Landscape Impacts on Bird Diversity in East Bay Urban Parks

Linqian Sheng

ABSTRACT

Birds, as one of the best indicators of urban habitat quality, have been threatened by urbanization due to the change of urban landscapes. Parks are potential solutions for bird species conservation. This study examined what landscape elements are significant to bird biodiversity in urban parks in East Bay, and aimed to build a model that might help to assist bird species conservation in future parks design. Among all the models we built, the best-fitting model was a combination of square-root area, percentage of impervious cover, mean greenness and mean temperature, which yielded a high r^2 of 0.6727 and the lowest AICc value of 308.8. All landscape parameters within the model were significant, as they all had p-values less than 0.05. Among these, mean greenness was the most significant parameter and had a negative impact on bird biodiversity. The size of the park was the most dominant parameters in all candidate models and had a positive impact on bird diversity. Both percentage of impervious land covers and mean temperature negatively affected the bird diversity.

KEYWORDS

urban landscape; bird biodiversity; urban parks; GIS; remote sensing; ANOVA; biodiversity conservation; ecology; habitat

INTRODUCTION

Birds, as one of the most commonly seen animals in urban areas, are good indicators of urban environment quality, not only because they have well-studied ecology, but also because they are very habitat selective (Suri et al., 2016). As an essential part of biodiversity, however, birds are severely threatened by continuing urban sprawl (Hardman, 2012). Studies as far back as the 70s have already pointed out that avian diversity declines with increasing urban development (Emlen, 1974; Zalewski, 1994; Clergeau et al., 1998). One major influence of urban expansion on birds is the change of surface landscape. Urban sprawl largely modifies wildlife habitats by replacing native vegetation cover with buildings, roads and other impervious surface (Pick`ett et al., 2011), meanwhile fragmentizing originally connected wild habitats (Crooks, 2004). In Southeast Asia, rapid development has increased the density of concrete buildings as well as human activities (Aida et al., 2016). In coastal southern California, urban fragmentation particularly threatens Mediterranean scrub habitats that lead to high probability of endangerment and extinction of those native and even endemic species in the region (Myers, 1990).

With the loss and human interruption of their natural habitats, bird communities have been adversely impacted in population, community integrity, and breeding and nesting activities. In the urban centers, where more human modifications take place and less landscapes are preserved, bird diversity is usually lower than in the suburbs (Aida et al., 2016). In northern New York State, bird community integrity was strongly affected by the appearance of roads and other artificial developments, represented as low index of biotic integrity (IBI) (Glennon and Porter, 2005). In Seattle metropolitan area, sensitive forest avian species suffered constraints on their adaptive breeding dispersal, including territory shift and mate change after reproduction failure, due to the development of the metropolitan area. (Marzluff et a., 2016).

Despite these challenges, there are still opportunities to conserve bird diversity along with urban development. Parks, as a crucial component in urban design, can provide high quality bird habitat in urbanized areas by offering large greenspaces. Parks and gardens are often developed in central business districts to improve biodiversity in the urban landscape (Aida et al., 2016). Urban parks that constitute a variety of habitats, composed by heterogeneous plant species, usually contain higher bird diversity compared to other urban areas (Fernández-Juricic, 2004). Parks with additional components, such as waterbodies, can also maintain a higher level of bird biodiversity (Schwartz et al., 2008) Thus, with dedicated design, urban can conserve bird species in urban areas with a great potential.

There are several landscape attributes that contribute to urban parks being bird friendly. Larger green space with more heterogeneous vegetation is positively related to bird species diversity, with a R^2 value of 0.78 (Chang and Lee, 2016). Roadlessness and less human interruptions in large urban parks strongly supports bird community integrity, which, in some cases, represented by high IBI (Glennon and Porter, 2005). Waterbodies within or near urban parks function as ecological corridors, as certain bird species and functional groups respond strongly to the existence of catchments and rivers. In a study with a total of 95 species recorded, 64 species were recorded on either catchments or rivers (Suri et al., 2016). Individual bird species, according to their physical traits, also respond to different vegetation characteristics (McElhinny et al. 2006). Furthermore, temperature change related to climate change has significant impacts ($P < 0.05$ in F-test) on avian turnover (Peterson et al., 2015), implying a possible correlation between surface temperature and avian diversity.

Nevertheless, most of the studies that raised concerns about bird conservation either concerned only one particular aspect during city planning, or just provided vague, general recommendations of considering several aspects during urban (Fernández-Juricic, 2000; Ikin, 2012). These studies did not specifically provide practical solutions to conserve avian diversity along with the urban development. This research, instead, aimed for a more practical and comprehensive way for bird diversity conservation and preservation in urban areas. Current urban parks are mostly designed for aesthetic and recreational purposes, but they are also a great opportunity to conserve bird species during planning process. Furthermore, most of the previous studies took place in tropical cities, like in southeastern Asia, or in specific metropolitan cities, like Vancouver and Seattle. There has been neither similar studies carried in larger scale in terms of county or region, nor specifically in East San Francisco Bay area. Compared to all other previous studies in smaller scale areas or less developed tropical cities, a study in East Bay, a larger and more developed area, can provide further inspirations on urban bird species conservation.

This research studied the effect of urban park landscape elements on avian diversity in East Bay. The research specifically focused on two questions: what landscape elements in parks have significant effects on bird species diversity? How do these elements contribute to bird biodiversity? Eventually, we built a model of landscape parameters that could be utilized to predict bird diversity and contribute to bird species conservation in future urban planning process. We collected species diversity data through e-Bird website, a citizen-scientist based bird data website, and therefore the sub-question of this study is to test the use of citizen science data, whether this data source can effectively assist similar studies in the future. Based on previous studies, I hypothesized that parks with larger size, higher mean greenness, and higher trees and grass coverage tend to have higher bird diversity, while those with higher impervious surface coverage and higher mean surface temperature tend to have lower bird diversity. I also hypothesized that parks that are along shoreline, near stream, contain waterbody tend to attract more diverse bird species, while parks that contain playgrounds or sport fields, due to extra human interruptions and activities, tend to attract less diverse bird species.

METHODS

Study site

This study focused on 34 urban parks in Alameda County and 2 urban parks in Contra Costa County, California, which have bird observation records on the citizen-scientist website, eBird. These two counties, Alameda and Contra Costa are the two major counties in San Francisco Bay Area along the western coast of the US. In this study, we defined urban parks as human-designed, open, green public spaces for recreational use. Our urban parks include public open parks, shoreline parks, green spaces established around museums and libraries, and sport fields. There are no private lands or urban farms included in our study.

Species diversity dataset

The number of bird species observed in each urban park over the past 10 years was our diversity index. This information is directly available on e-Bird (e-Bird has three time categories: All Years, Species Last 10 Years, Current Year), a citizen science website. The eBird team is based at the Cornell Lab of Ornithology in the Information Science and Macaulay Library programs. On eBird, each hotspot incorporates checklists that include information of date, name of observer, name of bird species, and number of the specific species observed. Users of the website can choose different types of checklists according to their needs. The one we used in our study is the checklist of *Species in Last 10 Years*, which can be directly selected at each hotspot. *Species Last 10 Years* can cover most of the possible species in each park. We selected data records of urban parks that are hotspots available on the website in Alameda County and Contra Costa County.

Landscape parameters

The landscape parameters of the dataset are comprised of landscape data including mean temperature, mean greenness, and the percent of coverage for different land types in each park. We utilized geographical information system (GIS) and remote sensing (RS) technology to collect these data and information. We first obtained data of each park size and different land coverage through digitization. We imported background maps of our studied parks through Google Earth Pro into ArcMap, projected the site under WGS 1984, and then digitized each park. After exporting the shapefiles, we calculated the geometry of these shapefiles and obtained the size of each park and different land cover percentages in each park. In order to assess mean greenness and mean temperature data for each park, we used satellite images from Landsat-9 Operational Land Imager (OLI, 30m spatial resolution) and Thermal Infrared Sensor (TIRS; 100m spatial resolution) for World Reference System-2 path/row 44/34. Mean greenness was evaluated as normalized difference vegetation index (NDVI).

Despite the RS and GIS data, we also ran binary tests to examine some characteristics of each park. These characteristics included whether the park is a shoreline park (abbreviated as "Shoreline Park"), whether the park is near stream ("Near Stream"), whether the park contain its own waterbody ("Waterbody"), and whether the park is a playground or sports field ("Playground/Sports Field"). To numerate the results, an answer of "yes" was represented as 1, and "no" as 0.

Data analysis

We ran linear regression test and ANOVA (Faraway, 2002) to analyze the relationship between bird diversity index and each landscape parameter and multi-parameter combinations. We first used linear regression test to examine the correlation between diversity index and each landscape parameter. We focused on their p-values (less than 0.05), multiple R^2 (larger than 0.5), and signs of their coefficients. P-value is the probability of a given model that, when the null hypothesis is true, the statistical summary would be more extreme than the actual observed results. A smaller p-value indicates a that null hypothesis is false (Wasserstein and Lazar, 2016).

We then combined different parameters for multivariable tests. Since 10 landscape parameters can yield extremely large number of different combinations, and since increasing number of parameter within a combination can rise up the possibility of multicollinearity, our multivariable combinations included up to four landscape parameters. We computed detailed information about multivariable regression by using *lm()* function (fitting linear model) and *summary()* function, and recorded the multiple R^2 value, p-value, and the sign of estimated coefficient of each variable, positive or negative. From there, we chose models with p-value < 0.05 and multiple R^2 >0.5 as our candidate models.

Next, we ranked these candidate models using AICc, an estimator of relative quality of statistic models for a given small set of data:

AICc = −2lnL + 2k + (2k(k+1))/((n-k-1)),

where lnL is the model likelihood, k is the number of variables in the model, n is the number of observations in the model, which is the number of parks in our case. We preferred smaller AICc as it indicated a higher rank of the model (Dronova, 2016).

From there, we checked the collinearity and multicollinearity of the best-fitting model, aka the model with the lowest AICc value. For collinearity, we plotted its residuals vs. fitted values graph, which displayed any possible heteroscedascity and nonlinearity between parameters combinations and species diversity (Faraway, 2002). For multicollinearity, we computed variance inflation factor (VIF) to see if there is any mutual impacts between any two of the parameters within the model. VIF is calculated as $\frac{1}{1-R^2}$, where R^2 is the one produced by the multivariable regression of any two of the parameters.

RESULTS

Our dataset included 36 urban parks, their diversity index, and their landscape parameter data. A total of 240 bird species were observed over the past 10 years in these parks. The average number of bird species observed in each park was 69.

Diversity index

Diversity index significantly differed between urban parks over the past 10 years (Table 1). The richness ranged from 176 in Aquatic Park as the highest, to 12 in Cerrito Vista Park as the lowest. The mean of species richness was 69, and the median was 58, with a standard deviation of 42.39.

Table 1. Number of species observed in the past 10 years of each park. I collected the number of bird species observed in the past 10 years through e-Bird.

Park Name	Diversity Index	Park Name	Diversity Index
Alameda Creek--Staging Area	83	Meek Park	27
Albany Bulb	157	Mirabeau Park	29
Albany Hill Park	51	Morcom Rose Garden	48
Aquatic Park	176	Niles Park	138
Arlington Park	24	Oakland Coliseum	42
Berkeley Rose Garden	30	Oakland Museum	64
Cerrito Vista Park	12	MLK Jr. Civic Center Park	51
Cesar Chavez Park	174	Peralta Park	71
Chabot Park	56	Port View Park	62
Codornices Park	39	Remillard Park	40
Creekside Park	94	Shepherd Canyon Park	77
Dimond Park	83	Shinn Park and Arboretum	39
Estuary Park	85	Shoreline Park (ALA Co.)	75
Grinnell Natural Area	59	Silliman Activity Center	35
Garber Park	38	Strawberry Creek Park	24
Joaquin Miller Park	83	Tule Pond	60
Knowland Park	57	Union City Library Pond and Park	56
McLaughlin Eastshore SP--Albany access	146	Washington Park (ALA Co.)	99

Landscape parameters

The area of our studied parks varied widely (Table 2). The average size of parks was 194911m^2, with a median of 59136m^2 and standard deviation of 417244 m^2. The largest park, Knowland Park, had a size of 1975057 m^2, while the smallest park, Peralta Park, only had a size of 624 m^2. Mean greenness varied from 0.2018 (Port View Park) to 0.7323 (Codornices Park). The mean of mean greenness was 0.5028, the median was 0.5370, and the standard deviation was 0.1596. Mean temperature of the parks ranged from 20.41°C to 36.40°C, with Joaquin Miller Park being the coolest park and Alameda Creek Staging Area being the warmest. For each park, the sum of percentages of tree, grass and impervious land cover equals to 1.

Table 2. Landscape parameters of each park. We collected landscape parameters of targeted parks by using GIS digitization and remote sensing.

Park Name	Size (m^2)	p_Tree	p_Grass	p_Impervious	Mean Greenness	Mean Temperature (°C)
Alameda Creek--Staging Area	3493	0.2365	0.3275	0.3365	0.2072	36.40
Albany Bulb	152078	0.07402	0.5443	0.2486	0.3024	24.41
Albany Hill Park	185227	0.9262	0.07393	0.006677	0.6614	27.58
Aquatic Park	189290	0.5228	0.2522	0.2250	0.3267	22.63
Arlington Park	22139	0.4062	0.3949	0.1661	0.6049	34.37
Berkeley Rose Garden	14935	0.6466	0.1815	0.1719	0.5359	22.99
Cerrito Vista Park	32234	0.2667	0.5352	0.2737	0.6079	33.50
Cesar Chavez Park	350242	0.03816	0.8696	0.09221	0.3983	24.10
Chabot Park	149458	0.8841	0.02089	0.09503	0.6381	22.15
Codornices Park	32713	0.8502	0.1290	0.02082	0.7323	21.12
Creekside Park	6459	0.5780	0.3094	0.1302	0.6809	27.71
Dimond Park	142340	0.9364	0.05365	0.009913	0.6461	24.78
Estuary Park	17556	0.05586	0.5544	0.3902	0.2274	30.58
Grinnell Natural Area	30890	0.8245	0.1540	0.03024	0.6011	32.47
Garber Park	710269	0.8189	0.1459	0.03515	0.6609	21.60
Joaquin Miller Park	1603314	0.9109	0.05177	0.03729	0.6951	20.41
Knowland Park	1975057	0.6132	0.3348	0.05201	0.5594	24.67
McLaughlin Eastshore SP--Albany access	321216	0.2082	0.7403	0.05176	0.5026	29.12
Meek Park	36418	0.3569	0.5027	0.1357	0.6124	34.51
Mirabeau Park	25391	0.5064	0.4483	0.04528	0.5350	35.58
Morcom Rose Garden	31730	0.6364	0.2061	0.1662	0.5521	34.48
Niles Park	130094	0.3874	0.2407	0.07938	0.4604	30.64
Oakland Coliseum	65450	0.01007	0.1934	0.7966	0.2746	35.07
Oakland Museum	67884	0.3079	0.1544	0.5376	0.3019	26.72
MLK Jr. Civic Center Park	11263	0.2504	0.4432	0.3087	0.5074	33.70
Peralta Park	623.5	0.4083	0	0.5917	0.2427	27.04
Port View Park	18315	0.04685	0.3607	0.5857	0.2018	27.87
Remillard Park	6826	0.9182	0.08178	0	0.6698	21.22
Shepherd Canyon Park	187255	0.8582	0.1108	0.03104	0.6883	21.30
Shinn Park and Arboretum	16594	0.4605	0.3107	0.2295	0.5663	35.92
Shoreline Park (ALA Co.)	174196	0.04357	0.9126	0.04327	0.2716	33.60
Silliman Activity Center	111400	0.1219	0.5115	0.3674	0.5002	35.87
Strawberry Creek Park	8374	0.6579	0.3086	0.03420	0.6001	32.59
Tule Pond	63647	0.4838	0.1493	0.08690	0.5381	33.44
Union City Library Pond and Park	67792	0.2916	0.4021	0.2683	0.4676	35.65
Washington Park (ALA Co.)	54625	0.2301	0.4853	0.2866	0.5233	33.82

Binary test

The binary test examined some specific characteristics of each park. Out of the total 36 parks, 14 parks were shoreline parks, 17 parks had stream nearby, three contained waterbodies within the parks, and 11 were or contained playgrounds or sports fields (Table 3). Six parks satisfied both “Shoreline Park” and “Near Stream”. Three parks satisfied both “Shoreline Park” and “Playground/Sports Field”. One park, Union City Library Pond and Park, was both near a stream and contained its own waterbody. Another park, Creekside Park, satisfied three characteristics: “Shoreline Park”, “Near Stream”, and “Playground/Sports Field”.

Table 3. Results of binary test of each park. We used Google Earth to determine some binary characteristics of each park.

Park Name	Shoreline Park	Near Stream	Waterbody	Playground/Sports Field
Alameda Creek--Staging Area	1	1	0	0
Albany Bulb	1	0	0	0
Albany Hill Park	1	1	0	0
Aquatic Park	1	0	1	0
Arlington Park	0	0	0	0
Berkeley Rose Garden	0	0	0	1
Cerrito Vista Park	0	0	0	0
Cesar Chavez Park	1	0	0	0
Chabot Park	1	1	0	0
Codornices Park	0	0	0	1
Creekside Park	1	1	0	1
Dimond Park	0	1	0	0
Estuary Park	1	0	0	1
Grinnell Natural Area	0	1	0	0
Garber Park	0	0	0	0
Joaquin Miller Park	0	1	0	0
Knowland Park	0	1	0	0
McLaughlin Eastshore SP--Albany access	1	0	0	0
Meek Park	0	1	0	0
Mirabeau Park	0	0	0	1
Morcom Rose Garden	0	0	0	1
Niles Park	0	1	0	0
Oakland Coliseum	0	0	0	1
Oakland Museum	0	1	0	1
MLK Jr. Civic Center Park	0	1	0	1
Peralta Park	1	1	0	0
Port View Park	1	0	0	0
Remillard Park	0	0	0	0
Shepherd Canyon Park	0	1	0	0
Shinn Park and Arboretum	0	1	0	0
Shoreline Park (ALA Co.)	1	0	0	0
Silliman Activity Center	1	1	0	1
Strawberry Creek Park	0	0	0	1
Tule Pond	0	0	1	0
Union City Library Pond and Park	0	1	1	0
Washington Park (ALA Co.)	1	0	0	0

Data analysis

Linear regression test

According to the results of linear regression tests, $\sqrt{Area}$, mean greenness and "Shoreline Park" were the three significant parameters that had p-values less than 0.05 (Table 4). $\sqrt{Area}$ had the lowest p-value of 0.0001281. The p-value of mean greenness was 0.02518, and the one of "Shoreline Park" was 0.0009819. Mean temperature had a p-value of 0.05287, which was slightly above the borderline of 0.05. All the other parameters had p-values above 0.05, therefore they were all insignificant if considered individually.

Table 4. P-value, sign of coefficient and R^2 of each individual landscape parameter. Significant parameters have p-values <0.05, coefficient sign tells the direction of correlation, and R^2 measures the fit of the model.

Parameters	P-value	Sign of Coefficient	R^2
$\sqrt{Area}$	0.0001281	+	0.3916
p_Tree	0.132	+	0.06741
p_Grass	0.08744	+	0.08594
p_Impervious	0.7427	-	0.00331
Mean Greenness	0.02518	-	0.1429
Mean Temperature	0.05287	-	0.1089
Shoreline Park	0.0009819	+	0.284
Near Stream	0.7896	-	0.002188
Waterbody	0.2205	+	0.04513
Playground/Sports Field	0.1087	-	0.0761

Among all the parameters, five parameters positively influenced the bird diversity: $\sqrt{Area}$, p_Tree, p_Grass, "Shoreline Park" and "Waterbody". The rest of the parameters, including p_Impervious, mean greenness, mean temperature, "Near Stream" and "Playground/Sports Field", had negative coefficients, therefore negatively affecting bird diversity.

None of the parameters yielded a good R^2 value larger than 0.5. Even only four out of the ten parameters had R^2 larger than 0.09. $\sqrt{Area}$ produced the highest R^2 of 0.3916, mean greenness had R^2 equaled to 0.1429, mean temperature had R^2 equaled to 0.1089, and "Shoreline Park" had R^2 equaled to 0.284. This result of low R^2 values was also visually obvious through correlation graphs of those individually significant parameters (Figure 1), indicating that none of our parameters could accurately model the diversity index. Hence, we needed to run multivariable regression to test the performance of combinations of different parameters.

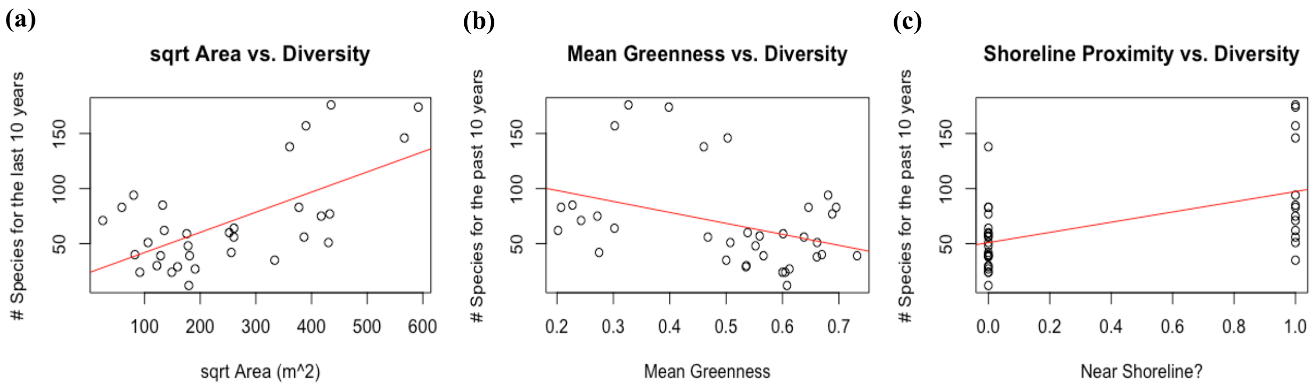

Figure 1. Individual correlations between diversity index and $\sqrt{\text{Area}}$, mean greenness, and "Shoreline Park" respectively. Although these three parameters were significant, their low R^2 suggested that they could not individually determine the bird biodiversity. Hence, we needed to run multivariable regression for combinations of different parameters to test the performance of these different models.

Multivariable regression and AICc ranking

Through multivariable regression, we yielded 38 combinations of parameters that had p-value < 0.05 and $R^2 > 0.5$. The average R^2 of these 38 models was 0.5919. We ranked these models with their AICc values, an estimator of the quality of models for a given small dataset, from low to high (Table 5). Smaller AICc value indicates a better quality of the model. Out of these 38 candidate models, 37 of them contained $\sqrt{Area}$. The only model that did not contain $\sqrt{Area}$ was the combination of p_Tree + p_Impervious + Mean Greenness + Mean Temperature. However, this model yielded the highest AICc value of 343.93 among all 39 models, being the model with the worst quality.

Table 5. P-values, R^2 and AICc rank, from low to high. A smaller AICc value means a higher quality of the model.

Rank	Model	p-Value	R^2	AICc
1	$\sqrt{Area}$ + p_Impervious + Mean Greenness + Mean Temperature	2.859e-06	0.6727	308.82
2	$\sqrt{Area}$ + Mean Greenness + Mean Temperature + Playground/Sports Field	6.883e-06	0.6498	310.98
3	$\sqrt{Area}$+ Mean Greenness + Mean Temperature	5.692e-06	0.6138	311.32
4	$\sqrt{Area}$ + p_Impervious + Mean Greenness	8.927e-06	0.6009	312.37
5	$\sqrt{Area}$ + Mean Greenness + Mean Temperature + Shoreline Park	1.464e-05	0.6289	312.84
6	$\sqrt{Area}$ + Mean Greenness + Mean Temperature + Waterbody	1.615e-05	0.6261	313.08
7	$\sqrt{Area}$ + p_Grass + Mean Greenness + Mean Temperature	1.617e-05	0.626	313.08
8	$\sqrt{Area}$ + Mean Greenness + Playground/Sports Field	1.315e-05	0.5895	313.27
9	$\sqrt{Area}$ + p_Impervious + Mean Greenness + Shoreline Park	1.767e-05	0.6235	313.30
10	$\sqrt{Area}$ + Mean Greenness + Mean Temperature + Near Stream	2.221e-05	0.6168	313.87
11	$\sqrt{Area}$ + p_Tree + Mean Greenness + Mean Temperature	2.342e-05	0.6152	314.00
12	$\sqrt{Area}$ + p_Grass + p_Impervious + Mean Greenness	2.567e-05	0.6124	314.22
13	$\sqrt{Area}$ + p_Tree	1.39e-05	0.5376	314.48
14	$\sqrt{Area}$ + p_Tree + Mean Greenness + Playground/Sports Field	2.94e-05	0.6084	314.60
15	$\sqrt{Area}$ + Mean Greenness + Shoreline Park	2.298e-05	0.5723	314.58
16	$\sqrt{Area}$ + p_Impervious + Mean Greenness + Near Stream	3.089e-05	0.6069	314.68
17	$\sqrt{Area}$ + p_Impervious + Mean Greenness + Waterbody	3.219e-05	0.6056	314.78
18	$\sqrt{Area}$ + p_Tree + p_Impervious + Mean Greenness	3.237e-05	0.6054	314.80
19	$\sqrt{Area}$ + Mean Greenness + Shoreline Park + Playground/Sports Field	3.255e-05	0.6053	314.81
20	$\sqrt{Area}$ + p_Tree + Mean Greenness + Playground/Sports Field	3.664e-05	0.6016	315.10
21	$\sqrt{Area}$ + p_Tree + Mean Temperature	3.343e-05	0.5605	315.46
22	$\sqrt{Area}$ + p_Tree + Mean Temperature + Shoreline Park	4.389e-05	0.596	315.55
23	$\sqrt{Area}$ + p_Grass + Mean Greenness + Playground/Sports Field	4.814e-05	0.5931	315.78
24	$\sqrt{Area}$ + Shoreline Park	2.647e-05	0.5166	315.90
25	$\sqrt{Area}$ + Mean Greenness + Near Stream + Playground/Sports Field	5.142e-05	0.591	315.95
26	$\sqrt{Area}$ + Shoreline Park + Playground/Sports Field	4.153e-05	0.5534	315.97
27	$\sqrt{Area}$ + Mean Greenness + Waterbody + Playground/Sports Field	5.208e-05	0.5906	315.98
28	$\sqrt{Area}$ + p_Tree + Mean Greenness	4.178e-05	0.5532	315.98
29	$\sqrt{Area}$ + p_Tree + Mean Temperature + Waterbody	5.701e-05	0.5877	316.20
30	$\sqrt{Area}$ + Mean Greenness + Playground/Sports Field + Waterbody	5.965e-05	0.5863	316.31
31	$\sqrt{Area}$ + p_Tree + Mean Greenness + Shoreline Park	6.277e-05	0.5846	316.44
32	$\sqrt{Area}$ + Mean Greenness + Near Stream	6.582e-05	0.5381	317.05
33	$\sqrt{Area}$ + Mean Temperature + Shoreline Park + Playground/Sports Field	8.211e-05	0.5759	317.11
34	$\sqrt{Area}$ + Mean Temperature + Shoreline Park + Waterbody	9.082e-05	0.5725	317.36
35	$\sqrt{Area}$ + Mean Greenness + Shoreline Park + Near Stream	9.131e-05	0.5724	317.37
36	$\sqrt{Area}$ + p_Tree + Mean Temperature + Near Stream	9.883e-05	0.5697	317.57
37	$\sqrt{Area}$ + p_Tree + p_Grass + Mean Greenness	0.0001075	0.5669	317.78
38	p_Tree + p_Impervious + Mean Greenness + Mean Temperature	0.0001178	0.5664	343.93

Besides $\sqrt{Area}$, mean greenness and mean temperature also appeared frequently in candidate models. The top five models included $\sqrt{Area}$, mean greenness, mean temperature, p_impervious, "Shoreline Park", and "Playground/Sports Field". Among these 38 candidate models, mean greenness appeared in 26 models, and mean temperature appeared in 15 models. "Shoreline Park" showed up 11 times, "Playground/Sports Field" showed up 10 times, and p_Tree, p_Grass and p_Impervious showed up 13, 4 and 8 times respectively.

Best-fitting model

The model with the highest quality was the combination of $\sqrt{Area}$, p_Impervious, mean greenness and mean temperature. It not only had the lowest AICc value of 308.8, but also the highest R^2 of 0.6727 among those 38 candidate models. Each parameter within this model had coefficients with different signs, in which the signs were consistent with those in linear regression. All of these parameters had p-values less than 0.05, implicating that they were all significant within this model (Table 6). The interception had the lowest p-value, 7.29e-05, being the most significant component in the model. Mean greenness was the parameter with the lowest p-value, 0.000227, which was even lower than $\sqrt{Area}$, 0.002443. P_Impervious had the highest p-value of 0.03628

Table 6. P-values, R^2 and AICc rank, from low to high. A smaller AICc value means a higher quality of the model.

	Coefficients	**P-value**
(Intercept)	220.63588	7.29e-05
$\sqrt{Area}$	0.12346	0.002443
P_Impervious	-83.32842	0.036281
Mean Greenness	-190.10556	0.000227
Mean Temperature	-2.43183	0.021893

Model check

To test the collinearity of the current best-fitting model, we plotted its residuals versus fitted values (Figure 2). The graph did not show any obvious trend of diverging out or diverging in, thus there was no conspicuous heteroscedasticity, a sign of non-constant variance. The graph neither displayed a trend of parabola, and the residuals bounced randomly around the 0 line, therefore our model was not nonlinear, and our assumption of a linear model was reasonable.

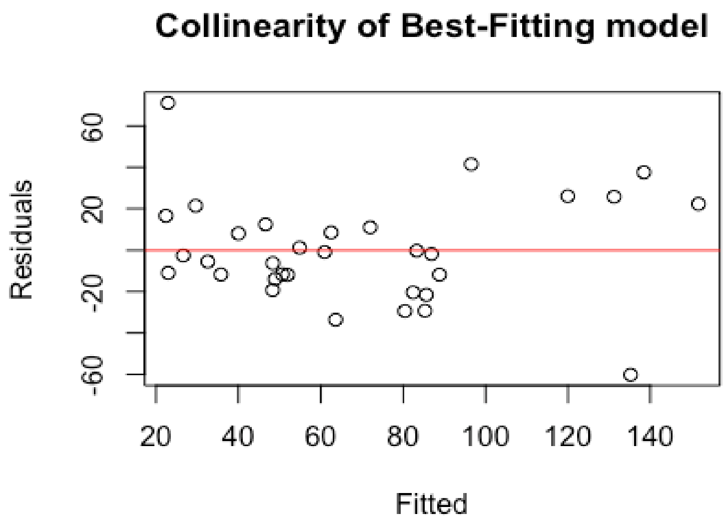

Figure 2. Residuals vs. Fitted Values of the best-fitting model. Fitted values stand for the predicted diversity. Residuals = 0 line corresponds to the estimated regression line.

We also computed variance inflation factor (VIF) between any two of the parameters within the model (Table 6). VIF is the ratio of parameters within the model, and quantifies the severity of multicollinearity among these parameters. None of any two parameters yielded VIF larger than 10. $\sqrt{Area}$ and mean greenness had the highest VIF of 2.163, while mean temperature and mean greenness had the lowest VIF of 1.562.

Table 6. VIF between different parameters in the best-fitting model. VIF larger than 10 suggests a sever multicollinearity between parameters.

Parameter Combinations	**VIF**
$\sqrt{Area}$ + p_Impervious	1.679
$\sqrt{Area}$ + Mean Greenness	2.163
$\sqrt{Area}$ + Mean Temperature	1.761
p_Impervious + Mean Greenness	1.635
Mean Greenness + Mean Temperature	1.562

DISCUSSION

Area, mean greenness and "Shoreline Park" were the three parameters that individually played significant roles on bird diversity. The 38 candidate models verified the dominant role of park size on species diversity, and also indicated the significant impacts of mean greenness, mean temperature, p_Impervious, "Shoreline Park" and "Playground/Sports Field" in multivariable models. The best-fitting model was *Diversity = 0.12346** $\sqrt{Area}$ *- 82.32842*p-Impervious – 190.10556*Mean Greenness – 2.43183*Mean Temperature*. This model provided us some hints on future urban parks design specifically in East Bay as a measure of bird species conservation. In addition, this study tested the use of citizen science data, revealing both its advantages and limitations.

