90:103295. https://doi.org/10.1016/j.ecoinf.2025.103295, URL https://linkinghub.elsevier.com/retrieve/pii/S1574954125003048

Somvanshi S, Das S, Javed S, et al (2026) A survey on tabular data: From tree-based methods to tabular deep learning. ACM Comput Surv p Article 3807777. . https://doi.org/10.1145/3807777, URL https://dl.acm.org/doi/10.1145/3807777

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(sup1):234–240. https://doi.org/10.2307/143141, URL https://www.jstor.org/stable/143141?origin=crossref

Weaver SC, Lecuit M (2015) Chikungunya virus and the global spread of a mosquito-borne disease. N Engl J Med 372(13):1231–1239. https://doi.org/10.1056/NEJMra1406035

Yang YF, Qiu YB, Xu Q, et al (2025) Mapping the global risk of chikungunya virus endemicity and autochthonous transmission following importation. Travel Med Infect Dis 67:102892. https://doi.org/10.1016/j.tmaid.2025.102892, URL https://linkinghub.elsevier.com/retrieve/pii/S1477893925000985

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process. During the preparation of this work the authors used DeepSeek V4 Flash in order to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Statements and Declarations

Funding. This study was supported by the Fundamental Research Funds for the Central Universities (Nos. 31920250035, 31920260101), the Scientific Research Project of the Education Department of Gansu Province (No. 2026QB-009), and the Innovation Team of Ecosystem Restoration Modeling Theory and Application of Northwest Minzu University.

Competing interests. The authors declare that they have no competing interests.

Data and code availability. All environmental and socioeconomic raster data and disease occurrence records used in this study are publicly available from the sources cited in Section 2.1. An open-source demo implementation of the core ASR-XGBoost method (with synthetic data for reproducible simulation-based validation) is publicly available at https://github.com/souserJ/ASR-XGBoost. The full implementation of the method and its post-processing extension used to produce the results of this study is available from the corresponding author upon reasonable request.

Author contribution. S. Ji: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing—original draft. J. Mu: Conceptualization, Methodology, Supervision, Funding acquisition, Writing—review and editing. L. Guo: Methodology, Writing—review and editing. R. Suonan: Data curation, Investigation. Z. Zhao: Data curation, Formal analysis. Q. Yuan: Data curation, Writing—review and editing.