International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Adaptive Urban Biodiversity Monitoring System: A Case Study of Mangalavanam Bird Sanctuary

Aneena Jacob[1], Dr. Shajeni Justin[2]

[1]Department of Computer Science, Sacred Heart College, Kochi
[2]Assistant Professor, Department of Computer Science, Sacred Heart College, Kochi,

Abstract
- Wetlands in urban areas are under immense pressure from rapid urbanization and human activities. This is a major problem in maintaining biodiversity in tropical cities. Our research proposes a step-by-step guide on how to identify where birds are likely to be at home within Kochi's Mangalavanam Bird Sanctuary using a mixed model approach. We used existing data from eBird Citizen Science and integrated environmental variables from Sentinel-2 satellite data and VIIRS Day/Night Band night light data from Google Earth Engine. To identify where birds are likely to be without assuming that we observe them every time, we used a hierarchical Bayesian occupancy model. We classified birds into functional groups such as birds living in wetlands, birds living in forests, and birds living in urban areas to understand how these groups react to environmental variables. Our research indicates that birds react differently to environmental variables and highlights areas within the sanctuary that need urgent conservation. We have also developed an interactive web application to display bird habitats and identify areas of high conservation value.

Keywords: Occupancy modeling, urban ecology, citizen science, remote sensing, conservation informatics, Bayesian inference, habitat suitability, biodiversity monitoring

1. Introduction

1.1 Urban Biodiversity Significance

In particular, urban biodiversity is most relevant to the extent to which cities expand into tropical Asia, where ecosystems are already pushed to the limits. As the city extends outward, natural ecosystems become smaller, while new urban ecosystems grow with the city's expansion, carrying with them the remnants of life that need to be protected. Cities' wetlands located inside the city's fringes are crucial biodiversity sites. These not only provide flood control services, groundwater recharge, stabilization of the local climate, but also support bird life despite the constant sounds of human activities. Mangroves, for instance, are a type of ecosystem dominated by trees that thrive in the intertidal zone due to their tolerance of high salinity levels. These areas are home to diverse species of animals that thrive at the boundary between land and sea life zones. A good example of an urban biodiversity hotspot is the Mangalavanam Bird Sanctuary located in the heart of Kochi City. The 7.86-hectare wetland dominated by Avicennia and Rhizophora species is completely surrounded by dense residential and commercial areas. Despite the isolation from other natural ecosystems, the area harbors 194 bird species. Indeed, the biodiversity hotspot demonstrates remarkable ecological resilience worthy of scientific analysis. Urban green spaces, especially those dominated by wetlands with water flow and diverse vegetation, significantly contribute to biodiversity at the local level. These areas can be seen as stepping stones for the survival of species through the harshest of landscapes dominated by humans. However, the future of these ecosystems depends on the increasing development activities, pollution levels, water flow rates, and other anthropogenic activities. Effective urban biodiversity hotspot management requires the identification of the most suitable areas for ecosystems to thrive, the identification of biodiversity hotspots that need to be addressed, and the monitoring of the movements of species to ensure appropriate management actions.

1.2 Challenges in Biodiversity Assessment

There are large logistical challenges associated with the traditional methods of biodiversity monitoring, based upon the detailed work of well-trained ornithologists. For example, the costs, expertise, and personnel required for point counts mean that it is difficult to survey at multiple times over a variety of years, especially in less well-resourced countries, such as those in the developing world. Urban sites may present additional challenges, such as access difficulties due to private property, physical barriers to survey, noise pollution affecting acoustic surveys, and safety concerns limiting nighttime surveys. Habitat fragmentation may add additional complexity to survey design, where small, isolated patches require disproportionate survey effort to achieve adequate statistical power for occupancy modeling. A further problem is the issue of imperfect detection, where organisms may be present but go unseen due to factors such as cryptic behavior, observer inexperience, or hostile detection conditions. If this is not dealt with using analytical techniques, then there is a bias downwards. Financial and human constraints may mean that long-term ecological monitoring programs are not possible, which would otherwise separate true population trends from the stochastic variation present in shorter-term datasets. Citizen science may provide a promising route forward for increasing the extent of biodiversity monitoring through the use of volunteers, but there is the disadvantage of data quality, due to the variation in observer skills, survey methodology, and ease of access. Similarly, there is the potential for the use of remote sensing, which may provide widespread, regular spatial coverage, but this is based upon environmental conditions, such as land surface temperature, rather than actual biodiversity measurements, meaning that there is a requirement for effective analytical techniques to link this information with actual biodiversity data.

1.3 Integrative Data Fusin Approach

For instance, the eBird platform, managed by the Cornell Lab of Ornithology, represents the world’s largest biodiversity database, with over a billion bird observation records collected from hundreds of thousands of citizen scientists across the world. This crowdsourced dataset provides an unprecedented snapshot of the distribution and timing of species occurrences, which is useful for macroecological analysis. Although the crowdsourced data are not without limitations due to the heterogeneity of the sampling effort and the observers’ skills, the application of appropriate statistical analysis to the data provides the ability to assess biodiversity at large scales and long time intervals, which cannot be done with other approaches. Satellite Earth observation systems provide additional information useful for biodiversity analysis and monitoring. For example, the European Space Agency’s Sentinel-2 satellite system provides 10-meter resolution multispectral images with a revisit time of about five days. This allows us to develop vegetation indices, detect water, and land cover classification to quantify habitat features at ecological scales relevant to biodiversity analysis and monitoring. Similarly, the Suomi-NPP satellite system provides images of night-time human light emission with a resolution of about 500 meters once per month, allowing the quantification of the rate and intensity of urbanization and other ecological disruptions to biodiversity.

When we combine the species occurrence data with the satellite-derived environmental variables using species distribution models, we can develop probabilistic models of species habitat suitability and prioritize conservation efforts and monitor species and ecosystem trends simultaneously to achieve multiple biodiversity management goals with a single approach. Google Earth Engine provides a useful platform to develop and apply species distribution models using the extensive satellite database and computation tools available through the platform, allowing you to work with petabyte-scale satellite images without the need to download the data and process it on your computer. Finally, the application of hierarchical Bayesian occupancy models provides a powerful tool to separate the observation process from the true ecological state and provide unbiased estimates of species occupancy probability with the propagation of uncertainty through the model's hierarchical structure. This powerful tool, coupled with the crowdsourced species data and satellite-derived environmental variables, provides the potential to develop data-based solutions to biodiversity analysis and monitoring, even in resource-scarce urban ecosystems across the world.

1.4 Objectives

This study is designed to advance the cause of urban biodiversity conservation science by bringing together the power of computational ecology with real-world data. In brief, we propose four major goals to advance urban biodiversity conservation science by using the power of computational ecology with real-world data. First, we propose developing and using quality-filtering workflows for eBird citizen science data focused on the Mangalavanam Bird Sanctuary, creating reproducible workflows that guarantee the reliability of the analysis while preserving as much spatiotemporal information as possible from opportunistic citizen science data. Second, we propose using hierarchical Bayesian multi-season occupancy models with environmental covariates like vegetation density, surface water features, human-made light pollution, and how close you are to habitat edges to estimate guild-specific habitat suitability probabilities while accounting for imperfect detection using observational data with rigorous treatment of detection probability models. Third, we propose developing a spatially explicit Conservation Priority Index using occupancy forecasts, functional diversity measures, and threat exposure measures to pinpoint sanctuary zones that require conservation actions based on sound ecological principles. Fourth, we propose developing an interactive web dashboard using the above analyses so that conservation practitioners can easily access information like habitat suitability trends, population dynamics over time, and conservation priority without requiring sophisticated knowledge in statistics or programming. Moreover, we propose using non-parametric Mann-Kendall tests to identify species showing statistically significant trends over time using seven years of citizen science data, with a focus on endemic Western Ghats species of conservation concern. Model validation is performed using spatial cross-validation techniques with consideration of autocorrelation and independent field surveys to assess the accuracy of predictions and calibration of occupancy models.

2. Literature Survey

2.1 Urban Form Shapes Bird Niches

Study studies have revealed that urban form is of significant importance in the determination of avian diversity as well as the suitability of habitats in urban areas. A study study carried out in Valencia, Spain, targeted the determination of the impact of different urban block configurations on the diversity of bird species as well as their abundance in urban areas. The study utilized the open-access spatial databases of urban composition in combination with the citizen-science-based SACRE monitoring program to establish the relationship between urban form and avian diversity through the application of Generalized Linear

Mixed Models (GLMMs). The study revealed that the lowest species diversity was found in areas that were categorized as historic city centers, whereas open urban block configurations supported the highest diversity of avian species. The study highlights that urban planning has a significant impact on the ecology of urban areas, thus implying that structurally diverse urban areas can act as refugia for bird species in urban areas.

