Harnessing iNaturalist to quantify hotspots of urban biodiversity: the Los Angeles case study

Joscha Beninde[1,2]*, Tatum W. Delaney[3†], Germar Gonzalez[3†] and
H. Bradley Shaffer[1,3]

[1]La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability,
University of California, Los Angeles, Los Angeles, CA, United States, [2]IUCN WCPA Connectivity
Conservation Specialist Group, Gland, Switzerland, [3]Department of Ecology and Evolutionary Biology,
University of California, Los Angeles, Los Angeles, CA, United States

Introduction: A major goal for conservation planning is the prioritized protection and management of areas that harbor maximal biodiversity. However, such spatial prioritization often suffers from limited data availability, resulting in decisions driven by a handful of iconic or endangered species, with uncertain benefits for co-occurring taxa. We argue that multi-species habitat preferences based on field observations should guide conservation planning to optimize the long-term persistence of as many species as possible.

Methods: Using habitat suitability modeling techniques and data from the community-science platform iNaturalist, we provide a strategy to develop spatially explicit models of habitat suitability that enable better informed, place-based conservation prioritization. Our case study in Greater Los Angeles used Maxent and Random Forests to generate suitability models for 1,200 terrestrial species with at least 25 occurrence records, drawn from plants (45.5%), arthropods (27.45%), vertebrates (22.2%), fungi (3.2%), molluscs (1.3%), and other taxonomic groups (< 0.3%). This modeling strategy further compared spatial thinning and taxonomic bias file corrections to account for the biases inherent to the iNaturalist dataset, modeling species jointly and separately in wildland and urban sub-regions and validated model performance using null models and a "test" dataset of species and occurrences that were not used to train models.

Results: Mean models of habitat suitability of all species combined were similar across model settings, but the mean Random Forest model received the highest median AUC_{ROC} and AUC_{PRG} scores in model evaluation. Taxonomic groups showed relatively modest differences in their response to the urbanization gradient, while native and non-native species showed contrasting patterns in the most urban and the most wildland habitats and both peaked in mean habitat suitability near the urban-wildland interface.

Discussion: Our modeling framework is based entirely on open-source software and our code is provided for further use. Given the increasing availability of urban biodiversity data via platforms such as iNaturalist, this modeling framework can easily be applied to other regions. Quantifying habitat suitability for a large, representative subset of the locally occurring pool of species in this

way provides a clear, data-driven basis for further ecological research and conservation decision-making, maximizing the impact of current and future conservation efforts.

Introduction

Increased urbanization in the Anthropocene has given rise to megacities, fundamentally transforming previously existing landscapes across much of the globe. The environment of cities is strikingly different from adjacent non-urban areas, with elevated levels of human population densities, impervious surfaces, roads, vehicular traffic, artificial light at night, pollution, urban heat, and many other factors shaping the microclimate, hydrology, and soil properties of cites (Groffman et al., 2014). Species living in urban areas must either cope with these altered environmental conditions (Johnson and Munshi-South, 2017) or be relegated to their fringing landscapes. The conservation of high levels of urban biodiversity has become a goal of many urban administrations (Waldrop, 2019) aiming to counteract the loss of human-nature interactions by city-dwellers (Soga and Gaston, 2016) and recoup the multifaceted benefits of biodiversity for human well-being (Fuller et al., 2007; Methorst et al., 2020), and more generally for sustainable urban footprints and biodiversity-ecosystem service synergies (Ziter, 2015; Schlaepfer et al., 2020).
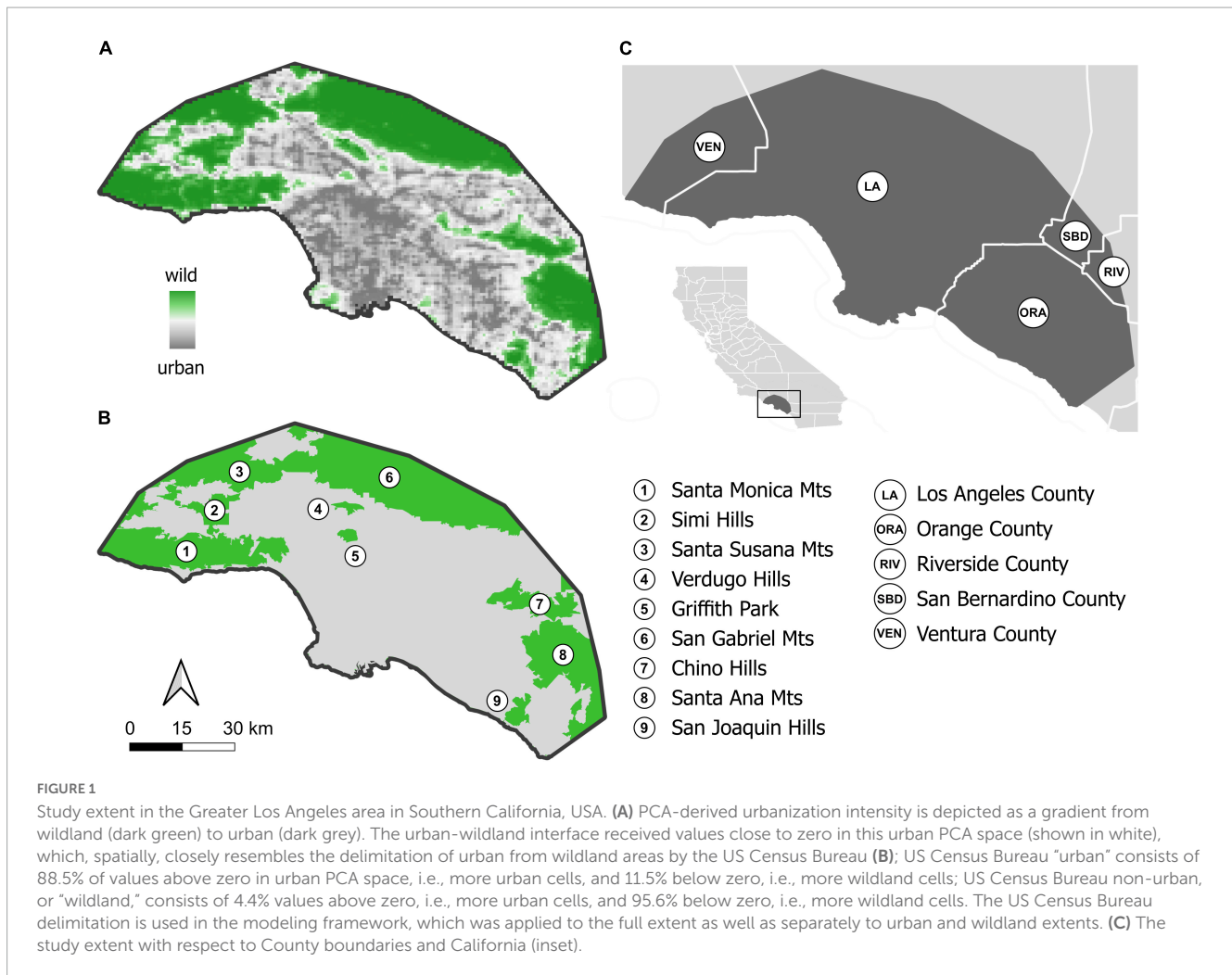
To plan for effective biodiversity conservation, we must first understand the geographical distribution of as much of the regional species pool as possible. The central position of "place" in Morrison's virtuous cycle framework for biodiversity conservation (Morrison, 2016) emphasizes this point—we need to know the places to focus conservation actions, followed by local community engagement. However, the available data on which conservation decisions hinge is typically restricted to only a small subset of all species in a given landscape (Hochkirch et al., 2021) and is often biased toward charismatic or endangered species. These species may serve as umbrella species, where the benefits of conservation efforts directed at them cascade across other, co-occurring species (Roberge and Angelstam, 2004). However, the umbrella functionality of a species can be difficult to quantify (Fleishman et al., 2001; Roberge and Angelstam, 2004) and may vary for species richness, abundance, or functional diversity (Branton and Richardson, 2011; Sattler et al., 2013). Despite these uncertainties, the traditionally limited availability of information on species distributions usually necessitates using a limited set of species as surrogates to guide decisions on the protection and restoration of habitats for conservation.

Urban areas are no exception to this general trend, and historically knowledge of patterns of distribution and abundance of local biodiversity in our cities has been limited (Kohsaka et al., 2013). However, this is changing. Emerging community-science projects, also referred to as citizen-science (Cooper et al., 2021),

have turned cities into hotspots of biodiversity monitoring, and this trend is only increasing over time (Devictor et al., 2010). Observations of species that are accessible on platforms such as eBird or iNaturalist frequently center on urban habitats and their surrounding landscapes, simply because those areas are the most accessible to large numbers of people (Spear et al., 2017). Unrestricted by the limited capacity of researchers for field observations, iNaturalist alone surpassed 33.7 million observations globally in 2022, up more than 10-fold from 3.3 million observations in 2017[1], rendering such datasets an unprecedented resource for urban conservation prioritization (Li et al., 2019; Callaghan et al., 2020a). Observations on iNaturalist are taxonomically diverse and unstructured with respect to survey methodology. The nature of how users utilize iNaturalist, which relies on proximity, ease of access, and human esthetics, presumably biases these datasets in predictable ways according to human preferences for certain species/taxonomic groups, species detectability, the reduced chance to take photographs of small, distant or fast-moving species (Di Cecco et al., 2021), site accessibility (Zizka et al., 2021), and varying sampling effort at sites (Beck et al., 2014). Of course, these biases are inherent to most datasets, including herbarium and museum records, which contain similar spatial, environmental, temporal, and taxonomic biases (Newbold, 2010; Martin et al., 2012; Kling et al., 2018).

In contrast to these more traditional data sources, observations on community-science platforms have two important advantages: They accumulate data orders of magnitudes faster than conventional research data (Spear et al., 2017; Callaghan et al., 2020b), and they include observations from private land, which conventional research data frequently cannot sample (Martin et al., 2012). Given the knowledge of potential biases in occurrence datasets, methodologies for adequately addressing them have been proposed for models that use occurrence records and landscape predictors to describe and predict the environmental niche space of species (Warren et al., 2010). These techniques are widely used in biodiversity assessments (Araújo et al., 2019) and we refer to them as habitat suitability modeling (they are also commonly called environmental niche modeling, ENM, or species distribution modeling, SDM). When the input data are managed appropriately, these models allow users to correct for sampling biases via spatial thinning (to reduce over-representation at hotspots of observer activity; Steen et al., 2021), by scaling background locations to the distribution of sampling effort in the

---

1 https://www.inaturalist.org/stats/"year"

FIGURE 1
Study extent in the Greater Los Angeles area in Southern California, USA. **(A)** PCA-derived urbanization intensity is depicted as a gradient from wildland (dark green) to urban (dark grey). The urban–wildland interface received values close to zero in this urban PCA space (shown in white), which, spatially, closely resembles the delimitation of urban from wildland areas by the US Census Bureau **(B)**; US Census Bureau "urban" consists of 88.5% of values above zero in urban PCA space, i.e., more urban cells, and 11.5% below zero, i.e., more wildland cells; US Census Bureau non-urban, or "wildland," consists of 4.4% values above zero, i.e., more urban cells, and 95.6% below zero, i.e., more wildland cells. The US Census Bureau delimitation is used in the modeling framework, which was applied to the full extent as well as separately to urban and wildland extents. **(C)** The study extent with respect to County boundaries and California (inset).

landscape (Kramer-Schadt et al., 2013; Merow et al., 2013; Kling et al., 2018), and/or by creating habitat-specific models (Fourcade et al., 2014). Following these precautionary measures to address spatial biases of occurrence datasets, bias-corrected models of habitat suitability can be parameterized and used to infer habitat suitability for sites that have not been sampled, which is frequently done across thousands of species simultaneously (Bradie and Leung, 2016; Kling et al., 2018). While many approaches to model habitat suitability exist, we use Maxent and Random Forests, which frequently rank among the top-performing modeling methods and require only a fraction of the computational resources of similarly high-performing methods, such as boosted regression trees or ensemble methods (Harrigan et al., 2014; Valavi et al., 2022).

