Received: 7 July 2024 | Accepted: 1 April 2025

DOI: 10.1111/1365-2664.70100

Journal of Applied Ecology

RESEARCH ARTICLE

Biodiversity, Planning and Development – Towards Best Practice

A predictive approach to assess urban biodiversity and plan for future development scenarios

Xiao Ping Song[1,2] | Edwin Y. W. Tan[1,3] | Rachel S. K. Lee[1] | Hong Jhun Sim[1] | Justin K. H. Nai[1,3] | Jie Yi Chan[1] | Sherry M. X. Hung[1] | Shao Hua Ng[1] | Emmanuel S. C. Goh[1,3] | Chloe Y. T. Tan[1] | Yong Kiat Chua[4] | Audrey Xu[4] | Leonard Cai[4] | Darren C. J. Yeo[1] | Hugh T. W. Tan[1] | Kwek Yan Chong[1,3]

[1]Department of Biological Sciences, National University of Singapore, Singapore, Singapore

[2]Department of Architecture, National University of Singapore, Singapore, Singapore

[3]National Parks Board, Singapore, Singapore

[4]Housing and Development Board, Singapore, Singapore

Correspondence
Kwek Yan Chong
Email: kwek@nus.edu.sg

Funding information
Ministry of National Development-Singapore

Handling Editor: Ian Thornhill

Abstract

1. Protecting and enhancing biodiversity in urbanized areas is recognized as an important priority. To achieve this through urban planning, there must be empirically derived tools to predict biodiversity at the appropriate spatial scales and resolutions, given various options in urban designs to compare the expected biodiversity outcomes and make optimal decisions.
2. We demonstrate how this can be done by developing models that predict the expected species densities or 'alpha diversity' in urban landscapes for four animal groups: birds, butterflies, odonates and amphibians, based on assemblage data from spatiotemporally replicated surveys conducted in the tropical city of Singapore. We demonstrate two use cases for these predictive models: citywide assessment and future scenario planning.
3. For citywide assessment, sub-city 'towns' (equivalent to districts or suburbs elsewhere) were compared and benchmarked relative to all other towns, based on the average species densities as indicators of habitat value for each of the four animal groups.
4. For future scenario planning, four development scenarios were compared and the compatibility of vector-type planning layers with the models was tested.
5. An open-source R package, *biodivercity*, was developed that would facilitate the use of the same workflow elsewhere: to build, apply and validate predictive models elsewhere given similar available empirical data.
6. *Synthesis and applications*. The models developed can also be examined to generate recommendations for further actions that can improve biodiversity across different spatial scales. These techniques can be incorporated into current planning practices to achieve a more quantitative and performance-based approach to enhancing biodiversity at fine spatial scales in human-dominated landscapes.

KEYWORDS
biodiversity conservation, heat maps, indicators, predictive modelling, re-development, urban ecology, urban planning

1 | INTRODUCTION

Given the impacts of urbanization on biodiversity, there has been increasing interest behind the conservation of biodiversity in cities (Garrard et al., 2018; Ikin et al., 2015). Urban biodiversity also presents unique opportunities to contribute to environmental education, improve ecological literacy and contribute to the well-being of urban residents as they interact with nature in cities (Carrus et al., 2015). This has led to enthusiasm in biodiversity-sensitive/friendly urban design (Garrard et al., 2018; Ikin et al., 2015) and blue-green infrastructure (Perrelet et al., 2024) that conserve and enhance biodiversity in urban areas while providing ecosystem services to people.

Recommendations of design principles and interventions to conserve and enhance biodiversity must be grounded in evidence. Numerous studies have examined the effect of urban landscapes on biodiversity, especially those of urban greenery. Animal groups that have often been studied in urban areas include birds (Marzluff, 2017), butterflies (Ramírez-Restrepo & MacGregor-Fors, 2017), odonates (i.e. dragonflies and damselflies; Villalobos-Jiménez et al., 2016) and anurans (i.e. frogs and toads; Hamer & McDonnell, 2008); for example, tree foliage and shrub cover are commonly associated with higher bird species richness and diversity (Brunbjerg et al., 2018; Chong et al., 2014, 2019; Dietzel et al., 2024; Ortega-Álvarez & MacGregor-Fors, 2009). Water-dependent animals such as odonates and amphibians are affected by both aquatic and terrestrial vegetation near water bodies, alongside other human factors such as traffic and pollution (Hamer & McDonnell, 2008; Villalobos-Jiménez et al., 2016). Different animal groups may respond differently to these aspects of the urban landscape, not just in terms of quantity or cover, but also in terms of aspects of spatial configuration, such as the degree of fragmentation in the types of land cover (Alberti & Wang, 2022; Beninde et al., 2015; Hamer & McDonnell, 2008). Beyond land covers that represent habitat needs such as vegetation and water, the built environment (i.e. roads and buildings) may also be associated with urban animals. For example, roads and traffic have been found to negatively impact many animal groups (Alberti & Wang, 2022; Chong et al., 2014, 2019; Hamer & McDonnell, 2008; Plummer et al., 2020; Ramírez-Restrepo & MacGregor-Fors, 2017). Buildings and roads are artificial structures with which animals will interact differently, and represent the degree of urbanization, such as the intensity of human activities that may either cause stress or create opportunities for animals, or sealed surfaces that affect hydrological regimes or are sources of deposition and subsequent runoff of pollutants (Alberti & Wang, 2022; Casanelles-Abella et al., 2021; Cooper et al., 2023, 2024; Dietzel et al., 2024; Hamer & McDonnell, 2008; Marzluff, 2017; Plummer et al., 2020; Villalobos-Jiménez et al., 2016). The challenge is to identify the most relevant scale at which each of these myriad factors operate to affect biodiversity (Brunbjerg et al., 2018; Chong et al., 2019), in order for urban planners to have the flexibility to juggle the needs of biodiversity and people.

To properly assess the performance of biodiversity conservation efforts in cities, an approach is needed that can generate spatially explicit assessments of the state of biodiversity across an urban landscape (Casanelles-Abella et al., 2021), at spatial scales or resolutions compatible with that used in urban planning. Remotely sensed landscape data is now frequently used as inputs to predictive spatial models, allowing changes in biodiversity to be predicted across time and over large areas and at a spatial granularity required to inform planning decisions (Cavender-Bares et al., 2022). With the use of publicly available remotely sensed landscape data, high-spatial granularity can be achieved with relatively limited resources. One common example of its use in predictive models of biodiversity are species distribution/occupancy models, which can provide spatial predictions of species occurrences across landscapes, based on models built from the known locations of observed species and surrounding landscape conditions (Baker et al., 2021; Casanelles-Abella et al., 2021). However, species distribution models typically focus on only a single species of interest. On the other hand, the total number of species that can be expected at a location as a metric of measuring habitat value for multiple species can be predicted in the same way by models built from the observed number of species and the characteristics of the surrounding landscape (e.g. Brunbjerg et al., 2018; Casanelles-Abella et al., 2021).

Other than assessing existing landscapes, there is also a need to plan for future scenarios, that is, evaluate the effectiveness of actions aimed at promoting conservation or mitigating the loss of biodiversity. Urban design and planning involve the evaluation of multiple design scenarios, where biodiversity is but one factor amidst a multitude of other constraints and considerations (Chakraborty & McMillan, 2015). Being able to produce spatially explicit predictions of the consequences to biodiversity for various urban development scenarios will better allow urban planners to include biodiversity in their design and decision-making process (Brunbjerg et al., 2018; Cooper et al., 2023, 2024; Plummer et al., 2020). However, because such future landscapes do not yet exist, snapshots of remotely sensed data cannot be used to generate the landscape elements required to make predictions of biodiversity. Since remotely sensed data cannot be used, models may instead be trained based on landscape elements compatible with those produced from prospective designs (e.g. planning layers of vegetation and built elements). This would result in separate types of data being available from urban planning versus that used in the predictive models. Compatibility between the two sources of data should thus be assessed to ensure that the claims made during the design stage align well with reality after implementation.