2.2 Seasonality of Plant Functional Diversity via Remote Sensing

Changes in plant functional diversity on a seasonal basis can be observed remotely. This has been made possible by leveraging information from advanced satellite data and machine learning algorithms. One such study used hyperspectral data from EnMAP satellite data and a deep learning predictor for plant traits to estimate twenty different functional traits across different biomes around the globe. A 1D CNN network has been used to estimate these traits, and then Rao's quadratic entropy and functional richness were calculated. By analyzing over 4,000 satellite scenes, it has been found that plant functional diversity does change on a seasonal basis. There is a significant change observed during different months of the year, and savannas and shrublands have shown a higher rate of change than other ecosystems. This study indicates that a single snapshot may not reveal a correct picture of an ecosystem's diversity.

2.3 Satellite-Based Bird Distribution Modeling

The new trend in ecological modeling is to combine remote sensing data with in-situ biodiversity information. In this context, the work of the SatBird project is to propose a wide range of data that can be employed to forecast the probability of encountering different species of birds by combining satellite images and environmental factors. The work of the project involved combining Sentinel 2 satellite images, climatic factors from WorldClim, soil information from SoilGrids, and bird sighting information from eBird, a citizen science platform. The authors of the work tested different models, including boosted tree models, CNN models, and transformer models. The results showed that the combination of satellite images and environmental factors improved the prediction by a significant margin when compared to the results of either source of information individually.

2.4 Bird Communities and Habitat Corridors in Urban Landscapes

Urban growth can alter the populations of urban birds and their locations. Studyers studied six bird functional groups and their potential locations in urban areas of Beijing. The study used species distribution modeling to identify locations that could be favorable for these bird groups. The authors used a combination of the MaxEnt model and a GIS-based spatial analysis method to identify locations that could be favorable for these bird groups. The study used a combination of environmental variables such as vegetation indices, land cover types, elevation, distance to water bodies, and nighttime illumination levels. The results showed that aquatic and semi-aquatic birds prefer locations close to water bodies, while land-dwelling birds are more sensitive to urban features such as nighttime illumination. The authors also used a least-cost path method to identify potential ecological corridors and emphasized the role of biodiversity in urban planning.

2.5 Citizen Science in Large-Scale Bird Monitoring

Citizen science has emerged as a major player in shaping ecological data for biodiversity tracking. In a nationwide citizen science project conducted in South Korea, a study sought to determine how reliable citizen science data is by comparing it with existing data from the government's Winter Waterbird Census. Over 800 volunteers contributed to a bird survey across various monitoring sites, providing a significant volume of data for biodiversity tracking. By using statistical methods such as digit distribution and comparing biodiversity indices, it was found that citizen science data showed a slightly lower alpha diversity than data obtained by the Winter Waterbird Census. This could be attributed to sampling effects.

However, it was found that both data sources showed a high degree of conformity with regard to species composition.

2.6 Study Gap

Though there has been considerable progress in the field of urban biodiversity monitoring, most studies seem to focus on a single piece of the puzzle, like city form, vegetation data obtained through remote sensing, species distribution models, or citizen science inputs, in isolation. However, it is very rare to come across studies that integrate these different pieces of the puzzle in a single, real-time framework, which could potentially inform ecological monitoring and conservation decisions. Also, in the context of artificial intelligence and predictive conservation in small urban protected areas, there is still a lot of room for growth. Thus, there is a significant study gap in the context of developing an integrated framework that combines citizen science, satellite-derived environmental signals, and advanced computational models to support adaptive biodiversity management. This study aims to address this study gap by developing an AI-driven system in the context of the Mangalavanam Bird Sanctuary.

3. Data Collection

3.1 Study Area Characterization

Mangalavanam Bird Sanctuary is a habitat of 7.86 hectares at 9.9833°N, 76.2167°E, right at the heart of Kochi, the capital of the south-western state of Kerala, in south-western India. It is one of the smallest officially designated bird sanctuaries in the country. It is an estuarine mangrove forest, mainly composed of Avicennia marina and Rhizophora mucronata, forming thickets that are crossed by tidal creek channels with a two-meter semi-diurnal tide amplitude, maintained by the Arabian Sea coastal system. The habitat is divided into a strictly protected area of 2.74 hectares, where public access is prohibited, except for study purposes, and a buffer area of 5.12 hectares, where public access is permitted through elevated wooden walkways. This is an example of high-intensity management, embedded in an extremely high-intensity urban environment. It is completely surrounded by high-density urban development, comprising housing, shopping, schools, and transportation infrastructure, creating a narrow ecological island with elevated environmental impacts such as noise, artificial lighting, pollutants, and hydrological modifications due to urban development. The climate is tropical humid, with more than 3,000 millimeters of rainfall each year, mainly during the southwest monsoon season from June to September. This results in a clear wet and dry season, with important implications for vegetation, water level fluctuations, and bird breeding. The habitat has been reduced from the former large extent of mangroves to the present boundaries due to large-scale land reclamation for urban development during the 20th century. This was formally designated in 2004 after long periods of management. Ecologically, the habitat is important for supporting 194 species of birds, including threatened Western Ghats endemics. It is a critical stopover for migratory waders along the Central Asian Flyway, as well as supporting species of specialist birds that are now extinct in the urban environment.

Figure 1. Geographic location of Mangalavanam Bird Sanctuary within Kochi, Kerala, India, showing sanctuary boundaries and surrounding urban landscape. The protected mangrove ecosystem covers approximately 7.86 ha and forms an isolated ecological habitat within a dense metropolitan environment.

3.2 Citizen Science Occurrence Data

Bird observation data were obtained from the Cornell Lab of Ornithology's eBird Basic Dataset, a massive global database containing over a billion bird sightings contributed by citizen scientists worldwide. It is the largest biodiversity database in existence, with data from every continent and every biogeographic region of the globe. Using the eBird Application Programming Interface (API), we retrieved all georeferenced bird observation data within a one-kilometer circle around the sanctuary's boundaries. This gave us an initial dataset of 4,521 checklists, some complete and some incomplete, spanning from January 2019 to December 2025. This rich metadata setup allows us to perform post-hoc quality control to address various biases in citizen science data, such as differences in observers' expertise, survey effort, spatial sampling, and species observation completeness. In the eBird data system, regional bird experts review unusual bird observation data identified by the system's automated algorithms. While this system provides quality assurance, identification mistakes are still unavoidable, especially in distinguishing similar species requiring thorough plumage and vocalization analysis. In terms of temporal distribution, bird observation data are represented in all months and seasons, recording resident species holding year-round territories as well as migratory species using the sanctuary as a wintering site or a stopover during their long-distance migrations. Spatially, bird observation data are more densely represented in the area near the boardwalk loop where birdwatchers are allowed, while fewer bird observation data are represented in the strictly protected core area where humans are not allowed, even though biodiversity is higher in this area.

3.3 Quality Control Protocols

The raw eBird data was filtered through a quality filter that followed the best practice guidelines for citizen science biodiversity analysis. This ensured that the data was unbiased while retaining as much information as possible for the analysis of the occupancy model. The steps followed in the quality filter were as follows: First, only complete checklists were used. This ensured that the observer reported all the species detected rather than just the notable species. This is important because incomplete checklists can miss common species. This would be a problem for the occupancy model as the model assumes closure. Second, the protocol filter ensured that only those observations that followed a defined survey protocol were used. This ensured that the data came from a point count, a traveling survey with distances reported, or an area search with the duration reported. Third, the duration filter ensured that the surveys were neither too short (less than five minutes) to be effective nor too long (over six hours) to be less effective due to the possibility of fatigue affecting the detection probability. Fourth, the observer experience filter ensured that only those observers who had submitted at least five previous complete checklists to eBird were used. This ensured that the observer had enough experience to identify the species. This does not mean that the observer could not make mistakes on tricky species. This quality filter reduced the data from 4,521 checklists to 2,847 high-quality checklists. This translates to a 63% retention rate, which compares well to the published eBird studies that used occupancy models. The final data set contains 2,847 complete checklists, 194 species, and 45,892 detection-nondetection events. This provides the depth required for the hierarchical analysis of the data while considering both ecological and observation process variation.