Bird communities

The total number of avian species in all targeted parks was 240, 37.7% out of the total 636 species observed in California within the past 10 years recorded on e-Bird. As Alameda County and Contra Costa County are just very small portions of the entire California State, this percentage actually suggested a relatively large bird species conservation potential through urban parks design. The discussion of the results below will provide some potential perspectives that we should pay attention in the future.

Significant landscape parameters

Individually, $\sqrt{Area}$, mean greenness and "Shoreline Park" were the three parameters that largely affect bird diversity, since they all had p-value less than 0.05. Area, p_Tree, p_Grass "Shoreline Park" and "Waterbody" had positive coefficients, therefore they were positively correlated with bird diversity. P_impervious, mean greenness, mean temperature, "Near Stream" and "Playground/Sports Field" had negative coefficients, hence they were negatively correlated to bird diversity. Therefore, surprisingly, the directions of impacts of mean greenness and "near stream" were against our hypothesis. Although previous studies by Marzluff and Donnelly (2004), Aida et al. (2016), and Chang and Lee (2016) suggested that more greenness could support greater biodiversity in urban areas, this did not correspond to our research. This might be because that our study scaled down the study sites into individual parks instead of the entire city.

Area was the most significant parameter that positively influencing avian diversity, not only because of its low p-value of 0.0001281, but also because it appeared in almost all the candidate models. In multivariable models, area presented a dominant role, as 37 out of 38 models contained $\sqrt{Area}$, and the only one that did not include $\sqrt{Area}$ was the model with the worst quality. Previous study of Chang and Lee in 2016 has verified the point that area played a primary role in positively affecting species diversity and number of nestings. Marzluff and Donnelly (2004) also emphasized the importance of reserve size in the context of urban bird conservation. Hence, size of park should be the primary concern during park design process.

Besides area, mean greenness and mean temperature were the other two parameters that appeared frequently among the 38 candidate models, implying their essential role in multivariable models to predict bird diversity. Although p_Tree was the fourth frequent parameters, none of the models that contained p_Tree was in top ten. With a high individual p-value of 0.06741, p_Tree played an essential role neither individually nor with other parameters.

Best-fitting models

Our best-fitting model was

*Diversity = 0.12346** $\sqrt{Area}$ *- 82.32842*p-Impervious – 190.10556*Mean Greenness – 2.43183*Mean Temperature*.

This model had a high R^2 of 0.6727 and a low p-value of 2.859e-06. The residuals vs. fitted values plot did not suggest any trends of heteroscedascity or nonlinearity, therefore the current linear model is reasonable. The VIF test revealed that none of the VIF values among any two of the parameters was larger than 10, therefore there was no severe multicollinearity among parameters within this best-fitting model.

The p-values of all parameters within this model were less than 0.05, suggesting that all parameters wree significant for this model. Mean greenness had a p-value even less than $\sqrt{Area}$, suggesting it was even more significant than $\sqrt{Area}$ in this model. P_impervious produced the highest p-value, hence it played a relatively less significant role. The intercept has the lowest p-value of 7.29e-05, even much lower than any other parameters. This not only suggested its significant role in the model, but also reminded us the need of using more data to refine the model.

Limitations and Future Directions

Use of eBird

eBird provided a lot of convenience of data collection for this study. First, it is very accessible and largely saves the time of data collection, without us going to the sites and observing birds for the whole day. Second, it broke the geographic limitations, so that we could easily obtain bird observation data at each hotspot in East Bay, and even globally if necessary. Last but not the least, eBird is direct and accessible. The records include detailed observations of species name, number of specific species observed, and dates, which helped us to better understand the records at each hotspot.

However, despite its convenience, eBird also largely limited the size of our dataset. Although we covered 36 urban parks in our study, this was only a little proportion of all urban parks in East Bay. This limitation was mainly because that not all parks had bird observation records on eBird. As we cannot control citizen scientists where they observed birds, it is hard to include all urban parks into observation hotspots.

eBird also has some other limitations. First, as pointed out by Sullivan et al. in 2009, species detactability is a bias in most bird-sampling techniques, as easily detected species are reported more frequently than those hard to detect. More specifically, citizen-science observations might be restricted to detection of changes in abundance, richness or similarity over space and time (Kremen, 2011). Even professionally trained scientists have these issues in bird sampling, thus we cannot ensure whether citizen scientists detected all species that were present. Second, the skill levels of eBird users vary widely; therefore it is impossible to assume all observations reported were correct (Sullivan et al., 2009). Finally, since birding community is not evenly distributed, there is a bias of birding effort (Ferrer et al., 2006). This is reflected by eBird dataset distribution, as dataset is more heavily concentrated in areas with high populations.(Sullivan et al., 2009).

GIS digitization error

Another limitation is the use of GIS digitization. Digitization can also be improved with more accurate operations. In the process of manual digitization, there were inevitably some errors, including dangling nodes, slivers, overshoots or undershoots, switchbacks, knots, and loops (Fisher, 1995). These errors could affect the accuracy of park sizes and land cover percentages. Although the effect of these errors might be minor compared to the large size of studied parks, it still affected the accuracy of the study.

Broader implications

A further refined model can be constructed by collecting more bird observations in more urban parks in East Bay and more accurate landscape parameters, and taking the collinearity of the model into account. This means that the model does not have to be linear but nonlinear instead. This refined model could help to predict bird species richness in urban parks in or even out of East Bay. Further factors need to be taken into consideration when we build future local models for different areas, including geographic locations, climates, surrounding population and vise versa. These refined models, based on complete dataset and accurate GIS analysis, could be developed into planning tools in the process of urban design for the sake of urban species conservation.

ACKNOWLEDGEMENTS

I acknowledge the Cornell Lab of Ornithology for their establishment and management of eBird website, which brought a huge convenience for this individual study. I would like to express my deepest appreciation for my mentor, Iryna Dronova, who assisted me to complete this individual study by providing all available guidance and resources even when she was extremely busy, gave me a lot of trust to encourage my research independence, and willing to write me a recommendation letter that eventually lead me into my dream school to pursuit my academic interests. I would like to thank Patina Mendez and Kurt Spreyer for teaching the art of writing a thesis and tirelessly contributed all they can to assist me finishing this thesis. I also want to thank Claire Neal, Delaney Inman, Kathleen Andy McCully and Weijie Dong for the doing all those peer reviews to refine my writing and organization. All these helps are precious and essential for me to finish this thesis.

REFERENCES

Aida, N., S. Sasidhran, N. Kamarudin, N. Aziz, C. L. Puan, and B. Azhar. 2016. Woody trees, green space and park size improve avian biodiversity in urban landscapes of Peninsular Malaysia. Ecological Indicators 69:176–183.

Chang, H.-Y., and Y.-F. Lee. 2016. Effects of area size, heterogeneity, isolation, and disturbances on urban park avifauna in a highly populated tropical city. Urban Ecosystems 19:257–274.

Clergeau, P., J.-P. L. Savard, G. Mennechez, and G. Falardeau. 1998. Bird Abundance and Diversity along an Urban-Rural Gradient: A Comparative Study between Two Cities on Different Continents. The Condor 100:413–425.

Crooks, K. R., A. V. Suarez, and D. T. Bolger. 2004. Avian assemblages along a gradient of urbanization in a highly fragmented landscape. Biological Conservation 115:451–462.

Donnelly, R., and J. M. Marzluff. 2004. Importance of Reserve Size and Landscape Context to Urban Bird Conservation. Conservation Biology 18:733–745.

Emlen, J. T. 1974. An Urban Bird Community in Tucson, Arizona: Derivation, Structure, Regulation. The Condor 76:184–197.

Fernández-Juricic, E. 2000. Bird community composition patterns in urban parks of Madrid: The role of age, size and isolation. Ecological Research 15:373–383.

Fernández-Juricic, E. 2004. Spatial and temporal analysis of the distribution of forest specialists in an urban-fragmented landscape (Madrid, Spain): Implications for local and regional bird conservation. Landscape and Urban Planning 69:17–32.

Ferrer, X., L. M. Carrascal, Gordo, O., and Pino, J. 2006. Bias in avian sampling effort due to human references: An analysis with Catalonian birds (1900 - 2002). Ardeola 53(2): 213-227

Fisher, P. 1995. Innovations In GIS. CRC Press.

Glennon, M. J., and W. F. Porter. 2005. Effects of land use management on biotic integrity: An investigation of bird communities. Biological Conservation 126:499–511.

Hardman S., (2011) How does urbanization affect biodiversity. Ecologica, https://ecologicablog.wordpress.com/2011/11/06/how-does-urbanization-affect-biodiversity-part-one/

Ikin, K., E. Knight, D. B. Lindenmayer, J. Fischer, and A. D. Manning. 2012. Linking bird species traits to vegetation characteristics in a future urban development zone: implications for urban planning. Urban Ecosystems 15:961–977.

Kremen, C., K. S. Ullman, and R. W. Thorp. 2011. Evaluating the Quality of Citizen-Scientist Data on Pollinator Communities. Conservation Biology 25:607–617.

McElhinny, C., P. Gibbons, C. Brack, and J. Bauhus. 2006. Fauna-habitat relationships: a basis for identifying key stand structural attributes in temperate Australian eucalypt forests and woodlands. Pacific Conservation Biology 12:89–110.

Myers, N. 1990. The biodiversity challenge: Expanded hot-spots analysis. Environmentalist 10:243–256.

Peterson, A. T., A. G. Navarro-Sigüenza, E. Martínez-Meyer, A. P. Cuervo-Robayo, H. Berlanga, and J. Soberón. 2015. Twentieth century turnover of Mexican endemic avifaunas: Landscape change versus climate drivers. Science Advances 1:e1400071.

Pick`ett, S. T. A., M. L. Cadenasso, J. M. Grove, C. G. Boone, P. M. Groffman, E. Irwin, S. S. Kaushal, V. Marshall, B. P. McGrath, C. H. Nilon, R. V. Pouyat, K. Szlavecz, A. Troy, and P. Warren. 2011. Urban ecological systems: Scientific foundations and a decade of progress. Journal of Environmental Management 92:331–362.

Shwartz, A., S. Shirley, and S. Kark. 2008. How do habitat variability and management regime shape the spatial heterogeneity of birds within a large Mediterranean urban park? Landscape and Urban Planning 84:219–229.

Sullivan, B. L., C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling. 2009. eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation 142:2282–2292.

Zalewski, A. 1994. Diet of urban and suburban tawny owls (Strixaluco) in the breeding season. J. Raptor Res. 28: 246–252.