We present a framework utilizing Maxent and Random Forest modeling in combination with the iNaturalist dataset for Greater Los Angeles (Figure 1), the largest metropolitan region in the US by area and the second largest by human population size. More than 1.5 million iNaturalist observations are available for this study extent (as of November 2021), including observations of 6,082 species whose identifications were verified by the iNaturalist community. Using only those terrestrial species with a minimum of 25 occurrence records, we generated habitat suitability models for 1,200 taxonomically diverse species composed of native taxa, ranging from velvet ants

(*Dasymutilla* spp.), swallowtail butterflies (*Papilio* spp.), tarantulas (*Aphonopelma* spp.), rattlesnakes (*Crotalus* spp.), coyotes (*Canis latrans*), mountain lions (*Puma concolor*), oak (*Quercus* spp.), maple (*Acer* spp.), sycamore (*Platanus racemosa*) and pine trees (*Pinus* spp.), poppy (*Eschscholzia* spp.) and Clarkia flowers (*Clarkia* spp.), manzanita shrubs (*Arctostaphylos* spp.), lichenized fungi (Lecanoromycetes), and non-native, human commensal species, including earthworms (*Lumbricus terrestris*), pill bugs (*Armadillidium vulgare*), honey bees (*Apis mellifera*), cockroaches (Blattidae), rats (*Rattus* spp.), house sparrows (*Passer domesticus*) and house cats (*Felis catus*; Supplementary Figure 1). By producing a mean model of habitat suitability of all species (also referred to as stacked models, Calabrese et al., 2014), as well as models for all native and non-native species separately, we created spatially explicit models at 1 km x 1 km raster cell resolution to summarize spatial biodiversity value. Land managers can use these models to guide spatial prioritization and protection, direct conservation and restoration efforts, or simply track current biodiversity, enabling ongoing efforts to create meaningful urban biodiversity indices (Kohsaka et al., 2013; Isaac Brown Ecology Studio and La Sanitation and Environment, 2018).

The question of what species to protect is controversial when discussing non-native species (Sax et al., 2022) and may be especially difficult to answer in cities (Gaertner et al., 2016). The

city of LA's ambitious goals toward a no-net-loss of biodiversity largely focuses on native species (City of Los Angeles, 2019) and some ecosystem services may be enhanced by focusing only on native species. For example, native California chaparral vegetation provides less fuel for fires and has better wind-stopping qualities than non-native vegetation (Keeley, 2020), and in the Greater Los Angeles area some species of native trees and shrubs attract higher densities of birds (Wood and Esaian, 2020; Smallwood and Wood, 2023) and insects (Adams et al., 2020) than non-native tree and shrub species. However, individuals of non-native species in urban areas also have recognized value when viewed through an ecosystem services lens; they provide habitat, disperse seeds, and take over the roles of native pollinators in some situations (Sax et al., 2022). Given the overarching role of urban biodiversity in conveying direct benefits to human health (Sandifer et al., 2015; Methorst et al., 2020), as well as generating an appreciation for biodiversity and avoiding the 'extinction of experience', non-native species in urban areas may sometimes surpass their native counterparts as important agents for developing a public understanding and personal motivation to conserve biodiversity (Schuttler et al., 2018). It has also been argued that urban individuals of endangered non-native species outside of their native range may qualify as a form of *ex situ* conservation; the endangered red-crowned parrot (*Amazona viridigenalis*) in Los Angeles is one example (Shaffer, 2018). We believe these to be important benefits of non-native species in urban areas which deserve consideration in urban conservation prioritization, as they are weighed against the equally important potential threats non-native species can impose on ecosystems and human health (Gaertner et al., 2016). As emphasized by Morrison (2016), identifying the benefits to biodiversity are often specific to place, species (including native versus non-native taxa), local human population and/or need; establishing a positive feedback cycle, or virtuous cycle in Morrison's language, between people and biodiversity conservation, is almost certainly different in urban Los Angeles than in the wilderness of the adjacent San Gabriel mountains. Our goal here is to provide the place-based biodiversity data, across types of species and geographic areas, that is required to establish such positive feedback cycles.

Given this variability in how to think about native vs. non-native biodiversity and the various modeling parameters that address sampling biases explored here, there are three key elements of our study: (1) identify the optimal habitat suitability modeling strategy to address the biases in iNaturalist occurrence data, (2) quantify mean habitat suitability across a gradient of urban intensity and identify hotspots of urban biodiversity, and (3) contrast mean habitat suitability for native and non-native species and different taxonomic groups along a gradient of urban intensity. We predict higher suitability for native species in wildland areas than in fully urbanized sites and the opposite for non-native species, following the conceptual framework of Cadotte et al. (2017) and based on observations across taxonomic groups that high levels of urbanization increase the ratio of non-native to native species (Celesti-Grapow et al., 2006; Ricotta et al., 2010). We further predict that the highest levels of mean habitat suitability in plants will occur where urban areas transition into wildland based on greater habitat heterogeneity, a pattern frequently found for plant species richness across urbanization gradients (Celesti-Grapow et al., 2006; McKinney, 2008). At the same time, the position of this peak likely varies with the taxonomic group, as levels of species richness along urbanization gradients often vary across many groups of plants and animals (McKinney, 2008; Piano et al., 2020; Theodorou et al., 2020). We end with a discussion of the utility of our modeling framework and highlight its potential application in future urban biodiversity conservation and research.

Methods

Study extent

Our study extent is located in Greater Los Angeles, totals 7,797 km^2 and fully encompasses the City of Los Angeles, large parts of the Los Angeles-Long Beach-Anaheim, CA Metropolitan Statistical Area, plus parts of adjacent Ventura County in the west, small parts of Riverside and San Bernardino Counties to the east and into Orange County in the south (Figure 1). Developed land use types predominate (60.3%; Supplementary Table 1) followed by mostly vegetated areas (37.2%), while highly managed, working landscapes, such as agricultural areas, are uncommon (0.7%). Using the US census bureau delineation (U.S. Census Bureau, 2018; Figure 1B), 65.1% of the study extent is urban. We refer to the remaining 34.9% of the area collectively as wildlands, as they entail vast expanses of native vegetation, with little development (Supplementary Table 1), including parts or all of the Santa Monica Mountains, Simi Hills, Santa Susana Mountains, Verdugo Hills, Griffith Park, San Gabriel Mountains, Chino Hills, Santa Ana Mountains, and the San Joaquin Hills (Figure 1B). The study extent thus encompasses extensive urban areas that are home to a dense human population of 13.4 million (2,640 people/km^2), framed by sparsely populated wildland areas with a modest combined population of 86,700 thousand humans and a density almost two orders of magnitude lower (32 people/km^2; calculated based on data by Rose et al., 2017). This study extent is thus uniquely suited to study species' distribution patterns within an urban megacity and the immediately adjacent wildlands, separated by a sharp urban-wildland interface (Figure 1A). Given the extreme environmental differences between urban and wildland areas and the potential for contrasting habitat association of species in urban and wildland areas (Fourcade et al., 2014), we modeled (1) the combined urban and wildland habitat (which we refer to as the full study extent) and (2) the urban (5,074 km^2) and wildland (2,723 km^2) areas separately to assess habitat suitability (see details below).

Occurrence datasets

Occurrence records from iNaturalist were downloaded via GBIF and directly from the iNaturalist site to include non-research grade observations, which are not integrated into GBIF, but necessary to control for observer bias (see section "Exploring different settings for occurrence and background point selection" in Supplementary material). iNaturalist is a rapidly-growing platform where the general public can submit observations of species and iNaturalist users can add species identifications to observations. As soon as two identical, independent species-level identifications of an observation are proposed, and the observation

contains a photo voucher, a location, and a date, it receives a "research grade" quality grade. This label, however, is dynamic and persists only as long as 2/3 of the proposed identifications agree; observations can toggle between research and non-research grade as a consequence.

We created two occurrence datasets from iNaturalist observations, one to train models and an independent dataset to test models. The raw training dataset contained all 1,537,123 iNaturalist occurrences falling within our Greater Los Angeles study extent and was downloaded using the search queries on the iNaturalist website. We filtered this dataset according to the following five criteria to include only: (1) research grade entries; (2) non-captive and non-cultivated individuals; (3) spatially unobscured records; (4) observations with a maximum inaccuracy of 100 meters, which equals 10% of the 1km raster cell edge length (or 1% of the area) used in analyses; and (5) species with 25 or more observations. This yielded a training dataset of highest quality observations (both spatially and taxonomically) with sufficient observations to train models accurately. It contained 388,793 occurrence records of 1,286 species with observations made between 1 January 2000 to 31 December 2021. We also excluded all fully marine and aquatic species, including all species of Actinopterygii, Elasmobranchii, and Bivalvia, and some species of Mollusca, Arthropoda, and Plantae (e.g., marine and freshwater slugs and snails, water striders, water scorpions, some Malacostraca, and the plant genus *Pistia*), reducing the number of species in the training dataset to 1,200.

The raw test dataset contained 130,640 iNaturalist observations downloaded from GBIF.org (2022) and was restricted to observations made in 2022 only, ensuring there is no overlap with the training dataset which ended in 2021. This test dataset was identically filtered using the same five criteria as the training dataset with the exception that we relaxed criterion (5) and included species with fewer than 25 observations. The final test dataset consisted of 113,729 occurrence records of 3,458 species, including occurrence data for 2,258 species that were not present in the training dataset. We think that using this test dataset is particularly suitable to evaluate the ability of mean models to predict habitat suitability of species not used for training. It serves to test whether our training dataset reasonably serves as a surrogate for the entire local pool of sampled and unsampled species (see details below).

For all 1,286 species of the training dataset, we determined California native or non-native status using Calflora (2022), iNaturalist species accounts, and the expert opinions of colleagues, primarily at the Natural History Museum of Los Angeles County (NHMLAC) and UCLA (see acknowledgments).
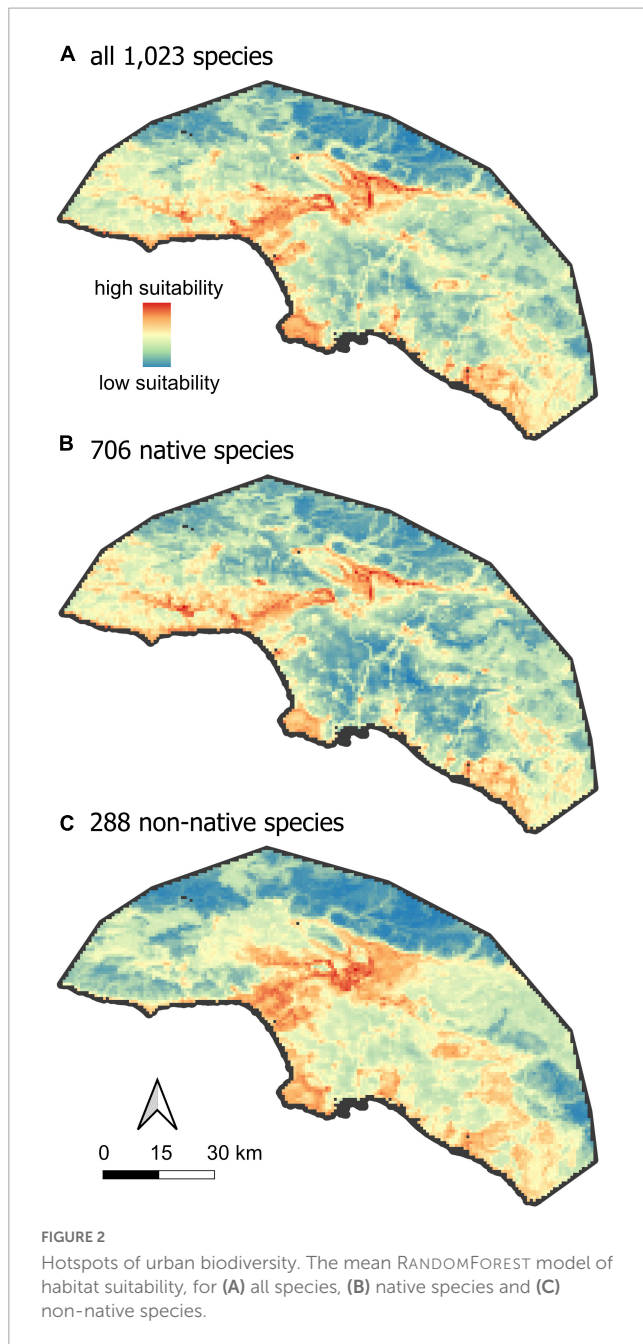
We ran the r-package "CoordinateCleaner" (v2.0-20; Zizka et al., 2019) on the training dataset to check for multiple errors in the coordinates using the function clean_coordinates(). This resulted in 26,580 coordinates flagged as being in the sea or too close to the coast and an additional 4,949 coordinates flagged because they were within a 10 km radius of recognized biodiversity institutions, such as museums or universities. We retained these flagged records in both cases. In the case of coast proximity, occurrence records that fall into non-terrestrial space would be dropped in the downstream modeling procedure as we only used terrestrial environmental predictors, and we wanted to retain all truly terrestrial records, including those close to shore,

following recommendations by Zizka et al. (2019). For institution proximity, we visually checked flagged records and found them to be proximate to the NHMLAC (2,330 observations), Occidental College (1,045 observations), or UCLA (645 observations), but represented reasonable species observations from a museum or college-associated green space. Historically, some traditional museum samples have erroneously received coordinates of the biodiversity institutions where they are housed rather than the coordinates of the actual sampling sites (Zizka et al., 2019), but this does not seem to be the case in the contemporary iNaturalist dataset for the region.

To assess how well our training dataset covers the environmental predictor space, we calculated a standardized spatial PCA using "RStoolbox" (v0.3.0; Leutner et al., 2022) on the set of landscape predictors chosen for Maxent modeling (see Section "Landscape variables" below) and quantified the distribution of occurrences across the available environmental PCA space. Adequate distribution of occurrences along this environmental PCA space is a key prerequisite for accurate habitat suitability modeling because unsampled, environmentally unique areas can limit the model's ability to correctly infer habitat suitability in such areas (Elith et al., 2011). We were particularly concerned about high-elevation (above 600 m) sites because these habitats make up only a small fraction of the entire sample extent, are located mostly in remote wildland habitats that may be undersampled by the public, and our initial analyses indicated that they had generally low multi-species predicted suitability. Given that these high-elevation sites often also had relatively few observations/species, we were concerned that undersampling may be contributing to this low suitability. We used a binomial test (prop.test function in base r-package "stats" v3.6.2) to compare the proportion of iNaturalist observations that were made in high-elevation habitats for species that were included in the training models (those with 25 or more observations) and species that were excluded from these models because they had fewer than 25 observations. This specifically tested whether species that are more closely associated with high-elevation habitats (those with a higher proportion of observations at higher altitudes) were more likely to drop out of the dataset due to an insufficient number of total observations (< 25), which could bias high-elevation inferences of mean habitat suitability.

Habitat suitability modeling

Many techniques can build habitat suitability models, also referred to as environmental niche models (ENMs) or species distribution models (SDMs), from occurrence data and a set of landscape predictors (Valavi et al., 2022). These models quantify the distribution of environmental niche space at species' presence locations and can be used to predict the suitability at unsampled locations, as a function of environmental predictors (Harrigan et al., 2014). The resulting predictions are best interpreted as depicting relative habitat suitability, circumventing issues of interpreting model output as the probability of presence or the relative occurrence rate, which is only valid in the rare cases of entirely random spatial sampling strategies for occurrence data or with complete knowledge of a species abundance in a

FIGURE 2
Hotspots of urban biodiversity. The mean RANDOMFOREST model of habitat suitability, for **(A)** all species, **(B)** native species and **(C)** non-native species.

given landscape (Merow et al., 2013). To generate these models, environmental predictors at user-specified presence points (the occurrence data) are contrasted with environmental predictors at a variable number of background points, which can be chosen from within the study extent following different strategies. By default, some methods, including the default settings of Maxent, randomly sample background points, such that each 1-km raster cell locality has an equal probability of being chosen. This approach includes a random sampling of cells containing and lacking species observations. Alternatively, it is often recommended to scale the background point distribution to match the sampling intensity dedicated to each location in the study extent, as one way to correct for sampling bias (Kramer-Schadt et al., 2013; Merow et al., 2013). This strategy increases the likelihood of

locations being chosen as background points if they contain many observations and decreases the likelihood for sites with fewer observations.

The number of occurrences necessary to parameterize a model is an important consideration. While models can be parameterized with as few as three occurrence records (Proosdij et al., 2016), model predictions derived from such small sample sizes are often highly variable and should be treated with caution; predictions generally converge with 25-30 occurrence records (Wisz et al., 2008). Based on this convergence, we filtered our training dataset to species with at least 25 observations.

Finally, it is important to recognize that using habitat suitability modeling on a spatial extent as small as in this study does not allow one to quantify a species' full environmental niche space. Rather, the model reflects the (limited) environmental niche space available within the study extent, and can be used to understand habitat suitability within this study extent but not necessarily beyond, in space or time (Araújo et al., 2019).

We used two different methods to model habitat suitability, the maximum entropy-based approach of Maxent (Phillips et al., 2004), and the tree-based approach of Random Forests (Breiman, 2001). Studies comparing different modeling techniques frequently identify these two methods as among the top-performing (Harrigan et al., 2014; Valavi et al., 2022), and they are considerably less computationally intensive than other high-performance models, such as boosted regression trees or Ensemble methods (Valavi et al., 2022). Maxent and Random Forests allow the user to account for potential biases in the occurrence dataset (Kramer-Schadt et al., 2013; Merow et al., 2013; Fourcade et al., 2014), have high predictive power across sample sizes (Wisz et al., 2008), and are comparable in accuracy to other techniques (Kaky et al., 2020; Valavi et al., 2022). We implemented Maxent (v3.4.4; Phillips et al., 2022) through the r-package "dismo" (Hijmans et al., 2017) and Random Forests using the r-package "randomForest" (v4.7-1.1; Liaw and Wiener, 2022) using R Statistical Software (v4.1.2; R Core Team, 2021). Maxent modeling was replicated (Sillero and Barbosa, 2021) and run with 10-fold cross-validation (CV), using AUC values calculated on independent test datasets as a measure of model fit. Random Forest models were fitted with a down-sampling procedure as recommended by Valavi et al. (2022) which uses equal numbers of occurrence and background points, preventing class imbalance issues. Random Forest models were fitted using the default values for mtry and ntree = 1,000.

Addressing biases in the modeling framework

We first explored different modeling strategies on a set of 15 plants and animal species using Maxent. For these species we made use of our knowledge of their distribution and relative abundance in the study extent, which we acquired during extensive fieldwork, to evaluate the performance of different modeling strategies to predict habitat suitability accurately. To address, and account for, sampling bias in the iNaturalist dataset, we tested several strategies to select occurrence records (Steen et al., 2021) and background points (Kramer-Schadt et al., 2013; Merow et al., 2013; Kling et al., 2018; section "Exploring different settings for occurrence and

background point selection" in the supplementary materials). To test for divergent habitat preferences and for species distributions that are in potential dis-equilibrium with the environment caused by rapid urbanization or recent introductions, (Fourcade et al., 2014; Searcy and Shaffer, 2014) we modeled each species for the full study extent, and separately in urban and wildland habitats (Supplementary Figure 2; section "Exploring separate modeling in urban and wildland extents" in Supplementary material). Based on our expert assessment of model performance for these 15 species, we identified the best model settings and spatial extents. We summarize the results here, as they are central to our methods (see comprehensive details in section "Exploring separate modeling in urban and wildland extents" in Supplementary material). We noticed considerable differences between models generated at the full, urban, and wildland extents but there was no extent that consistently returned the most realistic results, so we modeled in all three extents. The best models were generated with the following settings for occurrence records and background point selection: (1) THINNED, which uses a single occurrence record per species and raster cell as the occurrence dataset and a uniform prior for background point selection; (2) PHYLUMBIAS, uses all occurrence records of a species as the occurrence dataset and a prior for background point selection that scales to the number of iNaturalist observations of all species of the same Phylum as the target taxon; and (3) CLASSBIAS, also uses all occurrence records of a species as the occurrence dataset and a prior for background point selection that scales to the number of iNaturalist observations of all species of the same Class as the target taxon. These three model settings were adopted for all 1,200 species in our post-pruning dataset and all three spatial extents, resulting in a total of 6,953 models fitted for all combinations of species, model settings, and spatial extent that resulted in at least 25 occurrence records. All 6,953 single-species models were validated using a null modeling approach with randomly placed occurrences similar to Merckx et al. (2011; section "Null model validation") to ensure that model performance was significantly improved over that expected by chance (Raes and ter Steege, 2007; Gomes et al., 2018). Only models of species that exceeded null model expectations were used to calculate mean models.

We created mean models using the continuous predictions of habitat suitability, as recommended by Calabrese et al. (2014), and did so separately for each of the model settings and spatial extents. We rescaled all mean models to a 0—1 scale and evaluated them using the test dataset. Across THINNED, PHYLUMBIAS, and CLASSBIAS model settings, models generated at the full extent received higher AUC_{ROC} values than those generated separately for urban and wildland extents (Supplementary Figure 3; see section "Creating and validating mean models" in Supplementary material for comprehensive details). Therefore, we limited all further analyses and modeling to the full extent. We created an additional COMPOSITE model at the full extent, composed of the best-performing model of each species, which we identified as the model with the highest AUC_{ROC} value among model settings.

In summary, this generated four mean Maxent models that were all modeled at the full extent and that we will refer to as the THINNED, PHYLUMBIAS, CLASSBIAS, and COMPOSITE models.

Model evaluation

We used the test dataset (see above) to evaluate and rank the performance of each of the mean Maxent models using AUC. We randomly drew 5–100 occurrence records from the test dataset and used a varying number of background locations as absences. We chose the number of background locations to vary between 3 and 50 times that of the number of occurrence records, which means that the theoretically possible number of background locations varied between 15 and 5,000. This procedure was repeated 1,000 times for each of the mean Maxent models to generate a distribution of AUC_{ROC} and AUC_{PRG} values, which we calculated using the evalmod() function in the "precrec" r-package (v0.13.0; Saito and Rehmsmeier, 2017) and the calc_auprg() function in the "prg" r-package (v0.5.1; Kull, 2016). As outlined in Valavi et al. (2022), AUC_{ROC} and AUC_{PRG} are complementary and assess model performance either based on both presences and absences or based only on presences, respectively. AUC_{ROC} is computed using the number of true positives, or sensitivity, and the proportion of false positives, calculated as 1-the number of true negatives. AUC_{ROC} values vary between 0–1, where 1 indicates a perfect model, i.e., presence locations have higher habitat suitability values than absence locations and there is no overlap in their distributions. Values close to 0 indicate the unlikely, but theoretically possible, opposite case, where absence locations receive higher habitat suitability values than presence locations and there is no overlap in their distributions. A value of 0.5 indicates that a model is no better than randomly assigning habitat suitability values at presence and absence locations. However, there can be considerable variation around the AUC_{ROC} value of null models, which is why we included the additional null modeling step (see above and section "Null model validation" in Supplementary material). AUC_{PRG} is calculated based on the precision of predicted presences, calculated as the proportion of true and false positives, and the sensitivity, calculated as the proportion of true positives versus false negatives. The AUC_{PRG} metric is scaled so that perfect models also approach the value of 1, as in AUC_{ROC}, while negative values indicate that models are no better than randomly differentiating between presence and absence locations (Valavi et al., 2022).

Following these evaluations of mean Maxent models, we ranked them by AUC_{ROC} and AUC_{PRG} scores to identify the best-performing model. Using the same settings as for the best mean Maxent model, we generated new habitat suitability models for each species using a Random Forest modeling approach (see above), and created another mean model from these, which we refer to as the mean RANDOMFOREST model. In a final step, for comparison, we evaluated and ranked the mean RANDOMFOREST model in the same way as the mean Maxent models and identified the best-performing overall model.

Landscape variables

We assembled environmental layers that are relevant to the study region's biogeography, climate, and landscape. We assembled a total of 37 landscape variables, including all 19 bioclim variables (based on data compiled between the years 1960-1990; Hijmans et al., 2005), 9 soil variables (Walkinshaw et al., 2021), climatic

water deficit (Flint et al., 2013), elevation, slope (U.S. Geological Survey, 2017), surface imperviousness, land cover, tree canopy cover (U.S. Geological Survey, 2014), NDVI (ORNL DAAC, 2018), artificial lights at night (ALAN; The Earth Observatory Group, 2018), and water cover (U.S. Geological Survey, 2019). To reduce data collinearity, we reduced the number of landscape variables by checking pairwise correlation coefficients (calculated at the grid-cell level) and removing predictors until all correlation coefficients were below 0.7 (Dormann et al., 2013) using the raster.cor.matrix() function from the r-package "ENMTools" (v1.0.6; Warren et al., 2021). We chose the resulting final set of 10 landscape variables to preferentially include ones that, in our view, directly influence species across taxonomic groups, as recommended for creating habitat suitability models (Merow et al., 2013). These included: Bioclim02 (mean diurnal range), bioclim06 (mean temperature of the coldest month), bioclim09 (mean temperature of the driest quarter), bioclim16 (precipitation of the wettest quarter), soil bulk density, soil cation exchange capacity, climatic water deficit, imperviousness, NDVI, and percent water cover (**Supplementary Figure 4**). All of these data layers were available at a resolution of 1km or higher and we rasterized all spatial polygons or lines objects, or resampled rasters, to a 1 km^2 resolution using the r-package "raster" (v3.5-15; Hijmans and van Etten, 2022). To evaluate the importance of bioclimatic predictor sets averaged over different periods, we compared model performance using Maxent (based on the THINNED settings only) for the WorldClim dataset averaged for 1960-1990 (Hijmans et al., 2005) and 1970-2000 (Fick and Hijmans, 2017), the CHELSA dataset for 1980-2010 (Karger et al., 2017, 2018), and the ClimateNA dataset for 1990-2020 (Wang et al., 2016; AdaptWest Project, 2021).

Using the spatial PCA function from the R-package "RStoolbox", we generated a standardized spatial PCA based on only imperviousness and ALAN, the "urban PCA", to generate a gradient of urban intensity, which we used for analyses of species responses to the intensity of urbanization.

Further analyses

To evaluate and summarize the quality of the modeling framework presented here, we scored it following the 15 criteria established by Araújo et al. (2019), which provide minimum requirements of habitat suitability modeling for application in biodiversity assessments.

In addition to validation of mean habitat suitability models (see above), we compared Schoener's D, Warren's I, and rank correlation coefficients between mean models of habitat suitability using the raster.overlap() function in ENMtools (Warren et al., 2010). We provide all three metrics but place more emphasis on rank correlation coefficients for the interpretation of results because the other two measures tend to overestimate raster similarity (Warren et al., 2021). We categorized correlation coefficients > 0.9 as very similar, those between 0.7 and 0.9 as similar, and all coefficients below 0.7 as different (Dormann et al., 2013).

To quantify the influence of landscape variables on the highest-ranking mean model of habitat suitability, we used Random Forest models (Breiman, 2001) and the r-package "randomForest", with ntree = 10,000 and mtry = the number of predictor variables/3

(Liaw and Wiener, 2022). We used the best mean model of habitat suitability for (1) all species, and subsets of (2) only native and (3) only non-native species as response variables, and all 37 landscape variables as predictors.

By quantifying the influence of the number of iNaturalist observations on the highest-ranking mean model of habitat suitability, we tested for signatures of sampling bias persisting in the mean models of habitat suitability. As above, we used Random Forest models for this and the best mean model of habitat suitability for (1) all species, (2) only native and (3) only non-native species as response variables and four different summaries of the number of iNaturalist observations as predictors (the sum of iNaturalist observations of all species, and separately, the sum of iNaturalist observations of plants, vertebrates, and arthropods).

In another effort to evaluate whether sampling bias is driving the results of the mean habitat suitability models, we used density plots across the urban PCA space of all iNaturalist observations and of the highest predicted mean habitat suitability. For the latter, we included only those raster cells of the highest-ranking mean habitat suitability model that fell within the highest quartile of predicted values.

We used loess regressions to visualize changes in mean habitat suitability values across the urban PCA space, separately for plants, vertebrates, and arthropods, as well as for native and non-native species. We used GAM (generalized additive models) models to quantify the strength of these associations using adjusted R^2 values (r-package "mgcv" v1.8-40; Wood, 2017).

Results

Native and non-native status of species of the training dataset

For the training dataset, we generated habitat suitability models for 1,200 species and were able to determine the native/non-native status of 1,183 (98.6%) of those species. Of that set of taxa, 835 species (70.6%) are native and 348 (29.4%) are non-native. The majority of the 17 species for which we could not determine the native/non-native status were Fungi (6) and Insecta (6), followed by Plantae (3), Myxomycetes (1), and Platyhelminthes (1). The proportions of native species were higher for vertebrates (88%) and arthropods (77%) than for plants (59.3%; **Supplementary Table 2**).

Occurrence records across elevation

Testing the possibility that species restricted to remote, high-elevation sites may be underrepresented in our iNaturalist-based training dataset, we found a significantly higher proportion of observations from high-elevation areas (> 600 m) in species with a total number of observations below 25 than in species with a total number of records of 25 or more, for which habitat suitability models were created (χ^2 = 131.24, df = 1, p-value < 0.001). This suggests that species more strongly associated with high elevation, may be underrepresented in our analyses, either because of true rarity, sampling bias, or the delineation of the study extent.

Comparison of US Census Bureau defined urban and wildland regions to urban PCA

The US Census Bureau delimitation of urban areas was utilized to separate urban from non-urban (wildland) areas for separate Maxent modeling in these two study extents. However, urbanization also needs to be considered as a quantitative, rather than a qualitative, landscape attribute. To quantify the level of urbanization on a continuous scale, we generated a raster PCA of the levels of impervious surface and artificial lights at night, which we designate as the urban PCA. The first axis of this urban PCA explained 97.4% of the total variation in these two variables and is therefore a reasonable proxy for urban intensity as defined by artificial hardscapes and light. PC1 ranged from -0.7 – 4.7, and values at or near zero spatially resemble the borders of the US Census Bureau delimitated urban areas (which is a two-state, rather than continuous, delimitation) at the urban-wildland interface. US Census Bureau urban areas consist of 88.5% of values above zero and 11.5% below zero on PC1, while wildland regions (that is, non-urban areas defined by the Census Bureau) consist of 4.4% of values above zero and 95.6% of values below zero on PC1 (**Figures 1A, B**). This indicates that there is a strong correlation between these discrete and continuous measures of urbanization. Consistent with this result, the proportion of NLCD (National Land Cover Dataset) land cover types was also significantly different between these two classes, with developed land-use types dominating in urban areas (> 87%) and shrub/scrub dominating in wildland areas (> 70%; **Supplementary Table 1**).

Model performance using different sets of climatic variables

Comparisons of the WorldClim dataset for 1960–1990 and 1970–2000, the CHELSA dataset for 1980–2010, and the ClimateNA dataset for 1990–2020 returned very similar performances for Maxent (all comparisons were conducted using the Thinned settings; see section "Modeling using climatic variables across timespans" in supplements). Models using the 1960-1990 WorldClim predictors received marginally, but significantly, higher AUC_{ROC} values (**Supplementary Figure 5**) and we, therefore, conducted all additional modeling using this dataset.

Mean models of habitat suitability

The four mean Maxent models were composed of different numbers of species, given the additional spatial requirements for occurrence records of Thinned models and due to null model validation (section "Null model validation"; **Supplementary Figure 6**): Thinned = 1,023; PhylumBias = 1,196; ClassBias = 1,197; Composite = 1,199 species (derived from 399 Thinned models, 335 PhylumBias and 465 ClassBias models; all based on the full extent modeling). The Composite model contained single-species models ranging in AUC score from

0.629–0.998 (median = 0.843; **Supplementary Figure 7**), built with 25–13,346 occurrence records per species (median = 99); **Supplementary Figure 8**). The mean RandomForest model of habitat suitability was created with the same settings, and for the same set of species, as the Thinned Maxent model (see next paragraph).

Model evaluation

Using the test dataset, the Thinned Maxent model ranked higher for both AUC_{ROC} and AUC_{PRG} values than the three other Maxent models (ClassBias, PhylumBias, and Composite models; **Figure 3**). Random Forest modeling was therefore conducted for the same 1,023 species modeled with Maxent using the Thinned settings, i.e., based on a single observation per species per raster cell as the occurrence dataset and a uniform prior for background point selection. Following the same model evaluation procedure as for the Maxent models, the mean RandomForest model yielded the overall highest-ranking model, outperforming all Maxent models on both AUC_{ROC} and AUC_{PRG} scales (**Figure 3**). Based on this, all further analyses, quantifying hotspots of urban biodiversity, and the responses of different taxonomic groups and native and non-native species to levels of urban intensity are based on mean RandomForest models.

Raster comparisons of the RandomForest model and the four Maxent models (Thinned, ClassBias, PhylumBias, and Composite) were very similar, and pairwise comparisons using Schoener's D ranged from 0.899–0.983, Warren's I from 0.991–1 and correlation coefficients from 0.884–0.997 (**Supplementary Table 3**). Standard deviations between these five models, calculated at the level of each raster cell, ranged from 0.000698–0.165, with a median of 0.073 (**Supplementary Figure 9**).

Sampling bias

The first three axes of the environmental PCA explained 72.5% of the variation (PC1 = 36.7%; PC2 = 22.1%; PC3 = 13.7%) and
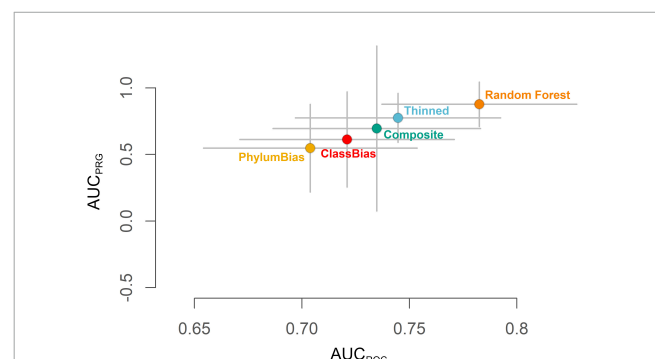

FIGURE 3

Evaluation of the mean RandomForest model of habitat suitability and the four mean Maxent models (Thinned, ClassBias, PhylumBias, and Composite). Grey bars indicate standard deviations around the mean estimates of AUC_{ROC} and AUC_{PRG} following model evaluation using the test dataset.

iNaturalist observations covered this environmental space well (**Supplementary Figure 10**). Random Forest models using the summed iNaturalist observations per raster cell (separately for all species of arthropods, vertebrates, plants, and as a sum of all species) as predictor variables and the mean RANDOMFOREST habitat suitability model as the response variable explained 8.15% of the variation, indicating that mean habitat suitability was not strongly associated with the spatial distribution of iNaturalist observations. Density plots of the highest quartile mean RANDOMFOREST habitat suitability values across the urban PCA space show a similar pattern, with some similarities but also marked differences to the density of all iNaturalist observations across the urban PCA (**Figure 4**).

Effects of the native status of species and different taxonomic groups

Using the mean RANDOMFOREST models of habitat suitability as the response variables and all landscape variables as predictor variables separately for all species, only native and only non-native species, Random Forest models explained 97.5, 97.4, and 98.6% of the variation, respectively. The landscape variables that explained most of the variation depended on the response variable being modeled (**Table 1**); water cover, soil bulk density, NDVI, and imperviousness explained most of the variation for the model containing all species as a response (**Supplementary Figure 11**).

Rank correlation coefficients showed the strongest differences in mean RANDOMFOREST habitat suitability models of only native and only non-native species (correlation coefficient = 0.505; Schoener's D = 0.81; Warren's I = 0.974). This difference persisted in the response of native and non-native species to urban PCA space; native species have higher mean suitability in wildland areas than non-natives, while in areas that are fully urbanized, non-native species have higher mean suitability than natives (**Figure 5** and **Supplementary Figure 12**).

GAM models testing the association of habitat suitability of native and non-native species to the continuous urbanization PCA space were highly significant ($p < 0.001$), with more variation explained for non-native species (adjusted $R^2 = 0.28$) than for native species (adjusted $R^2 = 0.07$). Within native species

FIGURE 4
Density plot of all iNaturalist observations of the training dateset and the highest quartal values of the mean RANDOMFOREST model of habitat suitability, plotted against urbanization PCA space.

(**Figures 6A–C**), GAM models testing the association of habitat suitability and urban PCA space explained most variation in plants (adjusted $R^2 = 43.6\%$), while the models for vertebrates and arthropods had low predictive power (adjusted $R^2 < 1\%$) and all models being highly significant ($p < 0.001$). For non-native species (**Figures 6D–F**), GAM models testing the association of habitat suitability and urban PCA space explained less variation in plants (adjusted $R^2 = 16.9\%$) than in vertebrates and arthropods (adjusted $R^2 = 45\%$ and 52.2%, respectively), again, all models being highly significant ($p < 0.001$).

Discussion

Large, human-dominated ecosystems like Greater Los Angeles differ from more ecologically pristine landscapes in at least two important ways. First, they often contain steep environmental gradients between urban and wildland habitats. In many cases, a single roadway or chain-link fence may separate largely intact natural habitats from areas characterized by high levels of imperviousness, extremely high human population densities, and a correspondingly high concentration of human commensal species. This habitat heterogeneity, over scales of a few hundred meters, presents both challenges and opportunities for different groups of organisms. Second, urban ecosystems tend to have high concentrations of non-native species. Whether these taxa should be considered unwanted pests, integral parts of novel ecosystems (Sax et al., 2022) or valuable elements of *ex situ* endangered species recovery (Shaffer, 2018) depends on the urban context, the goals of urban planners, the preferences of a diverse array of residents, and the conservation status of the species in question (Gaertner et al., 2016). Rather than enter into a debate on the roles of non-native species in urban ecosystems, our goal is to provide the biodiversity distributional data upon which decisions depend, based on best-practice standards for biodiversity assessments (Araújo et al., 2019; **Supplementary Table 4**). We provide models identifying hotspots of urban biodiversity jointly for a broad set of 1,023 commonly observed species, as well as separately for native and non-native species (**Figure 2**). We hope that the modeling framework outlined here for Los Angeles will provide a baseline for future research, and that the resulting data will allow planners to include more comprehensive appraisals of the distribution of biodiversity in their assessments and management plans for other urban centers.

Comparison of methods to model mean habitat suitability

Validation of models using the 2022 iNaturalist test dataset, which included records of more than 2,200 species that were not used to train our models, demonstrated that, across modeling settings, models generated at the full extent outperformed those that were combined from models generated separately in urban and wildland areas (**Supplementary Figures 2**, **3**). We were surprised by this result, but it may indicate that, while significant differences between urban and wildland models persist at the species level (section "Exploring separate modeling in urban and wildland extents" in **Supplementary material**), modeling at the full extent
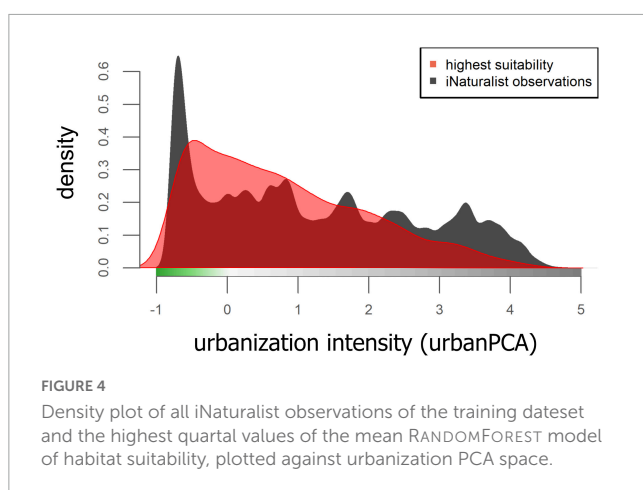
TABLE 1 Importance of landscape variables for mean habitat suitability (%incMSE: percent increase in mean squared error).

All species model		Only native species model		Only non-native species model	
%IncMSE	Landscape predictors	%IncMSE	Landscape predictors	%IncMSE	Landscape predictors
31.3	Water cover	38.5	water cover	30.4	bulk density
27.8	Bulk density	27.3	bulk density	28.9	NDVI
24.8	NDVI	25.0	NDVI	20.2	cation exchange capacity
23.5	Imperviousness	24.2	imperviousness	19.8	climatic water deficit
19.3	Cation exchange capacity	19.7	cation exchange capacity	18.7	organic matter

Models used the mean RANDOMFOREST model of habitat suitability as the response variable and landscape variables as predictors and were run separately for all, only native and only non-native species (explaining > 97% of the variation in all models). Soil cation exchange capacity, soil bulk density, and NDVI are among the five most important landscape variables of all three models, with water cover and imperviousness among the five most important landscape variables of models for all species and only native species.
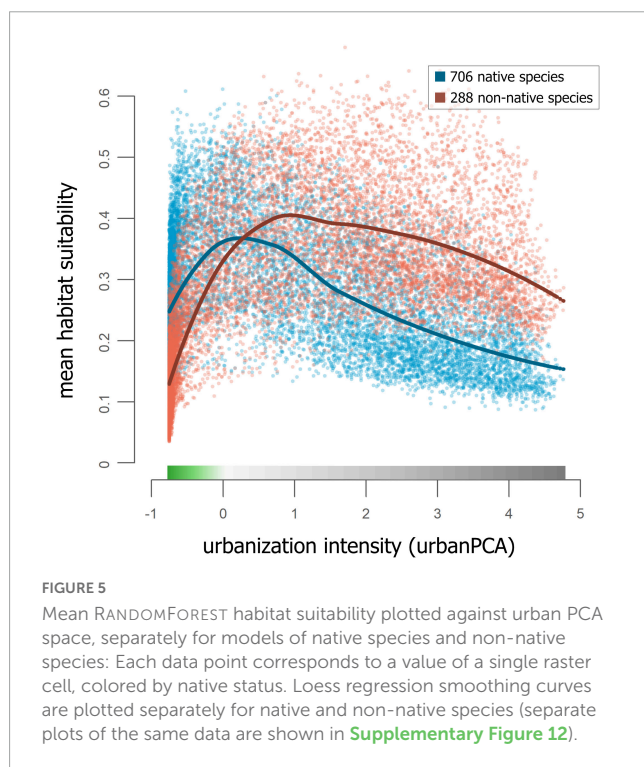

FIGURE 5
Mean RANDOMFOREST habitat suitability plotted against urban PCA space, separately for models of native species and non-native species: Each data point corresponds to a value of a single raster cell, colored by native status. Loess regression smoothing curves are plotted separately for native and non-native species (separate plots of the same data are shown in Supplementary Figure 12).

is the best (and fortunately, also the simplest) approach when generalizing across species.

Among those models generated at the full extent, the mean RANDOMFOREST model of habitat suitability, generated using the THINNED settings, ranked above all four mean Maxent models for both AUC_{ROC} and AUC_{PRG} metrics (Figure 3). Pairwise raster comparisons of all five mean models showed a fairly high similarity between models, with all values of Schoener's D > 0.89, Warren's I > 0.99, and correlation coefficients all > 0.88 (Supplementary Table 3).

By using the test dataset to generate AUC_{ROC} and AUC_{PRG} metrics, we can ask whether our mean RANDOMFOREST model is an accurate representation of unsampled biodiversity for the Greater Los Angeles ecosystem. The high values of AUC_{ROC} (median = 0.783) and AUC_{PRG} (median = 0.877) demonstrate that the mean RANDOMFOREST model of habitat suitability is suited to accurately predict hotspots of urban biodiversity more broadly, especially for species that occur but were not part of the modeling framework. At the same time, we recognize the need to explore

model validation further, particularly at the species level, and ideally using structured survey datasets that contain true absences (Valavi et al., 2022).

Habitat suitability of single species can be closely associated with the observed abundance of that species (Weber et al., 2017; de La Fuente et al., 2021), but it need not be. There is conflicting evidence on this association (Boyce et al., 2016; Dallas and Hastings, 2018), and it has been suggested that predicted habitat suitability values more closely reflect the upper limit of a species' abundance (VanDerWal et al., 2009). Habitat suitability for multiple species, including the mean habitat suitability model constructed here, can be interpreted as the cell-by-cell probability of encountering many species, analogous to interpreting a single species' habitat suitability model as the probability of that species' presence (Elith et al., 2011). Put another way, mean models of habitat suitability are commonly interpreted as reflecting species richness, or alpha diversity, across a modeling extent (Calabrese et al., 2014).

Can iNaturalist records be harnessed as valid indicators of species distributions?

A common, and reasonable critique of community science observational data is that it reflects where people go to observe nature, rather than the true distribution of biodiversity. To some extent, this must be true, just as it is for museum specimen records or sample sites in ecological studies—people tend to go where access is relatively easy (Newbold, 2010; Martin et al., 2012). However, four lines of evidence convince us that the post-filtered iNaturalist dataset paired with the analyses run here are a reasonable representation of true mean habitat suitability rather than a reflection of rates of human visitation. First, the vast majority of the raster cells encompassing the total environmental PCA space in our study extent contain iNaturalist observations (Supplementary Figure 10). While some cells have many observations and some relatively few, the iNaturalist dataset used in this study did not leave unique environmental conditions unsampled. Second, models using iNaturalist observations as predictors and the mean RANDOMFOREST model of habitat suitability of all 1,023 native and non-native species as the response variable explained a modest 8.15% of the variation. If visitation frequency and their associated iNaturalist observations were driving the mean model of habitat suitability, we would expect this to be much higher. Third, density plots of iNaturalist
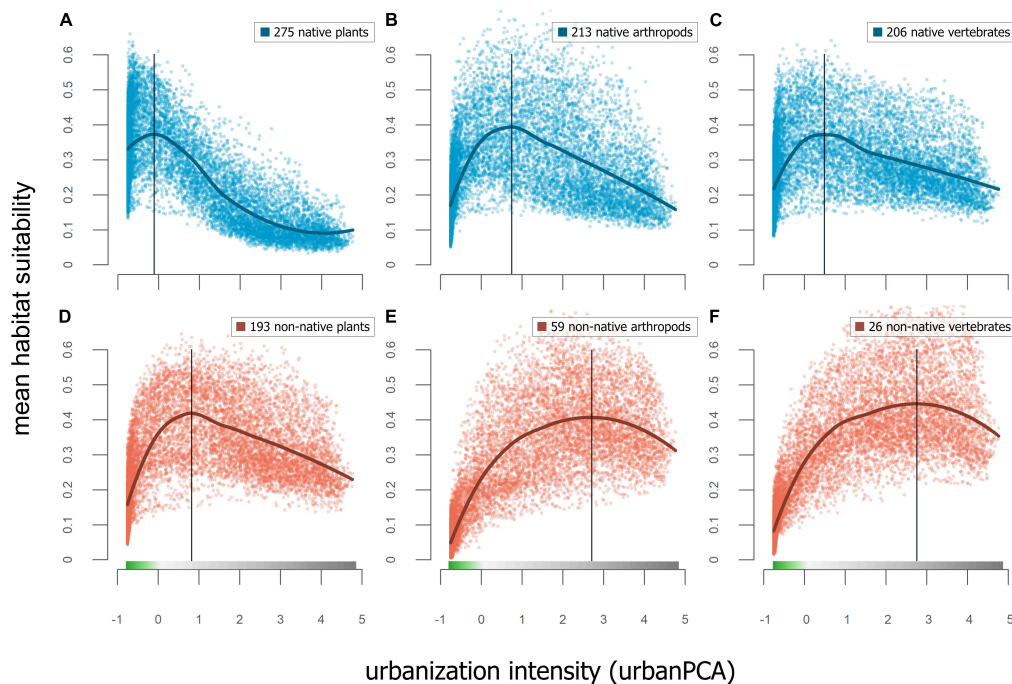
FIGURE 6
Top row shows mean RANDOMFOREST models of habitat suitability for **(A)** native plant species, **(B)** native arthropod species, and **(C)** native vertebrate species. The association to urban intensity were only strong in plants (adjusted R^2 = 43.6%) and low in vertebrates and arthropods (adjusted R^2 < 1%). Bottom row shows mean RANDOMFOREST models of habitat suitability for **(D)** non-native plant species, **(E)** non-native arthropod species, and **(F)** non-native vertebrate species. The association to urban intensity was less strong in plants (adjusted R^2 = 16.9%) and stronger for vertebrates and arthropods (adjusted R^2 = 45% and 52.2%, respectively). Vertical lines show the level of urban intensity with highest mean habitat suitability.

observations and the highest quartal mean habitat suitability values showed considerable mismatch (**Figure 4**), indicating that habitat suitability models are not driven by the sheer number of iNaturalist observations at a given locality. And last, the independent test dataset used to evaluate all mean models of habitat suitability produced median AUC$_{ROC}$ values > 0.7, similar to model validation in other studies (Valavi et al., 2022), indicating that it predicts the presence of many species accurately. Collectively, we interpret this as strong evidence that spatial variation in human sampling intensity was adequately addressed by the methodologies employed here, and that the resulting models can be interpreted as depictions of true multi-species habitat suitability, largely uninfluenced by the location and number of iNaturalist records alone. The one exception to this may be the low habitat suitability modeled for the highest elevation sites in our study extent, although this is likely due to the delineation of the study extent rather than the modeling itself, as discussed below.

Hotspots of urban biodiversity in the Greater Los Angeles ecoregion

In Greater Los Angeles, areas of the highest mean habitat suitability are distributed in a pattern that is closely aligned to, but not identical with, the spatial distribution of wildland habitat (compare **Figures 1**, **2**). The difference is a subtle offset, such that regions of high mean habitat suitability (orange-red in **Figure 2**) very closely align to the urban-wildland interface (the light gray regions of **Figure 1A**), while both very urbanized and very wild areas receive lower values of mean habitat suitability (**Figure 5**). This general result is similar to findings for species richness of plants (McKinney, 2008) and birds (Vale and Vale, 1976), and likely reflects the greater habitat heterogeneity at the interface of this steep environmental gradient. Beninde et al. (2015) found a similar increase in species richness as a function of habitat richness across taxonomic groups in globally distributed cities. The lowest mean habitat suitability was detected in heavily urbanized areas, but also in some of the wildest areas within the study extent, including the San Gabriel and Santa Susana Mountains (**Figure 1B**). The San Gabriel mountains reach the highest elevation (3,069 m) within the study extent and harbor unique environmental conditions, but only make up a limited area within the total study extent. Further analyses indicated that species restricted to high elevation (above 600 m) may have been under-represented in the training dataset, with too few occurrence records of these high-elevation species to pass our 25-observation filter, and this may reduce the apparent suitability of these habitats. However, independent studies from other parts of the world show a similar decrease in species richness with increasing elevation across taxonomic groups (Lee et al., 2004; Nogués-Bravo et al., 2008). Thus, our inferred low habitat suitability in these ecologically intact high-elevation areas could reflect insufficient sampling, true low habitat suitability, or both; increased sampling efforts are necessary to resolve this question.

The landscape variables that stand out in their importance to explain the mean RANDOMFOREST model of habitat suitability were, in decreasing order of importance, water cover, soil bulk

density, NDVI, imperviousness, and soil cation exchange capacity (**Table 1**). The positive effect of water cover and all types of vegetation, as captured by NDVI, on species richness is well known across taxonomic groups in globally distributed cities (Beninde et al., 2015), and may be even more pronounced in the relatively xeric conditions that characterize most of our southern California study extent. Cation exchange capacity, a measure of soil nutrient availability, has a negative impact on mean habitat suitability, similar to findings from non-urban systems that showed reduced plant species richness in soils with higher cation exchange capacity (Huston, 1980; Le Brocque and Buckney, 2003; Palmer et al., 2003). The positive effect of bulk density on habitat suitability deserves further inquiry. Bulk density is an indicator of soil compaction and is rarely used in analyses of species richness. Rather, it is often considered indicative of ecosystem functionality, since high levels of compaction decrease the water storage capacity of soils (Wang et al., 2018). Imperviousness has a negative association with mean habitat suitability, which is highest at low levels of imperviousness and decreases rapidly between 30 and 70% of impervious surface cover, beyond which levels of mean habitat suitability remain consistently low. This pattern is consistent with many other observations across plant and animal species that impervious surface cover reduces species richness, diversity, and abundance (Sattler et al., 2010; Geslin et al., 2016; Gillespie et al., 2017; Choate et al., 2018; Souza et al., 2019; Yan et al., 2019; Piano et al., 2020).

Responses to urbanization by native and non-native species

Many studies synthesizing data across taxonomic groups and scales have found very different responses to urbanization between generalist and specialist species (Callaghan et al., 2019), among taxonomic groups (McKinney, 2008), and between native and non-native taxa (Celesti-Grapow et al., 2006). In line with our most general predictions, native species had high mean habitat suitability in wildland areas, lowest habitat suitability in urban areas, and the highest suitability in the proximity of the transition from wildland to urban habitats (**Figure 5**). In contrast, and consistent with expectations (Cadotte et al., 2017), wildland areas received lower values of mean habitat suitability than urban areas for non-native taxa, and, like native species, non-native taxa peaked in mean habitat suitability at the transition from wildland to urban. Drivers of the difference in mean habitat suitability between urban and wildland areas for non-native species need to be explored further, as knowledge on this is limited (Cadotte et al., 2017; Spear et al., 2017). Such a difference in mean habitat suitability for non-native species may be higher in cities that have strongly seasonal and relatively arid Mediterranean climates, including Greater Los Angeles, than in less arid cities. Although this has not to our knowledge been explicitly examined, we suspect that supplemental watering may be a stronger environmental influence at urban-wildland interfaces in arid or extremely seasonal climates, leading to more severe environmental gradients and reduced spillover of non-native urban species into adjacent wildland areas. More generally, following the categorization of species based on their urbanization tolerance (Fischer et al., 2015), our results confirm findings from a global

analysis of urban bird and plant species (Aronson et al., 2014) and demonstrate that native species tend to be less urban tolerant than non-native species in Greater Los Angeles.

Similar responses to urbanization by different taxonomic groups

Our findings indicate that most taxonomic groups have hump-shaped responses in mean habitat suitability with respect to urbanization, peaking in the proximity of the urban-wildland interface. Comparisons between taxonomic groups, conducted separately for native and non-native species, showed similar responses (**Figure 6**). Native plant, arthropod, and vertebrate species show peaks at similar levels of urban intensity, with maxima at the transition from urban to wildland. In non-native species, these peaks shift toward higher levels of urban intensity, although this shift is strongest in arthropod and vertebrate species. These findings are consistent with other studies that found the highest levels of plant species richness at intermediate levels of urbanization across cities (McKinney, 2008). The higher habitat suitability values of native plants in relatively more wildland areas, in comparison to that of native arthropods and vertebrates, may be explained by the unique positioning of our study within the California Floristic Province biodiversity hotspot (Myers et al., 2000).

The decline in habitat suitability with increasing intensity of urbanization is gradual rather than showing an obvious threshold or step-cline pattern. However, the strengths of these associations for taxonomic groups were variable. The strongest associations were found for native plants (adjusted R^2 = 43.6%) and for non-native vertebrates and arthropods (adjusted R^2 = 45% and 52.2%, respectively). These results warrant further research into the responses of other taxonomic groups, for plants and animals, to levels of urban intensity. Highly variable responses in species richness of various insect taxa have been demonstrated by urban-rural comparisons in temperate European cities, with some taxa peaking in urban areas, others in rural areas, and some showing no significant differences between the two (Theodorou et al., 2020). A comparison of levels of avian species richness within multiple Mexican cities demonstrated that bird species richness was higher in green spaces than in areas with more impervious surfaces, although this varied with the functional group of species (MacGregor-Fors et al., 2021). Following this example, future studies could include the response of different functional groups, potentially including additional species traits such as aspects of life history or physiology, to explore the mechanisms underlying species' responses to urbanization across taxonomic groups.

Conservation efforts in Greater Los Angeles

The City of Los Angeles has the ambitious, and admirable, goal of no net loss of biodiversity by 2050 (City of Los Angeles, 2019). Given that many species in the region are negatively affected and threatened by urbanization (Vandergast et al., 2009; Thomassen et al., 2018; Gustafson et al., 2019), many existing and pending plans focus on protecting and enhancing existing

wildland areas. Efforts to mitigate the risks of future urbanization include the Annenberg Wildlife Crossing in Liberty Canyon, the Wildlife Pilot Study (City of Los Angeles, 2014), and the Rim of the Valley Corridor (Zellmer and Goto, 2022), which together aim to protect large habitat patches and existing and constructed connections between them to allow wildlife to achieve long-term persistence. Spatially, the extent of the Wildlife Pilot Study covers a portion of the eastern Santa Monica Mountains, while the Rim of the Valley extends around the San Fernando Valley to include portions of the Santa Monica Mountains, Simi Hills, Santa Susana, and San Gabriel Mountains, and Verdugo Hills (**Figure 1B**), including the Annenberg Wildlife Crossing. Many of these areas have received among the highest mean habitat suitability scores from our models and are thus rightful candidates for protection. However, our models also emphasize other regions with high suitability values, including the urban-wildland interface regions along the southern flanks of the San Gabriel Mountains, and pockets of urban open space dotted across the region. Many of these regions are relatively modest in size compared to large wildlands; key regions include the Sepulveda Basin, Baldwin Hills, Ballona Wetlands, Dominguez Gap Wetlands, Coyote Hills, Whittier Narrows, and Upper Newport Bay. These regions emphasize the well-established importance of urban green and open spaces for urban biodiversity, including sometimes-isolated or small patches (Beninde et al., 2015; Wintle et al., 2019).

Many of these highly suitable areas also encompass the most affluent areas in the region, including Bel Air, Beverly Glen, and the Hollywood Hills, while extensive areas of low habitat suitability often fall in low-income neighborhoods including Downtown and South Los Angeles. Both formal policy and broadly accepted equity concerns demand that the positive effects of nature and biodiversity should be accessible to, and impactful for, all, regardless of wealth (Schell et al., 2020). Our models highlight that providing equitable access to areas with high mean habitat suitability presents a real challenge and needs to become an integral goal for future biodiversity planning in Greater Los Angeles. While the habitat suitability models presented here can identify areas that have particularly low levels of biodiversity, corresponding efforts to restore sites and provide green-space access also need to take into account threats of green gentrification, further complicating such efforts (Maantay and Maroko, 2018). To put this in the context of Morrison (2016), achieving virtuous cycles that enhance biodiversity conservation will require different inputs, given the very different human constituencies interacting with that biodiversity, in different parts of Greater Los Angeles. The places, people, and benefits from and for nature are different, and require different approaches, even if the consistent goal is increased biodiversity.

Conclusion: Outlook on the application of iNaturalist for urban ecology and conservation

While data deficiency plagues biodiversity research globally (Hochkirch et al., 2021), and knowledge gaps in urban areas persist that range from the identification of habitat patch size thresholds, to evolutionary trap and population sink dynamics and the best landscape configuration to facilitate dispersal (Aronson et al., 2017), the accelerating number of species observations from urban community scientists is unprecedented and encouraging (Callaghan et al., 2020a). With growing confidence in adequately addressing the biases inherent in community science datasets using habitat suitability modeling techniques, iNaturalist and similar datasets have become invaluable resources allowing in-depth comparisons of thousands of species across cities globally. In the future, such data should allow tracking of changes in distributional patterns of taxa along urbanization gradients. These data have already been used to document shifts over the last decades for Los Angeles birds (Cooper et al., 2020), and the recent displacement of the region's native urban black widow spiders by the introduced congeneric brown widow spider (Kempf et al., 2021). The modeling framework outlined here can and should be expanded upon to include biotic interactions (Dormann et al., 2018). This can include methodological approaches, including linking them to macroecological models and comparisons of inferences to multi-species occupancy models (Calabrese et al., 2014; Devarajan et al., 2020), and empirically by integrating presence-absence modeling techniques and data (Isaac et al., 2020). A key goal should be to corroborate habitat suitability modeling from iNaturalist datasets with other, independent data sources, such as scientific monitoring surveys (Prudic et al., 2018) and field validation studies (Searcy and Shaffer, 2014). Furthermore, the increasing availability of high-resolution observations and data layers could allow for modeling fine-scale impacts of smaller patches of urban green spaces (Beninde et al., 2015) or scale-dependent effects across cities (Alberti and Wang, 2022). At its core, our study creates a resource for use by urban planners in Greater Los Angeles and provides a framework that other cities can implement to generate a more comprehensive understanding of the spatial distribution of biodiversity value in their region. Using this framework can provide policymakers with a spatially explicit tool for implementing planning strategies that are most appropriate for biodiversity conservation. The data exist and should be used.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The code to create Maxent and Random Forest models is provided at https://github.com/jbeninde/HotspotsUrbanBiodiversity.

Author contributions

JB and HBS designed the study. GG and TWD performed analyses on 15 species to explore Maxent settings. JB performed all subsequent analyses. JB wrote a first version of the manuscript, with assistance from GG and TWD. All authors contributed to initial reviews and editing with JB and HBS completing data revisions and edits.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor EW declared a shared committee with the authors JB and HBS at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2023.983371/full#supplementary-material

References

Adams, B., Li, E., Bahlai, C., Meineke, E., McGlynn, T., and Brown, B. (2020). Local- and landscape-scale variables shape insect diversity in an urban biodiversity hot spot. *Ecol. Appl.* 30:e02089. doi: 10.1002/eap.2089

AdaptWest Project (2021). *Gridded current and future climate data for North America at 1km resolution, interpolated using the ClimateNA v7.01 software (T. Wang et al., 2021).* Available online at: adaptwest.databasin.org (accessed September 13, 2022).

Alberti, M., and Wang, T. (2022). Detecting patterns of vertebrate biodiversity across the multidimensional urban landscape. *Ecol. Lett.* 25, 1027–1045. doi: 10.1111/ele.13969

Araújo, M., Anderson, R., Márcia Barbosa, A., Beale, C., Dormann, C., Early, R., et al. (2019). Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5:eaat4858. doi: 10.1126/sciadv.aat4858

Aronson, M. F. J., Lepczyk, C. A., Evans, K. L., Goddard, M. A., Lerman, S. B., MacIvor, J. S., et al. (2017). Biodiversity in the city: Key challenges for urban green space management. *Front. Ecol. Environ.* 15:189–196. doi: 10.1002/fee.1480

Aronson, M., La Sorte, F., Nilon, C., Katti, M., Goddard, M., Lepczyk, C., et al. (2014). A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proc. R. Soc. B* 281:20133330. doi: 10.1098/rspb.2013.3330

Beck, J., Böller, M., Erhardt, A., and Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* 19, 10–15. doi: 10.1016/j.ecoinf.2013.11.002

Beninde, J., Veith, M., and Hochkirch, A. (2015). Biodiversity in cities needs space: A meta-analysis of factors determining intra-urban biodiversity variation. *Ecol. Lett.* 18, 581–592. doi: 10.1111/ele.12427

Leutner, B., Horning, N., and Jakob, S. (2022). *Rstoolbox: Tools for remote sensing data analysis.* Available online at: https://github.com/bleutner/RStoolbox

Boyce, M. S., Johnson, C. J., Merrill, E. H., Nielsen, S. E., Solberg, E. J., and van Moorter, B. (2016). Review: Can habitat selection predict abundance? *J. Anim. Ecol.* 85, 11–20. doi: 10.1111/1365-2656.12359

Bradie, J., and Leung, B. (2016). A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *J. Biogeogr.* 44, 1344–1361. doi: 10.1111/jbi.12894

Branton, M., and Richardson, J. S. (2011). Assessing the value of the umbrella-species concept for conservation planning with meta-analysis. *Conserv. Biol.* 25, 9–20. doi: 10.1111/j.1523-1739.2010.01606.x

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Cadotte, M., Yasui, S., Livingstone, S., and MacIvor, J. (2017). Are urban systems beneficial, detrimental, or indifferent for biological invasion? *Biol. Invasions* 19, 3489–3503. doi: 10.1007/s10530-017-1586-y

Calabrese, J., Certain, G., Kraan, C., and Dormann, C. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* 23, 99–112. doi: 10.1111/geb.12102

Calflora (2022). *Information on California plants for education, research and conservation.* Available online at: https://www.calflora.org/ (accessed June 16, 2022)

Callaghan, C. T., Major, R. E., Wilshire, J. H., Martin, J. M., Kingsford, R. T., and Cornwell, W. K. (2019). Generalists are the most urban-tolerant of birds: A phylogenetically controlled analysis of ecological and life history traits using a novel continuous measure of bird responses to urbanization. *Oikos* 128, 845–858. doi: 10.1111/oik.06158

Callaghan, C. T., Ozeroff, I., Hitchcock, C., and Chandler, M. (2020a). Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework. *Biol. Conserv.* 251:108753. doi: 10.1016/j.biocon.2020.108753

Callaghan, C. T., Roberts, J. D., Poore, A. G. B., Alford, R. A., Cogger, H., and Rowley, J. J. L. (2020b). Citizen science data accurately predicts expert-derived species richness at a continental scale when sampling thresholds are met. *Biodivers. Conserv.* 29, 1323–1337. doi: 10.1007/s10531-020-01937-3

Celesti-Grapow, L., Pysek, P., Jarosik, V., and Blasi, C. (2006). Determinants of native and alien species richness in the urban flora of Rome. *Divers. Distrib.* 12, 490–501. doi: 10.1111/j.1366-9516.2006.00282.x

Choate, B. A., Hickman, P. L., and Moretti, E. A. (2018). Wild bee species abundance and richness across an urban–rural gradient. *J. Insect Conserv.* 22, 391–403. doi: 10.1007/s10841-018-0068-6

City of Los Angeles (2014). *Motion 14-0518: Wildlife Corridor/Santa Monica Mountains (Hillside Ordinance Zone)/Zone Change (ZC).* Available online at: https://cityclerk.lacity.org/lacityclerkconnect/index.cfm?fa=ccfi.viewrecord&cfnumber=14-0518 (accessed May 30, 2022).

City of Los Angeles (2019). *L.A.'s Green New Deal: Los Angeles Sustainable City pLAn.* Available online at: https://plan.lamayor.org/sites/default/files/pLAn_2019_final.pdf (accessed May 17, 2021).

Cooper, C., Hawn, C., Larson, L., Parrish, J., Bowser, G., Cavalier, D., et al. (2021). Inclusion in citizen science: The conundrum of rebranding. *Science* 372, 1386–1388. doi: 10.1126/science.abi6487

Cooper, D. S., Yeh, P. J., and Blumstein, D. T. (2020). Tolerance and avoidance of Urban cover in a southern California suburban raptor community over five decades. *Urban Ecosyst.* 24, 291–300. doi: 10.1007/s11252-020-01035-w

Dallas, T. A., and Hastings, A. (2018). Habitat suitability estimated by niche models is largely unrelated to species abundance. *Glob. Ecol. Biogeogr.* 27, 1448–1456. doi: 10.1111/geb.12820

de La Fuente, A., Hirsch, B. T., Cernusak, L. A., and Williams, S. E. (2021). Predicting species abundance by implementing the ecological niche theory. *Ecography* 44, 1723–1730. doi: 10.1111/ecog.05776

Devarajan, K., Morelli, T., and Tenan, S. (2020). Multi-species occupancy models: Review, roadmap, and recommendations. *Ecography* 43, 1612–1624. doi: 10.1111/ecog.04957

Devictor, V., Whittaker, R. J., and Beltrame, C. (2010). Beyond scarcity: Citizen science programmes as useful tools for conservation biogeography. *Divers. Distrib.* 16, 354–362. doi: 10.1111/j.1472-4642.2009.00615.x

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., and Hurlbert, A. H. (2021). Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *Bioscience* 71, 1179–1188. doi: 10.1093/biosci/biab093

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. doi: 10.1111/j.1600-0587.2012.07348.x

Dormann, C., Bobrowski, M., Dehling, D., Harris, D., Hartig, F., Lischke, H., et al. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Glob. Ecol. Biogeogr.* 27, 1004–1016. doi: 10.1111/geb.12759

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Fick, S., and Hijmans, R. (2017). WorldClim 2. New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086

Fischer, J. D., Schneider, S. C., Ahlers, A. A., and Miller, J. R. (2015). Categorizing wildlife responses to urbanization and conservation implications of terminology. *Conserv. Biol.* 29, 1246–1248. doi: 10.1111/cobi.12451

Fleishman, E., Blair, R. B., and Murphy, D. D. (2001). EMPIRICAL VALIDATION OF A METHOD FOR UMBRELLA SPECIES SELECTION. *Ecol. Appl.* 11, 1489–1501. doi: 10.1890/1051-0761(2001)011[1489:EVOAMF]2.0.CO;2

Flint, L. E., Flint, A. L., Thorne, J. H., and Boynton, R. (2013). Fine-scale hydrologic modeling for regional landscape applications: The California basin characterization model development and performance. *Ecol. Process.* 2, 1–21. doi: 10.1186/2192-1709-2-25

Fourcade, Y., Engler, J. O., Rödder, D., and Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS One* 9:e97122. doi: 10.1371/journal.pone.0097122

Fuller, R. A., Irvine, K. N., Devine-Wright, P., Warren, P. H., and Gaston, K. J. (2007). Psychological benefits of greenspace increase with biodiversity. *Biol. Lett.* 3, 390–394. doi: 10.1098/rsbl.2007.0149

Gaertner, M., Larson, B., Irlich, U., Holmes, P., Stafford, L., van Wilgen, B., et al. (2016). Managing invasive species in cities: A framework from cape town, South Africa. *Landsc. Urban Plan.* 151, 1–9. doi: 10.1016/j.landurbplan.2016.03.010

GBIF.org (2022). *GBIF occurrence download*. Available online at: https://doi.org/10.15468/dl.yqac5z (accessed October 14, 2022).

Geslin, B., Le Féon, V., Folschweiller, M., Flacher, F., Carmignac, D., Motard, E., et al. (2016). The proportion of impervious surfaces at the landscape scale structures wild bee assemblages in a densely populated region. *Ecol. Evol.* 6, 6599–6615. doi: 10.1002/ece3.2374

Gillespie, T. W., de Goede, J., Aguilar, L., Jenerette, G. D., Fricker, G. A., Avolio, M. L., et al. (2017). Predicting tree species richness in urban forests. *Urban Ecosyst.* 20, 839–849. doi: 10.1007/s11252-016-0633-2

Gomes, V. H. F., Ijff, S. D., Raes, N., Amaral, I. L., Salomão, R. P., de Souza Coelho, L., et al. (2018). Species distribution modelling: Contrasting presence-only models with plot abundance data. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-017-18927-1

Groffman, P. M., Cavender-Bares, J., Bettez, N. D., Grove, J. M., Hall, S. J., Heffernan, J. B., et al. (2014). Ecological homogenization of urban USA. *Front. Ecol. Environ.* 12:74–81. doi: 10.1890/120374

Gustafson, K. D., Gagne, R. B., Vickers, T. W., Riley, S. P. D., Wilmers, C. C., Bleich, V. C., et al. (2019). Genetic source–sink dynamics among naturally structured and anthropogenically fragmented puma populations. *Conserv. Genet.* 20, 215–227. doi: 10.1007/s10592-018-1125-0

Harrigan, R., Thomassen, H., Buermann, W., and Smith, T. (2014). A continental risk assessment of West Nile virus under climate change. *Glob. Change Biol.* 20, 2417–2425. doi: 10.1111/gcb.12534

Hijmans, R. J., and van Etten, J. (2022). *Raster: Geographic analysis and modeling with raster data. R package version 3.5-15*. Available online at: https://CRAN.R-project.org/package=raster

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hijmans, R. J., Phillips, S. J., Leathwick, J., and Elith, J. (2017). *Dismo: Species distribution modeling*. R package version 1.0-12. Available online at: http://CRAN.R-project.org/package=dismo

Hochkirch, A., Samways, M. J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P., et al. (2021). A strategy for the next decade to address data deficiency in neglected biodiversity. *Conserv. Biol.* 35, 502–509. doi: 10.1111/cobi.13589

Huston, M. (1980). Soil nutrients and tree species richness in costa rican forests. *J. Biogeogr.* 7, 147–157. doi: 10.2307/2844707

Isaac Brown Ecology Studio and La Sanitation and Environment (2018). *2018 biodiversity report: City of Los Angeles. Measurement of the Singapore index of Cities' biodiversity and recommendations for a customized Los Angeles index*. Available online at: https://www.lacitysan.org/cs/groups/public/documents/document/y250/mdi0/~edisp/cnt024743.pdf (accessed September 18, 2018).

Isaac, N., Jarzyna, M., Keil, P., Dambly, L., Boersch-Supan, P., Browning, E., et al. (2020). Data integration for large-scale models of species distributions. *Trends Ecol. Evol.* 35, 56–67. doi: 10.1016/j.tree.2019.08.006

Johnson, M. T. J., and Munshi-South, J. (2017). Evolution of life in urban environments. *Science* 358:eaam8327. doi: 10.1126/science.aam8327

Kaky, E., Nolan, V., Alatawi, A., and Gilbert, F. (2020). A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecol. Inform.* 60:101150. doi: 10.1016/j.ecoinf.2020.101150

Karger, D., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4:170122.

Karger, D., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R., et al. (2018). *Data from: Climatologies at high resolution for the earth's land surface areas*. Davis, CA: Dryad. doi: 10.1038/sdata.2017.122

Keeley, J. E. (2020). Protecting the WUI in California: Greenbelts vs thinning for wildfire threats to homes. *Bull. South. Calif. Acad. Sci.* 119, 35–47. doi: 10.3160/0038-3872-119.1.35

Kempf, J. K., Adams, B. J., and Brown, B. V. (2021). Urban spider diversity in Los Angeles assessed using a community science approach. *Urban Nat.* 40, 1–10.

Kling, M. M., Mishler, B. D., Thornhill, A. H., Baldwin, B. G., and Ackerly, D. D. (2018). Facets of phylodiversity: Evolutionary diversification, divergence and survival as conservation targets. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20170397. doi: 10.1098/rstb.2017.0397

Kohsaka, R., Pereira, H. M., Elmqvist, T., Chan, L., Moreno-Peñaranda, R., Morimoto, Y., et al. (2013). "Indicators for management of Urban biodiversity and ecosystem services: City biodiversity index," in *Urbanization, biodiversity and ecosystem services: Challenges and opportunities*, eds T. Elmqvist, M. Fragkias, J. Goodness, B. Güneralp, P. J. Marcotullio, R. I. McDonald, et al. (Dordrecht: Springer Netherlands), 699–718. doi: 10.1007/978-94-007-7088-1_32

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., et al. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19, 1366–1379. doi: 10.1111/ddi.12096

Kull, M. (2016). *Prg: R package for creating Precision-Recall-Gain curves and calculating area under the curve. V0.5.1*. Available online at: https://github.com/meeliskull/prg/tree/master/R_package (accessed October 21, 2022).

Le Brocque, A. F., and Buckney, R. T. (2003). Species richness-environment relationships within coastal sclerophyll and mesophyll vegetation in Ku-ring-gai Chase National Park, New South Wales, Australia. *Aust. Ecol.* 28, 404–412. doi: 10.1046/j.1442-9993.2003.01298.x

Lee, P. F., Ding, T. S., Hsu, F. H., and Geng, S. (2004). Breeding bird species richness in Taiwan distribution on gradients of elevation, primary productivity and urbanization. *J. Biogeogr.* 31, 307–314. doi: 10.1046/j.0305-0270.2003.00988.x

Li, E., Parker, S. S., Pauly, G. B., Randall, J. M., Brown, B. V., and Cohen, B. S. (2019). An Urban biodiversity assessment framework that combines an Urban habitat classification scheme and citizen science data. *Front. Ecol. Evol.* 7:277. doi: 10.3389/fevo.2019.00277

Liaw, A., and Wiener, M. (2022). Classification and regression by randomforest. *R News* 2, 18–22.

Maantay, J. A., and Maroko, A. R. (2018). Brownfields to greenfields: Environmental justice versus environmental gentrification. *Int. J. Environ. Res. Public Health* 15:2233. doi: 10.3390/ijerph15102233

MacGregor-Fors, I., Escobar-Ibáñez, J., Schondube, J., Zuria, I., Ortega-Álvarez, R., Sosa-López, et al. (2021). The Urban contrast: A nationwide assessment of avian diversity in Mexican cities. *Sci. Total Environ.* 753:141915. doi: 10.1016/j.scitotenv.2020.141915

Martin, L. J., Blossey, B., and Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Front. Ecol. Environ.* 10:195–201. doi: 10.1890/110154

McKinney, M. L. (2008). Effects of urbanization on species richness: A review of plants and animals. *Urban Ecosyst.* 11, 161–176. doi: 10.1007/s11252-007-0045-4

Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M., and Vanaverbeke, J. (2011). Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecol. Model.* 222, 588–597. doi: 10.1016/j.ecolmodel.2010.11.016

Merow, C., Smith, M. J., and Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography* 36, 1058–1069. doi: 10.1111/j.1600-0587.2013.07872.x

Methorst, J., Rehdanz, K., Mueller, T., Hansjürgens, B., Bonn, A., and Böhning-Gaese, K. (2020). The importance of species diversity for human well-being in Europe. *Ecol. Econ.* 181:106917. doi: 10.1016/j.ecolecon.2020.106917

Morrison, S. A. (2016). Designing virtuous socio-ecological cycles for biodiversity conservation. *Biol. Conserv.* 195, 9–16. doi: 10.1016/j.biocon.2015.12.022

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi: 10.1038/35002501

Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Prog. Phys. Geogr.* 34, 3–22. doi: 10.1177/0309133309355630

Nogués-Bravo, D., Araújo, M. B., Romdal, T., and Rahbek, C. (2008). Scale effects and human impact on the elevational species richness gradients. *Nature* 453, 216–219. doi: 10.1038/nature06812

ORNL DAAC (2018). *MODIS and VIIRS land products global subsetting and visualization tool.* Oak Ridge, TN: ORNL Distributed Active Archive Center.

Palmer, M. W., Arévalo, J. R., Del Cobo, M. C., and Earls, P. G. (2003). Species richness and soil reaction in a Northeastern Oklahoma landscape. *Folia Geobot.* 38, 381–389. doi: 10.1007/BF02803246

Phillips, S. J., Dudík, M., and Schapire, R. E. (2004). "A maximum entropy approach to species distribution modeling," in *Proceedings of the twenty-first international conference on Machine learning*, ed. C. Brodley (New York, NY: ACM), 655–662. doi: 10.1145/1015330.1015412

Phillips, S. J., Dudík, M., and Schapire, R. E. (2022). *Maxent software for modeling species niches and distributions (Version 3.4.4).* Available online at: http://biodiversityinformatics.amnh.org/open_source/maxent (accessed January 19, 2022).

Piano, E., Souffreau, C., Merckx, T., Baardsen, L. F., Backeljau, T., Bonte, D., et al. (2020). Urbanization drives cross-taxon declines in abundance and diversity at multiple spatial scales. *Glob. Change Biol.* 26, 1196–1211. doi: 10.1111/gcb.14934

Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., and Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39, 542–552. doi: 10.1111/ecog.01509

Prudic, K. L., Oliver, J. C., Brown, B. V., and Long, E. C. (2018). Comparisons of citizen science data-gathering approaches to evaluate Urban butterfly diversity. *Insects* 9:186. doi: 10.3390/insects9040186

Raes, N., and ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography* 30, 727–736. doi: 10.1111/j.2007.0906-7590.05041.x

R Core Team (2021). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Ricotta, C., Godefroid, S., and Rocchini, D. (2010). Patterns of native and exotic species richness in the urban flora of Brussels rejecting the 'rich get richer' model. *Biol. Invasions* 12, 233–240. doi: 10.1007/s10530-009-9445-0

Roberge, J.-M., and Angelstam, P. E. (2004). Usefulness of the umbrella species concept as a conservation tool. *Conserv. Biol.* 18, 76–85. doi: 10.1111/j.1523-1739.2004.00450.x

Rose, A. N., McKee, J. J., Urban, M. L., and Bright, E. A. (2017). *LandScan 2017.* Oak Ridge, TN: Oak Ridge National Laboratory.

Saito, T., and Rehmsmeier, M. (2017). Precrec: Fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* 33, 145–147. doi: 10.1093/bioinformatics/btw570

Sandifer, P. A., Sutton-Grier, A. E., and Ward, B. P. (2015). Exploring connections among nature, biodiversity, ecosystem services, and human health and well-being: Opportunities to enhance health and biodiversity conservation. *Ecosyst. Serv.* 12, 1–15. doi: 10.1016/j.ecoser.2014.12.007

Sattler, T., Duelli, P., Obrist, M. K., Arlettaz, R., and Moretti, M. (2010). Response of arthropod species richness and functional groups to urban habitat structure and management. *Landsc. Ecol.* 25, 941–954. doi: 10.1007/s10980-010-9473-2

Sattler, T., Pezzatti, G. B., Nobis, M. P., Obrist, M. K., Roth, T., and Moretti, M. (2013). Selection of multiple umbrella species for functional and taxonomic diversity to represent Urban biodiversity. *Conserv. Biol.* 28, 414–426. doi: 10.1111/cobi.12213

Sax, D., Schlaepfer, M., and Olden, J. (2022). Valuing the contributions of non-native species to people and nature. *Trends Ecol. Evol.* 37, 1058–1066. doi: 10.1016/j.tree.2022.08.005

Schell, C. J., Dyson, K., Fuentes, T. L., Des Roches, S., Harris, N. C., Miller, D. S., et al. (2020). The ecological and evolutionary consequences of systemic racism in urban environments. *Science* 369:eaay4497. doi: 10.1126/science.aay4497

Schlaepfer, M. A., Guinaudeau, B. P., Martin, P., and Wyler, N. (2020). Quantifying the contributions of native and non-native trees to a city's biodiversity and ecosystem services. *Urban For. Urban Green.* 56:126861. doi: 10.1016/j.ufug.2020.126861

Schuttler, S. G., Sorensen, A. E., Jordan, R. C., Cooper, C., and Shwartz, A. (2018). Bridging the nature gap: Can citizen science reverse the extinction of experience? *Front. Ecol. Environ.* 16:405–411. doi: 10.1002/fee.1826

Searcy, C. A., and Shaffer, H. B. (2014). Field validation supports novel niche modeling strategies in a cryptic endangered amphibian. *Ecography* 37, 983–992. doi: 10.1111/ecog.00733

Shaffer, H. B. (2018). Urban biodiversity arks. *Nat. Sustain.* 1, 725–727. doi: 10.1038/s41893-018-0193-y

Sillero, N., and Barbosa, A. M. (2021). Common mistakes in ecological niche models. *Int. J. Geogr. Inf. Sci.* 35, 213–226. doi: 10.1080/13658816.2020.1798968

Smallwood, N., and Wood, E. (2023). The ecological role of native-plant landscaping in residential yards to birds during the nonbreeding period. *Ecosphere* 14:e4360. doi: 10.1002/ecs2.4360

Soga, M., and Gaston, K. J. (2016). Extinction of experience: The loss of human-nature interactions. *Front. Ecol. Environ.* 14:94–101. doi: 10.1002/fee.1225

Souza, F. L., Valente-Neto, F., Severo-Neto, F., Bueno, B., Ochoa-Quintero, J. M., Laps, R. R., et al. (2019). Impervious surface and heterogeneity are opposite drivers to maintain bird richness in a Cerrado city. *Landsc. Urban Plan.* 192:103643. doi: 10.1016/j.landurbplan.2019.103643

Spear, D., Pauly, G., and Kaiser, K. (2017). Citizen Science as a Tool for Augmenting Museum Collection Data from Urban Areas. *Front. Ecol. Evol.* 5:86. doi: 10.3389/fevo.2017.00086

Steen, V. A., Tingley, M. W., Paton, P. W. C., and Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods Ecol. Evol.* 12, 216–226. doi: 10.1111/2041-210X.13525

The Earth Observatory Group (2018). *NOAA national centers for environmental information (NCEI): VIIRS Day/Night Band Nighttime Lights.* Available online at: https://www.ngdc.noaa.gov/eog/viirs/ (accessed November 2, 2018).

Theodorou, P., Radzevièiûtë, R., Lentendu, G., Kahnt, B., Husemann, M., Bleidorn, C., et al. (2020). Urban areas as hotspots for bees and pollination but not a panacea for all insects. *Nat. Commun.* 11:576. doi: 10.1038/s41467-020-14496-6

Thomassen, H. A., Harrigan, R. J., Semple Delaney, K., Riley, S. P. D., Serieys, L. E., Pease, K., et al. (2018). Determining the drivers of population structure in a highly urbanized landscape to inform conservation planning. *Conserv. Biol.* 32, 148–158. doi: 10.1111/cobi.12969

U.S. Census Bureau (2018). *Urban and rural classification.* Suitland, MD: U.S. Census Bureau.

U.S. Geological Survey (2014). *NLCD 2011 percent developed imperviousness (2011 Edition, amended 2014) – national geospatial data asset (n.d.) land use land cover.* Reston, VA: U.S. Geological Survey.

U.S. Geological Survey (2017). *3D elevation program 1-meter resolution digital elevation model (published 20170220).* Reston, VA: U.S. Geological Survey.

U.S. Geological Survey (2019). *National hydrography dataset (ver. USGS National hydrography dataset best resolution (n.d.) for hydrologic unit (HU) 4 – 2001 (published 20190221)).* Reston, VA: U.S. Geological Survey.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J., and Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecol.Monogr.* 92:e01486. doi: 10.1002/ecm.1486

Vale, T. R., and Vale, G. R. (1976). Suburban bird populations in west-central California. *J. Biogeogr.* 3, 157–165. doi: 10.2307/3038144

Vandergast, A. G., Lewallen, E. A., Deas, J., Bohonak, A. J., Weissman, D. B., and Fisher, R. N. (2009). Loss of genetic connectivity and diversity in urban microreserves in a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: Stenopelmatus n. sp. "santa monica"). *J. Insect Conserv.* 13, 329–345. doi: 10.1007/s10841-008-9176-z

VanDerWal, J., Shoo, L. P., Johnson, C. N., and Williams, S. E. (2009). Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance. *Am. Nat.* 174, 282–291. doi: 10.1086/600087

Waldrop, M. M. (2019). News Feature: The quest for the sustainable city. *Proc. Natl. Acad. Sci. U.S.A.* 116, 17134–17138. doi: 10.1073/pnas.1912802116

Walkinshaw, M., O'Green, A. T., and Beaudette, D. E. (2021). *Soil properties. California Soil Resource Lab.* Available online at: https://casoilresource.lawr.ucdavis.edu/soil-properties/ (accessed October 1, 2021).

Wang, P., Zheng, H., Ren, Z., Zhang, D., Zhai, C., Mao, Z., et al. (2018). Effects of urbanization, soil property and vegetation configuration on soil infiltration of Urban

forest in Changchun, Northeast China. *Chin. Geogr. Sci.* 28, 482–494. doi: 10.1007/s11769-018-0953-7

Wang, T., Hamann, A., Spittlehouse, D., and Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* 11:e0156720. doi: 10.1371/journal.pone.0156720

Warren, D. L., Glor, R. E., and Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography* 33, 607–611. doi: 10.1111/j.1600-0587.2009.06142.x

Warren, D. L., Matzke, N. J., Cardillo, M., Baumgartner, J. B., Beaumont, L. J., Turelli, M., et al. (2021). ENMTools 1.0: An R package for comparative ecological biogeography. *Ecography* 44, 504–511. doi: 10.1111/ecog.05485

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., and Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography* 40, 817–828. doi: 10.1111/ecog.02125

Wintle, B., Kujala, H., Whitehead, A., Cameron, A., Veloz, S., Kukkala, A., et al. (2019). Global synthesis of conservation studies reveals the importance of small habitat patches for biodiversity. *PNAS* 116, 909–914. doi: 10.1073/pnas.1813051115

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., and Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763–773. doi: 10.1111/j.1472-4642.2008.00482.x

Wood, E. M., and Esaian, S. (2020). The importance of street trees to urban avifauna. *Ecol. Appl.* 30, e02149. doi: 10.1002/eap.2149

Wood, S. N. (2017). *Generalized additive models: An introduction with R.* Boca Raton: CRC Press/Taylor & Francis Group. doi: 10.1201/9781315370279

Yan, Z., Teng, M., He, W., Liu, A., Li, Y., and Wang, P. (2019). Impervious surface area is a key predictor for urban plant diversity in a city undergone rapid urbanization. *Sci. Total Environ.* 650, 335–342. doi: 10.1016/j.scitotenv.2018.09.025

Zellmer, A., and Goto, B. (2022). Urban wildlife corridors: Building bridges for wildlife and people. *Front. Sustain. Cities* 4:954089. doi: 10.3389/frsc.2022.954089

Ziter, C. (2015). The biodiversity-ecosystem service relationship in urban areas: A quantitative review. *Oikos* 125, 761–768. doi: 10.1111/oik.02883

Zizka, A., Antonelli, A., and Silvestro, D. (2021). sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography* 44, 25–32. doi: 10.1111/ecog.05102

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., et al. (2019). CoordinateCleaner : Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* 10, 744–751. doi: 10.1111/2041-210X.13152