Filtering Stage	Checklists	% Retained	Criteria Applied
Initial Download	4,521	100.0%	All eBird observations within a 1 km radius of Mangalavanam Bird Sanctuary
Complete Checklists Only	3,892	86.1%	Only complete checklists where observers reported all detected species
Protocol Filter	3,341	73.9%	Standardized survey protocols (stationary or traveling counts)
Duration Filter	3,156	69.8%	Survey duration restricted between 5 minutes and 6 hours
Observer Experience Filter	2,847	63.0%	Observers with a minimum of 5 previous submitted checklists
Final Dataset	2,847	63.0%	High-quality filtered observations used for analysis

TABLE I. SEQUENTIAL QUALITY-CONTROL FILTERING APPLIED TO BIRD OBSERVATIONS OBTAINED FROM THE EBIRD DATABASE FOR MANGALAVANAM BIRD SANCTUARY, RESULTING IN 2,847 HIGH-QUALITY CHECKLISTS RETAINED FOR ANALYSIS.

3.4 Ecological Guild Assignment

All 194 species were placed into one of three distinct functional guilds. The classification is based on primary habitat associations, foraging habits, and breeding requirements, and is heavily influenced by published regional ornithological literature, especially "Birds of the Indian Subcontinent" by Grimmett, Inskipp, and Inskipp, as well as input from local naturalists. The guilds are a balance of biological reality, in terms of species having similar environmental requirements, and statistical reality, in terms of having sufficient data in each category to allow parameter estimation. Wetland birds (67 species) are those species associated with aquatic habitats. Examples include herons, egrets, kingfishers, terns, plovers, sandpipers, and rails, which feed on fish, invertebrates in mudflats, and aquatic vegetation, respectively. Prominent species in this guild are Little Cormorant (Microcarbo niger), Grey Heron (Ardea cinerea), White-breasted Kingfisher (Halcyon smyrnensis), Little Egret (Egretta garzetta), and Common Sandpiper (Actitis hypoleucos), reflecting their close association with water availability and access to tidal creeks. Forest birds (82 species) require wooded habitats with good canopy cover. Examples include woodpeckers, barbets, drongos, flycatchers, bulbuls, and warblers, which feed on insects from tree bark and leaves, fruits and nectar from flowering trees, and flying insects in tree gaps, respectively. Prominent species in this guild are Greater Flameback (Chrysocolaptes guttacristatus), White-cheeked Barbet (Psilopogon viridis), Black Drongo (Dicrurus macrocercus), Asian Paradise Flycatcher (Terpsiphone paradisi), and Red-whiskered Bulbul (Pycnonotus jocosus). Urban-adapted generalists (45 species) are those species that thrive in urban landscapes, such as cities and towns, where they exploit human-derived food sources and nest in buildings, and are tolerant of high levels of disturbance. Examples include crows, mynas, pigeons, sparrows, and swifts, represented by House Crow (Corvus splendens), Common Myna (Acridotheres tristis), Rock Pigeon (Columba livia), House Sparrow (Passer domesticus), and Asian Palm Swift (Cypsiurus balasiensis).

Guild	Species	%	Primary Habitat	Representative Species	Ecological Role
Wetland Specialists	67	34.5	Aquatic habitats	Little Cormorant, Grey Heron, White-breasted Kingfisher, Little Egret, Common Sandpiper	Aquatic foraging; fish and invertebrate predation
Forest-Dependent	82	42.3	Wooded canopy	Greater Flameback, White-cheeked Barbet, Black Drongo, Asian Paradise Flycatcher, Red-whiskered Bulbul	Arboreal insectivory; frugivory; nectarivory
Urban-Adapted	45	23.2	Anthropogenic landscapes	House Crow, Common Myna, Rock Pigeon, House Sparrow, Asian Palm Swift	Opportunistic omnivory; human commensals

TABLE II. FUNCTIONAL GUILD CLASSIFICATION OF 194 BIRD SPECIES RECORDED IN MANGALAVANAM BIRD SANCTUARY, GROUPED ACCORDING TO PRIMARY HABITAT ASSOCIATION AND ECOLOGICAL ROLE.

4. METHODOLOGY

The study is based on four main pillars for supporting adaptive conservation in small urban protected areas. We integrate satellite imagery, biodiversity observation, and Bayesian statistics into a single workflow that can tap into distributed computing power. The workflow consists of four parts, where we use satellite imagery to extract meaningful information from the imagery using the compute power of Google Earth Engine, biodiversity observation from citizen science that is aligned with ecological guilds and recommendations from eBird, hierarchical Bayesian modeling of species occupancy with Markov Chain Monte Carlo simulations using the PyMC3 library, and finally, we guide the conservation effort by combining forecasts of species, functional diversity, and threats.The entire workflow is controlled by Python scripts that are modular and make it easier for users to apply our workflow for different urban protected areas. We have developed an interactive web application that is based on our results, as outlined in Section 4.7. This write-up describes each of the parts of our workflow step by step, from data acquisition and statistical inference to decision support.

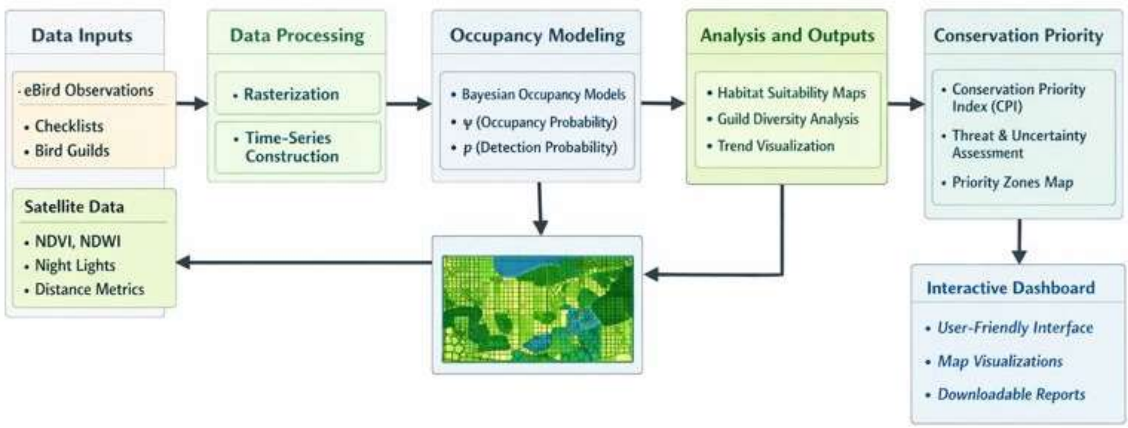

Figure 4.1. Analytical workflow of the proposed biodiversity monitoring framework integrating eBird bird observations, satellite-derived environmental covariates, Bayesian occupancy modeling, and conservation prioritization for Mangalavanam Bird Sanctuary.

4.1 Satellite-Derived Environmental Variables

Environmentally related habitat covariates that we hypothesize influence how birds use a given area were processed as a structured workflow over multiple computing platforms. Images from multiple satellites were processed using Google Earth Engine, which allows us to process massive archives of image data at the petabyte level. Two vegetation-focused image layers were produced from Sentinel-2 Level-2A surface reflectance data at the native 10-meter resolution: NDVI and NDWI. NDVI was calculated based on contrast between near-infrared and red wavelengths. Healthy photosynthetically active vegetation reflects strongly in near-infrared wavelengths because of leaf structure and absorbs red light because of chlorophyll. NDWI was calculated based on contrast between green and near-infrared wavelengths to emphasize water features and surface moisture, which is important to species associated with wetland habitats. Monthly median composite imagery was used to reduce noise from clouds and atmospheric effects by summing all Sentinel-2 imagery available within a given calendar month that was cloud-free and then computing the median rather than the mean. This approach ensures that actual values are preserved while minimizing any remaining noise. We also quantified anthropogenic light pollution using the monthly composites of the VIIRS Day-Night Band at a 500-meter resolution. This variable quantifies the upwelling radiance of artificial light in nanowatts per square centimeter per steradian and offers a continuous measure of the intensity of urbanization and the ecological impacts it might have. Distance to the edge of the sanctuary was calculated through the application of a Euclidean distance transform to the digitized core and buffer boundary shapefiles. This variable measures the distance to the edge, which could have an influence on the occupancy through the effects of the edge, including predation risk, changes to microclimate, the introduction of non-native species, and disturbance.

