

A Wrinkle in Space-Time: Spatial machine learning uncovers stage-specific drivers and hidden hotspots of illegal wildlife trade

Hannah Murray^{*1}, Alexander Perigon², Bienvenu M. Takem Mbi², Anne Stéphanie Kobla³,
Roméo Tene³, Meredith L. Gore², Matthew H. Shirley⁴, and Bistra Dilkina¹

¹University of Southern California, Department of Computer Science, Los Angeles, California, USA

²University of Maryland, Department of Geographical Sciences, College Park, Maryland, USA

³Zoological Society of London, Cameroon Country Programme, Yaoundé, Cameroon

⁴Florida International University, Global Forensics and Justice Center, Miami, Florida, USA

* Corresponding Author Email: hannahmu@usc.edu

Abstract

The illegal wildlife trade (IWT) persists despite intervention, demanding new approaches to allocate limited prevention and enforcement resources more effectively. Spatial machine learning has been applied to IWT for this purpose, but mainly along one dimension – place. We extend this to simultaneously address place, species, and time, applying machine learning to participatory mapping data analyzed through a crime script framework to identify drivers and undetected hotspots across the preparation, pre-activity, activity, and post-activity stages of the illegal pangolin scale trade in Cameroon. The level of biodiversity, distance to international borders, and tree cover emerged as shared drivers across all stages, while hunting zones, food insecurity, and banking infrastructure distinguished individual stages. Each location was further assigned a relative risk score of trafficking involvement, surfacing high-risk candidate hotspots absent from existing records. These findings demonstrate that machine learning can extend IWT analysis beyond space to include time and species, sharpening targets for interventions like enforcement.

1 Introduction

Most research on the illegal wildlife trade (IWT), a globally distributed crime affecting the environment, has been descriptive and place-based: counting seizures, mapping known trade routes, and characterizing historical trends [52, 62]. While this body of work has been essential for establishing the scale and geographic scope of the problem, it has not yet translated into desired improvements in enforcement effectiveness or IWT prevention [2, 8, 43, 44]. The quality and accessibility of existing data sources compound this limitation. Law enforcement seizures and incident records are a biased sample of actual IWT activity, reflecting where enforcement agencies looked rather than where IWT occurs, and are not frequently documented or shared [31, 42]. This bottleneck is not necessarily due to a lack of descriptive knowledge about IWT, but a lack of predictive capacity – the ability to anticipate not only where IWT will occur, but also *when* in the criminal process it will occur, before it is detected, rather than cataloging where it has already been observed. Machine learning models trained on such data risk reproducing the same blind spots that already limit the effectiveness of enforcement [38]. Building predictive capacity for IWT requires moving beyond retrospective incident data toward approaches that characterize the ecological and human-driven (e.g., infrastructure, governance, and socioeconomic) conditions that make specific locations conducive to IWT activity and surface high-risk locations before they appear in any seizure record. This is the gap our study helps fill.

Effective intervention against IWT is an urgent global priority, called for by international bodies including the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), the United Nations Office on Drugs and Crime (UNODC), and the International Criminal Police Organization (INTERPOL), yet existing approaches remain insufficiently targeted to produce meaningful reductions in trade [45, 65]. If IWT activity leaves distinct spatial signatures at each stage of the criminal process, then interventions targeted to those stage-specific locations — rather than only where seizures happen to occur — may prove more effective at disrupting the trade [40]. Crime script analysis provides a structured framework for decomposing complex criminal acts into four stages (preparation, pre-activity, activity, and post-activity) and enabling identification of the distinct locations associated with each [6]. Preparation locations are those where traffickers meet to strategize and gather supplies, pre-activity where they move into position, activity where the illegal acts of focus are committed, and post-activity where goods are processed, transported, or laundered. Accounting for this spatiotemporal variation adds a temporal and functional dimension to wildlife crime analysis that purely place-based approaches cannot provide.

Here, we demonstrate a machine learning-based predictive framework able to account for such spatiotemporal variation at the scale of individual triangular cells of size $\sim 4.72 \text{ km}^2$ across an entire country. The raw input data consisted of the locations of perceived illegal pangolin scale trade activities derived via participatory mapping with local community members and experts [22]. This participatory mapping data was categorized as being part of one of four crime stages and analyzed using a geographic information system to produce grid layers of hot spots [23]. These processed data were then used as the primary input for our predictive analysis. This is a form of spatial intelligence that goes beyond what enforcement-based records alone can provide, including informal spaces and places that official data systematically misses [22, 61]. These data should be interpreted as locations of perceived risk and community expertise rather than verified incident locations, which we elaborate on in Section 4.1.

Machine learning methods have been applied to many spatially explicit prediction tasks in conservation [16, 55, 64]. The application to IWT is more limited, but a growing body of work has begun to demonstrate how machine learning models can build risk maps of potential undetected hotspots of IWT activity and derive the key characteristics of those locations that make them targets of wildlife crime. For example, Murray et al. [46] applied machine learning to predict airport involvement in IWT globally. They identified airports that did not have any reported seizure activity, yet their model predicted them to be involved in the trade. They also uncovered the structural and geographic characteristics of airports that make them more likely targets for IWT. While their approach represents progress in utilizing advanced analysis for IWT, it was constrained to a single component of the trafficking problem, which is the location or place where the transport of illegally traded wildlife occurs. It did not address species-specific outputs or the temporal variation in drivers of criminal activity, from preparation through post-activity. This study extends this framework by using machine learning to consider socioeconomic, infrastructure, governance, and species-specific ecological features (i.e., covariates) measured at a fine geographic resolution country-wide and disaggregate our modeling across different stages of the crime, from preparation to post-activity.

This study addresses two research objectives. First, we show that machine learning models trained on crime scripting data can generate continuous risk surfaces across space and the temporal stages of the crime, surfacing candidate undetected locations and their associated degree of risk, which is information not represented in existing expert knowledge or reported incident records. Second, we show how machine learning can unveil the differing spatial drivers of IWT across crime script stages, and highlight what ecological and human-driven characteristics distinguish each stage from the others. Together, these objectives advance a spatiotemporal approach to machine learning for IWT that simultaneously integrates species, place, and stage of the crime, a combination not previously attempted in the IWT literature. We apply our approach to one of the most heavily trafficked species, pangolins, in Cameroon, using participatory mapping data gathered from community members and experts on their observations of the illegal pangolin scale trade in the country.

Our models achieved strong predictive performance across all four crime script stages. Our analysis identified biodiversity levels, distance to international borders, and open tree cover as consistent crime drivers across all stages. Distance to villages and protected areas distinguished preparation, distance to food insecurity zones distinguished pre-activity, and distance to banking infrastructure distinguished post-activity, revealing distinct geographic signatures for each temporal stage. Hunting zone distance was also an important driver shared across preparation, pre-activity, and activity, most pronounced in preparation.

Our spatiotemporal risk assessment further generated risk scores for each crime script stage, identifying potential high risk locations outside known illegal pangolin scale trade zones, including a corridor in the Centre, Littoral, and West administrative regions of Cameroon, and revealing how illegal pangolin scale trade risk varies in magnitude and spatial extent across stages. Together, these results demonstrated that crime script-informed spatial machine learning can reveal new and unique details of IWT across space, time, and species.

2 Results & Discussion

We present high-level results for four spatial machine learning models based on Balanced Random Forests, one for each stage of the crime script (i.e., preparation, pre-activity, activity, and post-activity). All four models demonstrated strong performance, as evaluated under 5-fold spatial block cross-validation. AUC-ROC values ranged from 0.803 to 0.914, indicating good discrimination between hotspot and non-hotspot cells across all models [15]. The preparation stage model achieved the highest overall performance (AUC-ROC = 0.914), while the pre-activity stage model showed the lowest (AUC-ROC = 0.803). The lower pre-activity performance could be due to the relative sparsity of pre-activity hotspot labels (11.5% of cells) relative to the other crime stages, or that the predictors included in our modeling did not represent the pre-activity stage to the same degree. Full performance evaluation for each crime stage model can be viewed in Table S.1.

2.1 Hidden Hotspots: Spatiotemporal Risk Assessment Modeling

Cells predicted by the model to be involved in the illegal pangolin scale trade in at least one stage of the crime, but not identified as a hotspot in the participatory mapping data, were mapped and labeled with our model’s risk score that the respective cell was implicated in the illegal pangolin scale trade (Figure 1). It is important to note that the participatory mapping hotspots represent locations where there was strong agreement among study participants. Cells outside these areas are not necessarily confirmed to be free of illegal pangolin scale trade activity, but simply are locations that were either not identified during data collection or where participant agreement fell below the hotspot threshold (Section 4.1). Alongside our model predictions, we mapped reported locations of pangolin-related IWT activity from the civil society groups TRAFFIC, the Environmental Investigation Agency (EIA), and EAGLE, who mainly document seizure locations [10, 12, 63]. This analysis was built directly on the prediction-error framework established by Murray et al. [46], which demonstrated that machine learning models could identify airports implicated in IWT that had never appeared in seizure records. This study advanced Murray et al. [46] in an important way. The previous work could only identify which nodes in an IWT network were likely involved. The fine-scale spatial framework applied here allowed us to characterize where, within a country, undetected activity is occurring, at what stage of the crime, and what ecological and non-ecological conditions define those stage-specific locations. These predicted high risk areas are mappable, characterizable places with identifiable ecological, infrastructural, governance, and socioeconomic signatures.

Predicted high risk areas were concentrated in the southern part of the country, predominantly in the Centre, East, Littoral, and South administrative regions and overlapped with many of the reported seizure locations (Figure 1, see Figure S.1 for administrative boundaries [29]). Many of these cells fell immediately adjacent to or outside the boundaries of confirmed hotspot locations identified by the participatory mapping study participants. This suggests that the spatial extent of illegal pangolin scale trade exceeds what was documented through participatory mapping.