This paper presents the development of models that enable present assessment and future planning of landscapes for urban biodiversity at fine spatial scales. We adopt an empirical approach that directly assesses the 'habitat value' (i.e. ability to support biodiversity) of landscapes, by performing predictive modelling of animal groups surveyed in the tropical city of Singapore. The objectives were to: (a) develop a workflow to build models that predict the species density (number of species) of four animal groups commonly studied and used as bioindicators in urban areas; (b) derive urban planning guidelines from the type of landscape attributes identified to be important predictors of species density in these models; (c) demonstrate how the predictive models can be used to compare biodiversity performance across all 'towns' (here referring to sub-city districts or suburbs) in the study city based on the predicted diversity of the animal groups; (d) demonstrate the use of predictive models to compare alternative scenarios in urban design and planning.

2 | MATERIALS AND METHODS

2.1 | Data collection

2.1.1 | Sampling sites

The study is conducted in the equatorial city-state of Singapore (1.3° N, 103.8° E) located in Southeast Asia. The climate is tropical and relatively aseasonal, with daily temperatures between 24 and 32°C throughout the year and mean annual rainfall of about 2100 mm. While Singapore would have been almost completely covered by tropical forest in the early 19th century, deforestation for timber, firewood and agriculture rapidly followed after the establishment of the British colony in 1819 (Corlett, 1992). By the end of the 19th century, less than 10% of the original forest remained. From the mid-20th century, Singapore gained independence and moved away from plantation agriculture. Today, Singapore is one of the most densely populated places in the world; in 2023, the total population is 5.9 million and the total land area is about 735 km^2. During the course of post-colonial urbanization, Singapore also established urban planning frameworks that sought to maintain high levels of greenery despite high population densities (Tan et al., 2013). About 22.1% of Singapore's land cover today is natural vegetation—almost entirely secondary regrowth vegetation except for small fragments of primary forest—while the remaining land is considered built-up and urban but with greenery and greenspaces consisting of cultivated vegetation (Gaw et al., 2019). Almost 80% of households are in high-rise public housing, organized in townships.

Four animal groups were surveyed at six towns (as officially delimited) in Singapore (Appendix S1: Figure S1) conducted with a research permit from the National Parks Board, NP/RP16-093. For each town, surveys were conducted every 2 months across a 1-year duration. Unless otherwise stated, all sampling design, data processing and analyses in this study were performed using the open-source statistical software R v. 4.1.1 (R Core Team, 2021). Point locations for surveys were randomly sampled across two land cover strata—natural vegetation and urban cover, following past studies that showed the key differences in urban bird and butterfly communities between these two main land cover types in Singapore (Chong et al., 2014, 2019). Before each survey year, 'natural vegetation cover' (forest patches) within the respective study towns was preliminarily delineated in Google Earth; large water bodies were then excluded from the remaining area to form 'urban cover' (which includes cultivated vegetation). The two land cover strata were used to randomly generate sampling points at a density higher than the target of 1 point per 50 ha, to provide backup points upon inspection and on-site reconnaissance. Each sampling point was checked for accessibility and the presence of water bodies nearby. Proximity to permanent or ephemeral water bodies were required for surveys of two out of the four animal groups—odonates and amphibians. Sampling points were adjusted up to 50 m to avoid safety hazards, or to increase their proximity to water bodies to be within 20 m, owing to the relatively low chance of points being randomly generated close to water bodies. For each survey round, the number of points close to water bodies was fixed at 10 per town.

2.1.2 | Animal surveys

Four animal groups were assessed in this study: Birds (Aves); butterflies (Insecta: Lepidoptera excluding moths); dragonflies and damselflies (Insecta: Odonata; hereafter 'odonates'); and frogs and toads (Amphibia: Anura; hereafter 'amphibians'). These animal groups were selected as they are well-studied in Singapore and hence the species can be sufficiently detected and easily identified with available field guides (e.g. Baker & Lim, 2008; Khew, 2015; Orr, 2005; Yong et al., 2016) and the help of relevant experts (e.g. authors of the respective sections of the latest Singapore Red Data Book; see Davison et al., 2024). They are also among the most studied urban bioindicators globally. Thirty-minute surveys were conducted by two observers for each animal group. The bird and butterfly surveys were conducted at all sampling points, while the odonate and amphibian surveys were only conducted at points that were close to water bodies (blue and pink points in Appendix S1: Figure S1). Counts of each species were recorded within a fixed radius of the sampling point; the radius of observations was set at 50 m for birds, and 20 m for the other three animal groups. No animals were captured; visual and aural surveys were conducted for birds (07:00–09:30 h) and amphibians (20:00–22:00 h); and visual surveys were conducted for butterflies (09:30–12:00 h) and odonates (14:00–16:00 h). Surveys were conducted during fair weather conditions. For points close to ephemeral water bodies, the odonate and amphibian surveys were conducted within 24 h after a rain event (recorded at the weather station nearest to the sampling point). To account for seasonal variation, surveys were repeated every 2 months at each sampling point, across a 1-year duration (six surveys). For some survey points, there were disruptions such as commencement of construction work that meant that the full set of six surveys could not be completed; only the points with six completed surveys were used in the analysis (sample size $n = 134$ for birds and butterflies; $n = 50$ for amphibians and odonates). For each animal group, the total number of species observed across the six repeated surveys was tallied at each sampling point,

Journal of Applied Ecology

referred to as 'species density' (see Gotelli & Colwell, 2001). No ethical approvals were required for this research.

2.1.3 | Remote sensing of landscapes

Remotely sensed data can reduce manual effort to derive landscape variables, streamline data collection and allow findings to be generalized across the wider landscape. Features of the built environment (i.e. buildings, roads) were derived based on OpenStreetMap (OSM) data (OpenStreetMap contributors, 2020), downloaded using the R package *osmextract* (Lovelace, 2019; Appendix S1: Table S1 and Figure S2c) and then ground-truthed on-site during the year of the animal surveys and amended if necessary for accuracy. Sentinel-2 satellite images were downloaded using the R package *sen2r* (Ranghetti et al., 2020), to derive classes of land cover such as vegetation and water cover at a 10-m pixel resolution (Appendix S1: Figure S2b). Vegetation cover was then sub-categorized into canopy (>2 m height) and short (≤2 m height) vegetation cover (Appendix S1: Figure S2d), based on height from ground level (Appendix S1: Figure S2a). The height from ground level was represented by the normalized digital surface model (nDSM) at a 0.5-m pixel resolution, derived from the LiDAR-based digital surface model (DSM) and digital terrain model (DTM) obtained from the Singapore Land Authority from airborne scans conducted in 2019. Landscape patterns were then quantified for each land cover class, summarized at each sampling point (Appendix S1: Figure S2 and Table S2). The R package *landscapemetrics* (Hesselbarth et al., 2019) was used to calculate each metric at the 'class' level.

