Using San Diego Bay Along the Pacific Flyway - A Spatial and Temporal Analysis of Bird Use using Systematic and Opportunistic Data

Corey C. Sheredy
GEOG 596B
Master of Geographic Information Systems (GIS) Capstone Project
Pennsylvania State University
Advisor: Dr. Jason Keagy

1. Introduction

Determining population trends for bird species is critical information for both conservation and environmental decision making. For most bird species, populations are typically monitored using systematically collected point count data (Howe and others 1989), where the overall species abundance is correlated with species detection (Horns and others 2018). In San Diego Bay, these systematic bird counts are conducted when resources are available to support the surveys, however, adequate temporal coverage is a continuous challenge (TierraData 2018) to generate reliable population trends. However, with the increase of citizen science data collection in recent years with databases such as eBird, there is opportunity to fill these temporal data gaps. Data from eBird has been successfully used to monitor population trends (Clark 2017). In this study, we evaluate the population trends of bird species observed in the San Diego Bay area using eBird data, compare the performance of eBird data population trends versus the population trends determined using the systematic survey data, and evaluate the connection between bird observations and habitat.

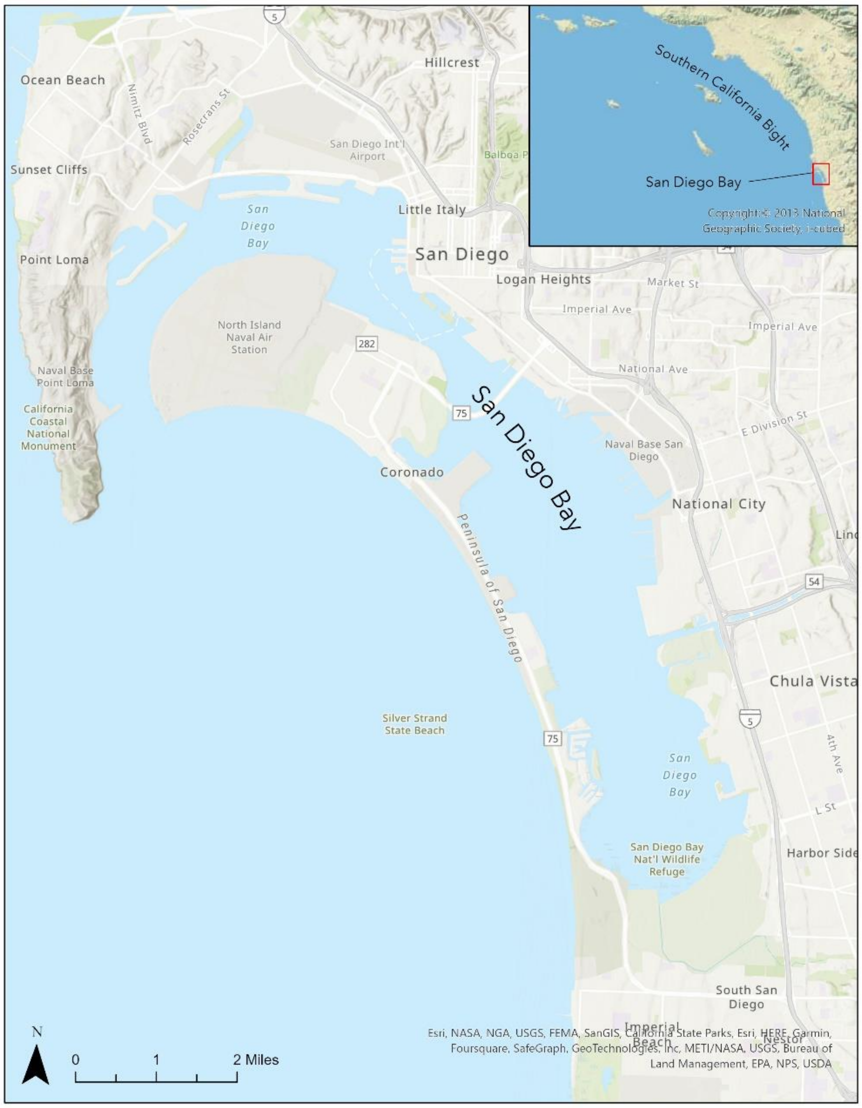
Background

Study Location

The San Diego Bay is a natural harbor and deepwater port, approximately 12 miles long and ranging from 1 to 3 miles wide, located in San Diego, California near the U.S./Mexico border. The shoreline and waters of the bay are managed by the Port of San Diego (referred to as “Port” below) and the U.S. Navy. An overview of San Diego Bay is provided in Figure 1-1.

San Diego Bay is a natural, nearly enclosed embayment with a deep entrance and sheltered waters. The bay originated from alluvial floodplains of the Otay, Sweetwater, and San Diego Rivers (Navy and Port 2013). San Diego Bay is part of the southern California Bight, a curve in the southwestern California coastline that extends from Point Conception to just south of the Mexican border. For several reasons, this ecological region is very diverse and productive -- the region is the northern extent of many tropical species and southern extent for many temperate species. The Bight’s embayments, including San Diego Bay, contain shallow and intertidal habitat needed by a variety of species. These ecological ‘edges’ are scarce in Southern California due to commercial development on the coast (Navy and Port 2013), as Southern California has lost an estimated 62 percent of its coastal wetlands (Lowe et al. 2018). However, while San Diego Bay has lost approximately 42 percent of its historic shallow and subtidal habitat, 84 percent of its intertidal mudflat habitat, and 70 percent of its salt marsh habitat due to development, recent efforts to restore and/or reestablish these habitats have been made. One example is the Port’s proposed coastal wetland mitigation bank on 110 acres of a site known as Pond 20, a former salt evaporation pond located in south San Diego Bay. Pond 20 is currently vacant, isolated from tidal influence, and provides little habitat value due to its salt-encrusted surface and invasive plants. The benefits of creating a wetland in south San Diego Bay include increased biodiversity; supporting migratory bird species and a suite of coastal wetland-dependent species; providing food, habitat, and spawning grounds to fish, including recreational and commercial fish species; improved water quality; increased carbon sequestration capacity; and others.

Figure 1-1. Study Location Overview Map.



San Diego Bay as Part of the Pacific Flyway

San Diego Bay, as part of the Pacific Flyway, is used by millions of birds traveling between northern breeding grounds and southern wintering sites. The bay provides the largest expanse of protected waters in southern California. It is one of a dwindling number of stopover sites used by migrant birds to replenish their energy during their long journeys. It also supports large populations of over-wintering birds that depend on its resources for food, shelter, resting, and staging before migration north (Navy and Port 2013).

More than 300 bird species have been documented using the bay (Port and Navy 2013) with close to half directly depending on it (Tierra Data Inc. 2009, 2011). Most bird species are migratory and use the bay as a winter stopover, while others come to nest or are resident species present year-round. San Diego Bay also supports many shorebirds and seabirds within its saltponds and wetland habitat; significant numbers of seabirds and shorebirds establish nests on the salt pond levees each spring and summer (USFWS 1994). These include the federally and state endangered California Least Tern (*Sternula antillarum browni*) and federally threatened Western Snowy Plover. Large multispecies breeding colonies include the Royal Tern (*Thalasseus maxima*), Elegant Tern (*T. elegans*) and among many others (TierraData 2018).

Avian Surveys in San Diego Bay

This study evaluates the population trends of bird species observed in the San Diego Bay area using opportunistically collected data (eBird data) as well as compares the performance of eBird data population trends versus the population trends determined using the systematic survey data. These two distinct datasets are further described below.

Systematic Data

The Port and the Navy jointly conduct a comprehensive, systematic survey of avian use of the bay that covers a year and contains focused methods to detect specific classes of birds (i.e., shorebirds, waterfowl, and seabirds). The goal of this program, henceforth referred to as the baywide surveys, is to establish a scientifically defensible baseline and conduct long-term trend monitoring to census water-dependent bird species of San Diego Bay (TierraData 2018). Initial surveys using this methodology were completed 1996-1997, 2006-2007, 2009-2010, and 2016-2017. It is recommended that these surveys are completed every three to five years with annual or biennial point count surveys to allow for the discernment of natural variation, however, this is not always possible due to resource constraints.

Opportunistic Data

The increase in ecotourism and the development of large citizen science programs, such as eBird, have resulted in a rapidly growing body of data on birds (Horns and others 2018). Opportunistic data have been previously employed to effectively answer questions about species occurrence at large geographic or temporal scales, however, many have cautioned against the use of this data, particularly when estimating population trends (Kamp and others 2016). However, while systematic monitoring programs require substantial resources, large volumes of opportunistic data have yielded results similar to those of formal bird-count surveys (Horns and others 2018). San Diego Bay has a large bird watching ecotourism base, providing a potential source of population trend data.