Of the 99,572 cells comprising the Cameroonian study area, 32,792 (32.9%) were predicted to be implicated in the illegal pangolin scale trade despite not being documented as a hotspot in the participatory mapping data in at least one stage-specific model. Among these predicted locations, 15,973 cells (48.7% of all predicted cells) were flagged by only a single stage-specific model, suggesting these locations have spatial conditions conducive to one particular stage of the crime but not others. A further 8,165 cells (24.9%) were flagged by two stage-specific models, 6,788 cells (20.7%) by three models, and 1,866 cells (5.7%) by all four stage-specific models simultaneously. This last group, cells predicted to be involved in illegal pangolin scale trade across every crime script stage but not implicated by the participatory mapping data, could be loca-

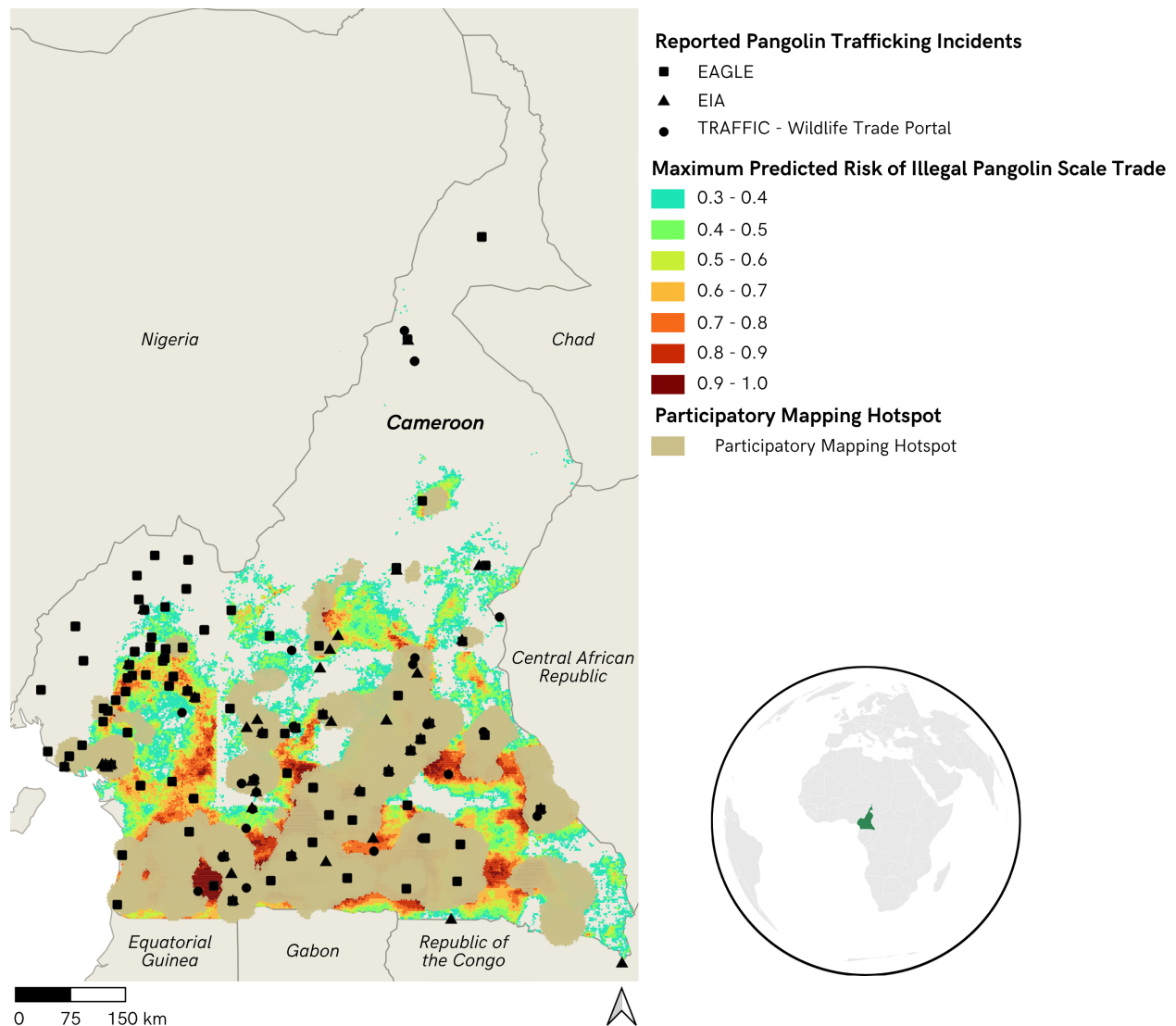


Figure 1: Maximum predicted risk of illegal pangolin scale trade across all four stage-specific pangolin trade crime script models for each cell in Cameroon. For each cell, the value shown represents the highest predicted risk assigned by any single stage-specific model, indicating locations where at least one stage of the trade crime script is predicted to be involved in the illegal pangolin scale trade. Only cells predicted to be involved by the model but not documented as a hotspot in the participatory mapping data are displayed, colored by maximum predicted risk of illegal pangolin scale trade (0.3-1.0), with warmer colors indicating higher relative risk scores. Participatory mapping hotspots (tan polygons) and reported trafficking incidents from EAGLE Network seizure records, the Environmental Investigation Agency (EIA), and the TRAFFIC Wildlife Trade Portal are overlaid for reference.

tions that may function as multi-purpose nodes within the illegal pangolin scale trade rather than locations implicated in a single stage of the crime.

Some meaningful variation was present in the spatial distribution of potential undocumented illegal pangolin scale trade at each crime script stage (Figure 2). The preparation-stage model (Figure 2.A) predicted high risk areas that were spatially concentrated along the edges and in connective corridors between areas identified through participatory mapping, rather than in entirely new geographic areas, expanding the perceived hotspot footprint by 53.4% (the smallest expansion of any stage-specific model). This implies that preparation activities, such as the sourcing of pangolins that are used for the illegal pangolin scale trade, are

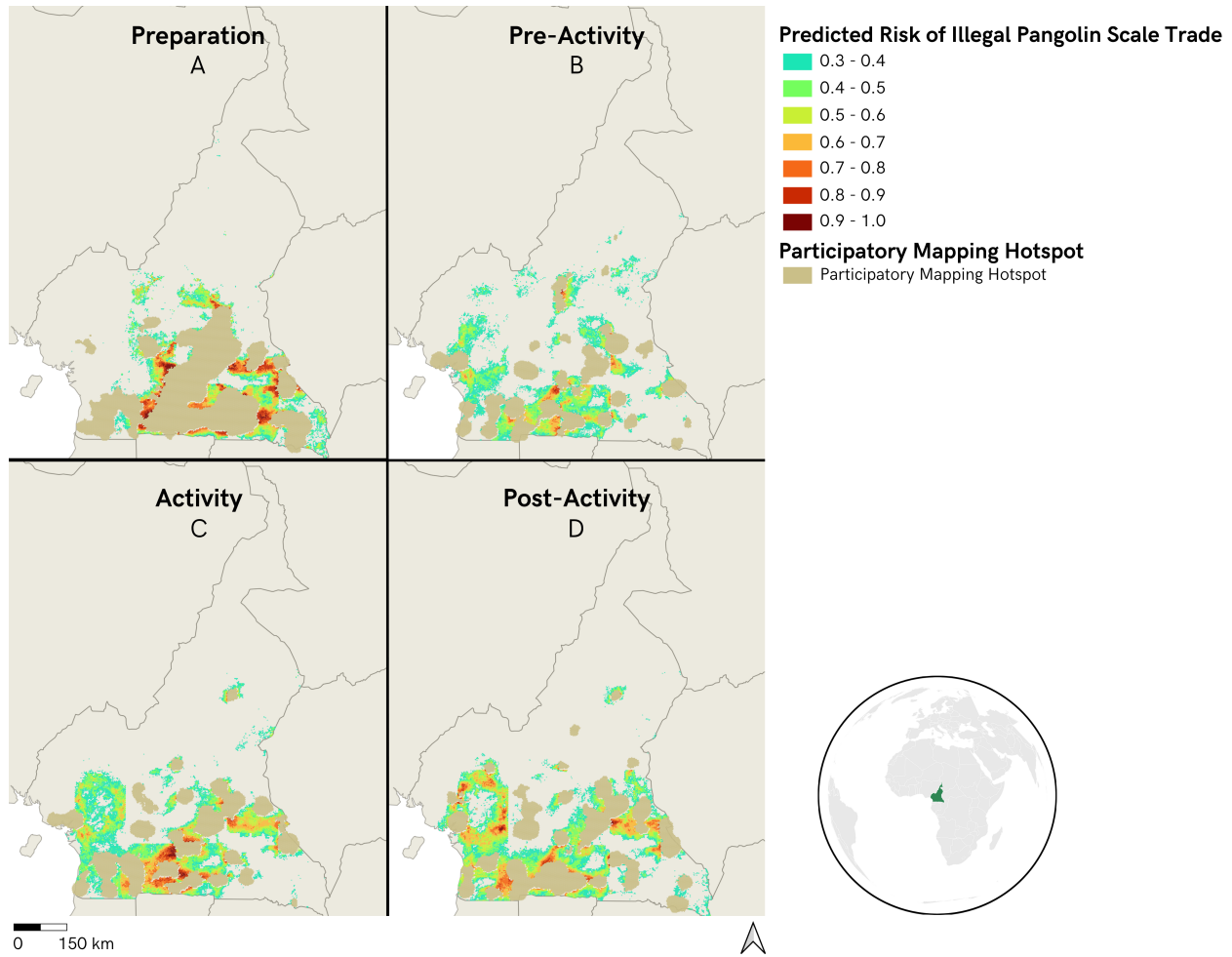


Figure 2: Crime-stage-specific risk mapping of predicted illegal pangolin scale trade: Predicted risk for each crime stage, including hotspots already identified by participatory mapping (tan polygons). Cells predicted as high risk by the model but not documented as hotspots in the processed participatory mapping data are colored by their relative model scores (0.3–1.0), with warmer colors indicating higher risk scores.

153 closely tied to the ecological conditions that define known hotspot areas. The predicted high risk areas here
 154 suggest that the preparation stage of the illegal pangolin scale trade operates across a more spatially contin-
 155 uous landscape than the discrete hotspot polygons derived from the participatory mapping study participant
 156 inputs imply, and that boundaries between participant-perceived activity spaces should not be considered
 157 truly empty space.

158 The pre-activity, activity, and post-activity models had different outcomes (Figure 2.B,C,D). In addi-
 159 tion to filling in connective gaps between documented hotspot areas from the participatory mapping data,
 160 these three stage-specific models consistently predicted a distinct cluster of locations in the western part of
 161 Cameroon, particularly in the Centre, Littoral, and West administrative regions, which overlaps with some of
 162 the documented seizures from EAGLE /seen in Figure 1. The proportional expansion of the known hotspot
 163 footprint is also considerably larger for these stages, at 124.8%, 126.0%, and 116.0% for pre-activity, activity,
 164 and post-activity, respectively, more than doubling the geographic extent identified by participatory map-
 165 ping alone. The presence of high risk cells not documented in the participatory mapping data across three
 166 functionally distinct crime script stages (pre-activity, activity, and post-activity) in the same geographic area
 167 in the western corridor may reflect the region supporting multiple stages of the trade network simultaneously.
 168 An alternative explanation is that participatory mapping data collection was less represented in this region,
 169 meaning fewer community members and experts documented these locations, and the corridor may represent

a data gap rather than a true gap in trade activity.

2.2 Uncovering Drivers of IWT across Time and Space

Feature importance analysis based on SHAP values was performed across all four models to identify the environmental, infrastructural, governance, and socioeconomic features most strongly associated with higher likelihood of illegal pangolin scale trade at each crime script stage (Figure 3) [39]. Notably, while ecological features, such as biodiversity levels and land cover type, appeared among the important predictors, many features with strong predictive influence fell outside of this category, suggesting the geography of the illegal pangolin scale trade is also shaped by human systems. Features related to distance to borders, enforcement presence, financial infrastructure, food insecurity, and land-use patterns collectively dominated the predictive signal across all four models, underscoring that illegal pangolin scale trade is human-driven [18]. The spatial variation of these features is visualized in Figure S.2.

Comparing the key drivers identified in each crime stage model, a subset of ecological and human-driven features emerged as strong predictors in every crime stage. High measurements of biodiversity were positively associated with higher likelihoods of illegal pangolin scale trade across all stages, reflecting the concentration of pangolin scale trade activity within the high-biodiversity forest zones of Cameroon. The distance to the closest open tree cover (areas with a relatively sparse or discontinuous tree canopy) and the distance to the nearest international border also appeared consistently, and cells further from these areas were more likely to be implicated in illegal pangolin scale trade by our models. These features are all accessible via openly available data sources. This has implications for the generalizability of our modeling framework. Even in countries with limited or no existing illegal pangolin scale trade detection data, these features could provide a spatially meaningful signal of illegal pangolin scale trade risk that could be used to prioritize prevention, intervention, or law enforcement efforts if illegal pangolin scale trade drivers are believed to be similar between Cameroon and the other countries.

Preparation The preparation stage model had the largest number of highly influential features (mean $|\text{SHAP}| \geq 0.020$, $n = 11$) and the highest proportion of features unique to this stage (Figure 3). This large number of predictors suggests that preparation areas for the pangolin trade in Cameroon are embedded in complex social-ecological environments that are simultaneously shaped by biodiversity, land use, population distribution, and proximity to administrative infrastructure. In addition to the features shared across all models, the preparation stage uniquely had biodiversity intactness, presence of meadow land use, presence and number of forests present, mean population density, distance to closest village, and distance to protected areas. The association with biodiversity intactness suggests that preparation activities concentrate in areas with relatively intact ecological communities, where pangolins could be more likely to be present, which is supported by bio-monitoring literature for the pangolin species that reside in Cameroon [25, 30, 33, 48]. The socioeconomic feature, human population density, the infrastructure feature, distance to closest village, and the ecological features, meadow land use (representing managed grassland typically used for agricultural activities such as hay production or grazing) and forest land use, may reflect the environments where hunters operate and animals are initially acquired before entering the trade network. The governance feature, distance to closest hunting zone, was also important for this stage, with over double the mean $|\text{SHAP}|$ values than it had for the other two crime stages that shared this feature, pre-activity and activity, suggesting that early-stage illegal pangolin scale trade activities (which include pangolin poaching) can occur in areas that allow hunting, even though pangolin hunting is illegal everywhere in Cameroon. This was also the only stage that did not have the infrastructure feature, distance to the closest parking facility, identified as influential. From a prevention standpoint, these results suggest that the involvement of a wide range of stakeholders and approaches beyond law enforcement can help with place-based efforts to reduce the conditions that make certain locations conducive to the earliest stages of the illegal pangolin scale trade. This includes community leaders, religious authorities, healthcare workers, teachers, sports coaches and players, local NGOs and governance structures working within their traditional approaches (e.g., demand reduction, alternative livelihood strategies).

Pre-Activity The pre-activity stage uniquely identified distance to food insecurity zones classified at severity level 2 as an important predictor, with cells closer to these zones more likely to be implicated

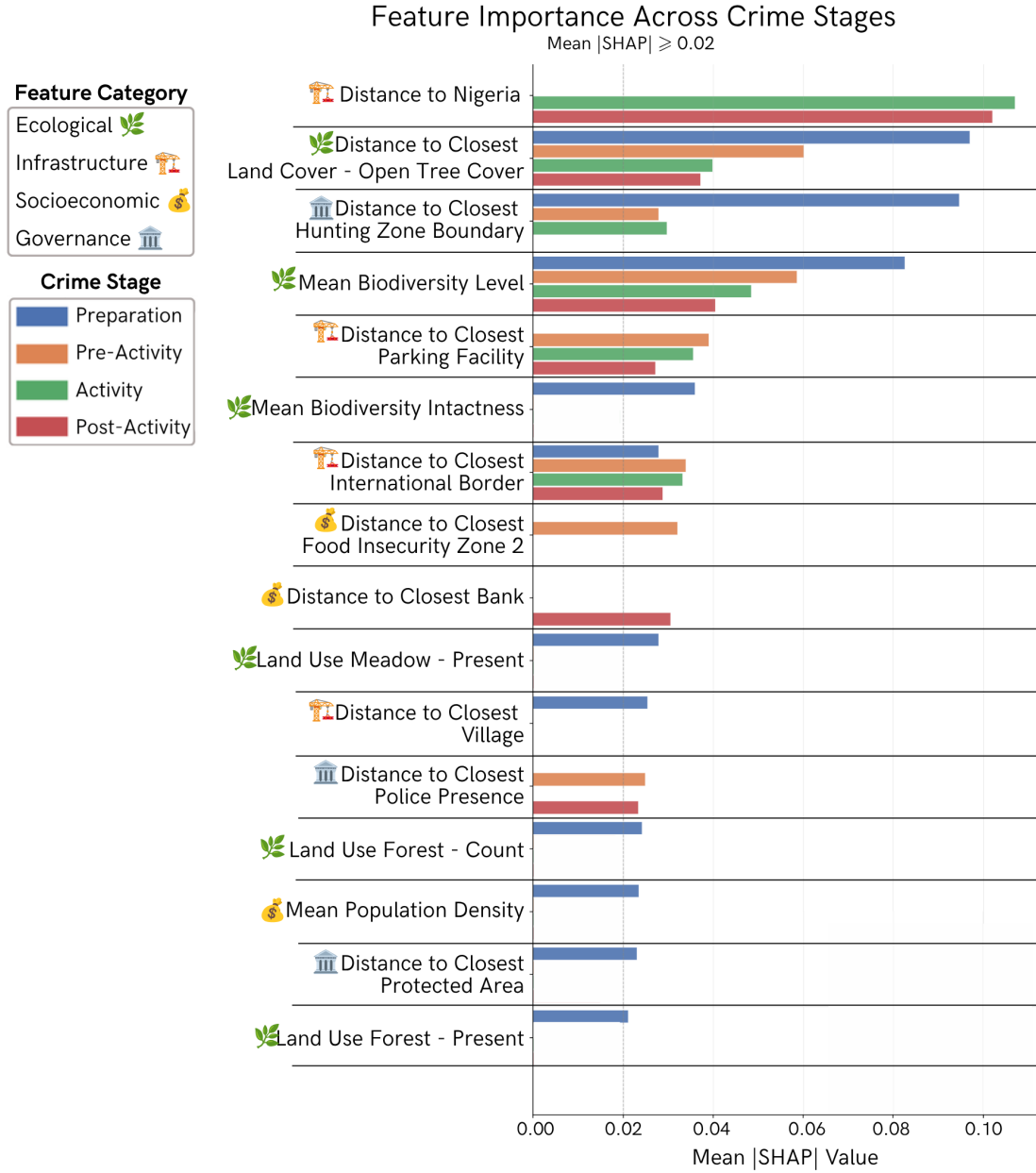


Figure 3: Feature Importance: Mean |SHAP| values for all features with mean |SHAP| ≥ 0.02 in at least one crime script stage model. Each bar represents the mean absolute SHAP value for a given feature in a given stage-specific model, with bar color indicating crime script stage (blue = preparation, orange = pre-activity, green = activity, red = post-activity). Features with no visible bar for a given stage were not selected by RFECV for that model or did not have a mean |SHAP| ≥ 0.02 . Features are ordered by maximum mean |SHAP| value across all stages. Icons to the left of each feature labels indicate feature category (🌿 = ecological, 🏠 = infrastructure, 💰 = socioeconomic, 🏛️ = governance), with the infrastructure, socioeconomic, and governance categories making up our human-driven features.