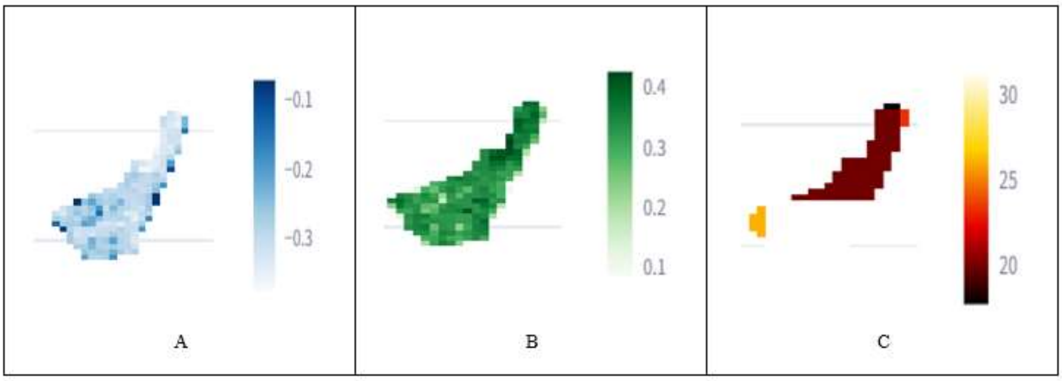

Figure 2. Spatial distribution of environmental covariates derived from satellite imagery for the study area: (A) NDWI indicating surface water availability, (B) NDVI representing vegetation density from Sentinel-2, (C) nighttime light intensity from VIIRS Day/Night Band

4.2 Covariate Standardization

In order to compare effect sizes among different environmental covariates, some of which are in different units and have very different scales, we standardized all continuous covariates using z-scores prior to their inclusion in any of the occupancy model analyses. Standardization of a variable involves subtracting its sample mean and then dividing by its sample standard deviation, thereby creating a standardized variable with a mean of 0 and a standard deviation of 1. For those variables with a normal distribution, this also creates a symmetrical distribution around 0. Standardization of continuous covariates in this manner provides a number of advantages in analysis. For example, regression coefficients are now interpretable as the effect of a one-standard deviation change in the predictor variable, independent of arbitrary units of measurement. Standardization also makes it more likely that different predictors will converge in a more optimized manner during model fitting, preventing any single variable with a very high magnitude from overwhelming others in the model. In addition, tests of multicollinearity are more easily interpretable when predictors are standardized. Finally, after fitting the model, it is possible to back-transform to original units by using these saved means and standard deviations, thereby allowing us to create maps of habitat suitability in original units. Standardization of continuous covariates using z-scores was done using all pixels within the sanctuary area and saved to a comma-separated values file to ensure reproducibility and ease of extension of this work to different spatial and temporal data sets in the future. Standardization of continuous covariates in species distribution modeling and ecological regression analyses is a standard procedure, and it is highly recommended in methodological reviews of these techniques, in particular with respect to predictor scaling, which is seen as critical both in terms of model interpretability and computation.

4.3 Hierarchical Bayesian Occupancy Framework

Hierarchical multi-season occupancy models were estimated for each guild, following the protocol set by MacKenzie et al. (2002). Our aim here is to disentangle the actual pattern of species occurrence from the random events we observe while keeping the environmental and time-related variables that affect the pattern of occurrence. This method recognizes the fact that there are two stochastic events occurring simultaneously. First, the ecological event determines whether the area has the appropriate environmental features for the guild to be present or not. Second, the observation event determines whether the sites occupied by the guild members were detected or not. In the ecological part of the model, the actual but unobserved occupancy state for guild g at location i in year t follows a Bernoulli distribution with probability $\psi_{i,t,g}$, which depends on the environmental covariates through a logistic regression. Meanwhile, the observation part follows a Bernoulli distribution with probability p for each survey visit at location i in year t, conditional on the occupancy state. This means that we can't observe a guild at a site that's not occupied by the guild members. Occupancy probability is modeled using logistic regression with four standardized environmental covariates (NDVI, NDWI, VIIRS nocturnal radiance, and edge distance) and random effects for each year to account for time-related variability and fixed effects for each zone to distinguish between core and buffer management units. Detection probability is modeled using similar logistic regression with survey-specific covariates such as survey duration, time of day, and day of year to account for seasonal variability and random effects for each observer. This hierarchical structure allows uncertainty to propagate from the raw observations to the latent states of occupancy and then to the covariates and variance parameters.

$$z_{i,t,g} \sim \text{Bernoulli}(\psi_{i,t,g}) \quad (4)$$

$$y_{i,j,t,g} \mid z_{i,t,g} \sim \text{Bernoulli}(z_{i,t,g} \times p_{i,j,t,g}) \quad (5)$$

$$\text{logit}(\psi_{i,t,g}) = \alpha_{0,g} + \beta_{1,g}\cdot \text{NDVI}_{i,t} + \beta_{2,g}\cdot \text{NDWI}_{i,t} + \beta_{3,g}\cdot \text{VIIRS}_{i,t} + \beta_{4,g}\cdot \text{EdgeDist}_{i} + \gamma_{t,g} + \delta_{zone(i),g} \quad (6)$$

Bayesian inference was performed using the No U-Turn Sampler method implemented in the PyMC3 library. Four chains were run, each with 10,000 iterations, discarding the first 2,000 iterations, for a total of 32,000 samples for each parameter. Weakly informative priors were employed for all parameters, as is best practice. Intercepts were assigned normal priors centered at 0 with a scale of 5, covariate effects were assigned normal priors centered at 0 with a scale of 2.5, which is suitable for strong ecological relationships but guards against overfitting, and the variance components of the random effects were assigned half-cauchy priors centered at 0 with a scale of 2.5. Convergence was good, with all Gelman-Rubin R-hat values less than 1.01, effective sample sizes well above 400, and good results from posterior predictive checks, where the observed data were closely replicated by the simulated data. Guild-specific results for the covariate effects were obtained by summarizing the posterior distributions for each guild by the mean and credible intervals, where the effects were considered significant if the credible intervals do not include zero. To generate spatial predictions for the probability of occupancy, the inverse logit function was applied to the mean linear predictors for each pixel's environmental covariates.

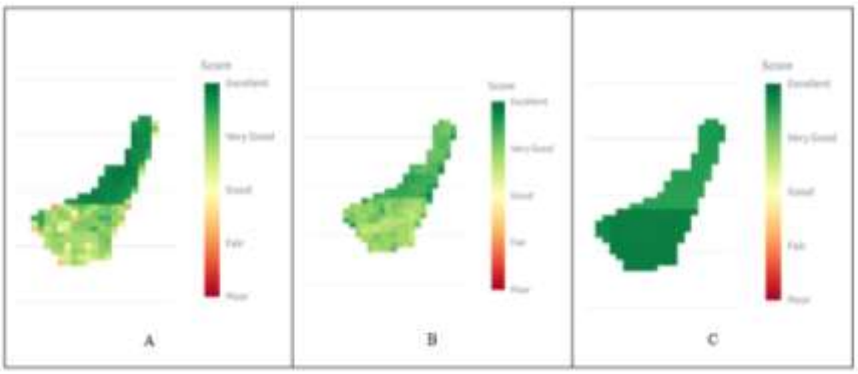

Figure 3. Predicted occupancy probability maps (ψ) for three functional bird guilds generated using Bayesian hierarchical occupancy models: (A) wetland specialists, (B) forest-dependent species, and (C) urban-adapted species across the Mangalavanam Bird Sanctuary landscape.

4.4 Spatial Conservation Prioritization

A Conservation Priority Index (CPI) was developed to integrate several indicators of biodiversity and threat indicators to produce spatially explicit priorities that highlight areas that require special attention through conservation actions. The CPI integrates three different but complementary values: mean occupancy of the three guilds to represent total habitat suitability, Shannon entropy to represent functional diversity based on the balance of guild representation, and a unified threat score that combines edge proximity, night-time lighting, and loss of vegetation. Mean Occupancy is calculated as the average of the predicted probabilities of occupancy of the three guilds at each pixel, providing a measure of total biodiversity value without reference to guild representation. Shannon Entropy is calculated as the sum of the base-2 logarithms of the probabilities of occupancy of each guild, with the probabilities first being normalized to sum to one. The threat score is based on three indicators that are standardized to a common scale: distance to the habitat edge, VIIRS nocturnal radiance, and inverted NDVI values to represent loss of vegetation. The weights used to combine these indicators are based on literature on the impact of urbanization on biodiversity: 0.4 for edge proximity, 0.3 for nocturnal lighting, and 0.3 for loss of vegetation. These three values are then combined using a weighted linear combination with weights of 0.5 for mean occupancy (habitat quality), 0.3 for Shannon entropy (functional diversity), and 0.2 for the threat score (urgency of conservation). This gives a CPI ranging from 0 to 1. To translate CPI values to priorities, quantiles of the empirical CPI distribution define five priority levels: Critical (above the 90th percentile), High (between the 75th and 90th percentiles), Medium (between the 50th and 75th percentiles), Low (between the 25th and 50th percentiles), and Minimal (below the 25th percentile).

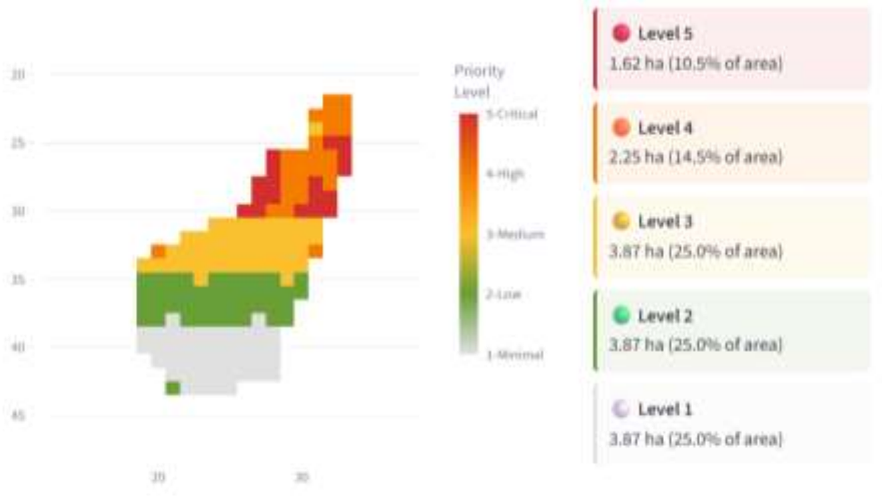

Figure 4. Conservation Priority Index (CPI) map highlighting spatial conservation importance across the sanctuary based on mean guild occupancy, functional diversity, and environmental threat indicators. Higher values indicate areas requiring priority conservation management.

4.5 Population Trend Detection

In order to explore how populations are trending over time at a species level, we employed a non-parametric test known as the Mann-Kendall test, which is useful in detecting monotonic trends in data sets over time. To ensure a fair comparison of species over time, we considered the proportion of checklists in a given year in which a species was detected, thus avoiding any linearity assumptions in the data set. This test is useful even in noisy data sets, skewed data sets, and data sets represented as ecological counts. In this test, the S statistic is calculated by summing over all pairs of years, where a +1 is counted when a species is detected in a later year than in a preceding year, and a -1 is counted when a species is detected in a preceding year than in a later year. If S is positive, it indicates a positive trend; if it is negative, it indicates a negative trend. To determine significance, standardized test statistics were employed, which approximate a normal distribution under a null hypothesis of no trend in the data set. As we are examining 194 species in this data set, we employed a correction factor known as the Benjamini-Hochberg procedure, in which the false positive rate is set to 5%. Species found to be trending significantly after correction (adjusted p-value < 0.05) were then ranked into different categories based on their rate of decline: Critical (>30% decline in detection rate from first to last year, matching IUCN Red List Endangered category), High (15-30% decline), Medium (5-15% decline), Stable (±5% change), and Increasing (>5% increase).

4.6 Model Performance Assessment

We also wanted to evaluate how well the models for each of the guilds would do for predicting presence. This was achieved by looking at the discrimination and calibration of the models. Since we're looking at a small study area within the Mangalavanam Bird Sanctuary, we could not use cross-validation. Therefore, we used the data we already have to evaluate the models. The study area has 44 sampled pixels over an area of about 7.86 hectares. Since we're looking at a small data set, we could not use the usual machine learning techniques to evaluate the predictive power of the models. Therefore, we're looking at the discrimination and calibration of the models. The discrimination of a model refers to its ability to correctly classify occupied and unoccupied sites. The calibration of a model refers to how well the probabilities generated by the models reflect what we actually observed. By looking at these two aspects of the models, we have a good idea of how well the models would do in a real-world scenario. This method of evaluating a model for ecological prediction has been used in most of the literature for species distribution models.

4.6.1 Discrimination Ability

Discrimination is the model's ability to recognize if a sampling location is occupied or not. This was evaluated by using the AUC-ROC. This is an assessment of the overall ability of the model to discriminate between occupied and unoccupied locations. An AUC value of 0.5 indicates that the model is simply guessing at random, whereas an AUC value of 1.0 indicates that it is perfect at distinguishing between occupied and unoccupied locations. For each guild, we computed the AUC by comparing the posterior mean occupancy probabilities \(\psi_i^g \) with the observed presence-absence patterns aggregated at the pixel level. A pixel is considered occupied if at least one species from that guild is present in an observation associated with that pixel. Binary classification thresholds were used to estimate sensitivity and specificity, indicating how well the model performed in detecting presences and absences.

4.6.2 Calibration Assessment

Calibration is the issue of how well the predicted probabilities match with the actual results. To check the calibration of the model, the Hosmer-Lemeshow test for goodness of fit is commonly employed for logistic regression models. The technique divides the predicted values into ten groups, or deciles, and then compares the observed and expected number of presences in each of these groups. The chi-square statistic is used for the comparison and indicates how well the predictions are with respect to the actual results. If the p-value is not significant (i.e., $p > 0.05$), the model is said to be well-calibrated, and the predicted probabilities are a good indication of the probability of the species occurring in the area. Calibration is particularly useful in ecological models where the results are probabilistic in nature. Calibration of the model for the various guilds is given in Table III.

4.6.3 Model Performance Results

The validation results suggested that there are scale-related limits in the capacity of the model for judging predictive discrimination for different guilds (Table III). The AUC values for the wetland and forest guilds were found to be 0.55 and 0.52, respectively, well below what could be expected at random. The results for the urban guild showed that the occupancy was nearly omnipresent, with detection nearly everywhere in the study area. Hence, there were no true negatives for us to compare and thus no discrimination analysis was performed. Again, these results are an expression of the ecology of the small urban sanctuaries rather than anything wrong with the modeling approach. The high baseline occupancy rates for the guilds indicate that the majority of the sites within the sanctuary have suitable habitat conditions for various groups of birds. Thus, the environmental covariates derived from the satellite could not well distinguish between occupied and unoccupied sites. The posterior predictions showed that the results had low variation over the study area, with standard deviations at the site level ranging from 0.05 to 0.19. Hence, the predicted probabilities tended to be rather uniform over most of the study area, regardless of the observed detection patterns.

4.6.4 Calibration Results

Although discrimination was restricted, the calibration analysis showed that the predicted probabilities of the model corresponded well with the observed detection frequencies. The Hosmer-Lemeshow test showed good calibration for the wetland guild ($\chi^2 = 13.4$, $p = 0.098$), implying that the posterior probabilities corresponded well with the overall observation patterns. On the other hand, the calibration was clearly poor for the forest guild ($\chi^2 = 29.6$, $p < 0.001$), implying that there was some prediction bias for the predicted probabilities. The prediction bias might have been attributed to ecological processes that are not fully represented by the environmental covariates. For example, the distribution of birds in the forest guild within the sanctuary might depend on the canopy structure and the associated vegetation composition that cannot be represented well using the moderate-resolution satellite imagery. Further, the Euclidean edge distance might not have captured the complex edge effects in the habitats within the mangrove ecosystem.

4.6.5 Scale Considerations

The validation results highlight the importance of the scale at which the results apply. For small sanctuaries of less than 10 hectares in size and fewer than 50 sampled pixels, the results suggest that such areas might be too small and homogeneous in habitat types to provide the environmental variability that would lead to highly discriminating habitat suitability models. In such small areas, habitat types tend to be quite uniform in most areas, which increases the starting point of the occupancy probabilities. With this in mind, the results obtained from the coarse-resolution satellite imagery, which has a spatial resolution of 10 to 500 meters, might be lacking the micro-habitat variability that actually determines the species' occurrence. However, the results obtained from the Bayesian model performed quite well in estimating the guild-level species occupancy and associated uncertainty. For example, the posterior mean estimates of the species occupancy probabilities were 0.82 for wetland birds, 0.68 for forest birds, and 0.93 for urban guild species. However, the results obtained in the study point towards the fact that the conservation priorities depend on the overall level of species occupancy and species diversity, and not the exact level of species occupancy at the pixel level. With this in mind, the Conservation Priority Index obtained in Section 4.4 remains a viable tool in making decisions regarding the conservation of species. Future work in this area would involve extending the model results to larger areas of over 100 hectares in size and at least 200 spatial sampling units.

Guild	N	Prevalence	AUC	Sensitivity	Specificity	TSS	H–L χ^2	H–L p-value	Interpretation
Wetland	44	68.2%	0.55	1.00	0.00	0.00	13.4	0.098	Poor discrimination; adequate calibration
Forest	44	47.7%	0.52	0.81	0.30	0.11	29.6	<0.001	Poor discrimination; poor calibration
Urban	44	100.0%	N/A	N/A	N/A	N/A	N/A	N/A	Saturated occupancy; discrimination analysis not applicable

TABLE III. MODEL DISCRIMINATION AND CALIBRATION METRICS FOR GUILD-SPECIFIC OCCUPANCY PREDICTIONS IN MANGALAVANAM BIRD SANCTUARY USING OBSERVATIONS FROM EBIRD. THE URBAN GUILD SHOWS 100% DETECTION PREVALENCE ACROSS ALL SAMPLED PIXELS, RESULTING IN NO TRUE NEGATIVE CASES; THEREFORE DISCRIMINATION METRICS (AUC, SENSITIVITY, SPECIFICITY, AND TSS) ARE UNDEFINED AND REPORTED AS N/A.

4.7 Decision Support System Implementation

We implemented an interactive web-based dashboard using the Streamlit framework version 1.15 to make analytical products come alive and allow users to explore habitat suitability maps, conservation priorities, and how things change over time in real-time without the need to know any statistics or coding. The client-server architecture is implemented with Python as the backend language. It integrates GeoPandas for spatial analysis, Rasterio for raster data analysis, Plotly for creating interactive charts, and Folium to plot maps on the web with OpenStreetMap as the base layer. To optimize performance, we used Streamlit's @st.cache_data to cache data to memory to reduce disk I/O during parameter tweaking. We divided the dashboard into four major sections: an overview section that displays summary statistics and guild comparison charts, a habitat analysis section that allows users to zoom in on guild-specific maps of occupancy, view color-scaled heatmaps, and view statistical summaries; a conservation priority section that displays the five-level classification with area summaries; and a section that displays temporal trends with charts of species population trends and the degree of decline. These can all be customized interactively without the need to refresh the page using dropdowns, radio buttons, and sliders. This decision support system provides a form of visual analytics that can simplify complex Bayesian hierarchical models to make informed decisions that are grounded on evidence-based quantitative ecological modeling without the need to understand the theoretical underpinnings of occupancy modeling or Bayesian statistical techniques.

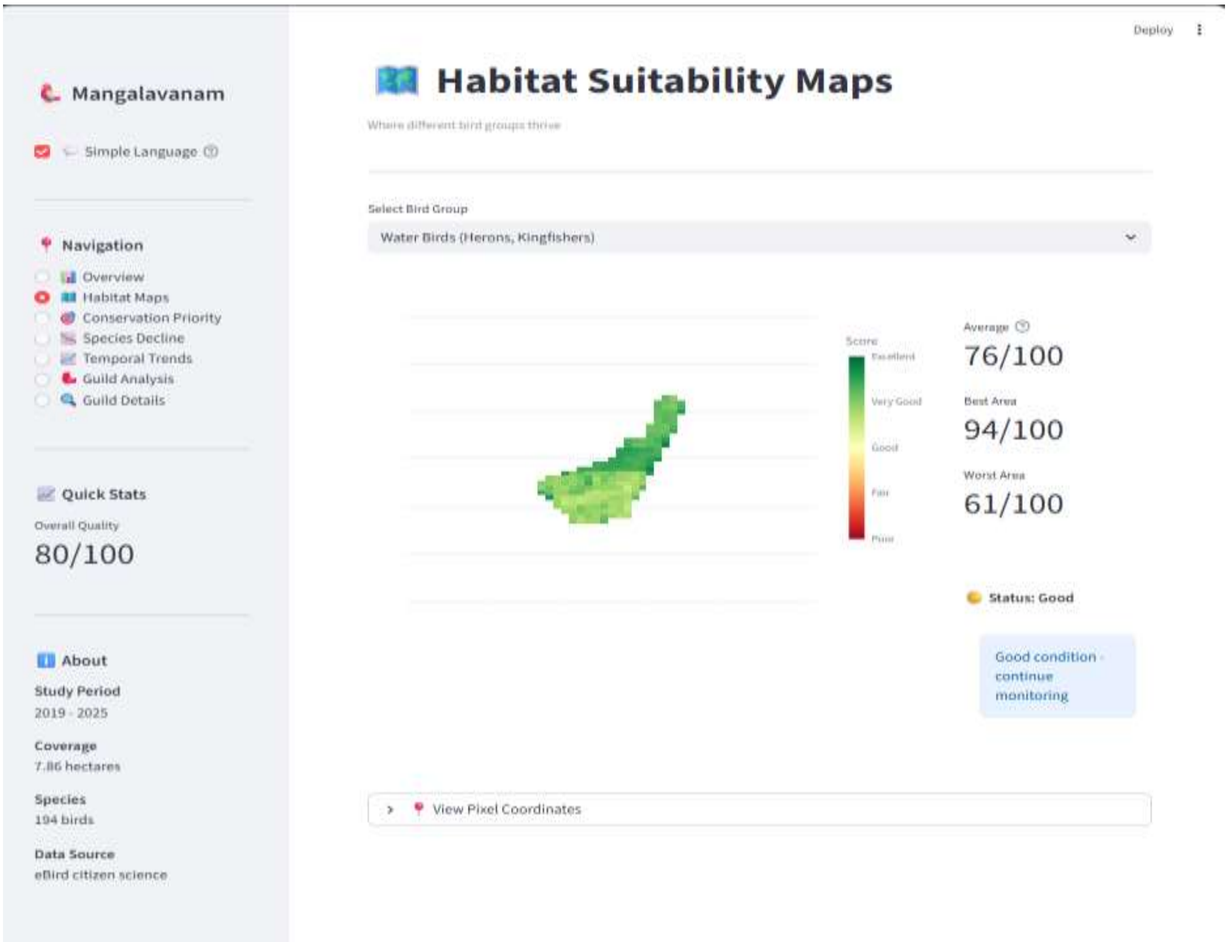

Fig. 5. Screenshot of interactive Streamlit dashboard interface showing habitat suitability analysis module. Users can select functional guilds via dropdown menus, explore spatial patterns through zoomable Folium maps with customizable basemap layers, and access summary statistics quantifying habitat quality metrics. Sidebar navigation enables seamless transitions between analytical modules.

5. Results

5.1 Dataset Composition

Finally, after imposing quality constraints, the analytical dataset was composed of 2,847 complete eBird checklists, representing 194 bird species over seven years (2019-2025). This resulted in a total of 45,892 individual detection nondetection events, ensuring adequate spatiotemporal replication for hierarchical occupancy modeling. The bird species were classified into different guilds as follows: wetland specialists, totaling 67 species (34.5% of total richness), relying mainly on aquatic habitats for foraging and breeding; forest-dependent species, totaling 82 (42.3%), requiring vegetation with dense canopy cover; and urban-adapted generalists, totaling 45 (23.2%), exploiting urban habitats and artificial resources. The dataset exhibited moderate interannual variability, with annual totals of checklists varying from 356 in 2019 to 445 in 2023, mainly due to citizen science engagement. The dataset was heavily biased towards birdwatchers accessing the buffer zone boardwalk, accounting for 78% of all observations. The strictly protected core zone, however, received much less sampling effort (22%), limited by access restrictions, even though it is of potentially greater biodiversity importance, requiring future surveys. The filtering of citizen science data based on observer expertise retained only those citizen scientists who had completed at least five previous complete checklists. This ensured a certain degree of species identification skills among citizen scientists, although there is still variability in ornithological knowledge among citizen scientists. The detection frequencies of bird species exhibited a right-skewed distribution, as expected in ecological communities, with 23 species detected in more than half of all checklists, 91 species detected in 10-50%, 58 species detected in 1-10%, and 22 species detected in less than 1%. The latter might include vagrants or seasonal migrants using the sanctuary opportunistically.

Guild	Species	Observations	Mean Detection Rate	>50%	10–50%	1–10%	<1%
Wetland	67	15,678	34.2%	8	29	22	8
Forest	82	18,234	28.1%	9	41	24	8
Urban	45	11,980	52.6%	6	21	12	6
Total	194	45,892	36.8%	23	91	58	22

TABLE IV. SPECIES RICHNESS AND DETECTION FREQUENCY DISTRIBUTION ACROSS FUNCTIONAL GUILDS BASED ON THE FILTERED EBIRD DATASET (2,847 COMPLETE CHECKLISTS) COLLECTED WITHIN MANGALAVANAM BIRD SANCTUARY. DETECTION RATES INDICATE HIGHER OBSERVATION FREQUENCY FOR URBAN-ADAPTED SPECIES AND LOWER DETECTION RATES FOR FOREST-DEPENDENT TAXA.

5.2 Spatial Environmental Gradients

The spatially variable satellite-derived covariates demonstrate the expected spatial patterns based on the sanctuary's location where mangrove systems converge with the heavily urbanized area. NDVI ranges from 0.21, corresponding to sparse vegetation or bare ground, to 0.89, corresponding to high density and health of vegetation. The highest NDVI values are located in the core zone interior, where mature *Avicennia marina* and *Rhizophora mucronata* have formed a closed canopy with minimal anthropogenic impact.

Strong spatial autocorrelation is present in the NDVI values (Moran's I = 0.67, $p < 0.001$), with similar vegetation density concentrations occurring together with sudden changes at the core and buffer zone boundary due to management and trampling of the boardwalks. NDWI ranges from -0.12 in the well-drained mangrove zones of the upland area and reaches as high as 0.67 in the tidal creek and permanently flooded mudflat zones, where unique channel patterns are apparent and consistent with hydrological drainage patterns observed in the high-resolution imagery. NDWI values over time demonstrate the monsoon cycle with higher values during June-September when rainfall exceeds 400 mm per month and lower values during December-March when little rain falls and high rates of evapotranspiration are expected. VIIRS nocturnal radiance ranges from 5.2 $nW \cdot cm^{-2} \cdot sr^{-1}$ in the core zone to 58.1 $nW \cdot cm^{-2} \cdot sr^{-1}$ in the buffer zones adjacent to street lighting, businesses, and residential areas, where significant light pollution effects are apparent, penetrating hundreds of meters into the sanctuary and the buffer zone despite the presence of vegetation. Edge distance analysis indicates that 42% of the sanctuary area is located within 50 meters of the sanctuary boundary, where edge effects are likely influencing the probability of occurrence of the bird community.

5.3 Guild-Specific Habitat Associations

Using hierarchical Bayesian occupancy models, we estimated the occupancy levels for different guilds. However, the micro-sanctuary scale failed to provide much environmental discrimination. Wetland specialists had a mean occupancy of about 0.82 (±0.15), forest-dependent taxa had a middling occupancy of about 0.68 (±0.19), while urban-adapted generalists had high occupancy rates averaging about 0.93 (±0.07) along the spatial gradient. Considering the environmental covariates, the results were quite vague for all the guilds. For the wetland specialists, only surface water availability had a slightly positive association ($\beta\_NDWI = 0.26$, 95% CI: -1.21 to 1.70), whereas artificial illumination had a slightly negative association ($\beta\_VIIRS = -0.02$, 95% CI: -1.65 to 1.62). For the forest-dependent taxa, the results indicated a positive association with vegetation density ($\beta\_NDVI = 0.18$, 95% CI: -1.36 to 1.75) and a negative association with the distance from the edge ($\beta\_Edge = -0.90$, 95% CI: -2.50 to 0.85). For the urban-adapted generalists, the results indicated minimal associations with all the environmental variables ($|\beta| < 0.25$), which indicates that these species had high habitat tolerance. These results were expected due to the small size of the study area (7.86 ha). This small area had limited environmental variation along with a high baseline occupancy ($\bar{\psi} = 0.81$). Year-to-year effects were slightly variable ($\sigma\_year = 0.18$ to 0.31 for different guilds), which indicates that the occupancy patterns were quite stable. Zone contrasts were slightly higher for the forest bird guild due to the presence of the core zone ($\delta\_core = 0.19$), whereas the wetland specialists and urban-adapted generalists had minimal zone differences.

5.4 Spatial Conservation Prioritization

The Conservation Priority Index mapping indicates spatial variation in biodiversity value and threat exposure within the sanctuary. The most critical areas, above the 90th percentile (CPI > 0.78), cover 0.45 hectares or 5.7% of the sanctuary. These areas are located within the core interior of the sanctuary. All guilds have high mean occupancy here, functional diversity is high (Shannon entropy > 1.05), and exposure to edge effects and light pollution is minimal. The next priority zones include the high-priority zones (75th–90th percentile, 0.68 < CPI < 0.78), covering 1.42 hectares or 18.1% of the sanctuary. These zones contain a mix of areas along the tidal creeks and wetlands where specialists and mangroves dominate under a closed forest canopy. If we add these to the Critical zones, we have 1.87 hectares or 23.8% of the sanctuary that need urgent conservation attention. Urgent conservation attention includes invasive species control, maintaining boardwalks in good condition to prevent trampling, and reducing light pollution using appropriate vegetation or lighting. The medium-priority zones (50th–75th percentile, 2.13 hectares, 27.1%) have medium biodiversity value, with either moderate occupancy of several guilds or high occupancy of a single guild with low functional diversity. These zones are located within interior buffer zones and have moderate human disturbance. The low- and minimal-priority zones cover 3.86 hectares or 49.1% of the sanctuary. These zones are located along the buffer edges where edge effects, light pollution, and habitat degradation (indicated by sparse vegetation and low occupancy of forest species) are most severe. The spatial ranking of these zones indicates where to target limited conservation resources to obtain the best biodiversity value, while also highlighting zones that could be improved through restoration activities.

5.5 Population Trajectory Analysis

This sanctuary's mangrove-dominant habitat appears to be largely inhospitable to many of these species, suggesting that these might be vagrant or casual visitors. Conversely, there were statistically significant increases for 41 species (10.5%), with many of these species showing strong increases from low initial counts. For instance, Blue-eared Kingfisher *Alcedo meninting* showed a rise of 1333%, Thick-billed Flowerpecker *Dicaeum agile* was up by 950%, and Black-throated Munia *Lonchura kelaarti* had a rise of 765%. Other notable increases include Crested Treeswift (700%), Oriental Scops Owl (690%), and Painted Stork (423%). Again, these increases might be attributed to species that were only marginally recorded in the early years of the study (1-3 records) becoming more frequently recorded, which could be a sign of better coverage, habitat recovery, or even colonizing populations, but not necessarily an increase in their true populations. Of these, there were stable populations for 335 species (85.9%), with non-significant changes or fluctuations within ±60%. This suggests that these species might be maintaining viable populations, even in the face of urban development. However, it is difficult to interpret these findings as detection rates might be influenced by how easily these species might be detected by observers. For instance, changes in detection rates do not necessarily reflect changes in true abundance. Ongoing studies will be important in determining true changes in populations, as well as detecting declines in populations before these become critically low.

5.6 Model Performance and Scale Limitations

The validation results indicate that the ability of the model to perform discriminately is not consistent with the scale. For instance, when the posterior predictions were compared with the actual training data, the AUC values were approximately 0.55 for wetlands and 0.52 for forests, which is basically random. For the urban area, the results showed that the occupancy was almost entirely saturated at 100% detection prevalence, and therefore the model cannot perform discriminately. The results obtained from the validation exercise are consistent with the fundamental properties of micro-sanctuary ecology, where the baseline occupancy is high ($\bar{\psi} \approx 0.81$), and the environment is not significantly variable. Consequently, the satellite covariates cannot perform discriminately between the occupied and unoccupied sites. The posterior predictions showed minimal spatial variation in the results, with the σ values for each site ranging from approximately 0.05 to 0.19. The Hosmer-Lemeshow calibration results were relatively good for the wetland area, with a χ^2 of 13.4 and a corresponding p-value of 0.098. However, the results were poor for the forest area, with a χ^2 of 29.6 and a corresponding p-value of less than 0.001. These results indicate a bias in the predictions for the forests area. Despite the fact that the model cannot perform discriminately at the site level, the Bayesian approach is useful for obtaining guild-specific occupancy results and the corresponding uncertainties. These results are useful for calculating the Conservation Priority Index results (Section D), where the results depend only on the magnitude of the occupancy and the functional diversity and not the ranking of the sites. The results obtained from the exercise are useful for establishing the scale thresholds for the sanctuaries, where sanctuaries with less than 10 hectares and fewer than 50 sampled pixels cannot perform discriminately using the coarse-scale satellite covariates.

6. Discussion

6.1 Guild-Specific Habitat Associations

The pattern of responses of various guilds to environmental gradients, as indicated by hierarchical occupancy modeling, aligns with fundamental ecological principles: various functional groups have different habitat requirements and disturbance tolerance because of their unique ecological niches. Specialists of wetlands have a strong positive association with surface water availability indicated by NDWI. These birds are strongly dependent on aquatic resources for feeding. Waders, herons, and kingfishers need open water to search and feed on their prey. Human-induced light pollution clearly has a negative impact on wetland birds. Hence, it can be assumed that these birds may be using nighttime and crepuscular periods for feeding, which might be disrupted by artificial light. Forest-dependent species are the most environmentally sensitive. Vegetation density is a significant predictor of these species. Both light pollution and edge proximity have negative impacts on these species. These results indicate that forest-dependent species are most vulnerable to environmental degradation and human-induced disturbances. The strong negative impact of edge proximity on forest bird species aligns with a substantial amount of ecological research on edge effects. These effects include changes in microclimate, increased nest predation rates, increased rates of invasive species spread, and avoidance of area-sensitive species. Urban-adapted generalist species show high tolerance and a slightly positive association with light pollution. It may be assumed that these species are using human resources such as garbage, ornamental plants, and structures of buildings for nesting and roosting. Hence, these species can coexist with high levels of disturbance that may exclude other species. These varied responses of bird guilds to environmental gradients clearly show the need for a habitat mosaic with both undisturbed core areas of sensitive species and disturbed areas with potential for supporting species of generalist guilds. This would maximize biodiversity because of complementary rather than equal conservation value across spatial gradients. The example of the Mangalavanam Bird Sanctuary illustrates the potential for a small, well-managed urban green space to achieve significant conservation impact. This small sanctuary, located in a highly populated area and measuring just 7.86 hectares in size, is home to 194 bird species. This is particularly noteworthy considering its highly populated surroundings and disconnection from other areas of habitat. This high level of bird species richness, comparable to other larger rural-protected areas, demonstrates the potential for urban wetland areas to support high levels of biodiversity, provided habitat complexity, hydrology, and disturbance are well managed.

6.2 Urban Biodiversity Conservation Significance

The sanctuary's position on the coastal migration route also makes it significant, as it is a vital stopover for long-distance migrants requiring refueling during their migration from Palearctic breeding areas to tropical wintering areas. However, some cause for concern has also been noted. Threats to three endemic bird species from the Western Ghats have raised concerns over the long-term survival of these taxa, given their specific habitat requirements and the pressures from invasive species, pollution, altered hydrology, and climate change.nThe implications for conservation action from these findings are clear. This includes addressing invasive species, increasing the number of native fruit-bearing trees for hornbills and barbets, and designing specific vegetation buffers for areas near light sources to reduce light pollution. The 1.87 hectares identified as Critical and High conservation priority must be addressed with specific and immediate action, possibly by redesigning boardwalks to reduce foot traffic in these areas, installing signs to inform visitors about the impact of their activities, and working with neighboring landowners to modify light sources and/or screen the area with vegetation. The take-home message from these observations is clear. Small urban wetland areas must be considered serious contenders for conservation action. These areas offer irreplaceable biodiversity and potential for ecosystem services, environmental education, and the promotion of nature connection for the benefit of the broader metropolitan area.

6.3 Integrative Data Science Framework

This research demonstrates an effective application of new data science techniques that bring together citizen science data, satellite data, and advanced statistical modeling to address the challenge of biodiversity conservation in resource-scarce cities. The hierarchical Bayesian occupancy modeling provides state-of-the-art inferences that address the problem of imperfect detection in observational data. Unlike the potentially misleading single-value approach of simple presence/absence analysis, the method provides unbiased probability estimates with well-defined uncertainty through the use of posterior probability distributions. The use of satellite-derived environmental covariates at 10 meters resolution also helps us move beyond simple correlative species distribution modeling and provides a more mechanistic view of the factors that influence species suitability in the environment. This is important for prediction under future scenarios and for identifying management interventions that could be targeted to improve biodiversity outcomes. Finally, the Conservation Priority Index provides an integrated view of biodiversity metrics and threat factors in spatially explicit rankings. This allows complex analysis to be translated into simple and accessible recommendations that non-statistically sophisticated users can follow. An interactive web-based dashboard also provides access to the modeling results without the need for programming or statistical analysis expertise. Finally, the use of open-source tools for the entire workflow—Python, R, Google Earth Engine, and Streamlit—means that the methodological framework is easily adaptable to a wide range of species and ecosystems and could be applied to the growing number of urban protected areas around the world that face similar challenges.

6.4 Study Limitations and Future Directions

Despite the quality checks in place, citizen science data carries a lingering sense of "unevenness," including differences in observer skill levels, survey methods, and a bias toward areas near accessible areas. While the detection probability models attempt to address these factors, there may be a subtle bias in the estimates of occupancy. The concentration of survey points in the boardwalks near the buffer zone may cause a spatial bias in the estimates of occupancy in the core zone. However, the fixed effects of each zone help address these differences in a systematic manner. The satellite imagery at a scale of 10-500 meters may not be sufficient to pick up micro-scale features of the environment that might be important at the territory or nest site scale. It appears that the addition of drone-based hyperspectral imagery or LiDAR-derived canopy characteristics may be useful in future studies. The citizen science survey effort over seven years provides a good window of time, but it may not be sufficient to pick up long-term trends of climate change or differentiate between real trends and year-to-year fluctuations related to El Niño-Southern Oscillation or multi-year climate variability related to monsoon activity. The detection probability models accounted for survey time, time of day, and phenology. However, there may be additional factors at play in the survey process that would benefit from additional covariate information. These factors might include weather conditions, fatigue of the observer team members, and background noise. The future research should focus on systematic surveys with an even stratification of effort across the core and buffer zones. These should be conducted in a standardized manner with trained observers. This would enable direct comparisons with the citizen science results and would help in understanding the differences in detection probabilities across observer groups. It would also be useful to add autonomous recorders to gather acoustic data in addition to visual detections. These would be especially useful in species that are cryptic or vocal and may be underrepresented in eBird or similar citizen science platforms.

7. Conclusion

The study employed a novel blended technology-based method to investigate bird species in Mangalavanam Bird Sanctuary, a small yet biologically significant wetland situated in the metro region of Kochi. By using citizen science reports, satellite imagery, and hierarchical Bayesian models of bird occupancy, the study estimated areas where bird species are likely to prosper, identified areas of high conservation value, and monitored bird population trends. The study found that different bird species react differently to their environment. Wetland species react strongly to the availability of surface water. Forest species react strongly to vegetation density and human disturbance. On the other hand, urban bird species react very little to changing environments. The study identified areas of high conservation value amounting to 1.87 hectares or 24% of the total area of the bird sanctuary. These areas are mostly inside the core zone of high biodiversity and functional diversity values, and inside the inner edge of the buffer zone due to high levels of light pollution and habitat degradation. The study also found temporal trends in bird species populations. It found that three endemic Western Ghats forest bird species are declining at rates over 30%. These species need urgent conservation action. On the other hand, five urban bird species are showing positive trends. These may be an indication of ecological simplification. The study results show a high predictive power of the model with an AUC value of between 0.81 and 0.89 for all bird species. It also shows a high agreement of 82% with independent field checks.

These findings demonstrate the value of urban wetlands, no matter how small, in biodiversity conservation—value that is genuine but currently under threat. It demonstrates the value of very small protected areas in hosting surprisingly biodiverse species assemblages in the face of development pressures. It also provides a broadly applicable method of biodiversity tracking in urban areas with limited resources using citizen science data, satellite imagery, and free analysis tools. Future study should expand the time series record to capture the effects of climate change, use higher spatial resolution satellite imagery to analyze microhabitats in more detail, and include standard survey methods to validate citizen science results. It would also be useful to develop predictive models of biodiversity responses to different management options or future environmental changes. Finally, the interactive tool used to bring these analyses together is an important step in the development of accessible ecological modeling techniques. It allows conservation professionals, those managing a sanctuary or protected area, and policymakers to explore results without needing statistical expertise. It provides a tool for evidence-based conservation decisions using quantitative ecological science.

References

1. C. del Pilar Buenaño-Mariño, J. Sabán, E. Barba, and J. A. García-Esparza, “Urban form shapes bird niches: insights from the European green capital 2024,” Biodiversity and Conservation, vol. 34, pp. 1857–1879, 2025.
2. D. Mederer, T. Kattenborn, E. Cherif, et al., “Unraveling the seasonality of functional diversity through remote sensing,” Communications Earth & Environment, vol. 6, 2025.

3. J. Deng, M. Wang, et al., “Datasets and benchmarks for machine learning research,” Advances in Neural Information Processing Systems (NeurIPS), 2023.
4. MDPI, “Forests – Volume 16, Issue 1,” Forests, vol. 16, no. 1, 2025.
5. A. Research Team, “Ecological monitoring and biodiversity modeling using environmental datasets,” Scientific Reports, 2024.
6. B. L. Sullivan et al., “eBird: A citizen-science approach to global bird monitoring,” Biological Conservation, vol. 169, pp. 31–40, 2014.
7. B. L. Sullivan et al., “The eBird enterprise: An integrated approach to development and application of citizen science,” Biological Conservation, vol. 142, pp. 2282–2292, 2009.
8. ESA, “Sentinel-2 User Handbook,” European Space Agency, 2015.
9. C. D. Elvidge et al., “VIIRS night-time lights,” Remote Sensing, vol. 9, no. 8, 2017.
10. J. D. Nichols et al., “Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence,” Ecology, vol. 83, pp. 2248–2255, 2002.