Problem Statement

As stated above, systematic surveys are completed periodically within San Diego Bay to assess the use of the bay for water-dependent avian species. Land managers and restoration project managers typically rely on species diversity monitoring programs, such as the baywide surveys, to assess spatial and temporal diversity trends. However, a common limitation of adequate data collection is the cost of ensuring sufficient temporal and spatial coverage (Callaghan and Gawlik 2015). Lapses in monitoring programs may inhibit the assessment of population trends, which are important for informing conservation policies. Additionally, the recommended annual efforts to better separate actual trends from inter-annual weather variation has not risen as a funding priority for the Port or the Navy. However, citizen science applications like eBird may provide an opportunity for members of the community to fill this data gap between major systematic surveys, such as the baywide surveys.

Additionally, the purpose of the baywide survey is to establish a baseline of avian use of San Diego Bay; however, the survey does not provide context on types of habitat or habitat use. It is important to analyze how populations of water-dependent bird species are changing and whether changing habitats/land use of the bay has influenced bird use, especially in the context of evaluating the effects of climate change or the success of habitat restoration or establishment efforts. The Port plans to create 110-acres of wetland habitat in a salt pond isolated from tidal influence within San Diego Bay, so understanding the connection between bird observations and habitat will be important in estimating the effects of this project.

Study Questions

Based on the problem statement, four study questions were developed:

1. Are the population trends of birds using San Diego Bay increasing or decreasing over time?
2. Does land/water use affect bird populations in San Diego Bay?
3. What species and observation numbers should we expect to see upon the establishment of a 110-acre wetland mitigation bank?
4. Is using opportunistic data a valid substitute for systematic data collection to monitor bird trends in San Diego Bay? Are there detectable changes between both datasets? Are they consistent with one another?

2. Methodology

The materials and methods needed to address the four study questions are discussed in this section.

Data

Data sources used in this study are summarized in Table 2-1, below.

Table 2-1. Data Sources.

Dataset	Source	Time Span	Purpose
eBird observation data	eBird.com	2006*-2021	Opportunistic bird observation data
San Diego Bay Avian Surveys point count location observation data	Port of San Diego and US Navy	2006-2007 2009-2010 2016-2017	Systematic bird observation data
San Diego Bay Avian Surveys point count locations	Port of San Diego and US Navy	2006-2007 2009-2010 2016-2017	Point count locations from baywide surveys
Coastline	SanDAG.org	2020	Used to create study boundary

Dataset	Source	Time Span	Purpose
Satellite imagery	SanDAG.org	2008, 2010, 2017, 2020	Used to delineate habitat within 500 meters of the point count locations
Baywide eelgrass surveys	Port of San Diego and US Navy	2008, 2011, 2017	Used to assist in shallow water delineations.

Notes:
*For this study, data collected before 2006 were excluded to maintain consistency with the baywide surveys.

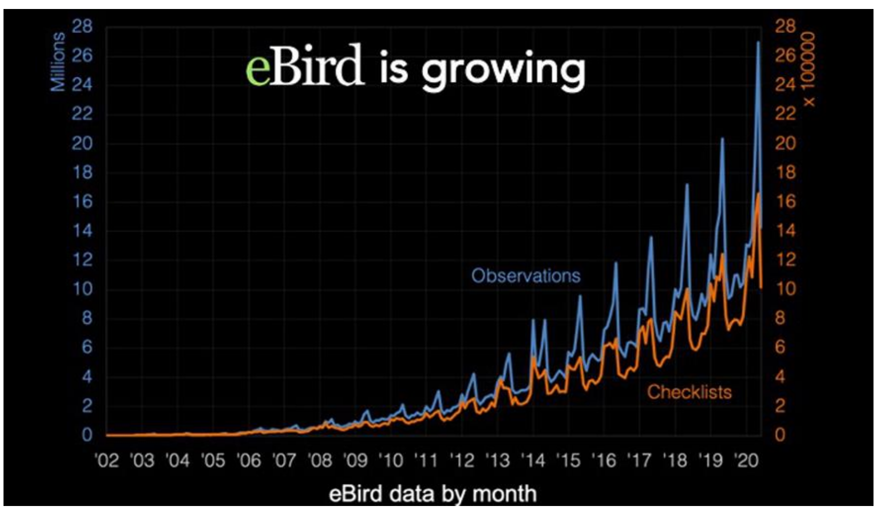
Analysis

Although the study questions are interrelated, methods for analysis are discussed separately, for each study question below:

Project Question 1: Are the trends of birds using San Diego Bay increasing or decreasing over time?

To answer the first study question, opportunistic bird observation data were used. We obtained observation data from the eBird database[1] within the San Diego Bay region for all data available at the time (observations ranged from 1947 through September 2021). Once the data were downloaded, we imported into R. Using R, we removed records of any observations that did not include a valid species identification (i.e., species that were identified to the group level, and not to the species level, such as tern species or *Peep sp*.). Next, any data that did not include a valid observation count were excluded (i.e., a count of "X", which indicates presence of a species, but the user did not specify count). Lastly, only eBird data that were considered complete checklists (i.e., all species observed were reported) were used. These complete checklists allow a zero count for any species that is not recorded (Strimas-Mackey and others 2020). Data prior to 2006 were also excluded for comparative purposes to the systematic surveys, which include three years of data collection spanning 2006-2007, 2009-2010, and 2016-2017. Post-2006 also corroborates with the increase of eBird data collection (see Figure 2-1).

Figure 2-1. eBird data submissions over time.



Source: eBird 2021.

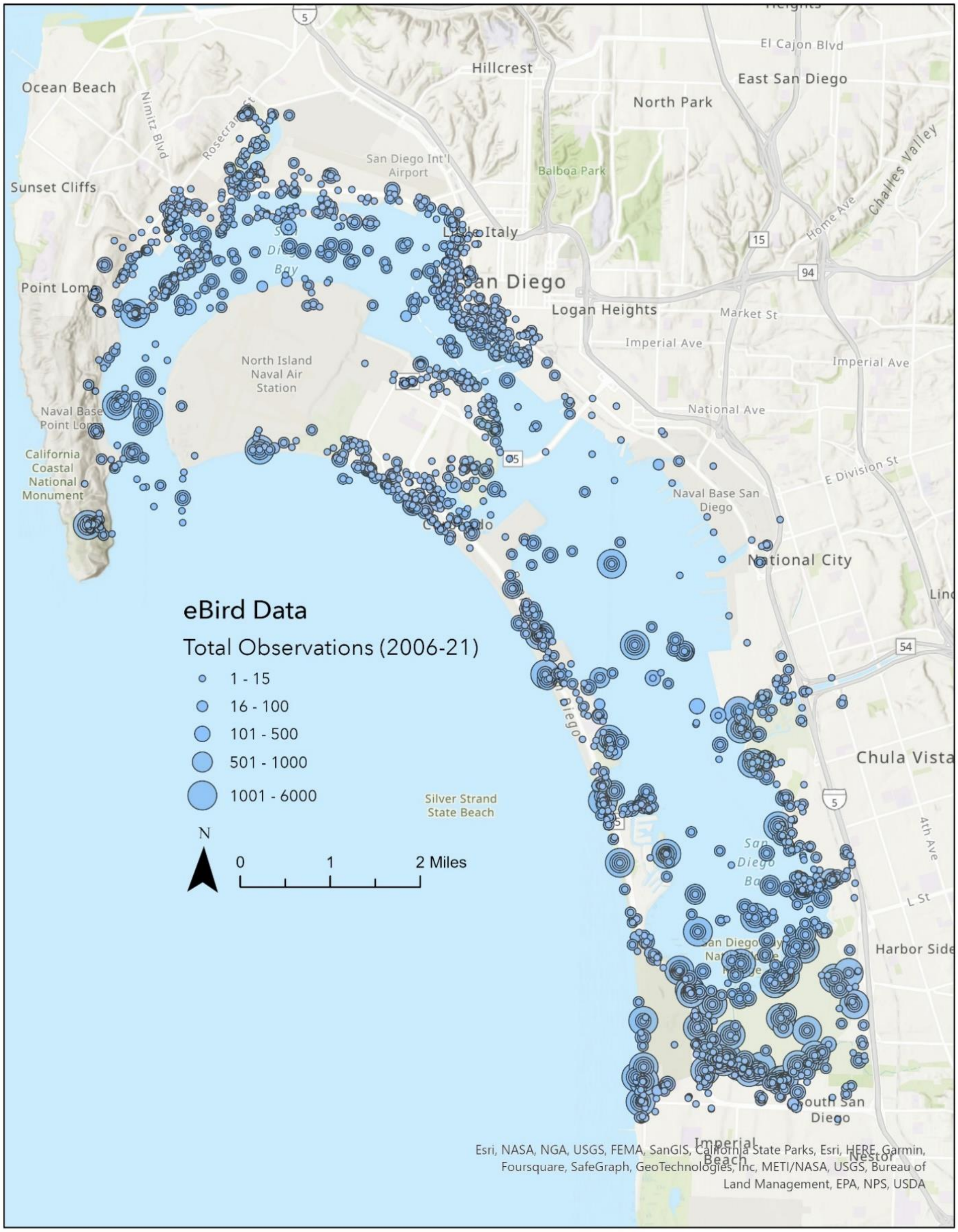
[1] https://science.ebird.org/en/use-ebird-data/download-ebird-data-products

Additional data columns were added to the eBird dataset as well in preparation for analysis. Such data included:

- Julian date
- Observation Year
- Observation Month
- Observation Season

The eBird dataset were then exported from R into a comma separated file (CSV). The CSV was uploaded into an ArcGIS Pro project as a table, and data were displayed using the XY Table to Point tool to create a feature class with the point data. A 500-meter (m) buffer was applied to the coastline within San Diego Bay, and all observations outside of San Diego Bay and its 500-m buffer were excluded, as shown in Figure 2-2. The attribute table of the observation data feature class was then exported as a CSV file.

Figure 2-2. eBird observation data clipped to 500-m of the coastline of San Diego Bay.



The updated eBird data CSV file was then reloaded into R for analysis.

To address the first study question, all data that met the criteria provided above (i.e., valid species, valid observation count value, within 500-m of San Diego Bay coastline) were used for analysis. Additionally, it was determined that a minimum of 150 individual records for each species were required to successfully run the model; therefore, species that had fewer than 150 occurrences within the dataset were excluded from this analysis. A Poisson generalized linear model was selected to estimate bird observations, using year as the explanatory variable, for each species. A Poisson distribution was selected because it is typically used to model count variables (Cameron and Trivedi 1998). The model used is as follows:

Species X Counts ~ Observation Year

We used the glm[2] function in the stats package (v 3.6.2) in R (v 4.2.1) with the family set to "Poisson" to estimate this model.

Once the predicted observations were calculated for each species, we then plotted these observation estimates, as well as their 95 percent confidence interval using the ggplot2[3] package in R.

To evaluate for seasonality, the data were split into seasons – spring, summer, fall, and winter, and the model was run to estimate observations for each species within these four subsets of data. As with the full dataset, we then plotted observation estimates for each season, as well as the 95 percent confidence intervals.

Project Question 2: Does land/water use affect bird populations in San Diego Bay?

One of the limitations of the Baywide avian surveys is that while it provides a rich dataset that is long-term and comprehensive, no habitat data are incorporated into this analysis. The study itself states:

While the high value of these surveys remains their long-term and comprehensive nature, much benefit could be extracted from the data sets by analyzing correlations between habitat use and types of habitat.

This study question aimed to incorporate habitat into analysis.

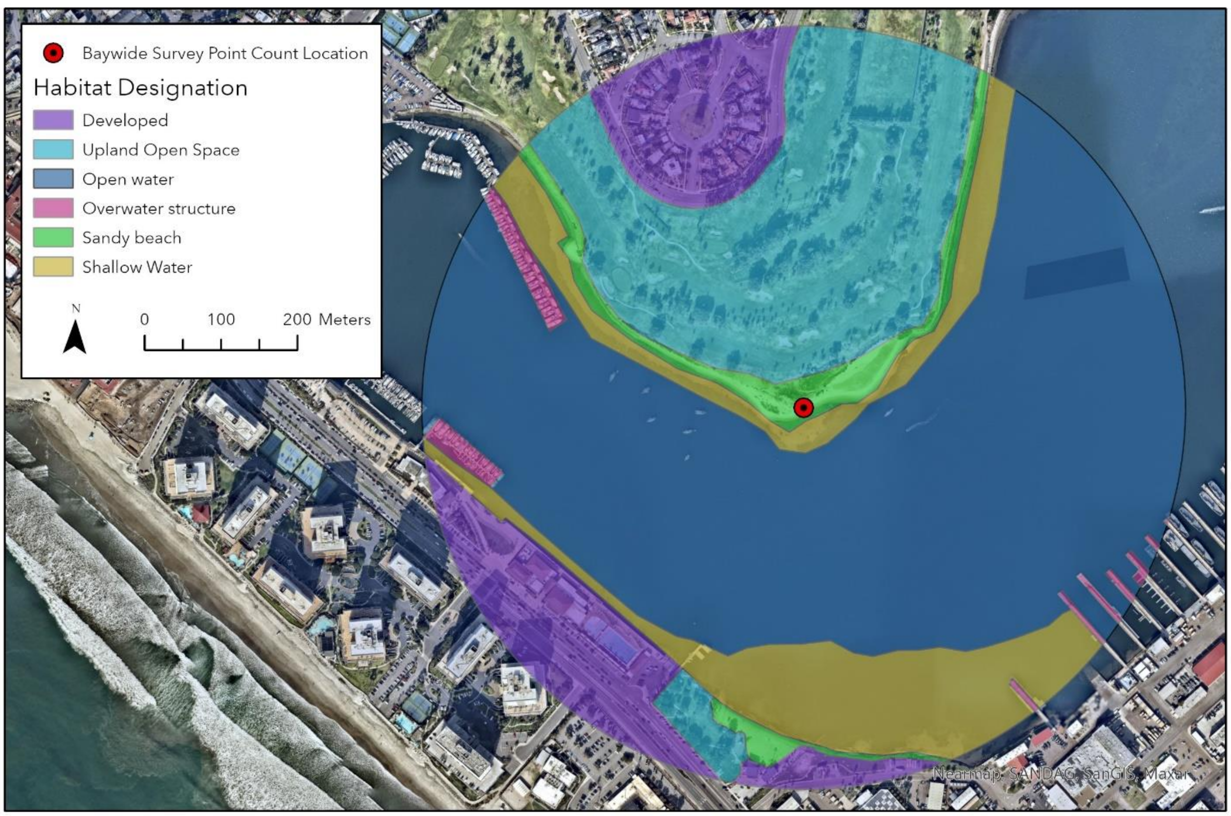
To analyze correlations between bird use and habitat type, analysis followed the methodology utilized in Desrochers and others (2008) and Hostetler and Knowles-Yanez (2003). We evaluated the habitat characteristics of the point count stations used during the Baywide surveys by creating a 500-m buffer around each station on high-resolution satellite imagery best corresponding with each Baywide survey year (2008, 2010, and 2017 images from SanDAG[4]). Habitat/land use were characterized by using a GIS-based desktop delineation. Categories for habitat/land-use were selected generally based on the habitat types used during the Baywide surveys, as well as guided by the San Diego Bay INRMP. They are as follows:

2 https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm
3 https://ggplot2.tidyverse.org/reference/ggplot.html
4 https://opendata.sandag.org/

- Developed
- Overwater structure
- Rocky riprap
- Marsh/wetland habitat
- Shallow water (approximately less than -12 ft mean lower low water [MLLW] depth)
- Open water (approximately deeper than -12 ft MLLW)
- Sandy beach
- Open space (non-developed upland, including open park space)

For each station, three sets of 500-m radius delineations were created. An example is provided below as Figure 2-3.

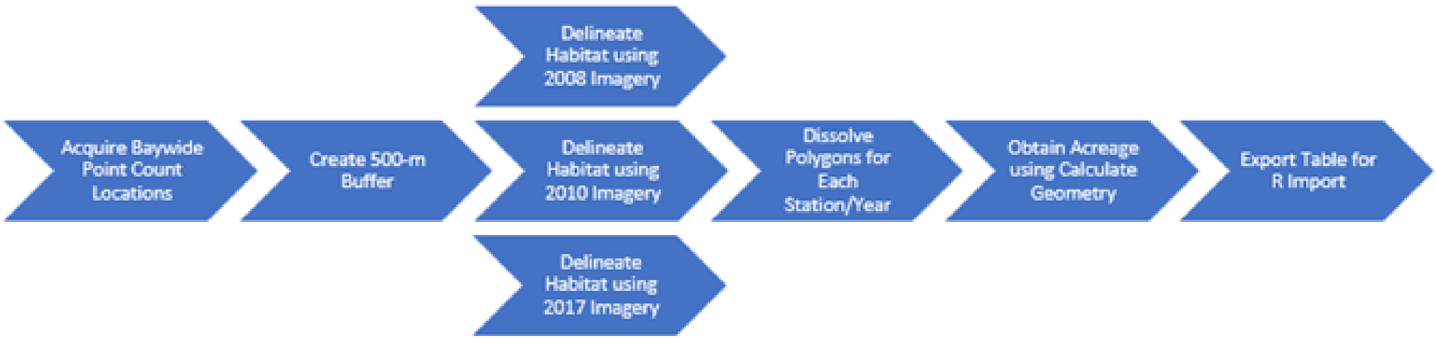
Figure 2-3. Habitat delineation of point count stations.



Habitat delineation for point count location Station 11, located on Coronado Island, using satellite imagery from 2017.

Once all stations were delineated for each survey year, the area of each habitat type within the 500-m buffer of the point count location were calculated using the Calculate Geometry tool. The process of this calculation is provided in Figure 2-4 below.