(Figure 3). While most of Cameroon fell in a severity level 1, this finding suggests that the closer a location is to a more food-insecure area, the more likely that location was to be involved in this stage of the illegal pangolin scale trade [17]. Exploring the relationship between the places and people where food insecurity occurs and conservation outcomes is an interesting area for future research, as food insecurity could be

driven by many factors, including poverty, geopolitical and human-wildlife conflict, or weather patterns. This stage also shared the governance feature, distance to hunting zones, with the preparation and activity stages, but had the lowest mean |SHAP| value across the three stages, suggesting it is less important than it is for the other two stages. The governance feature, distance to the closest police presence, was an important predictor in the pre-activity stage, with cells closer to documented police presence more likely to be implicated. This feature was shared with the post-activity stage and to a comparable degree, suggesting that both stages tend to occur in more densely populated, infrastructure-rich environments rather than remote areas — where proximity to law enforcement co-occurs with the urban and peri-urban conditions conducive to these activities rather than serving as a deterrent. Lastly, distance to the closest parking facility was more important for this stage than it was for the stages it was shared with, activity and post-activity.

Activity The activity stage had no features unique to it when compared with the other three stage-specific models (Figure 3). Distance to Nigeria was this model’s most important feature, which it shared with the post-activity stage. It also shared distance to parking facility with the pre-activity and post-activity stages. Lastly, it shared the distance to hunting zones feature with preparation and pre-activity. This suggests that the transaction of pangolin products likely occurs in locations that simultaneously serve as key locations for other stages of the crime. This means the spatial structure of the illegal pangolin scale trade is dictated by earlier or later stages of the crime. We know that when an activity stage requires specific spatial features/conditions, a crime is geographically bound. However, when it lacks unique spatial features, it is spatial opportunism [6, 60]. This tells us that offenders are acting wherever they encounter pangolins rather than seeking out specific geographic locations. This implies interdiction at the activity site will fail, as offenders can easily displace the activity stage to a different location. Additionally, hotspot policing relies on deploying patrols to the physical location where the crime occurs (e.g., activity location). If the activity stage has no unique spatial predictors, resources spent patrolling will be challenging to target and, therefore, yield low returns compared to other crime stages.

Post-Activity The post-activity stage uniquely identified distance to banking infrastructure as an important socioeconomic predictor, with cells closer to banks more likely to be implicated (Figure 3). The post-activity stage of wildlife trafficking encompasses the financial exchange of wildlife products following their acquisition and sale, including the laundering of proceeds from illegal sales [14]. Distance to closest bank appearing as a unique predictor at this stage is consistent with criminological research documenting that financial transactions and money laundering associated with wildlife crime tend to occur in urban and peri-urban environments with access to financial services [14]. Distance to police presence, shared with the pre-activity stage, and distance to parking facility, shared with both pre-activity and activity stages, remained important, reflecting that post-activity locations share the infrastructure-rich, accessible environments characteristic of the later stages of the crime script more broadly. Distance to Nigeria, shared with the activity stage, continued to be among the strongest predictors.

3 Conclusion

This study contributes to the growing field of GeoAI, which applies artificial intelligence to spatial data, by demonstrating that spatial machine learning when paired with spatialized crime scripting data derived via participatory mapping, can advance IWT research beyond retrospective mapping toward predictive capacity. By modeling preparation, pre-activity, activity, and post-activity stages of the illegal pangolin scale trade separately across Cameroon, we show that illegal pangolin scale trade risk is not a single, uniform geographic signal, but a set of distinct spatial processes, each shaped by a different mix of ecological, infrastructural, governance, and socioeconomic conditions. Biodiversity, distance to open tree cover, and distance to international borders emerged as consistent crime generators across all four stages, while features like distance to hunting zone, food insecurity, and banking infrastructure distinguished individual stages. This is evidence that the trade’s geography changes meaningfully as illicit goods move from sourcing to concealment.

The risk assessment modeling extends this contribution beyond feature interpretation. Building on the prediction-error framework established for global airport networks, this fine-scale spatial approach allows

potential undetected illegal pangolin scale trade activity to be not just flagged, but located, staged, and characterized. The consistent identification of a cluster of locations spanning three crime script stages in Cameroon’s Centre, Littoral, and West administrative regions, which was a corridor mainly absent from the processed participatory mapping data, illustrates the practical value of this approach. It surfaced specific, mappable locations that warrant field verification and targeted enforcement, intervention, and/or prevention attention.

Because the strongest and most consistent predictors across all stages rely on openly available environmental, governance, infrastructure, and socioeconomic data, this framework is not limited to contexts with rich trafficking-specific data. A model trained in Cameroon, where seizure records and participatory mapping data are available, could be applied to other countries as a starting point using only these public predictors, without requiring local trafficking data to fit the model. The resulting map would provide a spatially meaningful first estimate of illegal pangolin scale trade risk, provided the drivers of trafficking are assumed to be similar between Cameroon and the other location. Direct transferability and model calibration would still require evaluation using local data, but this has direct implications for resource-constrained enforcement agencies working across the broader Congo Basin and other pangolin range states.

More broadly, this work argues for treating the temporal crime script stages as a key dimension of spatial prediction in illegal wildlife trade research, alongside place and species. IWT is not a single event occurring at a single location. It is a distributed process unfolding across a landscape, with each stage leaving a different ecological, governance, socioeconomic, and infrastructural signature. Machine learning can help us derive these differences.

4 Methods

4.1 Data Gathering & Processing

Dependent Variables The dependent variable for each model is a binary indicator of whether a given cell was identified as a hotspot for that crime script stage, derived from spatialized participatory mapping data collected by a research team across Cameroon [22, 23, 61]. The participatory mapping dataset was created through data collection efforts targeting distinct participant groups: local community members ($n = 190$) living in and around protected areas with pangolin habitat in southern Cameroon, and subject matter experts ($n = 90$) with professional experience combating pangolin scale trade, including government agency specialists, national and international NGO representatives, and law enforcement personnel [22]. A total of 44 participatory mapping sessions were conducted from July 2024 to July 2025, with individuals participating across 27 communities [22]. Using a standardized base map of Cameroon that included political boundaries, roads, waterways, rail and sea and airports, and protected area boundaries, participants were asked to map 16 possible activities associated with pangolin hunting and trade (from pangolin harvest through scale storage, scale exchange, and wildlife markets) onto the base map [22]. These 16 activities span the four universal crime script stages (preparation, pre-activity, activity, and post-activity). Data were collected using projective questioning to reduce social desirability bias, in which participants described the behaviors of an external source rather than their own, encouraging honest reporting of knowledge about the trade [61]. Completed maps were digitally scanned and georeferenced in QGIS before being aggregated to the triangular grid used for modeling [22, 23].

To derive spatially consistent hotspot designations from the participatory mapping outputs, Gore et al. [23] conducted hotspot analysis in ArcGIS Pro using the **Optimized Hot Spot Analysis** tool and the **Getis-Ord Gi*** statistic [13]. This started with hexagonal grid cells, roughly of size $\sim 113.28 \text{ km}^2$. To mitigate spatial bias arising from the Modifiable Areal Unit Problem (MAUP), 21 unique analysis grids were generated through geometric translations of an initial hexagonal grid, and hotspot analysis was run iteratively across all 21 grid positions for each crime script stage. The spatial union of the 21 results was then computed for each stage using the Union geoprocessing tool in ArcGIS Pro, producing a triangular grid in which each $\sim 4.72 \text{ km}^2$ cell contains 21 confidence level determinations. A majority confidence level was assigned to each cell, and cells with a majority confidence level greater than or equal to 90% were coded as hotspots (binary = 1) for the corresponding crime script stage. All remaining cells were coded as non-hotspots (binary = 0). These triangular cells and their labels (1 or 0) were inputs to our machine learning modeling. For more details on the processing methods, refer to Gore et al. [23].

Independent Variables Candidate independent variables were selected based on expertise drawn from wildlife crime related literature regarding the ecological and non-ecological conditions associated with wildlife trafficking activity. Features were organized into four broad categories: infrastructure (e.g., roads, settlements, points of interest, law enforcement), socioeconomic (e.g., poverty, food insecurity), ecological (e.g., land cover, biodiversity levels, protected areas), and governance (e.g., community forests, hunting zones, mining permits). Infrastructure features were included based on evidence that trafficking networks exploit transportation corridors, urban centers, and border crossings to move wildlife products [36, 46, 67]. Economic features were included based on evidence that poverty, food insecurity, and livelihood vulnerability drive participation in wildlife trade [3, 9]. Environmental features were included based on the ecological requirement that pangolins must be present for poaching to occur [59]. Governance features were included based on evidence that formal and informal land governance structures shape both wildlife habitat and the opportunities and constraints for illegal extraction [5, 49, 54].

For each point layer (e.g., towns, law enforcement road checkpoints, airports, etc.), three features were generated per layer: (1) a count of features falling within each cell, (2) a binary presence/absence indicator, and (3) the Euclidean distance in meters from the cell centroid to the nearest point of that feature type. For each line layer (e.g., roads, waterways, oil and gas pipelines), the same three features were generated. Distance to nearest line was computed by first densifying each line geometry into evenly-spaced sample points at 500-meter intervals, then applying BallTree nearest-neighbor search from each cell centroid. For each polygon layer (e.g., protected areas, hunting zones, community forests, land use classes, food insecurity zones, informal settlements), count, presence, and distance features were also generated. Distance was computed from the cell centroid to the nearest polygon boundary, meaning cells whose centroids fell entirely inside a polygon received a distance of zero. For raster layers, two sampling strategies were employed depending on layer type. Categorical rasters (e.g., land cover and land use) were sampled at each cell centroid to extract the single pixel value. Continuous rasters (population density, biodiversity intactness, and biodiversity level) were sampled using the mean of all pixel values within each cell. Distance to the nearest international border and to each of Cameroon’s six bordering countries (Central African Republic, Chad, Equatorial Guinea, Gabon, Nigeria, and Republic of the Congo) was also computed as a set of candidate features using country boundary polygons from Natural Earth [47].

Prior to model fitting, two rounds of feature reduction were applied. First, pairwise Pearson correlations were computed between all candidate covariates. For each pair with $|r| > 0.70$, the feature less correlated with the dependent variable for that crime script stage was removed using a greedy elimination procedure, applied separately for each of the four binary outcome variables. Second, features where more than 95% of cells shared the same value were removed as near-zero variance predictors. Following these filters, the candidate feature set for each model comprised 48–53 features. A detailed list of the cleaned feature sets for each model can be found in the Supplementary Information files [7, 17, 19–21, 24, 27, 28, 32, 41, 47, 50, 51, 56–58, 66, 68].

4.2 Machine Learning Modeling

Model We utilized a Balanced Random Forest (BRF) binary classifier to predict whether each cell in Cameroon was implicated in each pangolin trade crime script stage or not [4]. Random Forest models are well documented for binary classification problems with class-imbalanced data, including for IWT classification [1, 4, 46]. The BRF model specifically addresses class imbalance, a common issue in wildlife crime-related analysis, where reported or experienced areas of wildlife trafficking activity usually constitute a minority of the study area. The model works by undersampling the majority class in each bootstrap sample, ensuring it is not biased toward predicting the more prevalent, non-hotspot class [4, 11]. Across our models, cells identified through the participatory mapping process comprised between 11.5% and 21.2% of the study area, making class balancing an essential component of our modeling framework. In total, we had four models, one for each of the individual crime stages (preparation, pre-activity, activity, and post-activity). All models were implemented using the `BalancedRandomForestClassifier` from the `imbalanced-learn` Python library [37], with 100 decision trees, which is the package’s default (`n_estimators = 100`). To implement class balancing consistent with the original Balanced Random Forest formulation, the majority class was undersampled with replacement to match the minority class size in each bootstrap sample (`sampling_strategy = 'all', replacement = True`).

Cross Validation Model performance was evaluated using 5-fold spatial block cross-validation (CV). Randomly assigning cells to folds would allow spatially adjacent cells with near-identical feature values to appear simultaneously in both training and test sets. This artificially inflated performance metrics due to spatial leakage. To combat this, we divided the study area into a 6×6 grid of 36 geographically contiguous blocks and randomly assigned blocks to one of five folds. For each fold, all cells in the held-out blocks were used as the test set, while the remaining cells served as the training set, ensuring that the model must generalize across geographic space rather than interpolating between spatially proximate training observations. Spatial fold assignments were implemented using the `PredefinedSplit` class from the `scikit-learn` Python library [53], with block boundaries defined by dividing the spatial extent of Cameroon into a regular 6×6 grid using cell centroid coordinates in the UTM Zone 33N coordinate reference system (EPSG:32633) [53].

Classification Threshold All models were evaluated using a classification threshold of 0.3, reduced from the default 0.5, to maximize the number of predicted high risk locations identified and the spatial intelligence available for enforcement and prevention targeting [34, 35]. This threshold places greater value on recall than precision, prioritizing the identification of as many potential high risk locations as possible over the certainty that each flagged location is a true hotspot (i.e., a location identified as a hotspot through the participatory mapping data). In the context of wildlife crime prevention, where the goal is to surface candidate locations for further investigation rather than to make definitive enforcement decisions, this trade-off is appropriate.

Feature Selection To identify the optimal subset of predictive features for each crime script stage, we applied Recursive Feature Elimination with Cross-Validation (RFECV), implemented using the `RFECV` class from the `scikit-learn` Python library [26]. RFECV iteratively removes the least important features, as ranked by the Random Forest’s internal impurity-based feature importance (mean decrease in Gini impurity), and evaluates model performance at each step using AUC-ROC as the scoring metric (`scoring = 'roc_auc'`). Features were eliminated one at a time (`step = 1`). This procedure was applied separately for each of the four binary classification targets, allowing each model to identify its own optimal feature subset rather than imposing a common feature set across all crime script stages. The number of features retained after RFECV ranged from 23 to 45, reflecting meaningful variation in the spatial complexity of each crime script stage (Table S.1).

4.3 Key Spatial Characteristics Predictive of IWT: SHAP Feature Importance

One output of our machine learning model is to identify which features or characteristics of locations influence the likelihood that they will be implicated in the illegal pangolin scale trade. To do this, we computed SHapley Additive exPlanations (SHAP) values across all five spatial cross-validation folds using the `TreeExplainer` class from the `shap` Python library [39]. SHAP values provide a game-theoretic measure of each feature’s marginal contribution to individual predictions, enabling both a global assessment of feature importance and an understanding of how each feature influences predictions for each individual point. SHAP has been used for other IWT classification studies [46]. To discuss features with the most meaningful predictive contribution, we applied a minimum mean $|\text{SHAP}|$ of 0.02, which retained the top 6 to 11 features per model for deeper analysis (Table S.1). Features below this threshold were present in the model but contributed less to predictions and were not discussed in detail.

4.4 Hidden Hotspots: Prediction-Error Analysis

Another output of the machine learning model allowed us to identify locations that are potentially high risk for IWT, despite not being documented or otherwise recorded. In machine learning, these are referred to as false positives. In our case, these are the cells predicted as high risk by the model that were not identified as such in the processed participatory mapping training data. False positives are usually treated as errors to be minimized. Following prior work on illicit wildlife crime, we reframed these false positives as potentially meaningful signals [46]. These errors may uncover areas where trafficking occurs but has not yet been observed, reported, or included in centralized databases.

To identify these false positives for each crime script stage’s model, each cell’s relative predicted risk for participating in that stage of the illegal pangolin scale trade was generated for each 4.72 km^2 cell in

Cameroon using the classification scores as output from the BRF models. Spatial block cross-validation was used to ensure that predictions for each cell were produced by a model that had not seen that cell during training. Cells can then be classified into three categories: *correct* (the processed participatory mapping data matched model prediction), *false negative* (processed participatory mapping data identified location as a hotspot, but the model predicted as not high risk), and *false positive* (participatory mapping data did not identify location as a hotspot but model predicted to be high risk). We analyzed false-positive cells in greater depth, as those represent potential “undetected hotspots”.

Acknowledgments

We wish to thank the following individuals: the community members and subject matter experts who were involved in the participatory mapping, Delong He, the Operation Pangolin team members, and the Cameroonian Ministry of Forestry and Wildlife (MINFOF). Lastly, we thank Allen Family Philanthropies for funding this research under the Operation Pangolin grant.

References

- [1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi:10.1023/A:1010933404324.
- [2] Pedro Cardoso, Kofi Amponsah-Mensah, João P. Barreiros, Jamie Bouhuys, Hubert Cheung, Alisa Davies, Sabrina Kumschick, Stuart J. Longhorn, Carlos A. Martínez-Muñoz, Thais Q. Morcatty, Gretchen Peters, William J. Ripple, Emmanuel Rivera-Téllez, Oliver C. Stringham, Adam Toomes, Patricia Tricorache, and Caroline S. Fukushima. Scientists’ warning to humanity on illegal or unsustainable wildlife trade. *Biological Conservation*, 263:109341, November 2021. ISSN 0006-3207. doi:10.1016/j.biocon.2021.109341.
- [3] Daniel W. S. Challender and Douglas C. MacMillan. Poaching is more than an enforcement problem. *Conservation Letters*, 7(5):484–494, 2014. ISSN 1755-263X. doi:10.1111/conl.12082.
- [4] Chao Chen, Andy Liaw, and Leo Breiman. Using random forest to learn imbalanced data. *University of California, Berkeley*, January 2004.
- [5] Chloe Clifford Astbury, Anastassia Demeshko, Eduardo Gallo-Cajiao, Ryan McLeod, Mary Wiktorowicz, Cécile Aenishaenslin, Katherine Cullerton, Kirsten M. Lee, Arne Ruckert, A. M. Viens, Peter Tsasis, and Tarra L. Penney. Governance of the wildlife trade and the prevention of emerging zoonoses: a mixed methods network analysis of transnational organisations, silos, and power dynamics. *Globalization and Health*, 20:49, June 2024. ISSN 1744-8603. doi:10.1186/s12992-024-01055-7.
- [6] Derek B. Cornish. The procedural analysis of offending and its relevance for situational prevention. *Crime Prevention Studies*, 3(1):151–196, 1994.
- [7] A. De Palma, S. Contu, G. E. Thomas, C. Duffin, S. Nix, and A. Purvis. The Biodiversity Intactness Index developed by The Natural History Museum, London, v2.1.1 (Open Access, Limited Release), 2024. URL <https://data.nhm.ac.uk/dataset/bii-developed-by-nhm-v2-1-1-limited-release/resource/c4c281c4-befa-4e1b-a162-ba2f25e5ae82>.
- [8] Enrico Di Minin, Neil D’Cruze, Angie Elwin, Annette Hübschle, Riley Kemp, Ritwik Kulkarni, Haozhong Si, Andrea Soriano-Redondo, and Miquel Torrents-Ticó. Global trends and impacts of the online illegal wildlife trade. *Nature Reviews Biodiversity*, page 1–12, August 2026. ISSN 3005-0677. doi:10.1038/s44358-026-00192-8.
- [9] Rosaleen Duffy, Freya A. V. St John, Bram Büscher, and Dan Brockington. Toward a new understanding of the links between poverty and illegal wildlife hunting. *Conservation Biology*, 30(1):14–22, 2016. ISSN 1523-1739. doi:10.1111/cobi.12622.

- [10] EAGLE (Eco Activists for Governance and Law Enforcement). Pangolin seizure records for Central Africa [unpublished dataset]. Provided via personal communication to [REDACTED], [REDACTED], for [REDACTED], 2023.
- [11] Bradley Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, New York, May 1994. ISBN 978-0-429-24659-3. doi:10.1201/9780429246593.
- [12] Environmental Investigation Agency. Pangolin incident data for Cameroon, 2015–2023 [unpublished dataset], 2023.
- [13] Esri. ArcGIS Pro (version 3.0), 2022. URL <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>.
- [14] FATF. Money laundering and the illegal wildlife trade. Technical report, Financial Action Task Force (FATF), Paris, France, 2020. URL <https://www.fatf-gafi.org/publications/methodandtrends/documents/money-laundering-illegal-wildlife-trade.html>.
- [15] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ISSN 0167-8655. doi:10.1016/j.patrec.2005.10.010.
- [16] Paul Fergus, Carl Chalmers, Steven Longmore, and Serge Wich. Harnessing artificial intelligence for wildlife conservation. *Conservation*, 4(4):685–702, December 2024. ISSN 2673-7159. doi:10.3390/conservation4040041.
- [17] FEWS NET (Famine Early Warning Systems Network). Cameroon acute food insecurity classification, April 2026 – September 2026, 2026. URL <https://fews.net/cameroon-acute-food-insecurity-classification-april-2026-september-2026>.
- [18] Caroline S. Fukushima, Patricia Tricorache, Adam Toomes, Oliver C. Stringham, Emmanuel Rivera-Téllez, William J. Ripple, Gretchen Peters, Ronald I. Orenstein, Thais Q. Morcatty, Stuart J. Longhorn, Chien Lee, Sabrina Kumschick, Marco Antonio de Freitas, Rosaleen V. Duffy, Alisa Davies, Hubert Cheung, Susan M. Cheyne, Jamie Bouhuys, João P. Barreiros, Kofi Amponsah-Mensah, and Pedro Cardoso. Challenges and perspectives on tackling illegal or unsustainable wildlife trade. *Biological Conservation*, 263:109342, November 2021. ISSN 0006-3207. doi:10.1016/j.biocon.2021.109342.
- [19] Geofabrik GmbH and OpenStreetMap Contributors. Cameroon OpenStreetMap data extract, 2024. URL <https://download.geofabrik.de/africa/cameroon.html>. Data © OpenStreetMap contributors, available under the Open Database License (ODbL): <https://www.openstreetmap.org/copyright>.
- [20] Global Forest Watch. Cameroon community forests, 2024. URL <https://data.globalforestwatch.org/datasets/gfw::cameroon-community-forests/explore>.
- [21] Global Forest Watch. Cameroon mining permits, 2024. URL <https://data.globalforestwatch.org/datasets/cameroon-mining-permits>.
- [22] Meredith L. Gore, Anne Stephanie Kobla, Bienvenu Takem Mbi, Hannah Murray, Romeo Tene, Bistra Dilkina, Ashleigh N. Price, and Matthew H. Shirley. Descriptor: Participatory mapping of trade in pangolin scales and other wildlife in cameroon dataset (pmpwsowc). *IEEE Data Descriptions*, page 1–5, 2026. ISSN 2995-4274. doi:10.1109/IEEEDATA.2026.3733730.
- [23] Meredith L. Gore, Alexander Perigon, Hannah Murray, Bienvenu Takem Mbi, Anne Stephanie Kobla, Carla Louise Mousset Mounbolou, Stephanie Bourgeois, Matthew H. Shirley, and Bistra Dilkina. Wildlife crime moves in spatial pulses. *EcoEvoRxiv*, 2026. URL <https://ecoevorxiv.org/repository/view/15508/>. Preprint.
- [24] H S Grantham, A Shapiro, D Bonfils, V Gond, E Goldman, F Maisels, A J Plumptre, T Rayden, J G Robinson, S Strindberg, E Stokes, A I T T Tulloch, J E M Watson, L Williams, and O Rickenbach. Spatial priorities for conserving the most intact biodiverse forests within central africa. *Environmental Research Letters*, 15(9):0940b5, September 2020. ISSN 1748-9326. doi:10.1088/1748-9326/ab9fae.

- [25] Maja Gudehus, Darren W. Pietersen, Michael Hoffmann, Rod Cassidy, Tamar Cassidy, Olufemi Sodeinde, Juan Lapuente, Brou Guy-Mathieu Assovi, and Matthew H. Shirley. *Black-bellied pangolin Phataginus tetradactyla (Linnaeus, 1766)*, page 123–138. Biodiversity of World: Conservation from Genes to Landscapes. Academic Press, January 2020. ISBN 978-0-12-815507-3. doi:10.1016/B978-0-12-815507-3.00008-3. URL <https://www.sciencedirect.com/science/article/pii/B9780128155073000083>.
- [26] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422, January 2002. ISSN 1573-0565. doi:10.1023/A:1012487302797.
- [27] Matthew C Hansen, Peter V Potapov, Amy H Pickens, Alexandra Tyukavina, Andres Hernandez-Serna, Viviana Zalles, Svetlana Turubanova, Indrani Kommareddy, Steve V Stehman, Xiao-Peng Song, and Anil Kommareddy. Global land use extent and dispersion within natural land cover using landsat data. *Environmental Research Letters*, 17(3):034050, March 2022. ISSN 1748-9326. doi:10.1088/1748-9326/ac46ec.
- [28] HeiGIT (Heidelberg Institute for Geoinformation Technology). Cameroon road surface data, 2024. URL <https://data.humdata.org/dataset/cameroon-road-surface-data>.
- [29] Robert J. Hijmans. First-level administrative divisions, Cameroon, 2015, 2015. URL <https://pur1.stanford.edu/db005yd3959>. Global Administrative Areas, Version 2.8. Polygon shapefile.
- [30] Michael Hoffmann, Stuart Nixon, Daniel Alempijevic, Sam Ayebare, Tom Bruce, Tim R. B. Davenport, John Hart, Terese Hart, Martin Hega, Fiona Maisels, David Mills, and Constant Ndjassi. *Giant pangolin Smutsia gigantea (Illiger, 1815)*, page 157–173. Biodiversity of World: Conservation from Genes to Landscapes. Academic Press, January 2020. ISBN 978-0-12-815507-3. doi:10.1016/B978-0-12-815507-3.00010-1. URL <https://www.sciencedirect.com/science/article/pii/B9780128155073000101>.
- [31] Alice C. Hughes, Benjamin Michael Marshall, David Edwards, Daniel Challender, Julie L. Lockwood, Oscar Morton, Kelly Malsch, Caroline Sayuri Fukushima, Amy Hinsley, Vincent Nijmans, Sharon Baruch-Mordo, Michael Tlusty, Rodrigo Oyanedel, Jacqueline Jurgens, Nazli Demirel, Neil Burgess, Evan Eskew, Shai Meiri, Samuel Pironon, Alyson T. Pavitt, Oliver Tallowin, and Molly Brown. The true scope of global wildlife trade is obscured by data gaps. January 2026. URL <https://ecoevorxiv.org/repository/view/11272/>.
- [32] Humanitarian OpenStreetMap Team (HOT). Cameroon OpenStreetMap exports, 2024. URL <https://data.humdata.org/organization/hot>. Data © OpenStreetMap contributors, available under the Open Database License (ODbL): <https://www.openstreetmap.org/copyright>.
- [33] Raymond Jansen, Olufemi Sodeinde, Durojaye Soewu, Darren W. Pietersen, Daniel Alempijevic, and Daniel J. Ingram. *White-bellied pangolin Phataginus tricuspis (Rafinesque, 1820)*, page 139–156. Biodiversity of World: Conservation from Genes to Landscapes. Academic Press, January 2020. ISBN 978-0-12-815507-3. doi:10.1016/B978-0-12-815507-3.00009-5. URL <https://www.sciencedirect.com/science/article/pii/B9780128155073000095>.
- [34] Nathalie Japkowicz and Mohak Shah. In *Evaluating Learning Algorithms: A Classification Perspective*, chapter Performance Measures II, page 111–160. Cambridge University Press, Cambridge, 2011. ISBN 978-0-511-92180-3. doi:10.1017/CBO9780511921803.005. URL <https://www.cambridge.org/core/books/evaluating-learning-algorithms/performance-measures-ii/5619AD7A547F5BA3595A8313A203B244>.
- [35] Max Kuhn and Kjell Johnson. Measuring performance in classification models. In *Applied Predictive Modeling*, page 247–273. Springer, New York, NY, 2013. ISBN 978-1-4614-6849-3.
- [36] William E. Laurance, Barbara M. Croes, Landry Tchignoumba, Sally A. Lahm, Alfonso Alonso, Michelle E. Lee, Patrick Campbell, and Claude Ondzeano. Impacts of roads and hunting on central african rainforest mammals. *Conservation Biology: The Journal of the Society for Conservation Biology*, 20(4):1251–1261, August 2006. ISSN 0888-8892. doi:10.1111/j.1523-1739.2006.00420.x.

- [37] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*, 18(1):559–563, January 2017. ISSN 1532-4435.
- [38] Kristian Lum and William Isaac. To predict and serve? *Significance*, 13(5):14–19, 2016. ISSN 1740-9713. doi:10.1111/j.1740-9713.2016.00960.x.
- [39] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 4768–4777, Red Hook, NY, USA, December 2017. Curran Associates Inc. ISBN 978-1-5108-6096-4. URL <https://dl.acm.org/doi/10.5555/3295222.3295230>.
- [40] Nicholas R. Magliocca, Ridvan Aksu, Kendra McSweeney, Hashir Tanveer, Kevin M. Curtin, and Burcu Keskin. Towards spatially disaggregated cocaine supply chain modeling. *Socio-Economic Planning Sciences*, 96:102086, December 2024. ISSN 0038-0121. doi:10.1016/j.seps.2024.102086.
- [41] N. Marchio and L. M. A. Bettencourt. Million Neighborhoods Africa database, 2023.
- [42] Benjamin Michael Marshall, Aubrey L. Alamshah, Pedro Cardoso, Phillip Cassey, Sebastian Chekunov, Evan A. Eskew, Caroline S. Fukushima, Pablo García-Díaz, Meredith L. Gore, Julie L. Lockwood, Andrew L. Rhyne, James S. Sinclair, Colin Thomas Strine, Oliver C. Stringham, Michael F. Tlusty, Jose W. Valdez, Freyja Watters, and Alice C. Hughes. The magnitude of legal wildlife trade and implications for species survival. *Proceedings of the National Academy of Sciences of the United States of America*, 122(2):e2410774121, January 2025. ISSN 1091-6490. doi:10.1073/pnas.2410774121.
- [43] Dominic Meeks, Oscar Morton, and David P. Edwards. Wildlife farming: Balancing economic and conservation interests in the face of illegal wildlife trade. *People and Nature*, 6(2):446–457, 2024. ISSN 2575-8314. doi:10.1002/pan3.10588.
- [44] E. J. Milner-Gulland. Documenting and tackling the illegal wildlife trade: change and continuity over 40 years. *Oryx*, 52(4):597–598, October 2018. ISSN 0030-6053, 1365-3008. doi:10.1017/S0030605318001047.
- [45] Annika Mozer and Stefan Prost. An introduction to illegal wildlife trade and its effects on biodiversity and society. *Forensic Science International: Animals and Environments*, 3:100064, December 2023. ISSN 2666-9374. doi:10.1016/j.fsiae.2023.100064.
- [46] Hannah Murray, Meredith L. Gore, and Bistra Dilkina. Encoding and decoding illegal wildlife trade networks reveals key airport characteristics and undetected hotspots. *Communications Earth & Environment*, 6(1):399, May 2025. ISSN 2662-4435. doi:10.1038/s43247-025-02371-5.
- [47] Natural Earth. Natural Earth: Free vector and raster map data [dataset], 2024. URL <https://www.naturalearthdata.com/downloads/>.
- [48] Tim Newbold, Lawrence N. Hudson, Andrew P. Arnell, Sara Contu, Adriana De Palma, Simon Ferrier, Samantha L. L. Hill, Andrew J. Hoskins, Igor Lysenko, Helen R. P. Phillips, Victoria J. Burton, Charlotte W. T. Chng, Susan Emerson, Di Gao, Gwilym Pask-Hale, Jon Hutton, Martin Jung, Kattia Sanchez-Ortiz, Benno I. Simmons, Sarah Whitmee, Hanbin Zhang, Jörn P. W. Scharlemann, and Andy Purvis. Has land use pushed terrestrial biodiversity beyond the planetary boundary? a global assessment. *Science*, 353(6296):288–291, July 2016. doi:10.1126/science.aaf2201.
- [49] OECD. *Strengthening Governance and Reducing Corruption Risks to Tackle Illegal Wildlife Trade: Lessons from East and Southern Africa*. Illicit Trade. OECD, October 2018. ISBN 978-92-64-30649-3. doi:10.1787/9789264306509-en.
- [50] OFAC (Observatoire des Forêts d’Afrique Centrale). Données SIG [GIS Data Library], 2024. URL https://www.observatoire-comifac.net/monitoring_system/gis_library.
- [51] OpenStreetMap contributors. OpenStreetMap data for Cameroon (amenity=police) [data set]. retrieved via Overpass API, 2026. URL <https://www.openstreetmap.org>.

- [52] Kumar Paudel, Amy Hinsley, Diogo Veríssimo, and Ej Milner-Gulland. Evaluating the reliability of media reports for gathering information about illegal wildlife trade seizures. *PeerJ*, 10:e13156, April 2022. ISSN 2167-8359. doi:10.7717/peerj.13156.
- [53] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12(null):2825–2830, November 2011. ISSN 1532-4435.
- [54] J. R. Poulsen, C. J. Clark, G. Mavah, and P. W. Elkan. Bushmeat supply and consumption in a tropical logging concession in northern congo. *Conservation Biology: The Journal of the Society for Conservation Biology*, 23(6):1597–1608, December 2009. ISSN 1523-1739. doi:10.1111/j.1523-1739.2009.01251.x.
- [55] Yogita Yashveer Raghav, Aditi Chauhan, Pallavi Pandey, and Surbhi Bhatia Khan. *AI and Machine Learning Techniques for Wildlife Conservation*. IGI Global Scientific Publishing, January 2025. ISBN 979-8-3693-6935-7. URL <https://www.igi-global.com/book/machine-learning-techniques-wildlife-conservation/346237>.
- [56] Sukanya Randhawa, Eren Aygün, Guntaj Randhawa, Benjamin Herfort, Sven Lautenbach, and Alexander Zipf. Paved or unpaved? A deep learning derived road surface global dataset from Mapillary street-view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024. URL <https://www.sciencedirect.com/science/article/pii/S0924271625000784>.
- [57] F. Schug, A. Bar-Massada, A. R. Carlson, H. Cox, T. J. Hawbaker, D. Helmers, P. Hostert, D. Kaim, N. K. Kasraee, S. Martinuzzi, M. H. Mockrin, K. A. Pfoch, and V. C. Radeloff. The global wildland–urban interface. *Nature*, 621(7977):94–99, 2023. doi:10.1038/s41586-023-06320-0.
- [58] F. Schug et al. Map of the Global Wildland-Urban Interface, version 1.1, 2023.
- [59] Franklin T. Simo, Ghislain F. Difouo, Sévilor Kekeunou, Ichu G. Ichu, Daniel J. Ingram, and David Olson. Pangolin hunting and trafficking in the forest–savannah transition area of cameroon. *Oryx*, 57(6):704–713, November 2023. ISSN 0030-6053, 1365-3008. doi:10.1017/S0030605322001429.
- [60] G. Stummvoll. Environmental criminology and crime analysis. *Crime Prevention and Community Safety*, 11:144–146, 2009. doi:10.1057/cpcs.2008.22.
- [61] Bienvenu M. Takem Mbi, Anne Stéphanie Kobla, Roméo Tene, Matthew H. Shirley, and Meredith L. Gore. Using community-based crime scripting to elucidate pangolin scale trafficking. *Ambio*, July 2026. ISSN 1654-7209. doi:10.1007/s13280-026-02437-4. URL <https://doi.org/10.1007/s13280-026-02437-4>.
- [62] Jia Hao Tow, William S. Symes, and Luis Roman Carrasco. Economic value of illegal wildlife trade entering the usa. *PLOS ONE*, 16(10):e0258523, October 2021. ISSN 1932-6203. doi:10.1371/journal.pone.0258523.
- [63] TRAFFIC. Wildlife Trade Portal, 2026. URL <https://www.wildlifetradeportal.org/>.
- [64] Devis Tuia, Benjamin Kellenberger, Sara Beery, Blair R. Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W. Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D. Couzin, Grant van Horn, Margaret C. Crofoot, Charles V. Stewart, and Tanya Berger-Wolf. Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1):792, February 2022. ISSN 2041-1723. doi:10.1038/s41467-022-27980-y.
- [65] United Nations Office on Drugs and Crime. *World Wildlife Crime Report 2024: Trafficking in Protected Species*. Stylus Publishing, LLC, Vienna, 2024. URL <https://www.unodc.org/unodc/en/data-and-analysis/wildlife.html>.

- 648 [66] WorldPop (www.worldpop.org — School of Geography and Environmental Science, University of
649 Southampton). Africa continental population datasets, version 2.0, 2020. URL [https://hub.worldpop.](https://hub.worldpop.org/geodata/summary?id=139)
650 [org/geodata/summary?id=139](https://hub.worldpop.org/geodata/summary?id=139).
- 651 [67] Stanislas Zanzo, Sylvestre C. A. M. Djagoun, Fortuné A. Azihou, Bruno Djossa, Brice Sinsin, and
652 Philippe Gaubert. Ethnozoological and commercial drivers of the pangolin trade in benin. *Journal*
653 *of Ethnobiology and Ethnomedicine*, 17(1):18, March 2021. ISSN 1746-4269. doi:10.1186/s13002-021-
654 00446-z.
- 655 [68] J. Zwerts, F. A. E. de Zwart, K. W. S. Norbart, M. Kruise, and L. M. Oudejans. Targeted FSC
656 certification to increase conservation impact in the Congo Basin [data set], 2026.

Supplementary Information

Metric	Prep	Pre-A	A	Post-A
<i>Model overview</i>				
% Hotspot cells	21.2%	11.5%	13.4%	15.3%
N features (RFE)	23	45	45	44
N features ($ \text{SHAP} \geq 0.020$)	11	7	6	7
<i>Overall Performance</i>				
Accuracy	0.837	0.806	0.795	0.785
Macro F1	0.780	0.643	0.681	0.690
Weighted F1	0.845	0.829	0.820	0.809
AUC-ROC	0.914	0.803	0.850	0.861
<i>Hotspot Class</i>				
Precision	0.590	0.312	0.368	0.395
Recall	0.768	0.566	0.733	0.757
F1-score	0.667	0.402	0.490	0.519
<i>Non-Hotspot Class</i>				
Precision	0.932	0.937	0.951	0.947
Recall	0.856	0.838	0.805	0.790
F1-score	0.892	0.885	0.872	0.861

Table S.1: Model performance summary across the four stage-specific binary classification models, evaluated using 5-fold spatial block cross-validation (36 geographic blocks, classification threshold = 0.3). Prep = preparation stage, Pre-A = pre-activity stage, A = activity stage, Post-A = post-activity stage. N features (RFE) = number of features retained after recursive feature elimination with cross-validation. N features ($|\text{SHAP}| \geq 0.020$) = number of features with mean $|\text{SHAP}|$ value ≥ 0.020 .

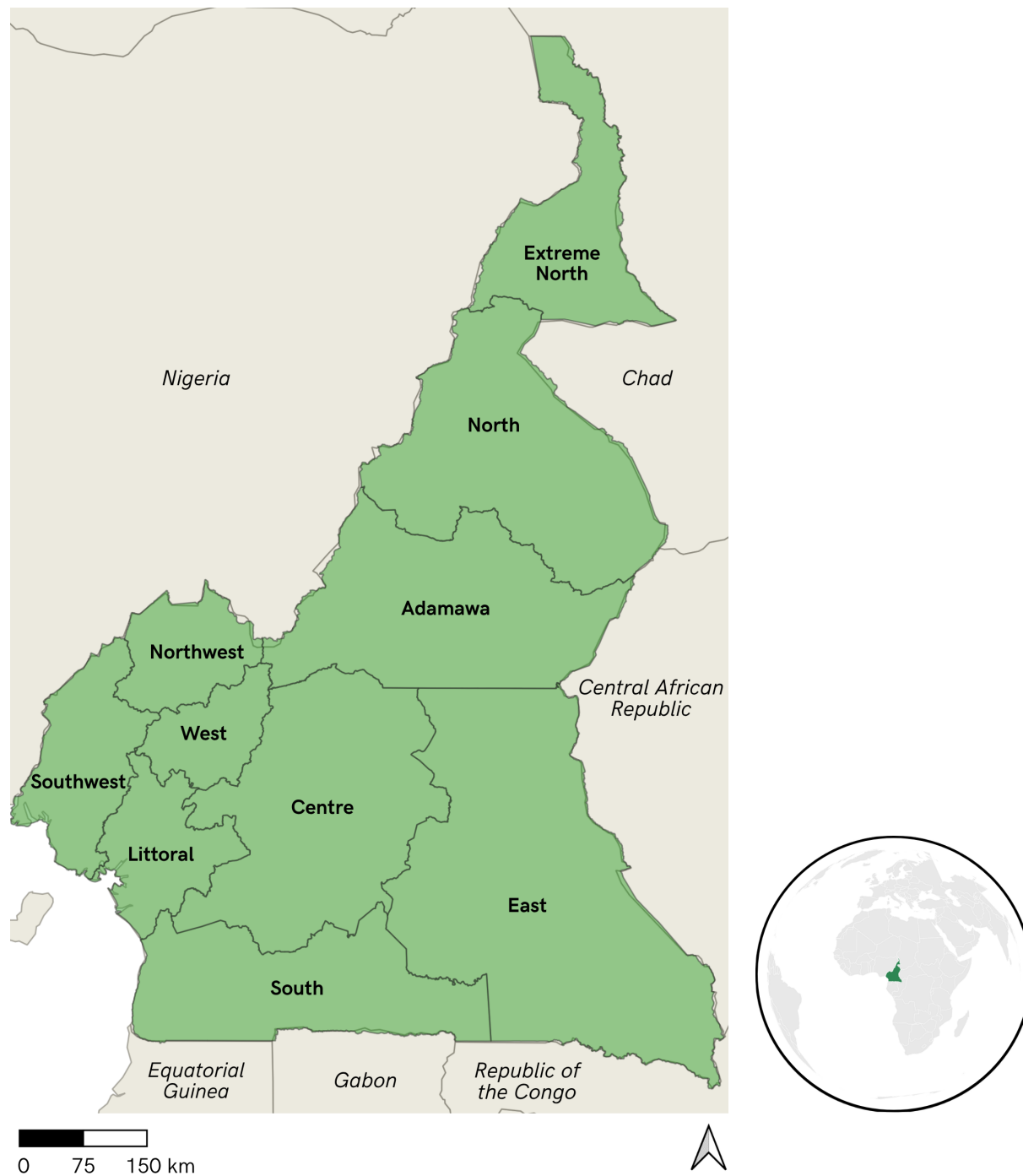


Figure S.1: Cameroon's ten administrative regions, with bordering countries (Central African Republic, Chad, Equatorial Guinea, Gabon, Nigeria, and Republic of the Congo) labeled.

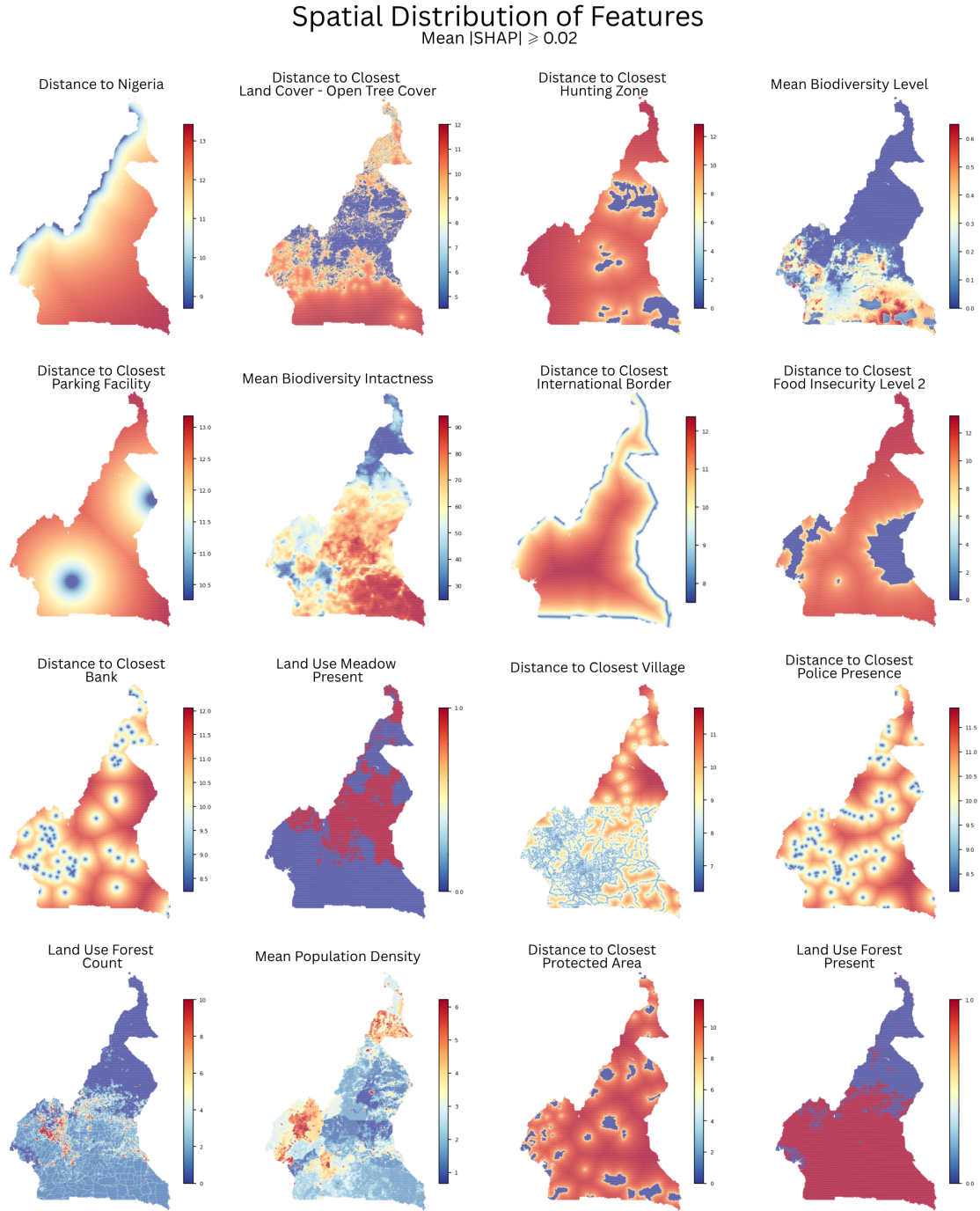


Figure S.2: Spatial distribution of the 16 features with mean |SHAP| value ≥ 0.02 in at least one crime script stage model across Cameroon, ordered from top to bottom and left to right from highest mean |SHAP| to lowest. Each panel shows the value of a single feature for every triangular grid cell within the country, colored on a blue-to-red gradient from low (blue) to high (red) values. Distance features and population density are displayed on a logarithmic scale to better visualize the spatial gradient. These maps illustrate the geographic variation in the ecological, infrastructural, socioeconomic, and governance conditions that the models identify as most strongly associated with the illegal pangolin scale trade in Cameroon across crime script stages.