2.2 | Predictive modelling

2.2.1 | Relationships between landscapes and species density

For each animal group, species density (number of species) at sampling points was modelled as counts, with the attributes of the surrounding urban landscape as predictors. With the large number of landscape predictors at multiple buffer radii (e.g. Appendix S1: Table S2), random forest models were first used to select useful predictors, based on the relative importance of variables in their ability to improve model performance, that is, averaged variable importance (Arthur et al., 2010; Li et al., 2017). The random forest algorithm is a machine learning technique that can handle and reduce the effects of overfitting, as well as the collinearity and selection bias of predictors (Hothorn et al., 2006). It has been shown to select useful predictors for generalized linear models, particularly using averaged variable importance values (Arthur et al., 2010; Yeo et al., 2014). Two random forest hyperparameters were tuned via 10-fold cross-validation of the training data (75% of the full dataset), using the grid search technique: (a) number of predictors randomly sampled per split and (b) the number of observations needed to continue splitting nodes. Model performance was based on root-mean-square error (RMSE). The workflow followed the *tidymodels* framework implemented in R (Kuhn & Wickham, 2020), using the R package *ranger* (Wright & Ziegler, 2017) to fit the random forest models. For each predictor variable, the average importance value across 200 random forest scenarios was calculated and used to explore the effect of spatial scale (i.e. relative importance of each buffer radius) for each landscape predictor; the sampling point buffer radius with the highest average importance value was selected. Then, among these predictor variables at the selected buffer radius, only those with 95% confidence intervals of the importance value not overlapping with zero were considered as useful predictors (see Appendix S2).

Next, generalized linear mixed-effects models (GLMMs) with Poisson error structures and natural logarithm link function (Bolker et al., 2009) were fitted with species density as counts and the 'town' specified as a random intercept to control for spatial non-independence between points in the same town. GLMM selection was based on an information theoretic multi-model inference approach; different combinations of predictors derived from remotely sensed datasets (Appendix S1: Tables S1 and S2) pre-selected after random forest analyses were supplied as possible candidate models using the '*dredge*' function from the R package *MuMIn* v. 1.43.6 (Bartón, 2019). The predictors were centred and scaled prior to model fitting. Combinations of predictors that represented similar landscape characteristics were avoided during the model fitting process. For example, for each land cover class, only one of each metric type could be present in the fitted model (see Appendix S1: Table S2) and similar landscape predictors at different buffer radii were not allowed. In addition, to avoid overfitted GLMMs given the sample size, the maximum number of predictors (k) in any GLMM was restricted to nine (i.e. an n:k ratio of 14.9 for birds and butterflies and 5.6 for amphibians and odonates). Akaike's Information Criterion corrected for small sample sizes (AIC_c) was used to rank the models. The differences in AIC_c value between each model and the model with the lowest AIC_c was calculated (ΔAIC_c). The best performing models were defined as those with $\Delta AIC_c < 2.0$ (Burnham & Anderson, 2002). From ΔAIC_c, the relative likelihood of the model ($rL_i = e^{-0.5 \times \Delta AIC_c}$) and the model weight ($w_i = \frac{rL_i}{\sum rL}$) was calculated.

2.2.2 | Cross-validation

The predictive power of the best performing models was assessed with cross-validation, that is, randomly divide the data into two sets for model training and testing (Fielding & Bell, 1997; Franklin, 2010). In each of 1000 iterations, the training data (75% of the original data) was first used to select and calibrate the best models, while the testing data (remaining 25% of the data) was used to validate the predictions made by the models. The model-averaged predictions were made using the R package *AICcmodavg* v2.3.3 (Mazerolle, 2023). The 95% confidence interval of the mean errors of the iterations (from 1000 bootstrapped resamplings) were checked if it overlapped zero (indicating no bias) while the proportion of RMSEs from the iterations that were less than the observed standard deviation of species densities (indicating improvement in precision) was calculated.

2.2.3 | Spatial predictions across the landscape

Model-averaging across the set of best performing models for each of the four animal groups was used to predict species density across regions where landscape data were available. Landscape variables that were important predictors within these models were mapped continuously across the area of interest at a specified spatial resolution. A grid was then overlaid across the landscape and the respective landscape variables used as predictors were calculated at the appropriate radius (spatial scale) from the centre of each grid square (Appendix S1: Figure S3a). For each animal group, predictions at each centre were made using the best performing models (e.g. Figure S3b).

With the availability of citywide remotely sensed data, predictions from the model were made directly at a pixel resolution of 100 m and pixel values were averaged within 23 towns in Singapore; the frequency distribution of the resulting averaged values was used to compare the performance of each town relative to others in the city.

Predictions were also made for a future urban re-development plot within the town of Queenstown in Singapore, based on four urban design scenarios. Scenario 1 was sparsely planted; Scenario 2 was lushly planted; Scenario 3 was lushly planted with rooftop greenery; and Scenario 4 was moderately planted with a connecting strip of natural vegetation (examples in Figure 2). Relevant vector layers generated in the design scenarios were rasterized and thereafter used to calculate the respective landscape predictors (Appendix S1: Figure S2e and Table S2). Turf and shrub were both classified as short vegetation, while trees and natural vegetation were classified as vegetation canopy. Since this rasterization is expected to introduce error, accuracy assessments were also performed for predictions generated based on this conversion process. During each survey round, the geolocation and species of cultivated trees, palms, shrubs and turf within a 50 m radius of each sampling point were mapped manually on-site. Trees and palms were recorded as individual points, while shrubs and turf were recorded as polygons. These vector (manually mapped) data were rasterized and then replaced the remotely sensed data within the buffer region (see Appendix S1: Figure S2e). The amended landscape components were then summarized at each sampling point and used with the predictive models (trained with remotely sensed data) to predict the number of species. These predictions were compared with the actual number of species recorded at the sampling points.

3 | RESULTS

A total of 118 bird taxa, 98 butterfly taxa, 34 odonate taxa and 13 amphibian taxa were encountered and identified to species. Most animal encounters could be identified to species (birds: 92.7%; butterflies: 90.3%; odonates: 99.9%; amphibians: 100%). Swifts (Family: Apodidae) and grass yellows (*Eurema* spp.) constituted the lion's share of the unidentified encounters of birds (94.3%) and butterflies (78.7%) respectively. The mean ± standard deviation in species densities for the four groups were: birds—24.9 ± 7.9; butterflies—13.9 ± 6.0; odonates—6.2 ± 3.8; amphibians—4.0 ± 2.3.

3.1 | Relationships between landscapes and species density

The predictors selected in the best sets of GLMMs (Tables S3–S6; Figures S4–S11; summarized in Table 1) can be interpreted to understand the landscape characteristics associated with animal diversity and hence derive a variety of recommendations for urban design and planning for various animal groups at specific spatial scales (Table 2; Appendix S1). We limited ourselves to the variables with sum-of-weights >0.5; the sum-of-weights of a variable is the sum of the w_i's of all models that contain that variable and can be interpreted as the probability that the variable is in the true best model, that is, sum-of-weights >0.5 means that the variable is more likely to be in the true best model than not.

Total building volume was negatively associated with birds at 100 m scale (Figures S4 and S5) and with odonates at 400 m scale (Figures S8 and S9), while building floor area ratio was positively associated with butterflies at 400 m scale (Figures S6 and S7; Table 1); taken together, we interpret that total building volume should be minimized across scales, while building floor area can be maximized at the larger scale of 400 m (Table 2). Bird diversity was positively associated with larger mean area of vegetation patches (at 200 m scale) and simpler patches of canopy patches (at 400 m scale; Figures S4 and S5), while the diversities of other animal groups generally showed positive relationships with continuous, compact, simpler vegetation at smaller scales (100 m for butterflies and odonates; 200 m for amphibians) but were predicted to decrease with aggregation of vegetation at larger scales (200 to 400 m for butterflies; 400 m for amphibians; Figures S6–S11; Table 1). Again, taken together, we interpret that vegetation should be in larger and less fragmented patches at smaller scales of 100–200 m but can be disaggregated at larger scales of 200–400 m (Table 2).

3.2 | Cross-validation

Model validation using 3:1 train-test split showed that the 95% confidence intervals of prediction errors centred around zero for all animal groups, indicating that the models were unbiased (Appendix S1: Figure S12a). The proportions of RMSE values smaller than observed standard deviations were: birds—97%, butterflies—81.6%, odonates—84.3%, amphibians—91.1% (Figure S12b), generally indicating acceptable precision of the predictions viz. observed variation in the raw data.

TABLE 1 Summary of the important predictors of species densities of each animal group at specific spatial scales.

Landscape feature/cover class	Metric	Scale	Birds	Butterflies	Odonates	Amphibians
Building	Volume	≤100 m	–			
		≤400 m			–	
	Floor area ratio	≤400 m		+		
Road	Lane density	≤200 m	–		+	
Vegetation	Patch area (mean)	≤200 m	+			
	Landscape shape index	≤400 m		–		–
Canopy vegetation	Perimeter-area ratio (mean)	≤200 m				+
		≤400 m	+			
	Largest patch index	≤100 m		+		
	Contiguity index (mean)	≤200 m		–		
	Division index	≤400 m		+		
Short vegetation	Edge density	≤100 m				–
		≤200 m		+		
	Related circumscribing circle (mean)	≤100 m			–	

Note: Based on the predictors within the best models with sum-of-weights >0.5 (Appendix S1: Figures S1–S11, Tables S3–S6); '+' indicates a positive relationship while '–' indicates a negative relationship.

TABLE 2 Design recommendations to achieve a higher species density for each animal group.

Animal group	Recommendations (and scale)
Birds	Total building volume should be minimized (≤100 m) Area of each vegetation patch should be maximized (≤200 m) Road lane density should be minimized (≤200 m) The shape of each patch of vegetation canopy should be simple (≤400 m)
Butterflies	Patches of short vegetation should be disaggregated (≤200 m) Vegetation cover should be disaggregated (i.e. less compact in arrangement) (≤400 m) Building floor area can be maximized across a larger region (≤400 m) Vegetation canopies should be continuous and cover a large area (i.e. dominate the landscape) (≤100 m) Vegetation canopies should be disaggregated (≤400 m) Vegetation canopies should be disconnected from each other (≤200 m)
Odonates	Total building volume should be minimizsed (≤400 m) Patches of short vegetation should each be compact in shape (i.e. circular) (≤100 m) Road lane density can be maximized across a larger region (≤200 m)
Amphibians	Patches of short vegetation should be disaggregated (≤100 m) Vegetation cover should be disaggregated (≤400 m) The shape of each patch of vegetation canopy should be simple (≤200 m)

Note: Appendix S1: The numbers within the parentheses denote the specific buffer radius applicable to each recommendation.

3.3 | Citywide assessment

Spatial predictions of species density (i.e. number of species per pixel) were made within 23 towns across the city, at a pixel resolution of 100 m (Figure 1). The town Punggol with numerous forest patches during the year of animal surveys had high mean species densities for birds (27 species), butterflies (10 species), odonates (7 species) and amphibians (2 species) (see Appendix S1: Table S7). Conversely, highly developed towns were predicted to have fewer species across the four animal groups (Appendix S1: Table S7). To compare across all towns by summing up the mean pixel value for the four animal groups may bias comparisons between towns toward animal groups that have a larger number of species (e.g. birds), compared to those that have fewer species (e.g. amphibians). Our

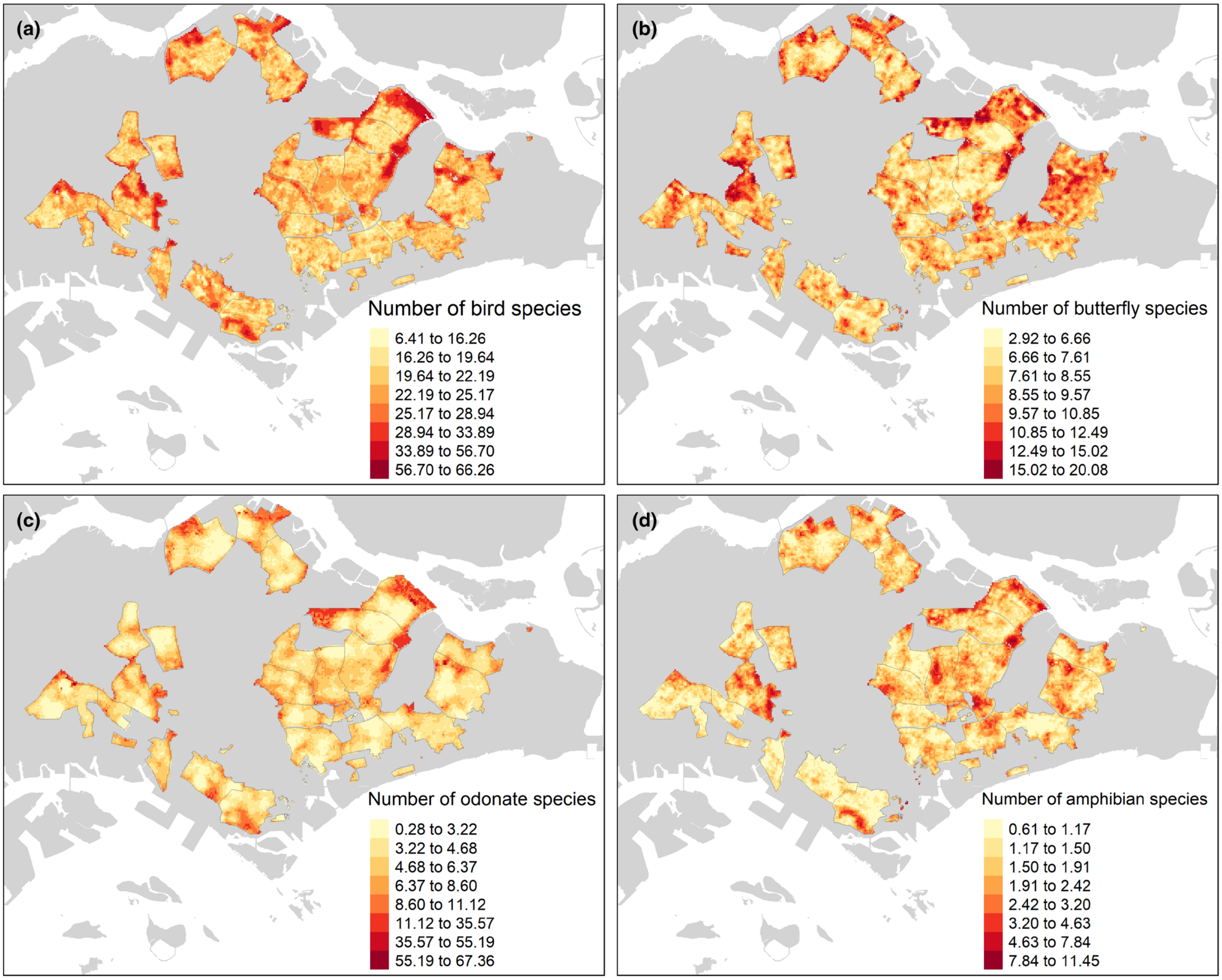

FIGURE 1 Spatial predictions of species density for (a) birds; (b) butterflies; (c) odonates; (d) amphibians within towns in Singapore, based on models trained with remotely sensed data (2019–2022). The pixel resolution is 100 m. The colour palettes were binned based on the Jenks natural breaks classification. The summarized and distribution of values per town can be found in Appendix S1: Table S7 and Figure S13.

approach assigns equal representation and weightage (importance) across all four animal groups, by assigning each town a score of 0–4 per animal group (Appendix S1: Table S7), based on the deviation of species density from the 'mean' value across all towns (Appendix S1: Figure S13). For each town, the total species densities across all four animal groups and their corresponding total scores are shown in Appendix S1: Table S7; towns that have a higher total species density may not necessarily have a higher total score.