Figure 2-4. Calculation of Habitat Areas Process Workflow.



Once the areas were exported as a table, they were loaded into R. Principal component analyses (PCA) were completed using the prcomp[5] function in R to reduce the dimensionality of the habitat data and eliminate multicollinearity in our statistical models (i.e., habitat types are not independent). Since the habitat/land use between each of the aerial images (images from 2008, 2010, and 2017) did not vary greatly (i.e., habitat did not vary greatly within stations between years, habitat acreage values between the three years were averaged at each station prior to PCA.

The habitat PCA results and interpretations are shown in Table 2-2. Table 2-3 provides the eigenvectors (loadings) for the PC1 and PC2, while Figure 2-5 provides a visual representation of the correlation matrix of the habitat types.

Table 2-2. Importance of Principal Components.

	PC1*	PC2*	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	1.80	1.40	1.20	0.79	0.64	0.59	0.32	0.01
Proportion of Variance (%)	40%	23%	18%	8%	5%	4%	1%	0%
Cumulative Proportion of Explained Variance	40%	63%	81%	89%	94%	99%	100%	100%

Notes: (*) selected for analysis

Table 2-3. PCA loadings for selected PCs.

Habitat	PC1	PC2
Developed	-0.41	-0.31
Marsh	0.46	-0.21
Open Space	0.01	-0.57
Open Water	-0.28	0.48
Overwater Structure	-0.40	0.14
Riprap	-0.42	-0.02

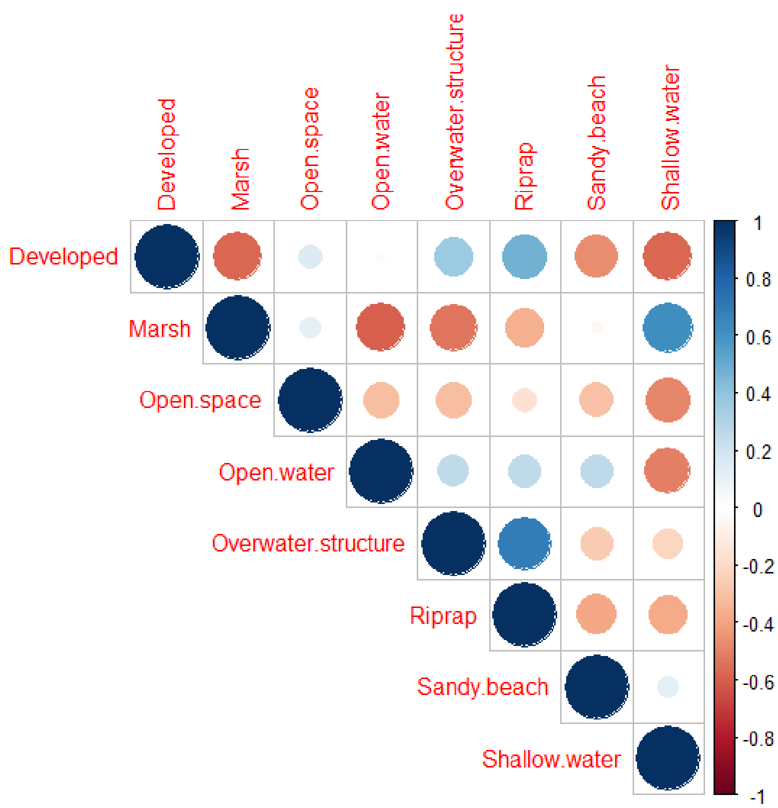
[5] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp

Habitat	PC1	PC2
Sandy Beach	0.18	0.52
Shallow Water	0.41	0.14

For use in our modeling, we selected the two top principal components (PCs), as these two PCs explained approximately 63% of the variance in the data and were easily interpretable.

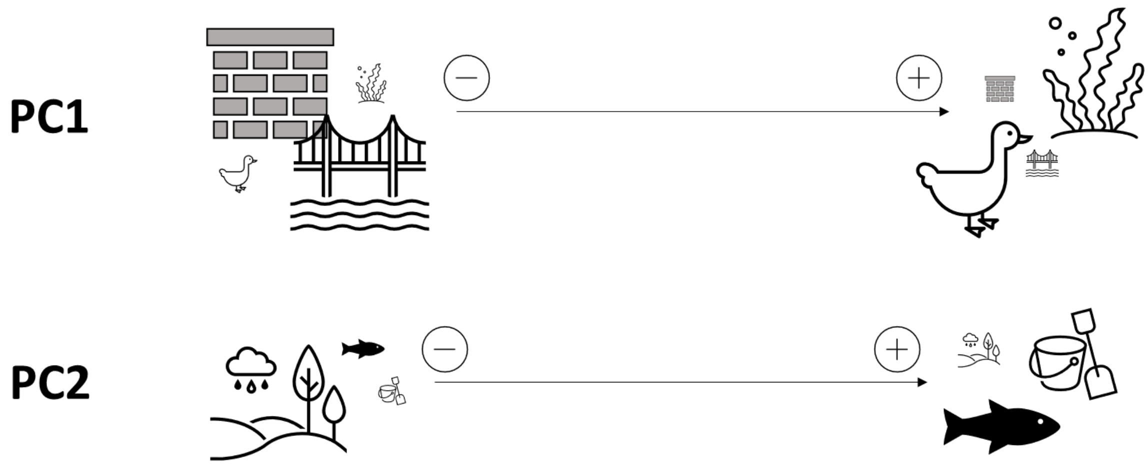
Based on the eigenvectors, we see that PC1 is positively correlated with natural habitat variables, such as marsh/wetlands and shallow water. PC1 is negatively correlated with development variables, such as developed areas, overwater structures, and riprap. PC2 is positively correlated with sandy beach and open water, and negatively correlated with open space and developed areas. A visual representation the interpretations of PC1 and PC2 are provided in Figure 2-6.

Figure 2-5. Correlation matrix for habitat variables.



Notes: this matrix provides a visual representation of how each habitat variable is correlated with another. Blue dots indicate a positive relationship, while red indicates a negative relationship, the larger the dot, the stronger the relationship. For example, areas that are developed and marsh habitat are negatively correlated (i.e., typically not found within the same area), while marsh and shallow water areas have a strong positive correlation (i.e., typically found together).

Figure 2-6. Variable Significance for PC1 and PC2.



Notes: For PC1, the brick wall represents Developed, bridge represents Overwater Structures/Riprap, Duck represents Marsh/Wetland, and algae represents Shallow Water. For PC2, Open Space is represented by the field with trees, Open (Deep) Water is the fish, and Sandy Beach is represented by the shovel and pail.

The next step was to incorporate PC scores into the dataset. For this, eBird data was loaded back into the ArcGIS Pro document; a spatial selection was performed to identify records that were within the 500-m buffer of a point count location. A spatial join was then completed to add station IDs to the eBird dataset for records that corresponded to a station. The data were again exported as a CSV file and loaded into R. In R, the calculated PC scores were merged with eBird records with corresponding station IDs. Predicted bird observations were estimated using a Poisson general linear mixed-model (GLMM), as described in methodology outlined in Boersch-Supan and others (2019). The model equation is as follows:

Species X Counts ~ Observation Year + PC1 + PC2 with Location nested within Observation Year as a random effect.

Counts were modeled as a Poisson distribution. The glmer[6] function in the lme4 package (v 1.1-31) in R.

The model was run for species that were selected based on the top species observed during the Baywide surveys, species of importance, and trying to garner a range of species assemblages. The species selected for analysis are provided in Table 2-4.

[6] https://www.rdocumentation.org/packages/lme4/versions/1.1-31/topics/glmer

Table 2-4. Avian Species Selected for Analysis.

Species Assemblage	Species
marshbird	snowy egret
seabird	Brandt's cormorant
	brown pelican
	California least tern
	elegant tern
	elegant tern
	osprey
	royal tern
	western gull
shorebird	black-necked stilt
	dowitcher
	killdeer
	marbled godwit
	western sandpiper
	western snowy plover
	willet
terrestrial	horned lark
waterfowl	American coot
	brant
	Clark's/ western grebe
	eared grebe
	surf scoter

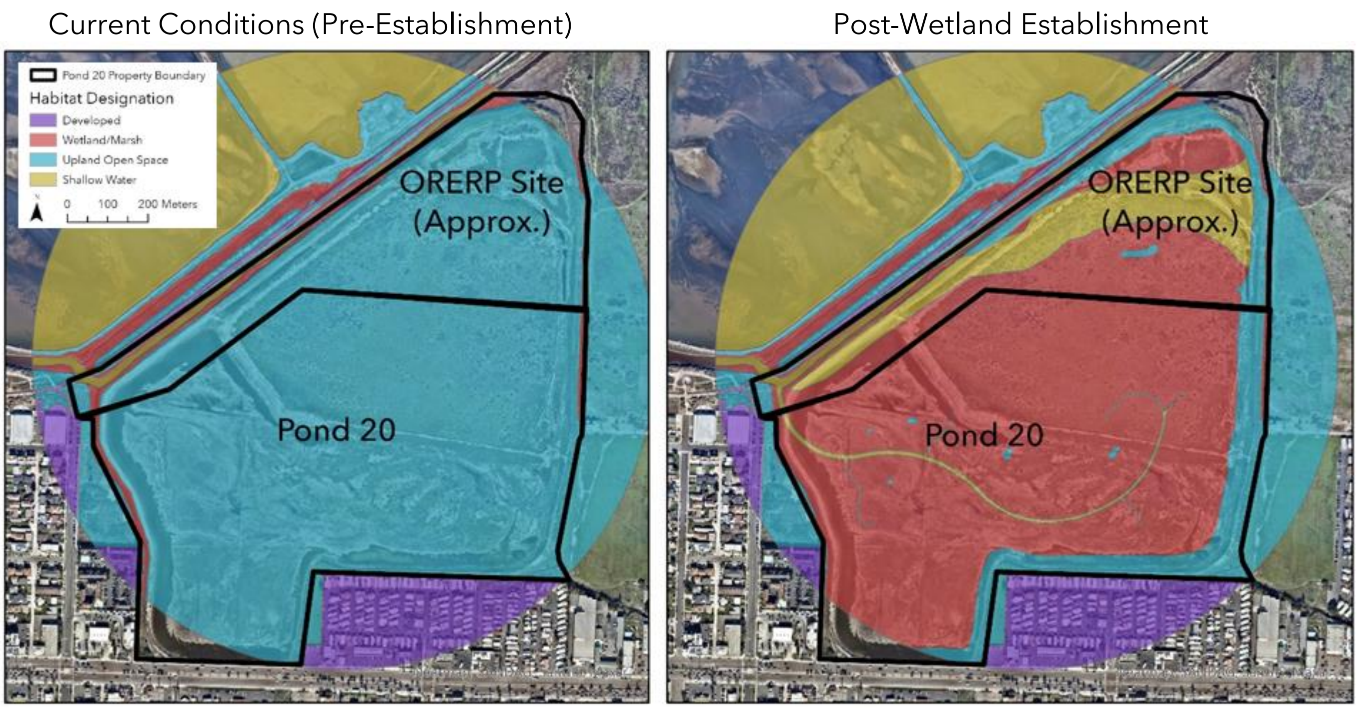
Project Question 3: What species and observation numbers should we expect to see upon the establishment of a 110-acre wetland mitigation bank?

The Port is planning on establishing a 110-acre wetland that will be used as a mitigation bank. To forecast changes in bird abundance upon the establishment of the wetland mitigation bank at Pond 20, we used the Poisson GLMM to predict bird observations assuming current and post-mitigation habitat matrix. To do this, we created a 500-m radius circle buffer around the approximate centroid of Pond 20, and a desktop delineation was completed using the most recent high-resolution satellite imagery available. Then, using the engineering design GIS files, as well as a georeferenced image to delineate the approximate north restoration site of Pond 20, part of the Otay River Estuary Restoration Project[7] (USFWS n.d.), the post-establishment conditions were assessed. The delineations for the pre- and post-Pond 20 wetland establishment are provided as Figure 2-8. The PC scores were then calculated for the Pond 20 site assuming these habitat values for pre- and post-wetland establishment. Lastly, we used the

[7] The Otay River Estuary Restoration Project is a combined effort with USFWS National Wildlife Refuges and Poseidon Water to restore the northern portion of Pond 20. This is not a part of the Port's proposed Pond 20 wetland mitigation bank, however, a Memorandum of Understanding (MOU) between the parties but will complement the efforts and result in the transformation of approximately 145 acres of salt pond to wetland habitat.

predict[8] function in R and the Poisson GLMM model output from Study Question 2 to estimate the observations expected if no mitigation was done versus if Pond 20 is improved as planned, giving us a quantitative measure of the change in bird numbers expected.

Figure 2-8. Habitat Delineations, Pre- and Post-Wetland Establishment at Pond 20 in South San Diego Bay.



[Project Question 4:] *Is using opportunistic data a valid substitute for systematic data collection to monitor bird trends in San Diego Bay? Are there detectable changes between both datasets? Are they consistent with one another?*

Determining population trends is critical for conservation and policy decisions. However, given resource constraints and the fact that these surveys are not a regulatory requirement, the structured Baywide surveys are not necessarily completed on the recommended scale of every three to five years. San Diego Bay is a popular bird-watching destination and as a result, has seen a large amount of opportunistic data from applications such as eBird and iNaturalist. Therefore, an important component of this study is to evaluate whether opportunistically collected data can be a substitute or supplement for the Baywide surveys.

To evaluate the differences between the two datasets, we ran the Poisson GLMM model from Study Question 2 separately on both the systematic and opportunistic datasets. Species selected for this analysis are the same as those provided in Table 2-4. The direction (positive vs. negative) of effects as

8 https://www.rdocumentation.org/packages/car/versions/3.1-1/topics/Predict

well as effect sizes (significantly different from zero versus not) were compared to understand the similarities and differences between the two datasets.

3. Results

Results for each of the four study questions are discussed below.

Project Question 1: Are the trends of birds using San Diego Bay increasing or decreasing over time?

To address the first study question, all eBird data observed within San Diego Bay were utilized, as discussed in Section 2. For each species within the eBird dataset[9], the predicted trend and its 95 percent confidence interval were plotted onto a graph, which allows for an overview of the general observation trend for each species from 2006 through 2021, as shown in Figure 3-1. Species that were selected for further analysis in this study are highlighted in light blue. Approximately 400 unique species were identified in this dataset.

Additionally, due to its position along the transition zone between cold subarctic waters and warmer subtropical water, San Diego Bay experiences a large variability in the structure of its bird communities throughout the seasons (TierraData 2018). To evaluate for differences in trends between seasons, we split the data by season and evaluated trends for each species during each season. As in Figure 3-1, species trends are plotted with the 95 percent confidence interval. These results are shown in Figures 3-2 through 3-5. As with Figure 3-1, species that were selected for further analysis are highlighted in light blue. Approximately 365 unique species were identified within the eBird data during the fall season; about 300 during both winter and spring; and around 200 unique species during the summer season.

[9] Note: a minimum of 150 data points was required for each species to be included in analysis.

Figure 3-1. Predicted Observation Trend for All Species

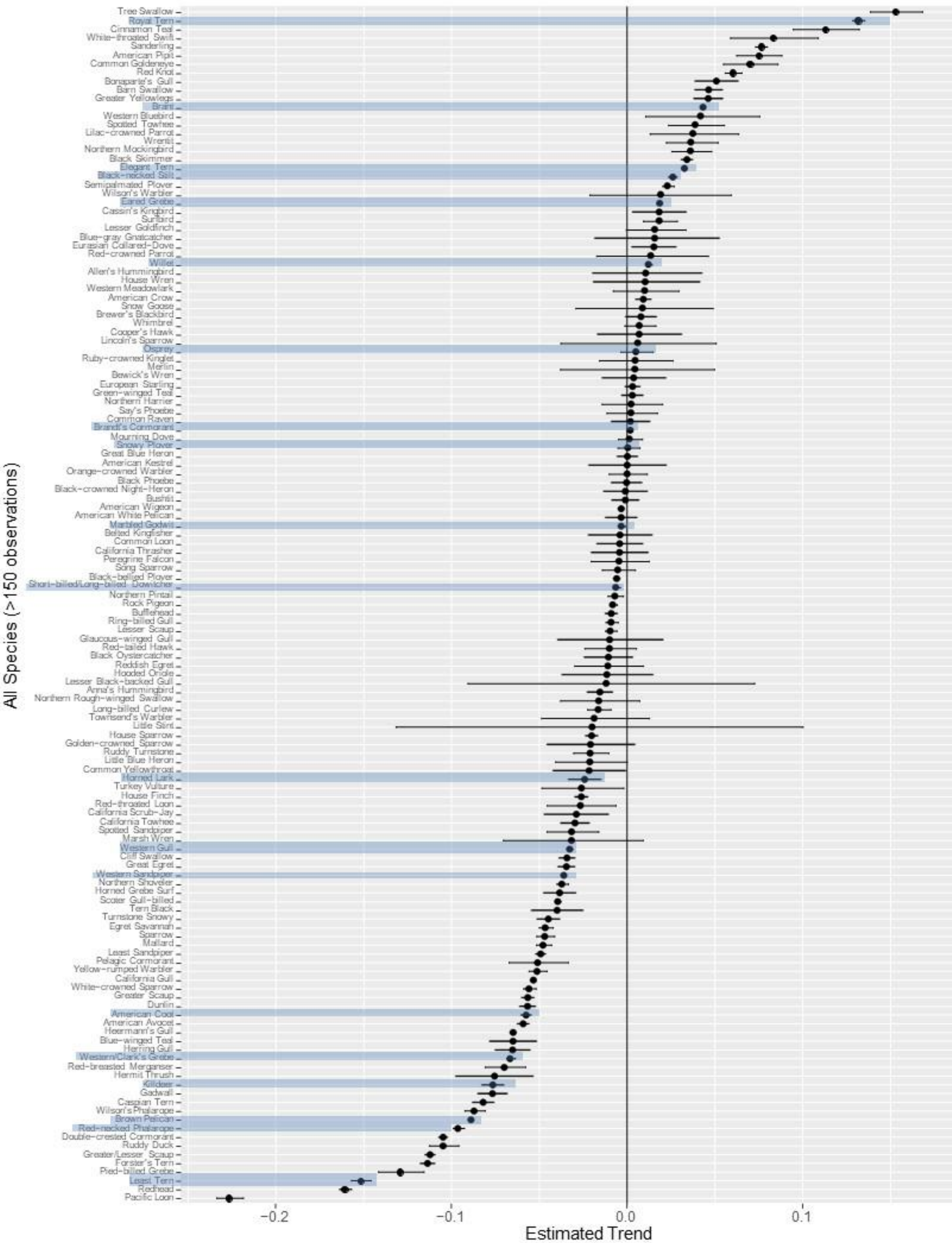


Figure 3-2. Predicted Observation Trend for All Species during Spring

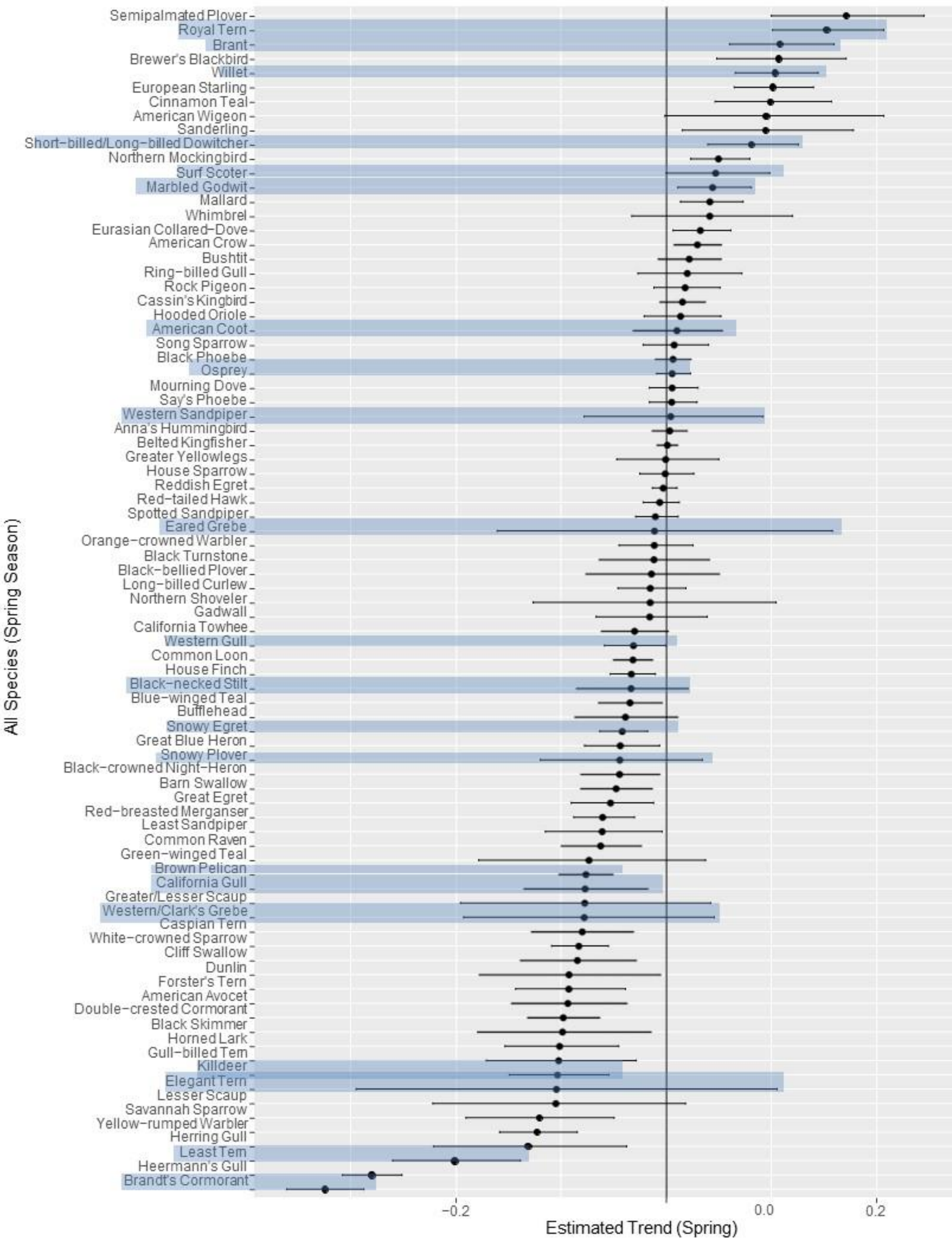
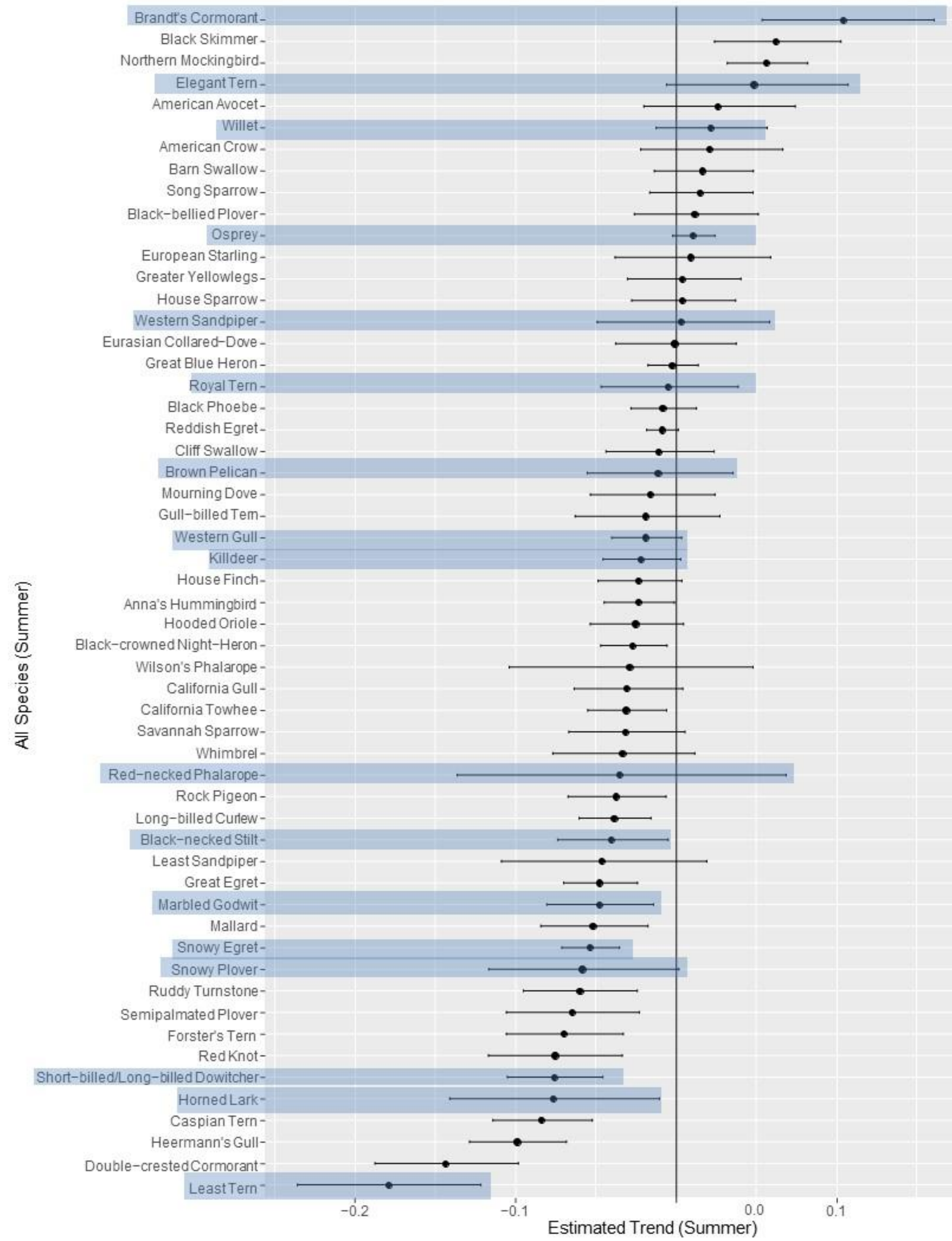


Figure 3-3. Predicted Observation Trend for All Species during Summer



This page intentionally left blank.

Figure 3-4. Predicted Observation Trend for All Species during Fall

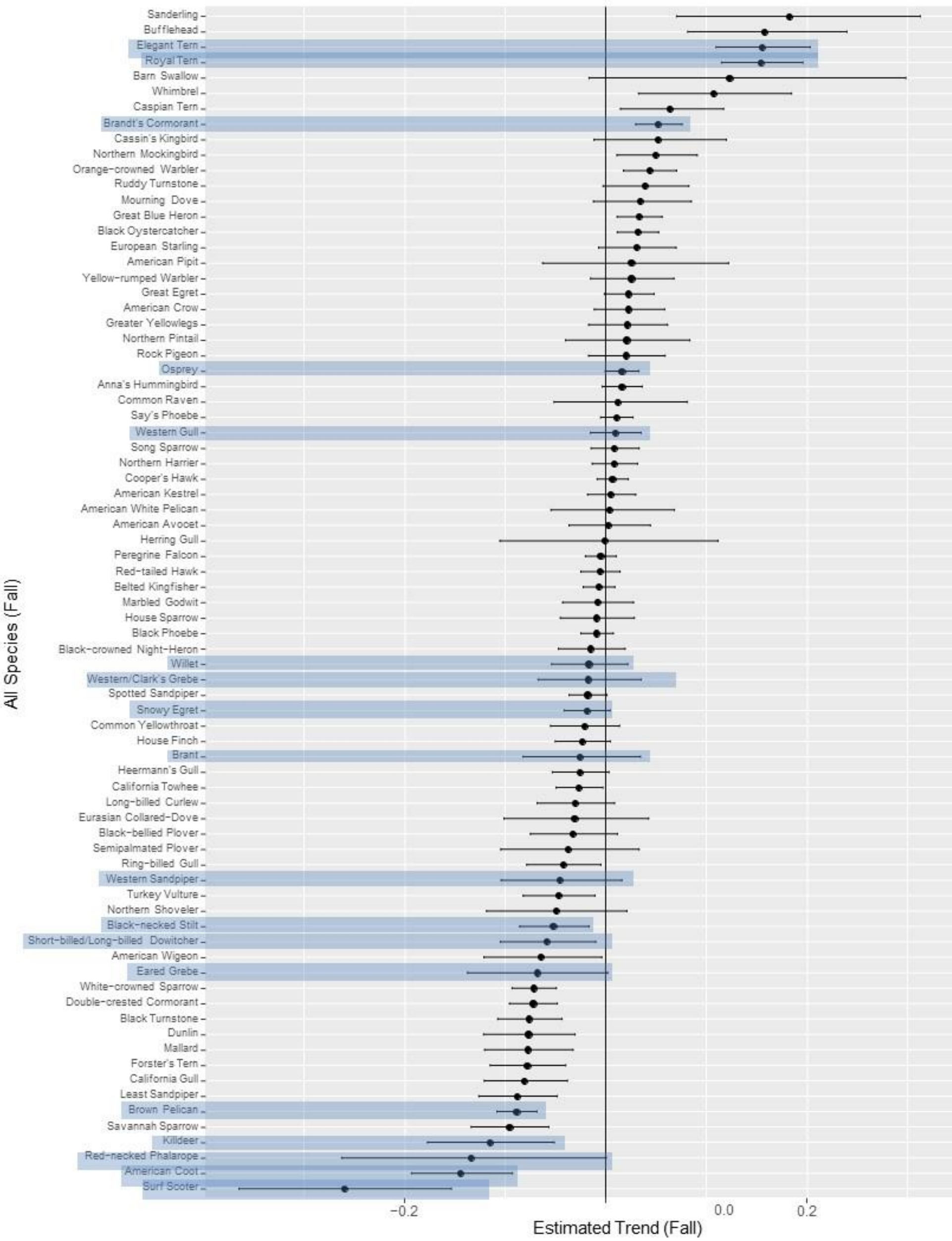
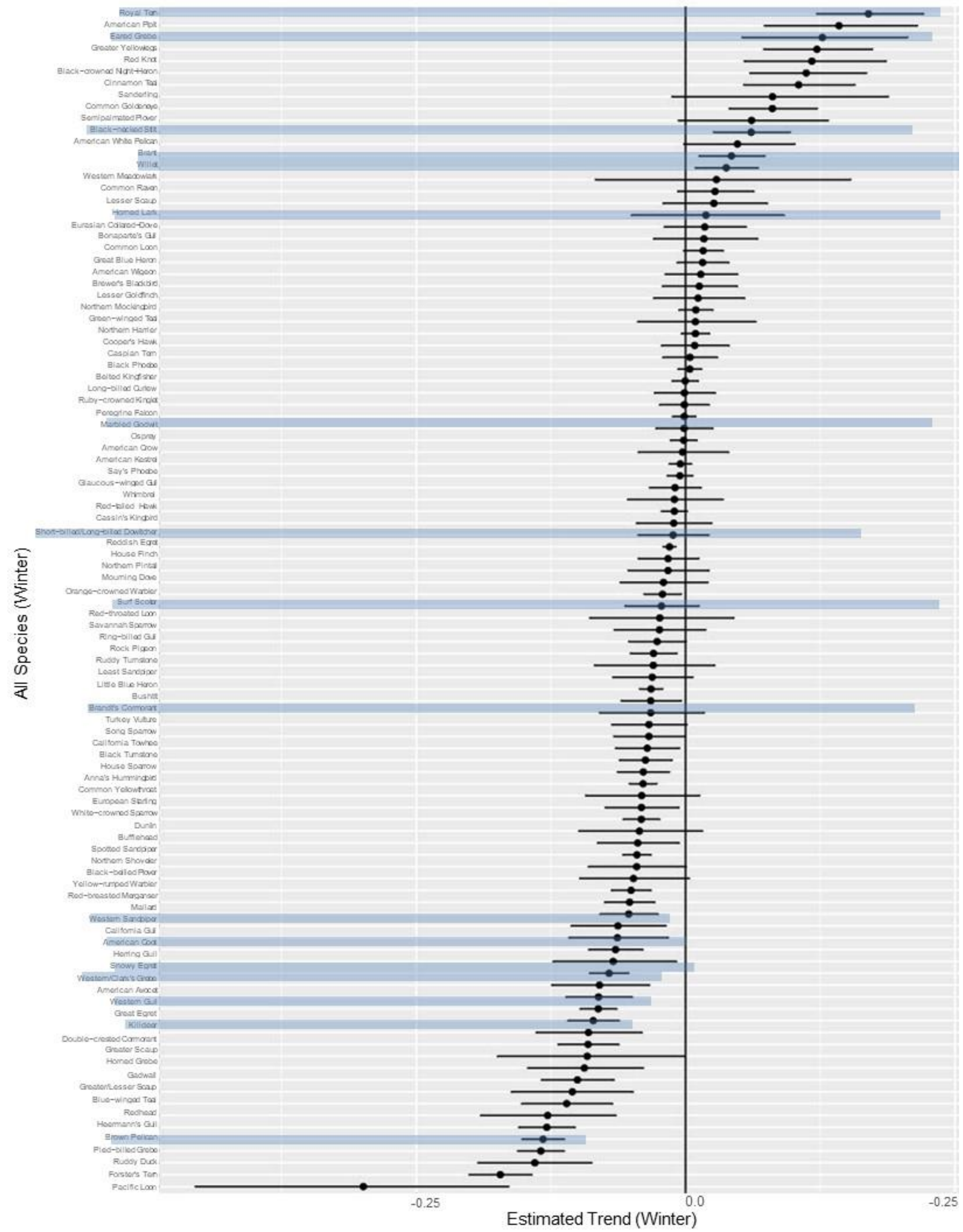


Figure 3-5. Predicted Observation Trend for All Species during Winter



This page intentionally left blank.

Project Question 2: Does land/water use affect bird populations in San Diego Bay?

The mixed effects model reveals a strong association between availability of wetland and shallow water availability and observations of several bird species, particularly wetland-depended species. The model also revealed an association between availability of beach and shallow water for several species. Lastly, the model revealed an overall decrease in predicted observations for some species over time, particularly within seabird and waterfowl assemblages.

Table 3-5. Model Results for Bird Observations ~ Time + PC1 + PC2

Species Assemblage	Species	Variable	Estimate (Trend)	Std. Error	Z-Value	P-Value	Model Interpretation
seabirds	brown pelican	*Year*	-0.05	0.05	-0.97	0.33	--
		PC1	0.06	0.09	0.62	0.54	--
		PC2	0.22	0.13	1.71	**0.09**	increased predicted observations as open water availability increases
	Brandt's cormorant	*Year*	0.34	0.20	1.69	**0.09**	increased predicted observations over time
		PC1	-0.33	0.21	-1.55	0.12	--
		PC2	0.24	0.23	1.05	0.29	--
	California least tern	*Year*	-0.09	0.04	-2.02	**0.04**	decreased predicted observations over time
		PC1	0.52	0.15	3.39	**<0.001**	increased predicted observations as wetland/shallow water availability increases
		PC2	0.32	0.18	1.82	**0.07**	increased predicted observations as open water/sandy beach availability increases
	elegant tern	*Year*	-0.56	0.24	-2.37	**0.02**	decreased predicted observations over time
		PC1	0.56	0.25	2.25	**0.02**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.35	0.27	-1.30	0.20	--
	royal tern	*Year*	0.06	0.06	0.97	0.33	increased predicted observations over time
		PC1	0.40	0.11	3.60	0.00	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.28	0.13	-2.15	0.03	decreased predicted observation as sandy beach/open water increases
	osprey	*Year*	0.01	0.01	0.92	0.36	--
		PC1	0.13	0.05	2.33	**0.02**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.01	0.07	-0.14	0.89	--
	western gull	*Year*	0.01	0.02	0.59	0.56	--
		PC1	0.05	0.12	0.45	0.66	--
		PC2	0.01	0.15	0.10	0.92	--
shorebirds	black-necked stilt	*Year*	0.06	0.02	3.66	**<0.001**	increased predicted observations over time
		PC1	0.08	0.35	0.23	0.82	--
		PC2	0.25	0.38	0.66	0.51	--
	dowitcher	*Year*	-0.01	0.04	-0.34	0.73	--
		PC1	0.29	0.13	2.23	**0.03**	increased predicted observations as wetland/shallow water availability increases

Species Assemblage	Species	Variable	Estimate (Trend)	Std. Error	Z-Value	P-Value	Model Interpretation
		PC2	-0.21	0.14	-1.48	0.14	--
shorebirds (cont'd)	killdeer	*Year*	-0.07	0.03	-1.97	**<u>0.05</u>**	increased predicted observations over time
		PC1	1.12	0.06	18.95	**<u><0.001</u>**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.84	0.08	-9.84	**<u><0.001</u>**	increased predicted observations as upland open space availability increases
	marbled godwit	*Year*	0.02	0.03	0.57	0.57	--
		PC1	0.26	0.11	2.36	**<u>0.02</u>**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.44	0.15	-2.99	**<u>0.003</u>**	increased predicted observations as upland open space availability increases
	western sandpiper	*Year*	0.09	0.10	0.92	0.36	--
		PC1	0.59	0.28	2.12	**<u>0.03</u>**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.35	0.35	-1.00	0.32	--
	western snowy plover	*Year*	-0.10	0.05	-1.93	**0.05**	increased predicted observations over time
		PC1	0.30	0.21	1.45	0.15	--
		PC2	-0.01	0.22	-0.06	0.95	--
	willet	*Year*	0.00	0.02	-0.14	0.89	--
		PC1	0.52	0.07	7.67	**<u><0.001</u>**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.52	0.09	-5.91	**<u><0.001</u>**	increased predicted observations as upland open space availability increases
marshbirds	snowy egret	*Year*	0.01	0.01	0.51	0.61	--
		PC1	0.03	0.03	0.92	0.36	--
		PC2	-0.13	0.04	-2.91	**<u>0.004</u>**	increased predicted observations as upland open space availability increases
waterfowl	American coot	*Year*	-0.05	0.02	-2.50	**<u>0.01</u>**	decreased predicted observations over time
		PC1	0.14	0.15	0.91	0.36	--
		PC2	-0.38	0.21	-1.78	**0.08**	increased predicted observations as upland open space availability increases
	brant	*Year*	-0.10	0.05	-1.93	**0.05**	decreased predicted observations over time
		PC1	0.30	0.21	1.45	0.15	--
		PC2	-0.01	0.22	-0.06	0.95	--
	Clark's/ western grebe	*Year*	-0.29	0.53	-0.55	0.58	--
		PC1	0.40	0.42	0.95	0.34	--
		PC2	0.43	0.57	0.77	0.44	--
	eared grebe	*Year*	-0.10	0	-2.20	**<u>0.03</u>**	increased predicted observations over time
		PC1	1	0	5.48	**<u><0.001</u>**	increased predicted observations as wetland/shallow water availability increases
		PC2	0	0	-0.03	0.97	--
	surf scoter	*Year*	-0.02	0.08	-0.29	0.77	--

Species Assemblage	Species	Variable	Estimate (Trend)	Std. Error	Z-Value	P-Value	Model Interpretation
		PC1	0.17	0.17	0.98	0.33	--
		PC2	0.29	0.23	1.28	0.20	--
terrestrial	horned lark	*Year*	-0.04	0.06	-0.71	0.48	--
		PC1	-0.11	0.04	-2.58	**<u>0.01</u>**	increased predicted observations as developed area increases
		PC2	0.18	0.03	6.23	**<u><0.001</u>**	increased predicted observation as sandy beach/open water increases

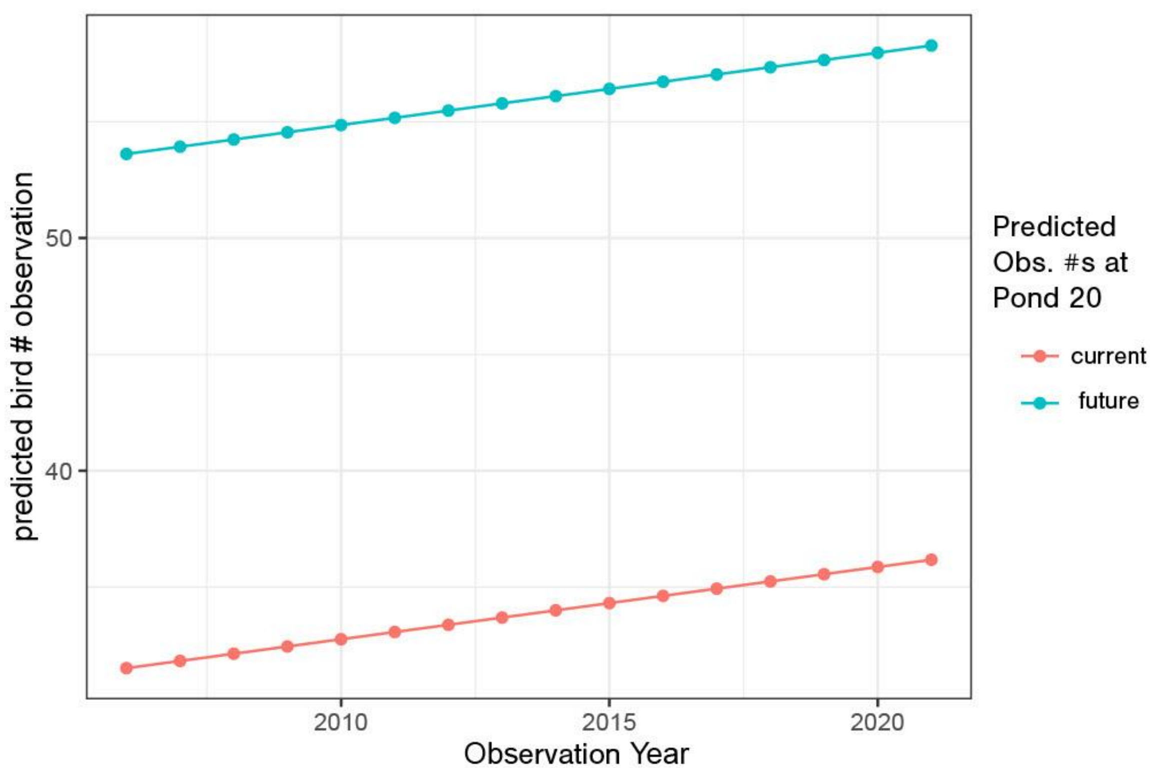
Notes: p-values **underlined in bold** indicate a significance level of <0.05. P-values in **bold** indicate a lesser significance level (<0.1, but greater than 0.05).

Project Question 3: What species and observation numbers should we expect to see upon the establishment of a 110-acre wetland mitigation bank?

To address this question, we utilized the GLMM model from Study Question 2 to compare predicted outcomes of no change to Pond 20 (i.e., current conditions) versus post-wetland establishment. The model results are shown in Figure 3-6.

As compared to the current conditions (red line shown in Figure 3-6), the model indicates an estimated increase in bird observations over time post-wetland establishment (blue line). The model, on average, approximates a 67 percent increase in observations as compared to the current conditions.

Figure 3-6. Pre- versus Post-Pond 20 Wetland Establishment – Predicted Bird Observations



Project Question 4: Is using opportunistic data a valid substitute for systematic data collection to monitor bird trends in San Diego Bay? Are there detectable changes between both datasets? Are they consistent with one another?

Determining population trends is critical for conservation and policy decisions. However, given resource constraints and the fact that these surveys are not a regulatory requirement, the structured Baywide surveys are not necessarily completed on the recommended scale of every three to five years. