Constructing ecological indices for urban environments using species distribution models

Ariel Levi Simons (levisimons@gmail.com)
University of California Los Angeles https://orcid.org/0000-0003-0492-5492

Stevie Caldwell
University of California Los Angeles

Michelle Fu
University of California Los Angeles

Jose Gallegos
University of California Los Angeles

Michael Gatheru
University of California Los Angeles

Laura Riccardelli
University of California Los Angeles

Nhi Truong
University of California Los Angeles

Valeria Viera
University of California Los Angeles

Research Article

Keywords: Species distribution modeling, Maxent, random forest, ecological assessments, ecological indices

Posted Date: October 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-894926/v1

Abstract

In an increasingly urbanized world, there is the need for a framework to assess ecological conditions in these anthropogenically dominated environments. Using species observations from the Global Biodiversity Information Facility (GBIF), along with remotely sensed environmental layers, we used MaxEnt to construct species distribution models (SDMs) of native and non-native species in Los Angeles. 25 native and non-native Indicator species were selected based on the sensitivities of their SDM, as measured by the Symmetric Extremal Dependence Index (SEDI), to environmental gradients. These SDMs were summarized to produce ecological indices of native and non-native biodiversity in Los Angeles. We found native indicator species to have a greater sensitivity to environmental conditions than their non-native counterparts, with the mean SEDI score of native and non-native species MaxEnt models being 0.72 and 0.71 respectively. While both sets of species were sensitive to land use categories and housing density, native species were more sensitive to natural landscape variables while non-native ones were more sensitive to measures of water and soil contamination. Using random forest modeling we also found our native index could be more reliably predicted, given environmental conditions, than its non-native counterpart. The mean Pearson correlation between actual and predicted index values were 0.86 and 0.84 for native and non-native species. From these results we conclude that using SDMs to predict the biodiversity of environmental species is a suitable approach towards evaluating ecological conditions in urban environments, with the environmental sensitivity of native SDMs outperforming non-native ones.

Introduction

The expansion of cities, both in their geographic scope as well as their use of resources, often result in ecological degradation (Elmqvist et al., 2016). However, it is not inevitable that increased urbanization will lead to ecological destruction and collapse. Densely populated landscapes can contain resilient ecosystems (Beller et al., 2019), prompting the study and management of urban environments as ecosystems in their own right (Montero, 2020).

As a result of ecological transformations brought on by urbanization, many municipalities and regional agencies have begun quantifying the impacts of urbanization on ecological metrics such as biodiversity (von der Lippe et al., 2020) and ecological functions (Eötvös et al., 2018). These efforts have led to the popularization of environmental report cards: a means of assessing and summarizing the environmental performance of urban systems based on metrics such as the proportion of native to non-native species and area of habitat dedicated to pollinators (Logan et al., 2020).

One city actively involved in the development of urban ecological assessments is Los Angeles (Rauser, 2021). The Los Angeles area is situated within the California Floristic Province, one of the 36 most biodiverse terrestrial ecosystems in the world (Myers et al., 2000). Being a heavily urbanized area within a designated biodiversity ‘hotspot’ (Gillespie et al., 2018), Los Angeles served as a model city for evaluating a framework for assessing ecological conditions in an urban context.

Although various urban environmental report cards use measures of taxonomic and functional richness of local communities as part of their assessments, it remains an open question as to how indicative these values are of underlying environmental quality. Though various groups of species are used as ecological indicators, the determination of which species to use in constructing such assessments is often not well defined (Siddig et al., 2016). As urban environments are traditionally described as disturbed (Johnson and Munshi-South, 2017), there is the additional difficulty in performing ecological assessments as these often depend on the use of undisturbed reference conditions (Utz et al., 2009).

Here we propose the use of machine learning to bypass the issues of defining reference conditions in urban environments, as well as select species to use as ecological indicators. The use of machine learning has been used to select species which are sensitive to a number of environmental gradients, and which could act as ecological indicators (Moraitis et al., 2018). Such methods can also predict- given a set of species locations and maps of environmental conditions- the likely spatial distribution of a given species. These species distribution models (SDMs), representing suitable habits for a given species, have increasingly been used in conservation biology and environmental management (Angelieri et al., 2016).

We used the machine learning technique of Maximum Entropy (MaxEnt) in order to select indicator species and construct indices of urban ecological conditions from their SDMs. We selected MaxEnt for constructing our SDMs as it can work with presence-only data for species observations (Elith et al., 2011), which are commonly found in databases such as the Global Biodiversity Information Facility (GBIF) (Edwards, 2004), and which has been used in assessing species as ecological indicators (Jose V, 2020). Here we will combine SDMs from multiple indicator species in order to construct maps of their predicted richness as our measure of ecological assessment. In order to evaluate our indices we then used a random forest model, which has been used to evaluate the reliability of other ecological assessments (Rodriguez-Galiano et al., 2012).

Methods

Scope of data

Our initial set of species observational data was obtained from the GBIF with the following query: (1) observations made within the period 2010 - 2020, (2) within the spatial extent of Los Angeles County, and (3) with a spatial uncertainty of less than 30 m. This initial set of 120,713 observations was then filtered using the R package *sp* (Pebesma et al., 2012) to contain only observations within the city boundaries of Los Angeles and with a spatial uncertainty of less than 10 m. To ensure sufficient data to run a SDM for each species we then only retained those with at least 30 records. This filtering produced a list of 179 species (SI: File 1) containing 18,504 observations. This initial list was then split into native and non-native species using the CalFlora list of plant species native to Los Angeles County (http://www.calflora.org/), and a corresponding list for animals provided by the Los Angeles Sanitation and Environment (LASAN) (SI: File 2), as well as a list of invasive plant species provided by Cal-IPC

(https://www.cal-ipc.org/plants/inventory/). Following this split, we found 120 native and 59 non-native spaces with 12,320 and 6,184 observations respectively (SI: Fig. 4).

We obtained our initial set of 32 environmental layers from a variety of sources (https://doi.org/10.5068/D1W988), which were then clipped and aligned to the city boundaries of Los Angeles using GDAL (Warmerdam, 2008) and QGIS 3.4 (QGIS Development Team, 2018) with a 30 ft resolution. We then removed 9 of these layers with a collinearity of $r > 0.5$ (Meier et al., 2010) using the function *removeCollinearity* within the R package *virtualspecies* (Leroy et al., 2016) with 100,000 randomly selected points (SI: Fig. 4).

Generating ecological indices

To select species for our ecological indices we first generated 100 SDMs, hereafter referred to as initial SDMs, for each species from both our non-native and native lists. Each SDM was created using the *maxent* function within the R package *dismo* (Hijmans et al., 2017). As input, we randomly selected 25 presence and 500 background points using the function *spsample* within the R package *sp*. To account for the geographic clustering in species observations we generated a bias file for both our native and non-native species lists (SI: Fig. 5). These two files were generated using the function *kde2d* from the R package *MASS* (Ripley et al., 2013), with species density rasters generated using the function rasterize within R package *raster* (Hijmans and van Etten., 2015) as input.

The rates of true and false positives and negatives were found for these initial SDMs using the function *evaluate* within *dismo*. These rates were then used to calculate the Symmetric Extremal Dependence Index (SEDI) (Wunderlich et al., 2019) for each SDM, along with the mean SEDI per species. We selected the SEDI, as opposed to other metrics such as the Area Under the Curve (AUC), to evaluate our initial SDMs as it has been found to provide more consistent evaluation of the sensitivity and specificity of models over a wider range of study extents and species prevalence (Wunderlich et al., 2019). We then selected the 25 species from both lists with the highest mean SEDIs as indicator species (SI: Fig. 4). As indicator species are often selected based on their sensitivity to environmental conditions (Caro, 2010), our selections of SDMs with a greater SEDI score are expected to be associated with species whose occupancy is more strongly associated with a given set of environmental variables.

To further reduce the effects of spatial clustering of our observational data we then spatially thinned our data, using the function *thin* within the R package *spThin* (Aiello-Lammens et al., 2015) to set our minimum separation distance to 500 m between observations within species (SI: Fig. 4).

The relative importance of the 23 environmental variables used to generate these SDMs was calculated as a percent contribution to each model using the function *var.importance* within the R package *ENMeval* (Muscarella et al., 2014). The mean relative importance of each environmental variable was then calculated for both the native and non-native lists (Table 2). We then removed variables with a mean relative importance of less than 3% for the generation of our habitat suitability maps, leaving us with two lists of variables for further modeling with native and non-native species (Table 3, SI: Fig. 4).

For both lists, we then generated habitat suitability maps for each species using the function *predict* within the R package *raster*. Each species' habitat suitability map was generated using the *maxent* function with all available spatially thinned presence points, 20 times that number of background points, and a model threshold set as the maximum sum of the specificity and sensitivity. The habitat suitability maps for the 25 non-native (Fig. 1A) and native (Fig. 1B) indicator species were then summarized using the *calc* function within the R package *raster*. These two summary maps are the Los Angeles Ecological Index (LAEI) and the native Los Angeles Ecological Index (nLAEI), which predict non-native and native species richness respectively (SI: Fig. 4).

Evaluating ecological indices

To evaluate our indices we built and evaluated 100 random forest models of the LAEI and nLAEI as a function of our environmental parameters (SI: Fig. 4). Each model was constructed using 20 randomly selected locations within our study area. We chose 200 locations as this was found to be the minimum number needed to consistently model significant ($p < 10^{-4}$) predictions of both indices. Each of these groups of sampled points was split into training and testing sets using a 5-fold partitioning with the function *kfold* within *dismo* (Hijmans et al., 2017). We then constructed random forest models using our training data and the function *tuneRF*, within the R package *randomForest* (Liaw and Wiener, 2002), with stepFactor set to a value of 1 and doBest set as 'true'. Predicted index values were calculated using the function *predict* within *randomForest*, which were then compared against their actual counterparts with a Pearson correlation coefficient.

We calculated the mean and standard deviation on our 100 Pearson correlation coefficients under a Fisher transformation using the *FisherZ* function within the R package *DescTools* (Signorelli, 2020). We used a Fisher transformation as it has been found to produce less biased summary statistics for a set of Pearson correlation coefficients (Corey et al., 1998). The average and standard deviation were then inverse Fisher transformed using the function *FisherZInv* within the *DescTools* package.

To compare the relative importance of environmental variables within our models we used the function *partial* within the R package *pdp* (Greenwell, 2017). The relative importance of our model variables was calculated as their mean decrease in node impurity as quantified by their Gini indices, with a larger value denoting greater relative importance. To generate individual partial dependence plots we used the function *response* within the R package *dismo*.

To visualize a heat map of the 100 iterations of the partial dependence plots for our continuous variables we then used the function *geom_bin2d* within the R package *ggplot2* (Wickham et al., 2016). For all of our partial dependence heat maps, we divided our axes into 20 bins and generated a best-fit curve using the function *stat_smooth* within the R package *ggplot2*. For our categorical variables, we visualized the partial dependences of our models with violin plots generated using the function *geom_violin* within the R packages *ggplot2*. Using this procedure, we were able to visualize the partial dependences of our models against their constituent variables.

Results

Ecological indices and their responses to environmental gradients

We identified 25 separate native and non-native species as ecological indicators (Table 1), the SDMs of which we used to construct our ecological indices (Fig. 1). We found our SDMs for both sets of species to be strongly influenced by a number of the same environmental variables (Tables 1 and SI: Table 4), and that our indices responded in similar fashions to these same variables (Figs. 2A-G, 3A-F,H). For both indices we observed a greater richness of indicator species in areas where land is categorized as open space or public facilities (Figs. 2A and 3A), shallow (Figs. 2C and 3E), with a low housing density (Figs. 2B and 3B), near streams (Figs. 2D and 3F), in coastal zones within climate zone 24 (Figs. 2F and 3H), where tree cover is predominantly deciduous (Figs. 2E and 3C), or where land cover is predominantly mixed rangeland or barren land (Figs. 2G and 3D). We tended to observe a lower richness of indicator species in both our indices in areas where land experiences only occasional ocean influence within climate zone 21 (Figs. 2F and 3H), is dominated by built-up urban areas instead of trees (Figs. 2E and 3C), or where land cover is dominated by buildings or pasture (Figs. 2G and 3D).

We also observed a number of differences in the responses of both indices to various environmental conditions. For the nLAEI we tended to find an increase in the richness of indicator species associated with an increase in elevation (Fig. 3G), proximity to lakes (Fig. 3J), or areas dominated by shrubland or canopy forming tree cover (Fig. 3I), while the LAEI has no significant responses to these variables. With the LAEI we found lower indicator species richness associated with an increase in either the density of sites with threatened groundwater quality (Fig. 2H), or requiring a cleanup of hazardous materials (Fig. 2I), while nLAEI was largely unaffected by these variables.

Evaluation of ecological indices

We found a small, but significant ($p < 0.05$, Mann-Whitney-Wilcoxon test), higher set of mean SEDI scores for the MaxEnt models of native versus non-native indicator species (SI: Fig. 6). This difference suggests a greater degree of sensitivity of our native over our non-native SDMs to our environmental gradients. In evaluating both indices as random forest models of their constituent environmental variables, we found the Pearson correlations between the predicted and actual ecological index values (SI: Fig. 7) were slightly, but significantly ($p < 0.05$, Mann-Whitney-Wilcoxon test), higher for the nLAEI versus LAEI.

Discussion

Indicator species

As with previous research using community science observations of species (Petersen et al., 2021), we found geographic clustering in the locations of both our native and non-native indicators (SI: Fig. 5). While we were able to account for these spatial biases in generating our indices, we do note a geographic bias in our observations of our native indicator species towards large open spaces (SI: Fig. 5B) versus the

more heavily urbanized areas where we found observations of our non-native indicator species (SI: Fig. 5A).

We found that membership in our list of indicator species, particularly those native to Los Angeles, were biased towards birds (Table 1). This reflects a bias in the community science data sets found in GBIF towards observations of birds (Petersen et al., 2021). Despite this bias, birds have been useful as ecological indicators in a variety of studies (Mekonen, 2017), including assessments of urban environments (Callaghan et al., 2021).

Response of indices to environmental variables

For both indices, we found similar responses to seven of our environmental variables: land use category, housing unit density, slope, proximity to streams, predominant tree leaf type, climatic zone, and land cover category. We also observed that our indices were responsive to unique sets of environmental variables. For the LAEI this included the density of sites threatening groundwater quality and those requiring the cleanup of hazardous materials. For the nLAEI we observed significant responses to the proximity to lakes, elevation, as well as the dominant type of canopy forming plant cover.

Response to land use

We tended to find our indices predicting an elevated richness of indicator species in areas with land use categorized as either open space or containing public facilities (Figs. 2A and 3A). Prior observations have indicated greater species richness in open spaces, such as parks and unbuilt land, with comparatively lower levels of human activity (Paker et al., 2014). For both indices, we found land use to be the most important variable (Table 2), reflected in elevated indicator species richness in areas designated as parks or open space reserves such as the western Santa Monica mountains or Griffith Park (Fig. 1).

Response to housing density

Although species vary in their tolerances to the built density of urban environments (Godefroid and Ricotta, 2018), previous studies have indicated a broadly negative relationship between the richness of species and housing density (Clarke et al., 2013). We also found our indices to predict a lower richness of native, as compared to non-native indicator species richness, in response to housing density (Figs. 2B and 3B). These results may reflect the apparent role of cities acting as ecological filters for more urban-tolerant non-native species (Blouin et al., 2019).

Response to slope and elevation

We found both indices had a clear negative response to slope (Figs. 2C and 3E). Across various landscapes in the Los Angeles area, it has been found that rates of soil erosion tend to increase with slope (DiBiase et al., 2011). This observation indicates that steep hillsides will tend to have lower rates of soil stability, in turn supporting a reduced species richness (Hubbert et al., 2006). This relationship

appears to be particularly strong in semi-arid Mediterranean climates, such as Los Angeles (Smith et al., 2007).

The nLAEI predicted a trend of greater native indicator species richness with increased elevation (Fig. 3G). This may in part reflect a historic trend towards greater plant species richness at higher elevations in California (Seabloom et al., 2002), as well as the effects of greater human disturbance at lower elevations in the Los Angeles area (Syphard et al., 2018).

Response to stream and lake proximity

We found both our indices tended to decline with distance from streams, although this decline was sharper in native versus non-native species (Figs. 2D and 3F). This relationship between species richness and stream proximity has been observed in previous studies of urban habitats, at least with regards to bird diversity (Chin and Kupfer, 2020). For plants, especially those in areas with a Mediterranean climate such as Los Angeles, proximity to water has also been found to be a significant contributor to local diversity (Hawkins et al., 2003). This decline in indicator species richness with distance from water also appears to be the case with proximity to lakes, although this only appears to be significant for our native index (Fig. 3J).

Response to predominant tree leaf type

For both indices, the same dominant leaf cover type determined the same maximum and minimum indicators for species richness (Figs. 2E and 3C). In areas dominated by deciduous vegetation, our predicted indicator species richness was at its maximum. The native vegetation of Los Angeles was dominated by the coastal sage scrub and oak woodlands which are mainly seasonally dry, deciduous plants (Avolio et al., 2020). As the Los Angeles area urbanized this pattern in vegetation held, with many of the introduced tree species were deciduous as well (Avolio et al., 2020), especially in heavily managed landscapes such as parks and backyards (Clarke et al., 2013). Though these landscapes are heavily altered, they do tend to enhance local biodiversity of both native and non-native species (Clarke et al., 2013; Avolio et al., 2020).

In areas dominated by built structure, as opposed to tree cover, both our indices predicted a reduction in indicator species richness (Figs. 2E and 3C). This result may be due to urban areas containing large tracts of transformed, fragmented habitat (Elmqvist et al., 2016). Heavily built areas in urban environments also tend to experience biotic homogenization, tending to create landscapes containing a few remaining common native species, along with a relatively small number of ubiquitous non-native species (Concepción et al., 2016). As a result, urban species richness tends to be significantly lower in densely developed areas than in green spaces (Kondratyeva et al., 2020).

Response to climatic zones

Both indices had the same minimum indicator species richness within climatic zone 21 (Figs. 2F and 3H), designated as ‘Thermal Belts of SoCal w/Occasional Ocean Influence’, which tends to be further inland on

steeply sloped land. Within the landscapes of Los Angeles, this climatic zone tends to be drier, and with lower minimum temperatures, than other climatic zones such as those dominated by marine influence in zone 24. Across the landscapes of southern California there tends to be greater species richness in such moderate, and relatively wetter environments (Schoenherr, 2017), which may in part reflect the minima of our indices in drier areas with larger temperature ranges.

Response to land and predominant canopy cover

As with prior observations of urban landscapes, we found the differences in vegetation described by land cover to be influential in predicting species richness (Roy et al., 1999; Petersen et al., 2020). Both indices predict indicator species richness minima and maxima associated with the same categories of land cover (Figs. 2G and 3D). We predict indicator species richness maxima in areas with land cover types associated with either chaparral or dune vegetation, both associated with the pre-urbanized landscapes of Los Angeles (Anderson et al., 2009), and which support local diversity even when adjacent to built environments (Chace and Walsh, 2006). We found a similar response, although only for the nLAEI, for indicator species richness and the predominant canopy cover in an area (Fig. 3I). Similar to the responses of indicator species to land cover, we predicted a maxima in species richness in areas with shrubland, and a minima where no significant canopy forming vegetation is present.

With areas where the vegetative cover was replaced with buildings, or dominated by cropland and pasture, we predicted minima in indicator species richness (Figs. 2G and 3D). While human activity does not impact native and non-native diversity equally (Wania et al., 2006), anthropogenically dominated environments such as built and agricultural landscapes tend to have the reduced species richness we observed in the response of our indices (Simmonds et al., 2019).

Responses to environmental contamination

With the LAEI we predicted declines in non-native indicator species richness associated with increases in the density of contaminated sites involving both groundwater (Fig. 2H) and soil (Fig. 2I). Both measures of contamination are associated with degraded landscapes in Los Angeles (Campbell et al., 2021), which have been found to reduce local biodiversity (Guilland et al., 2018).

We found no strong responses from the nLAEI to either measure of environmental contamination. Independent of the pollution tolerances of our native versus non-native species, we propose that this aspect of the behavior of the nLAEI to be due in part to the preponderance of highly mobile native indicator species, which can readily avoid contaminated areas. Of our 25 non-native indicator species 10 were plants and 8 birds, while for our native indicator species these values were 3 and 18 respectively, potentially reflecting a taxonomic bias towards birds in community science (Troudet et al., 2017). Successful non-native species also tend to be generalists, with relatively low sensitivities to a variety of environmental gradients (Callaghan et al., 2019). With a reduced sensitivity to various environmental gradients, we propose that our non-native indicator species are also more likely to be found in all urban areas, and thus encounter additional anthropogenic stressors than their native counterparts.

Evaluation of ecological indices

Using random forest, we found a small, but significant, greater level of reliability in predicting the nLAEI over the LAEI (SI: Fig. 7). We propose that this difference is due in part to the lower sensitivity of our non-native to native indicator species to environmental conditions (SI: Fig. 6), which has been observed in other ecosystems (Davey et al., 2012). This reduced sensitivity, especially to anthropogenic stressors, tends to select species that are tolerant of various disturbances across an urban landscape (Callaghan et al., 2019). Various human activities in Los Angeles, such as the importation of large amounts of water to cultivate ornamental plants, further support the widespread distribution of non-native species (Avolio et al., 2020). These differences in the environmental responses of native and non-native indicator species are likely reflected in both the more uniform spatial distributions of the LAEI (Fig. 1A) over the nLAEI (Fig. 1B), as well as greater reliability of the latter (SI: Fig. 7).

Conclusions

Our results support the potential use of SDMs in building an assessment framework for urban ecosystems. Using MaxEnt, we selected a native and non-native set of indicator species based on the sensitivity of their SDMs. We then used these sets of SDMs to generate a native and non-native assessment of ecological conditions in Los Angeles, based on the predicted richness of our indicator species. While both indices can be reliably predicted using a relatively small number of environmental variables, we did find our index utilizing only native species performed better than its non-native counterpart.

However, the method by which we generated our indices remains constrained by the current limitations of community science-based observations of species. These observations have a well-established taxonomic bias towards vertebrates, and in particular birds, which is likely reflected in the number of avian species we selected as indicators. For future work we therefore recommend the addition of other sources of data on species observations, ranging from camera traps to the sequencing of DNA from environmental samples, in order to build more robust assessments of ecological conditions in urban environments.

Declarations

Acknowledgements: We would like to acknowledge Michelle Barton, Dr. Travis Longcore, Dr. Rachel Meyer, and Dr. Sabrina Drill for their insights in landscape ecology used in this project.

Funding: Not applicable

Conflicts of interest/Competing interests: Not applicable

Availability of data and material: All map layers used in this analysis are available at https://doi.org/10.5068/D1W988 . All species observations were obtained from https://www.gbif.org.

Code availability: All scripts used in this analysis are available here: https://github.com/LASanitation/LASAN

Authors' contributions: All authors contribution to the original draft, as well as subsequent edits. Conceptualization of this project was performed by Ariel Levi Simons. Methodology was performed by Ariel Levi Simons, Jose Gallegos, and Michael Gatheru. Software development was performed by Ariel Levi Simons, Laura Riccardelli, and Jose Gallegos. Validation was performed by Valeria Viera. Data curation was performed by Michelle Fu. Supervision was performed by Stevie Caldwell. Visualization was performed by Nhi Truong.

Literature Cited

Aiello-Lammens ME, Boria RA, Radosavljevic A, et al (2015) spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography (Cop) 38:541–545. https://doi.org/10.1111/ecog.01132

Anderson JR, Hardy EE, Roach JT, Witmer RE (2009) LAND USE AND LAND COVER CLASSIFICATION SYSTEM FOR USE WITH REMOTE SENSOR DATA.

Angelieri CCS, Adams-Hosking C, Paschoaletto KM, et al (2016) Using species distribution models to predict potential landscape restoration effects on puma conservation. PLoS One 11:e0145232. https://doi.org/10.1371/journal.pone.0145232

Avolio M, Pataki DE, Jenerette GD, et al (2020) Urban plant diversity in Los Angeles, California: Species and functional type turnover in cultivated landscapes. Plants, People, Planet 2:144–156. https://doi.org/10.1002/ppp3.10067

Beller EE, Spotswood EN, Robinson AH, et al (2019) Building Ecological Resilience in Highly Modified Landscapes. Bioscience 69:80–92. https://doi.org/10.1093/biosci/biy117

Blouin D, Pellerin S, Poulin M (2019) Increase in non-native species richness leads to biotic homogenization in vacant lots of a highly urbanized landscape. Urban Ecosyst 22:879–892. https://doi.org/10.1007/s11252-019-00863-9

California. Fisheries Programs Branch, Geological Survey (U.S.) and CBDB (2012) Lakes, Reservoirs and Ponds (Polygons): California, 2012. https://maps.princeton.edu/catalog/stanford-zx543xm6802

Callaghan CT, Major RE, Wilshire JH, et al (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. https://doi.org/10.1111/oik.06158

Callaghan CT, Sayol F, Benedetti Y, et al (2021) Validation of a globally-applicable method to measure urban tolerance of birds using citizen science data. Ecol Indic 120:106905.

https://doi.org/10.1016/j.ecolind.2020.106905

Campbell H, Eckerd A, Kim Y (2021) Administration of Community Participation in Small-Scale Projects: Brownfield Remediation in Los Angeles. Adm Soc 53:378–409. https://doi.org/10.1177/0095399720944064

Caro T (2010) Conservation by proxy: indicator, umbrella, keystone, flagship, and other surrogate species. Island Press

Chace JF, Walsh JJ (2006) Urban effects on native avifauna: A review. Landsc Urban Plan 74:46–69. https://doi.org/10.1016/j.landurbplan.2004.08.007

Chin EY, Kupfer JA (2020) Identification of environmental drivers in urban greenway communities. Urban For Urban Green 47:126549. https://doi.org/10.1016/j.ufug.2019.126549

Clarke LW, Jenerette GD, Davila A (2013) The luxury of vegetation and the legacy of tree biodiversity in Los Angeles, CA. Landsc Urban Plan 116:48–59. https://doi.org/10.1016/j.landurbplan.2013.04.006

Concepción ED, Obrist MK, Moretti M, et al (2016) Impacts of urban sprawl on species richness of plants, butterflies, gastropods and birds: not only built-up area matters. Urban Ecosyst 19:225–242. https://doi.org/10.1007/s11252-015-0474-4

Corey DM, Dunlap WP, Burke MJ (1998) Averaging correlations: Expected values and bias in combined pearson rs and fisher's z transformations. J Gen Psychol 125:245–261. https://doi.org/10.1080/00221309809595548

Davey CM, Chamberlain DE, Newson SE, et al (2012) Rise of the generalists: Evidence for climate driven homogenization in avian communities. Glob Ecol Biogeogr 21:568–578. https://doi.org/10.1111/j.1466-8238.2011.00693.x

DiBiase RA, Whipple KX (2011) The influence of erosion thresholds and runoff variability on the relationships among topography, climate, and erosion rate. J Geophys Res Earth Surf 116:F4. https://doi.org/10.1029/2011JF002095

Edwards JL (2004) Research and societal benefits of the global biodiversity information facility. Bioscience 54:485–486. https://doi.org/10.1641/0006-3568(2004)054[0486:rasbot]2.0.co;2

Elith J, Phillips SJ, Hastie T, et al (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17:43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Elmqvist T, Zipperer WC, Güneralp B (2016) Urbanization, habitat loss, biodiversity decline: solution pathways to break the cycle. In: Routledge Handbook of Urbanization and Global Environmental Change. Routledge, London and New York, pp 139–151

Eötvös CB, Magura T, Lövei GL (2018) A meta-analysis indicates reduced predation pressure with increasing urbanization. Landsc Urban Plan 180:54–59. https://doi.org/10.1016/j.landurbplan.2018.08.010

Faust J, August L, Bangia K (2017) CalEnviroScreen3

Gillespie TW, Ostermann-Kelm S, Dong C, et al (2018) Monitoring changes of NDVI in protected areas of southern California. Ecol Indic 88:485–494. https://doi.org/10.1016/j.ecolind.2018.01.031

Godefroid S, Ricotta C (2018) Alien plant species do have a clear preference for different land uses within urban environments. Urban Ecosyst 21:1189–1198. https://doi.org/10.1007/s11252-018-0792-4

Greenwell BM (2017) pdp: An R package for constructing partial dependence plots. R J 9:421. https://doi.org/10.32614/rj-2017-016

Guilland C, Maron PA, Damas O, Ranjard L (2018) Biodiversity of urban soils for sustainable cities. Environ Chem Lett 16:1267–1282. https://doi.org/10.1007/s10311-018-0751-6

Hawkins BA, Field R, Cornell H V., et al (2003) Energy, water, and broad-scale geographic patterns of species richness. Ecology 84:3105–3117. https://doi.org/10.1890/03-8006

Hijmans RJ, van Etten J (2015) raster: Geographic analysis and modeling with raster data. R Packag. version 2.5-2

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) Package 'dismo" - Species Distribution Modeling.' CRAN Repos.

Hubbert KR, Preisler HK, Wohlgemuth PM, et al (2006) Prescribed burning effects on soil physical properties and soil water repellency in a steep chaparral watershed, southern California, USA. Geoderma 130:284–298. https://doi.org/10.1016/j.geoderma.2005.02.001

Johnson MTJ, Munshi-South J (2017) Evolution of life in urban environments. Science (80-) 358:6363. https://doi.org/10.1126/science.aam8327

Jose V S, Nameer PO (2020) The expanding distribution of the Indian Peafowl (Pavo cristatus) as an indicator of changing climate in Kerala, southern India: A modelling study using MaxEnt. Ecol Indic 110:105930. https://doi.org/10.1016/j.ecolind.2019.105930

Kondratyeva A, Knapp S, Durka W, et al (2020) Urbanization Effects on Biodiversity Revealed by a Two-Scale Analysis of Species Functional Uniqueness vs. Redundancy. Front Ecol Evol 8:. https://doi.org/10.3389/fevo.2020.00073

Leroy B, Meynard CN, Bellard C, Courchamp F (2016) virtualspecies, an R package to generate virtual species distributions. Ecography (Cop) 39:599–607. https://doi.org/10.1111/ecog.01388

Liaw A, Wiener M (2002) Classification and Regression by randomForest. R News 2:18–22

Logan B (2006) A Statistical Examination of the Climatic Human Expert System, The Sunset Garden Zones for California. Virginia Tech

Logan M, Hu Z, Brinkman R, et al (2020) Ecosystem health report cards: An overview of frameworks and analytical methodologies. Ecol Indic 113:105834. https://doi.org/10.1016/j.ecolind.2019.105834

Los Angeles (2020) General Plan Land Use (GPLU). https://geohub.lacity.org/datasets/lahub::general-plan-land-use-gplu/about

Matyas WJ, Parker I (1980) CALVEG mosaic of existing vegetation of California

Maune DF (2006) Vertical accuracy assessment report, 2006 LiDAR Bare-Earth Dataset for Los Angeles Region Imagery Acquisition Consortium (LAR-IAC). Fairfax, Virginia

Meier ES, Kienast F, Pearman PB, et al (2010) Biotic and abiotic variables show little redundancy in explaining tree species distributions. Ecography (Cop) 33:1038–1048. https://doi.org/10.1111/j.1600-0587.2010.06229.x

Mekonen S (2017) Birds as Biodiversity and Environmental Indicator. Adv Life Sci Technol 7:

Montero S (2020) Leveraging Bogotá: Sustainable development, global philanthropy and the rise of urban solutionism. Urban Stud 57:2263–2281. https://doi.org/10.1177/0042098018798555

Moraitis ML, Tsikopoulou I, Geropoulos A, et al (2018) Molluscan indicator species and their potential use in ecological status assessment using species distribution modeling. Mar Environ Res 140:10–17. https://doi.org/10.1016/j.marenvres.2018.05.020

Muscarella R, Galante PJ, Soley-Guardia M, et al (2014) ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol Evol 5:1198–1205

Myers N, Mittermeler RA, Mittermeler CG, et al (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858. https://doi.org/10.1038/35002501

Paker Y, Yom-Tov Y, Alon-Mozes T, Barnea A (2014) The effect of plant richness and urban garden structure on bird species richness, diversity and community structure. Landsc Urban Plan 122:186–195. https://doi.org/10.1016/j.landurbplan.2013.10.005

Pebesma E, Bivand R (2012) Package ‘sp’

Petersen TK, Speed JDM, Grøtan V, Austrheim G (2021) Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. Ecol Solut Evid 2:e12048. https://doi.org/10.1002/2688-8319.12048

Petersen TK, Speed JDM, Grøtan V, Austrheim G (2020) Urban aliens and threatened near-naturals: Land-cover affects the species richness of alien- and threatened species in an urban-rural setting. Sci Rep 10:1–14. https://doi.org/10.1038/s41598-020-65459-2

QGIS Development Team (2018) QGIS Geographic Information System 3.0 Girona. Open Source Geospatial Found. Proj.

Radeloff VC, Helmers DP, Anu Kramer H, et al (2018) Rapid growth of the US wildland-urban interface raises wildfire risk. Proc Natl Acad Sci U S A 115:3314–3319. https://doi.org/10.1073/pnas.1718850115

Rauser C (2021) UCLA Sustainable LA Grand Challenge: Innovatively Applying Research to Create the First Sustainable Megacity. In: World Scientific Encyclopedia of Climate Change: Case Studies of Climate Risk, Action, and Opportunity Volume 1. pp 105–114

Ripley B, Venables B, Bates DM, et al (2013) Package “mass.” 113–120

Rodriguez-Galiano VF, Ghimire B, Rogan J, et al (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens 67:93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Roy DB, Hill MO, Rothery P (1999) Effects of urban land cover on the local species pool in Britain. Ecography (Cop) 22:507–517. https://doi.org/10.1111/j.1600-0587.1999.tb01279.x

Schoenherr AA (2017) A natural history of California: Second edition

Seabloom EW, Dobson AP, Stoms DM (2002) Extinction rates under nonrandom patterns of habitat loss. Proc Natl Acad Sci U S A 99:1129–1134. https://doi.org/10.1073/pnas.162064899

Siddig AAH, Ellison AM, Ochs A, et al (2016) How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in Ecological Indicators. Ecol Indic 60:223–230. https://doi.org/10.1016/j.ecolind.2015.06.036

Signorelli A (2020) DescTools: Tools for Descriptive Statistics. R package version 0.99.37. Cran

Simley JD, Carswell Jr. WJ (2009) The national map — hydrography

Simmonds JS, van Rensburg BJ, Tulloch AIT, Maron M (2019) Landscape-specific thresholds in the relationship between species richness and natural land cover. J Appl Ecol 56:1019–1029. https://doi.org/10.1111/1365-2664.13320

Smith S V., Bullock SH, Hinojosa-Corona A, et al (2007) Soil erosion and significance for carbon fluxes in a mountainous Mediterranean-climate watershed. Ecol Appl 17:1379–1387. https://doi.org/10.1890/06-0615.1

Syphard AD, Brennan TJ, Keeley JE (2018) Chaparral landscape conversion in southern California. In: Valuing Chaparral. pp 323–346

Troudet J, Grandcolas P, Blin A, et al (2017) Taxonomic bias in biodiversity data and societal preferences. Sci Rep 7:1–14. https://doi.org/10.1038/s41598-017-09084-6

Utz RM, Hilderbrand RH, Boward DM (2009) Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. Ecol Indic 9:556–567. https://doi.org/10.1016/j.ecolind.2008.08.008

von der Lippe M, Buchholz S, Hiller A, et al (2020) CityScapeLab Berlin: A research platform for untangling urbanization effects on biodiversity. Sustain 12:2565. https://doi.org/10.3390/su12062565

Wania A, Kühn I, Klotz S (2006) Plant richness patterns in agricultural and urban landscapes in Central Germany - Spatial gradients of species richness. Landsc Urban Plan 75:97–110. https://doi.org/10.1016/j.landurbplan.2004.12.006

Warmerdam F (2008) The geospatial data abstraction library. In: Open source approaches in spatial data handling. Springer Berlin, Heidelberg, pp 87–104

Wickham H, Chang W, Wickham MH (2016) Package ‘ggplot2.’ 1–189

Wunderlich RF, Lin YP, Anthony J, Petway JR (2019) Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. Nat Conserv 35:97. https://doi.org/10.3897/natureconservation.35.33918

Tables

Table 1: List of 25 native and non-native indicator species, along with species category.

Non-Native indicator species	Native indicator species
Alopochen aegyptiaca (Avian)	*Anas platyrhynchos* (Avian)
Anas americana (Avian)	*Ardea alba* (Avian)
Anas cyanoptera (Avian)	*Ardea herodias* (Avian)
Arundo donax (Plant)	*Baccharis salicifolia* (Plant)
Cairina moschata (Avian)	*Branta canadensis* (Avian)
Cenchrus setaceus (Plant)	*Bucephala albeola* (Avian)
Columba livia (Avian)	*Butorides virescens* (Avian)
Deroceras reticulatum (Mollusc)	*Charadrius vociferus* (Avian)
Dichelostemma capitatum (Plant)	*Egretta thula* (Avian)
Eleodes osculans (Insect)	*Euphagus cyanocephalus* (Avian)
Foeniculum vulgare (Plant)	*Falco sparverius* (Avian)
Lithobates catesbeianus (Amphibian)	*Fulica americana* (Avian)
Marrubium vulgare (Plant)	*Larus occidentalis* (Avian)
Nicotiana glauca (Plant)	*Libellula saturata* (Insect)
Otala lactea (Mollusc)	*Megaceryle alcyon* (Avian)
Passer domesticus (Avian)	*Melospiza melodia* (Avian)
Phloeodes diabolicus (Insect)	*Nycticorax nycticorax* (Avian)
Raphanus sativus (Plant)	*Phalacrocorax auritus* (Avian)
Ricinus communis (Plant)	*Pseudacris hypochondriaca* (Amphibian)
Rumina decollata (Mollusc)	*Quiscalus mexicanus* (Avian)

Sambucus cerulea (Plant)	*Rosa californica* (Plant)
Sturnus vulgaris (Avian)	*Sciurus griseus* (Mammal)
Syrmatium glabrum (Plant)	*Tyrannus verticalis* (Avian)
Trachemys scripta (Reptile)	*Uta stansburiana* (Reptile)
Vinca major (Plant)	*Venegasia carpesioides* (Plant)

Table 2: Mean relative importance of the most important environmental variables for Maxent SDMs of non-native and native species lists.

Variable	Mean percent relative importance (Non-native species list)	Mean percent relative importance (Native species list)
LandUse	22.7	10.2
HousingDensity	12.3	15.8
ClimateZones	9.0	6.8
DLC	8.0	10.2
Slope	7.3	8.7
LAStreamsProximity	4.8	8.9
gwthreats	4.5	NA
cleanups	3.4	NA
LandCover	3.1	5.8
Elevation	NA	4.3
DominantCanopyCover	NA	3.7
LALakeProximity	NA	3.3

Table 3: The sources, and descriptions, of environmental layers used in constructing both ecological indices.

cleanups (LAEI): The sum of weighted EnviroStor cleanup sites within buffered distances to populated blocks of census tracts (Faust et al., 2017).

ClimateZones (nLAEI/LAEI): Climatic zones (Logan, 2006).

0 - Ocean

7 - Southern California Mountains

18 - Above & Below Thermal Belt SoCal Interior Valley

19 - Thermal belts around SoCal Interior Valleys

20 - Cool Winters in SoCal w/ Occasional Ocean Influence

21 - Thermal Belts of SoCal w/Occasional Ocean Influence

22 - Cold-Winter Portions of SoCal Coastal Climate

23 - Thermal Belts of SoCal Coastal Climate

24 - Maritime Influence along SoCal Coast

DLC (nLAEI/LAEI): Dominant leaf cover: The predominant leaf phenology of classes defined by tree, shrub, or dwarf shrub stratum, and the average vegetation height for the herbaceous stratum (tall, medium, short) (Matyas & Parker, 1980).

1 - Hydromorphic rooted vegetation

2 - Annual graminoid and/or forb

3 - Non-Vegetated

4 - Deciduous vegetation

5 - Perennial forb

6 - Urban or built-up

7 - Mixed evergreen-deciduous vegetation

8 - Not determined

9 - Evergreen vegetation

10 - Perennial graminoid

11 - Unconsolidated material

DominantCanopyCover (nLAEI): The relative percent canopy cover of the tree, shrub, dwarf shrub, herb, and nonvascular life form in the uppermost strata during the peak of the growing season (Matyas & Parker, 1980).

1 - Shrubland

2 - Non-vegetated

3 - Open tree canopy

4 - Group of sparsely vegetated and non-vegetated

5 - Closed tree canopy

6 - Sparse tree canopy

7 - Herbaceous - grassland

8 - Class Not Determined

Elevation (nLAEI): Elevation in feet (Maune, 2006).

gwthreats (LAEI): Groundwater threats, sum of weighted GeoTracker leaking underground storage tank sites within buffered distances to populated blocks of census tracts (Faust et al., 2017).

HousingDensity (nLAEI/LAEI): Number of housing units per census tract (Radeloff et al., 2018).

LALakeProximity (nLAEI): Distance (ft) to the nearest open body of freshwater (California. Fisheries Programs Branch, 2012).

LandCover (nLAEI/LAEI): Areas classified using the Anderson Land Use/Land Cover Classification system (Matyas & Parker, 1980).

10 - Urban or build-up land	45 - Hardwood type forest land
11 - Residential	50 - Water
12 - Commercial and services	51 - Stream or canal
13 - Industrial	52 - Perennial Lake or Pond
14 - Transportation, communications, utilities	53 - Reservoir
15 - Industrial and commercial complexes	54 - Bay or estuary
16 - Mixed urban or built-up land	55 - Playa
17 - Other urban or built-up land	56 - Intermittent Stream Channel
20 - Agricultural land	57 - Ocean
21 - Cropland and pasture	58 - Intermittent Lake or Pond
22 - Orchard, vineyard, nursery, horticultural areas	59 - High Water Line/Gravel/Sand Bar
23 - Confined feeding operations	60 - Wetland
24 - Other agriculture land	61 - Forested wetland
30 - Rangeland	62 - Non-forested wetland
31 - Herbaceous rangeland	70 - Barren land
32 - Shrub and brush rangeland	71 - Dry salt flats
33 - Mixed rangeland	72 - Beaches
40 - Forest land	73 - Sandy area other than beaches
41 - Deciduous forest land	74 - Bare exposed rock
42 - Evergreen forest land	75 - Strip mines, quarries, and gravel pits
43 - Mixed conifer hardwood type forest land	76 - Transitional areas
44 - Conifer type forest land	77 - Mixed barren land

LandUse (nLAEI/LAEI): Categories of land use within Los Angeles (Los Angeles, 2020).

1 - Neighborhood Office Commercial
2 - LAX Airport Northside
3 - Low I Residential
4 - Regional Commercial
5 - High Medium Residential
6 - Light Industrial
7 - Heavy Manufacturing
8 - Low Residential
9 - Public/Quasi-Public Open Space
10 - Low Medium Residential
11 - Parking Buffer
12 - General Commercial
13 - Recreation and Commercial
14 - Limited Commercial - Mixed Medium Residential
15 - Low II Residential
16 - Industrial Commercial
17 - Minimum Residential
18 - Low Medium II Residential
19 - Regional Mixed Commercial
20 - General/Bulk Cargo (Non-hazardous)
21 - Public Facilities-Freeway
22 - Public Facilities
23 - Very Low I Residential
24 - Heavy Industrial
25 - Commercial Fishing
26 - Hybrid Industrial
27 - Very High Residential
28 - Regional Center Commercial
29 - Commercial Manufacturing
30 - Neighborhood Commercial
31 - Airport Airside
32 - General/Bulk Cargo (Hazardous)
33 - Other Public Open Space
34 - Limited Commercial
35 - Very Low II Residential
36 - Highway Oriented and Limited Commercial
37 - High Density Residential
38 - Community Commercial
39 - Light Manufacturing
40 - Airport Landside Support
41 - Very Low Residential
42 - Highway Oriented Commercial
43 - Highway Oriented Commercial - High Med Residential
44 - Medium Residential
45 - Intermodal Container Transfer Facility Site
46 - Limited Industrial
47 - Community Commercial - Mixed High Residential
48 - Low Medium I Residential
49 - Limited Manufacturing
50 - Airport Landside
51 - Low III Residential
52 - Open Space

LAStreamsProximity (nLAEI): Distance in feet to nearest stream (Simley et al., 2009).

Slope (nLAEI/LAEI): % land slope (Maune, 2006).

Figures

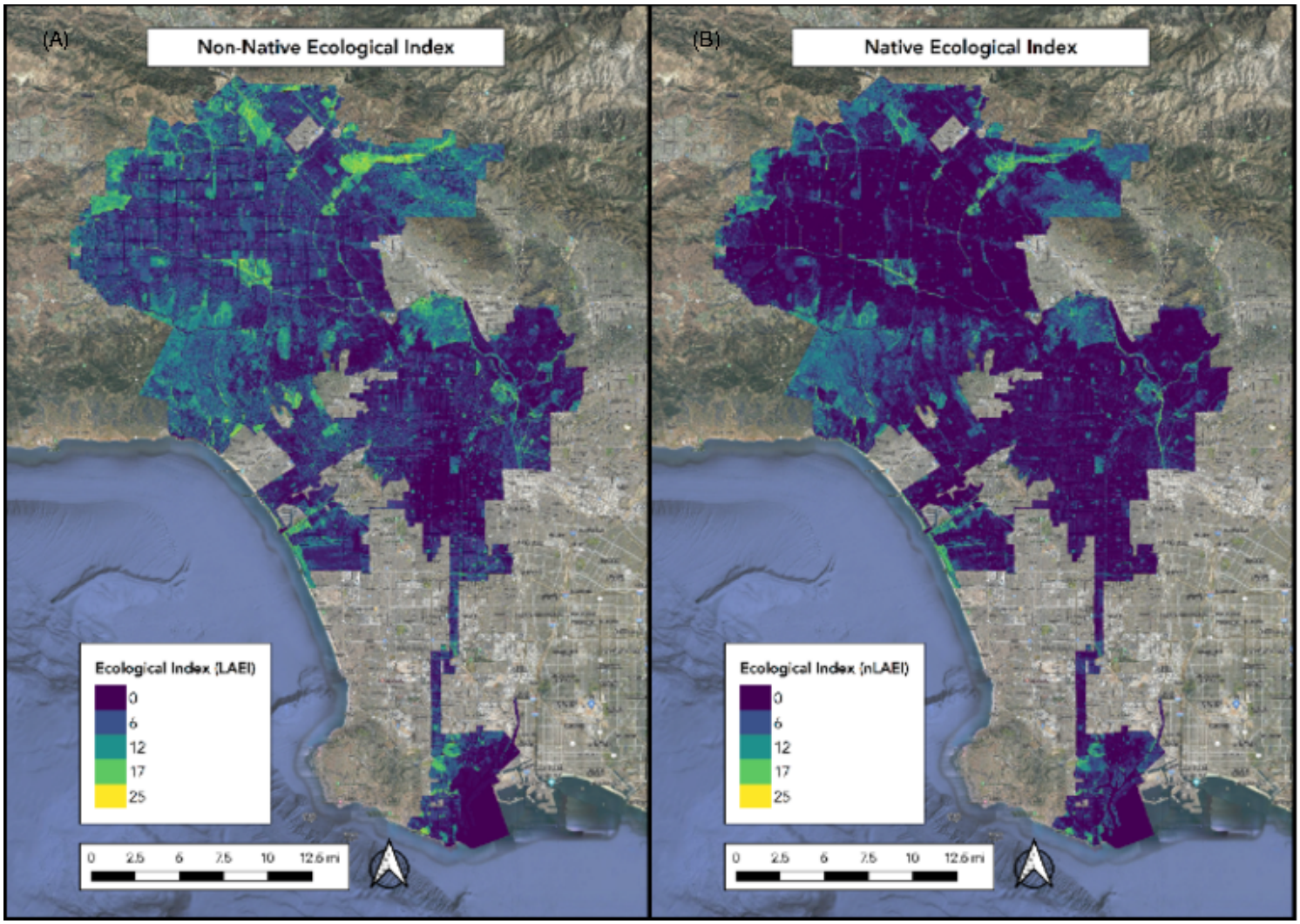

Figure 1

The LAEI (A) and nLAEI (B).

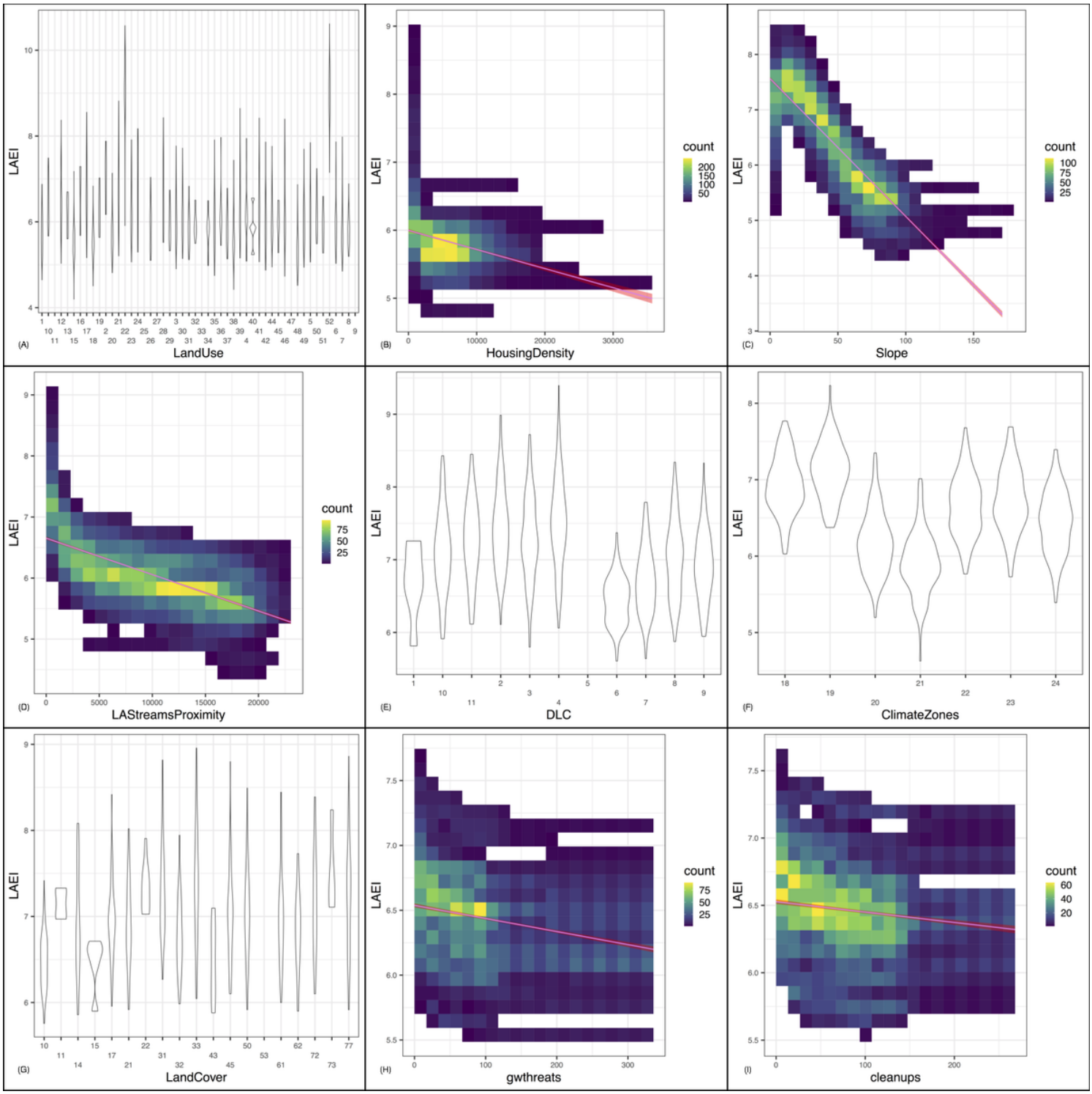

Figure 2

Density of 100 partial dependence plots for random forest models of the LAEI. The pink line represents a nonparametric loess curve with an associated 95% confidence interval.

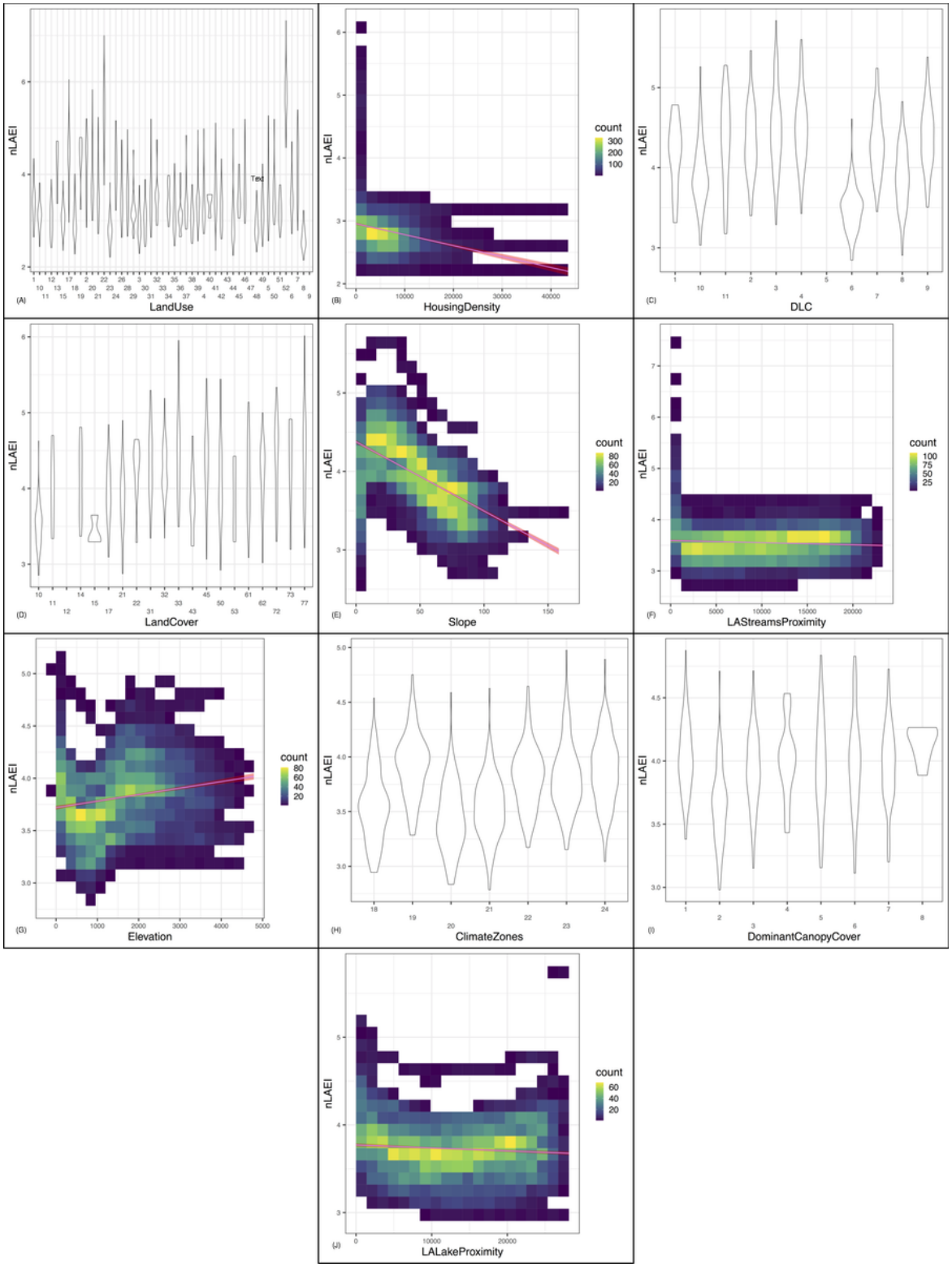

Figure 3

Density of 100 partial dependence plots for random forest models of the nLAEI. The pink line represents a nonparametric loess curve with associated 95% confidence interval.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- UEEcologicalIndexSI.docx
- Fig4.eps
- Fig5.eps
- Fig6.eps
- Fig7.eps