3.4 | Comparisons between multiple future scenarios

For each animal group, spatial predictions can also be made for future scenarios, using vector data supplied from proposed designs and converting them to rasters for use with the predictive models. Lush planting with rooftop greenery resulted in the highest species density across all four animal groups (Scenario 3 in Figure 2). Sparse planting generally resulted in lower numbers of species (Scenario 1 in Figure 2).

The conversion of manually mapped vegetation to rasters retained the correspondence with observed species densities at the sampling points, except for butterflies (Appendix S1: Figure S14). The 95% confidence band for the best fit line still encompassed the 1:1 line after rasterization for birds, odonates and amphibians, but for butterflies the model-averaged predictions no longer corresponded with observed species densities (Appendix S1: Figure S14).

4 | DISCUSSION

We demonstrated a workflow for developing predictive models for four animal groups: birds, butterflies, odonates and amphibians, using landscape metrics derived from remote sensing. We then drew

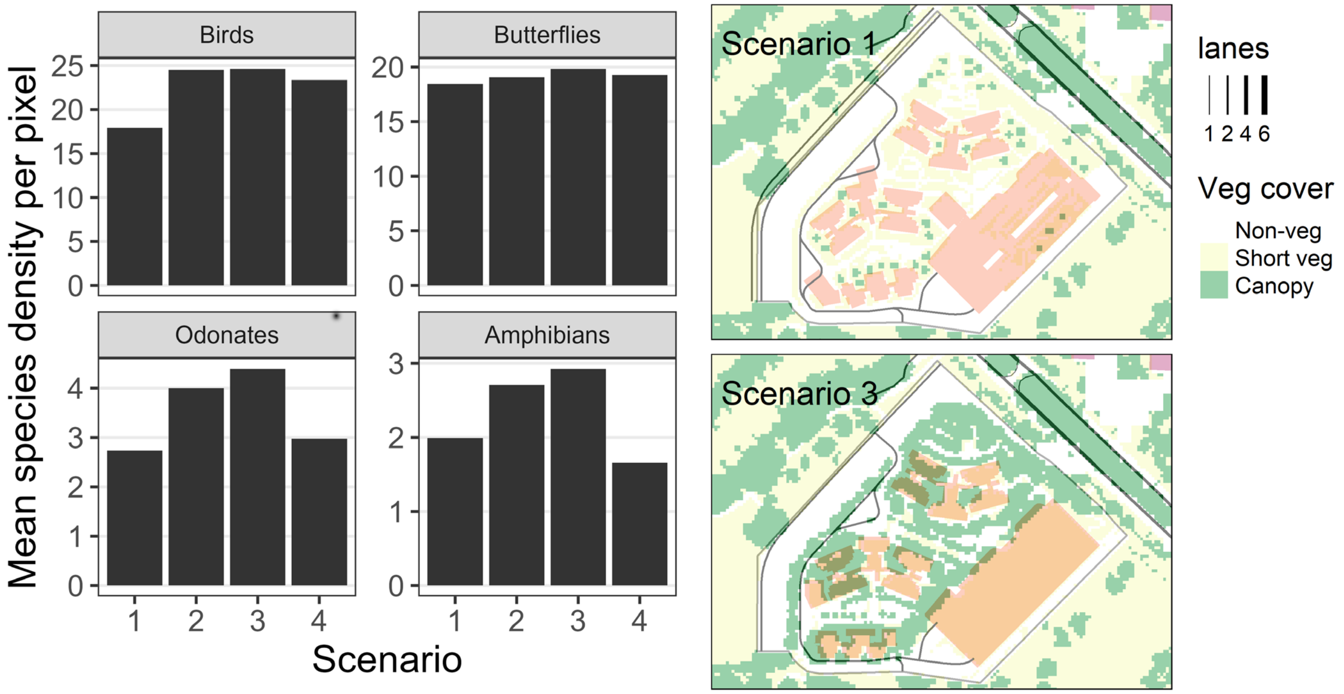

FIGURE 2 The mean species density per pixel for four future development scenarios. Contrasting landscape configurations of Scenarios 1 (sparse planting) and 3 (lush planting) are shown.

three sets of applications from these predictive models. First, the generation of a set of recommendations for urban design based on the most important predictors at their most predictive spatial scale, that is, evidence-based design recommendations to enhance biodiversity in urban areas. Second, we produced citywide predictions of present-day species densities as a possible metric for urban planners to quantitatively assess and benchmark the relative performance when summarized for different jurisdictions—in our case, residential towns—in terms of how biodiversity-friendly the landscape configurations were. Third, we demonstrated how predictions of future species densities can be used with these models to compare alternative urban design and planning scenarios, and at the same time highlighted the potential impacts of converting between the data types available from design and planning versus those used in building models and generating predictions from the models.

4.1 | Urban design recommendations

In the model selection process, metrics of spatial configuration of land covers were found to be more important predictors than simple proportional cover. Such measures of landscape configuration, while perhaps challenging to interpret, encompass more nuances on how the arrangement of each land cover class may influence species density (Appendix S1 Table S5). In addition to a large number of possible metrics to generate, each metric can also be derived at multiple spatial scales, of which some may be of better predictive value than others. This results in a large total number of possible predictors to be selected from while avoiding overfitted models (Casanelles-Abella et al., 2021; Dietzel et al., 2024). To do this, we adopted a two-step process, using machine learning in the first step to select the best spatial scale of each predictor and eliminate the likely least important predictors, and then fitting interpretable generalized linear mixed-effects models in the second step of model selection that also has the advantage of offering interpretability to the direction of the relationship to provide recommendations for urban design and planning.

However, it is worth noting that some recommendations may appear to be at odds with one another. For instance, even though canopy cover should be maximized at fine spatial scales (100 m), it may also be necessary to keep it disconnected/disaggregated at wider scales (200–400 m) (Table 1). This may be because continuous canopy cover at wider scales may not provide enough diversity in habitat structure for certain animal groups. For example, the homogeneous low-light environment may not provide the diversity of habitats (e.g. undergrowth, shrubbery) needed to support a range of amphibian and butterfly species (see recommendations in Table 2). To promote biodiversity in urban areas, it is thus crucial to consider the impact of spatial scale and the ecological requirements of different animal groups on the adoption of these design recommendations (Brunbjerg et al., 2018; Chong et al., 2019).

4.2 | Biodiversity assessment and benchmarking

Citywide predictions across all towns in Singapore aligned with expectations, both in terms of the maximum possible number of species per pixel (see legends in Figure 1), as well as the relative performance (ranking) of towns (Appendix S1: Table S7). Even though the predictors we used did not differentiate between natural and cultivated vegetation, towns with more forested areas at the time of the animal surveys showed a much higher species density across all four animal groups, while 'mature', densely built-up towns showed relatively lower species densities (Appendix S1: Table S7).

Landscape variables derived from remotely sensed data are readily available for download and processing as inputs to predictive models. Changes in biodiversity can thus be predicted across time and over large areas for monitoring purposes (Figure 2a–d). The tabulation of species densities and scores, if performed for multiple time periods, provides snapshots of the 'performance' of these city planning units (towns) across time. This provides a cost-efficient proxy of expected changes in biodiversity, alongside periodic field surveys for the testing and validation of model output. This framework can be extended to other types of human-dominated landscapes and contribute toward national reporting and monitoring of progress toward international biodiversity goals (Cavender-Bares et al., 2022).

The ranking of towns based on their total species densities or scores may also be used to prioritize towns for further actions to conserve and enhance urban biodiversity (Appendix S1: Table S7). However, it may not be obvious whether to enhance towns with already high scores, that is, through further reducing traffic or built intensity, or low scores, for example, by increasing the number of green spaces and the intensity of greenery. This is equivalent to land sparing, where development is concentrated in limited areas while setting aside areas dedicated for biodiversity conservation, versus land sharing, which involves interspersing natural habitats with human development (Soga et al., 2014). This land sparing versus land sharing approach is a central debate in biodiversity conservation, including in urban landscapes. While recent work (e.g. Soga et al., 2014) has suggested that a land sparing strategy (i.e. prioritize development of towns with low scores) may work best in compact and highly urbanized cities such as Singapore, urban development must also consider biodiversity assessments alongside other social, political and economic factors. Research suggests that a combination of both approaches may be most effective while considering specific context and characteristics of the area (Kremen, 2015). Our spatially explicit predictive model allows one to determine the relative performance of different areas and provide flexibility in adjusting the level of nuance that is crucial in fine-scale planning for biodiversity conservation and urban development.

4.3 | Model and data compatibility for future scenario planning

Among the four development scenarios in our study, lush planting with rooftop greenery resulted in the most species predicted across all animal groups (Figure 2). If higher granularity is required, scenarios may also be assessed at the level of individual pixels, allowing planners to identify specific areas with higher potential to support biodiversity and wildlife. In our predictions, we also assumed rooftop greenery to be equivalent to ground-level greenery, while a previous study in Singapore found that species richness of birds and butterflies decreased with increasing height of rooftop gardens to 50 m (Wang et al., 2017). Future studies can extend such results by quantifying the moderation of the effects of individual greenery elements with height so that corrections for the height of rooftop greenery can be incorporated into the predictions.

Urban planning often generates layers in the form of vector data (e.g. building polygons, road lines, tree point locations, shrub polygons). While it is possible to convert vectors from planning layers to raster data and incorporate the spatially broader context, we found that the data conversion process resulted in a loss of predictive ability for butterflies (Appendix S1: Figure S14). To overcome this, one solution could be to train models directly with manually mapped data, which can also incorporate information that cannot typically be captured by remote sensing, for example, differences in the types of vegetation (e.g. natural versus cultivated vegetation) and plant species information, as candidate predictors. This set of models would provide a more nuanced approach when considering trade-offs between development scenarios.

It was also surprising that water was not retained as an important variable for any of the taxa when odonates (Dietzel et al., 2024; Villalobos-Jiménez et al., 2016), amphibians (Dietzel et al., 2024; Hamer & McDonnell, 2008) and also birds (Wong et al., 2023) are expected to be associated with water bodies; this could have been because of the current limitations in using remote sensing to detect and quantify water bodies that are obscured by vegetation, small or only detectable after a major rainfall event such as stormwater channels, which could be overcome by additional manual mapping. However, manual mapping is manpower-intensive, and incorporating potential influences at a larger landscape scale will require information far beyond the boundaries of the scenario, which may not be typically available and hence incur additional costs to obtain. The models trained with remotely sensed data, on the other hand, easily extract information from the broader area around the development plot. Therefore, models trained with remotely sensed data would be preferred if the larger regional context is important, or if it is a priority to compare these future scenarios with existing landscapes across the city. Future developments in remote sensing such as new products at finer resolution from new satellites launched or unmanned aerial vehicle surveys may eventually be able to detect and quantify more features relevant for the monitoring and predictive modelling of biodiversity patterns.

4.4 | Future work and applications

An area of future work is the consideration of species relative abundances and community composition in fine-scale assessments. While species density (also known as *alpha* diversity) is an important aspect of biodiversity, urban landscapes can often be highly heterogeneous and span regions with unique environmental characteristics. Different sites with similar numbers of species may each have unique compositions of species communities, also known as *beta* diversity (Whittaker, 1972). Accounting for complementarity in species compositions at different sites will avoid the perverse outcome of creating urban ecological communities that may be locally diverse but are homogeneous across wider spatial scales (Chong et al., 2014; Sepkoski, 1988) thereby leading to lower total (*gamma*; Whittaker, 1972) diversity of the city. Beta diversity prediction and integration with species density will therefore be the next step to promote effective evaluation and decision-making that relates to biodiversity conservation.

A shortcoming of our current modelling approach is that we modelled observed species density as the response, but it does not take into account imperfect detection of species at each point. This is a problem addressed by hierarchical occupancy modelling in species distribution models, where data from repeated surveys within the same survey season or from distance sampling are used to estimate and account for how detectability of a species may vary together with predictors, thereby more accurately estimating the true probability of occurrence of the species (Devarajan et al., 2020; Iknayan et al., 2014). Recent extensions of single-species hierarchical occupancy models to joint species distribution or multi-species/community occupancy models, whereby multiple species or even whole communities are modelled together in addition to accounting for imperfect detection, will be able to produce estimates of species density through summing individual species occurrence probabilities, that is, a 'predict (species probabilities of occurrence) first, assemble later' approach as opposed to the 'assemble (species density) first, predict later' approach used in our study (Devarajan et al., 2020; Iknayan et al., 2014). Such models can also directly predict community composition and therefore simultaneously provide the functionality of incorporating beta diversity into design and planning of landscapes such as urban areas.

The methods developed in this study enable species density to be compared between sites and across time. Pixel-based spatial predictions of species density enable fine-grained assessments of the landscape by providing continuous values at a specified spatial resolution. Pixel values can also be summarized within the zones used in city planning to allow comparisons and benchmarking to be performed across these planning units. Our future work seeks to integrate the use of these predictive models and their specific recommendations into the larger toolbox of methods used to assess biodiversity in cities. In line with these efforts, we have begun development on an R package *biodivercity* (Song et al., 2022) to allow users to develop and apply such models for their own use cases and validate model results based on the data that they collect.

AUTHOR CONTRIBUTIONS

Kwek Yan Chong, Hugh T. W. Tan, Darren C. J. Yeo, Leonard Cai, Audrey Xu and Yong Kiat Chua conceived and designed the research; Xiao Ping Song, Edwin Y. W. Tan, Rachel S. K. Lee, Hong Jhun Sim, Justin K. Nai, Jie Yi Chan, Sherry M. X. Hung, Shao Hua Ng, Emmanuel S. C. Goh, Chloe Y. T. Tan collected the data; Xiao Ping Song, Edwin Y. W. Tan, Rachel S. K. Lee, Hong Jhun Sim, Justin K. Nai and Kwek Yan Chong performed the analysis; Xiao Ping Song and Kwek Yan Chong wrote the paper.

ACKNOWLEDGEMENTS

This research was funded by the Ministry of National Development (MND) Research Fund for the Built Environment. It was awarded to the National University of Singapore, in partnership with the Housing and Development Board.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available from the Dryad Digital Repository https://doi.org/10.5061/dryad.2fqz6131p (Song et al., 2025).

ORCID

Rachel S. K. Lee https://orcid.org/0000-0002-8358-8182
Kwek Yan Chong https://orcid.org/0000-0003-4754-8957

REFERENCES

Alberti, M., & Wang, T. (2022). Detecting patterns of vertebrate biodiversity across the multidimensional urban landscape. *Ecology Letters*, *25*, 1027–1045.

Arthur, A. D., Li, J., Henry, S., & Cunningham, S. A. (2010). Influence of woody vegetation on pollinator densities in oilseed Brassica fields in an Australian temperate landscape. *Basic and Applied Ecology*, *11*(5), 406–414. https://doi.org/10.1016/j.baae.2010.05.001

Baker, D. J., Maclean, I. M., Goodall, M., & Gaston, K. J. (2021). Species distribution modelling is needed to support ecological impact assessments. *Journal of Applied Ecology*, *58*(1), 21–26. https://doi.org/10.1111/1365-2664.13782

Baker, N., & Lim, K. (2008). *Wild animals of Singapore: A photographic guide to mammals, reptiles, amphibians, and freshwater fishes*. Draco Publishing and Distribution Pte. Ltd. and Nature Society (Singapore).

Bartón, K. (2019). *MuMIn: Multi-model inference*. https://cran.r-project.org/package=MuMIn

Beninde, J., Veith, M., & Hochkirch, A. (2015). Biodiversity in cities needs space: A meta-anlaysis of factors determining intra-urban biodiversity variation. *Ecology Letters*, *18*, 581–592.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*, 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Brunbjerg, A. K., Hale, J. D., Bates, A. J., Fowler, R. E., Rosenfeld, E. J., & Sadler, J. P. (2018). Can patterns of urban biodiversity be predicted using simple measures of green infrastructure? *Urban Forestry & Urban Greening*, *32*, 143–153.

Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. In *Model selection and multimodel inference* (2nd ed.). Springer.

Carrus, G., Scopelliti, M., Lafortezza, R., Colangelo, G., Ferrini, F., Salbitano, F., Agrimi, M., Portoghesi, L., Semenzato, P., & Sanesi, G. (2015). Go greener, feel better? The positive effects of biodiversity on the well-being of individuals visiting urban and peri-urban green areas. *Landscape and Urban Planning*, *134*, 221–228. https://doi.org/10.1016/j.landurbplan.2014.10.022

Casanelles-Abella, J., Chauvier, Y., Zellweger, F., Villiger, P., Frey, D., Ginzler, C., Moretti, M., & Pellissier, L. (2021). Applying predictive models to study the ecological properties of urban ecosystems: A case study in Zürich, Switzerland. *Landscape and Urban Planning*, *214*, 104137.

Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G. C., Schimel, D., Townsend, P. A., Ustin, S. L., Wang, Z., & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology & Evolution*, *6*, 506–519.

Chakraborty, A., & McMillan, A. (2015). Scenario planning for urban planners: Toward a practitioner's guide. *Journal of the American Planning Association*, *81*(1), 18–29. https://doi.org/10.1080/01944363.2015.1038576

Chong, K. Y., Teo, S., Kurukulasuriya, B., Chung, Y. F., Giam, X., & Tan, H. T. W. (2019). The effects of landscape scale on greenery and traffic relationships with urban birds and butterflies. *Urban Ecosystems*, *22*(5), 917–926. https://doi.org/10.1007/s11252-019-00871-9

Chong, K. Y., Teo, S., Kurukulasuriya, B., Chung, Y. F., Rajathurai, S., & Tan, H. T. W. (2014). Not all green is as good: Different effects of the natural and cultivated components of urban vegetation on bird and butterfly diversity. *Biological Conservation*, *171*, 299–309. https://doi.org/10.1016/j.biocon.2014.01.037

Cooper, J. E. J., Plummer, K. E., Middlebrook, I., & Siriwardena, G. M. (2024). Using butterfly survey data to model habitat associations in urban developments. *Journal of Applied Ecology*, *61*, 773–783.

Cooper, J. E. J., Plummer, K. E., & Siriwardena, G. M. (2023). Using species-habitat models to predict bird counts from urban development plans. *Landscape and Urban Planning*, *230*, 104629.

Corlett, R. T. (1992). The ecological transformation of Singapore, 1819–1990. *Journal of Biogeography*, *19*, 411–420.

Davison, G. W. H., Gan, J. W. M., Huang, D., Hwang, W. S., Lim, S. K. Y., & Yeo, D. C. J. (2024). *The Singapore red data book: Red lists of Singapore biodiversity* (3rd ed.). National Parks Board.

Devarajan, K., Morelli, T. L., & Tenan, S. (2020). Multi-species occupancy models: Review roadmap, and recommendations. *Ecography*, *43*, 1612–1624.

Dietzel, A., Moretti, M., & Cook, L. M. (2024). Shrinkage-based Bayesian variable selection for species distribution modelling in complex environments: An application to urban biodiversity. *Ecological Informatics*, *81*, 102561.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*(1), 38–49. https://doi.org/10.1017/S0376892997000088

Franklin, J. (2010). Model evaluation. In *Mapping species distributions: Spatial inference and prediction* (pp. 209–234). Cambridge University Press. https://doi.org/10.1017/CBO9780511810602.014

Garrard, G. E., Williams, N. S. G., Mata, L., Thomas, J., & Bekessy, S. A. (2018). Biodiversity sensitive urban design. *Conservation Letters*, *11*, 1–10.

Gaw, L. Y.-F., Yee, A. T. K., & Richards, D. R. (2019). A high resolution map of Singapore's terrestrial ecosystems. *Data*, *4*, 116.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*, 379–391.

Hamer, A. J., & McDonnell, M. J. (2008). Amphibian ecology and conservation in the urbanising world: A review. *Biological Conservation*, *141*, 2432–2449. https://doi.org/10.1016/j.biocon.2008.07.020

Hesselbarth, M. H. K., Sciaini, M., With, K. A., Wiegand, K., & Nowosad, J. (2019). Landscapemetrics: An open-source R tool to calculate landscape metrics. *Ecography*, *42*(10), 1648–1657. https://doi.org/10.1111/ecog.04617

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Ikin, K., Le Roux, D. S., Rayner, L., Villaseñor, N. R., Eyles, K., Gibbons, P., Manning, A. D., & Lindenmayer, D. B. (2015). Key lessons for achieving biodiversity-sensitive cities and towns. *Ecological Management & Restoration*, *16*, 206–214.

Iknayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, *29*, 97–106.

Khew, S. K. (2015). *A field guide to the butterflies of Singapore* (2nd ed.). Ink On Paper Communications Pte Ltd.

Kremen, C. (2015). Reframing the land-sparing/land-sharing debate for biodiversity conservation. *Annals of the New York Academy of Sciences*, *1355*(1), 52–76. https://doi.org/10.1111/nyas.12845

Kuhn, M., & Wickham, H. (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. https://www.tidymodels.org

Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L., Howard, F., & Nichol, S. (2017). Application of random forest and generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness. *Environmental Modelling & Software*, *97*, 112–129.

Lovelace, R. (2019). *osmextract*.

Marzluff, J. M. (2017). A decadal review of urban ornithology and a prospectus for the future. *Ibis*, *159*, 1–13. https://doi.org/10.1111/ibi.12430

Mazerolle, M. J. (2023). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). https://cran.r-project.org/package=AICcmodavg

OpenStreetMap contributors. (2020). *Planet dump*. https://planet.openstreetmap.org

Orr, A. G. (2005). *Dragonflies of peninsular Malaysia and Singapore*. Natural History Publications (Borneo).

Ortega-Álvarez, R., & MacGregor-Fors, I. (2009). Living in the big city: Effects of urban land-use on bird community structure, diversity, and composition. *Landscape and Urban Planning*, *90*(3–4), 189–195. https://doi.org/10.1016/j.landurbplan.2008.11.003

Perrelet, K., Moretti, M., Dietzel, A., Altermatt, F., & Cook, L. M. (2024). Engineering blue-green infrastructure for and with biodiversity in cities. *Npj Urban Sustainability*, *4*, 27.

Plummer, K. E., Gillings, S., & Siriwardena, G. M. (2020). Evaluating the potential for bird-habitat models to support biodiversity-friendly urban planning. *Journal of Applied Ecology*, *57*, 1902–1914.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Ramírez-Restrepo, L., & MacGregor-Fors, I. (2017). Butterflies in the city: A review of urban diurnal Lepidoptera. *Urban Ecosystems*, *20*(1), 171–182. https://doi.org/10.1007/s11252-016-0579-4

Ranghetti, L., Boschetti, M., Nutini, F., & Busetto, L. (2020). sen2r: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Computers and Geosciences*, *139*, 104473. https://doi.org/10.1016/j.cageo.2020.104473

Sepkoski, J. J. (1988). Alpha, beta, or gamma: Where does all the diversity go? *Paleobiology*, *14*(3), 221–234. https://doi.org/10.1017/S0094837300011969

Soga, M., Yamaura, Y., Koike, S., & Gaston, K. J. (2014). Land sharing vs. land sparing: Does the compact city reconcile urban development and biodiversity conservation? *Journal of Applied Ecology*, *51*(5), 1378–1386. https://doi.org/10.1111/1365-2664.12280

Song, X. P., Tan, E. Y. W., Lee, R. S. K., Sim, H. J., Nai, J., Chan, J. Y., Hung, S. M. X., Ng, S. H., Tan, C., Chua, Y. K., Xu, A., Cai, L., Yeo, D. C. J., Tan, H. T. W., & Chong, K. Y. (2025). Data from: A predictive approach to assess urban biodiversity and plan for future development scenarios. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.2fqz6131p

Song, X. P., Tan, E. Y. W., Lee, R. S. K., Sim, H. J., Nai, J., & Chong, K. Y. (2022). biodivercity: An R package for spatial assessment of biodiversity across city landscapes. https://doi.org/10.5281/zenodo.7410414

Tan, P. Y., Wang, J., & Sia, A. (2013). Perspectives on five decades of the urban greening of Singapore. *Cities*, *32*, 24–32.

Villalobos-Jiménez, G., Dunn, A. M., & Hassall, C. (2016). Dragonflies and damselflies (Odonata) in urban ecosystems: A review. *European Journal of Entomology*, *113*, 217–232. https://doi.org/10.14411/eje.2016.027

Wang, J. W., Poh, C. H., Tan, C. Y. T., Lee, V. N., Jain, A., & Webb, E. L. (2017). Building biodiversity: Drivers of bird and butterfly diversity on tropical urban roof gardens. *Ecosphere*, *8*(9), e01905. https://doi.org/10.1002/ecs2.1905

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, *21*(2), 213–251.

Wong, J. S. Y., Soh, M. C. K., Low, B. W., & Er, K. B. H. (2023). Tropical bird communities benefit from regular-shaped and naturalised urban green spaces with water bodies. *Landscape and Urban Planning*, *231*, 104644.

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Yeo, H. T. T., Chong, K. Y., Yee, A. T. K., Giam, X., Corlett, R. T., & Tan, H. T. W. (2014). Leaf litter depth as an important factor inhibiting seedling establishment of an exotic palm in tropical secondary forest patches. *Biological Invasions*, *16*, 381–392.

Yong, D. L., Lim, K. C., & Lee, T. K. (2016). *A Naturalist's guide to the birds of Singapore* (2nd ed.). John Beaufoy Publishing.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Figure S1. Sampling points within the six towns in Singapore.

Figure S2. Overview of the data processing workflow for landscape variables.

Figure S3. Example showing predictions for the number of bird species by first summarizing landscape variables across (a) center points within a regular grid laid over the region of interest; and (b) filling the grid squares according to the predicted species richness at each center point.

Figure S4. Landscape predictors within the best models ($\Delta AICc < 2$) that predict bird local diversity, ranked based on their sum of weights.

Figure S5. Relationship between bird local diversity and landscape predictors in the best performing models ($\Delta AIC < 2$) and with a sum-of-weights value above 0.5 are shown, arranged according to decreasing importance.

Figure S6. Landscape predictors within the best models ($\Delta AICc < 2$) that predict butterfly local diversity, ranked based on their sum of weights.

Figure S7. Relationship between butterfly local diversity and landscape predictors in models fit using remotely sensed data.

Figure S8. Landscape predictors within the best models ($\Delta AICc < 2$) that predict odonate local diversity, ranked based on their sum of weights.

Figure S9. Relationship between odonate local diversity and landscape predictors in models fit using remotely sensed data.

Figure S10. Landscape predictors within the best models ($\Delta AICc < 2$) that predict amphibian local diversity, ranked based on their sum of weights.

Figure S11. Relationship between amphibian local diversity and landscape predictors in models fit using remotely sensed data.

Figure S12. Results from cross-validation with train-test split (75%/25%), in terms of (a) mean error in predicting species density; and (a) root mean squared error (RMSE) of predictions, across 1000 iterations.

Figure S13. Histograms showing the distribution of values for the predicted diversity of (a) bird; (b) butterfly; (c) odonate; and (d) amphibian species per pixel within each of the 23 towns in Singapore.

Figure S14. The correspondence between predicted and observed species density when remotely sensed landscape data was used again with the trained models (left: 'before') and after manually mapped landscape data was converted and used to replace remotely sensed landscape data at sampling points (right: 'after').

Table S1. Landscape predictor variables derived from vector data.

Table S2. Landscape predictor variables derived from each land cover raster.

Table S3. Summary of the best models ($\Delta AIC < 2$) that predict bird local diversity, ranked based on decreasing $\Delta AICc$ value.

Table S4. Summary of the best models ($\Delta AIC < 2$) that predict butterfly local diversity, ranked based on decreasing $\Delta AICc$ value.

Table S5. Summary of the best models ($\Delta AIC < 2$) that predict odonate local diversity, ranked based on decreasing $\Delta AICc$ value.

Table S6. Summary of the best models ($\Delta AIC < 2$) that predict amphibian local diversity, ranked based on decreasing $\Delta AICc$ value.

Table S7. Detailed results and scoring for each of the 23 towns in Singapore.

Appendix S2. Mean and 95% confidence intervals of the importance value of each predictor variable at each spatial scale from random forests.

How to cite this article: Song, X. P., Tan, E. Y. W., Lee, R. S. K., Sim, H. J., Nai, J. K. H., Chan, J. Y., Hung, S. M. X., Ng, S. H., Goh, E. S. C., Tan, C. Y. T., Chua, Y. K., Xu, A., Cai, L., Yeo, D. C. J., Tan, H. T. W., & Chong, K. Y. (2025). A predictive approach to assess urban biodiversity and plan for future development scenarios. *Journal of Applied Ecology*, *62*, 2882–2893. https://doi.org/10.1111/1365-2664.70100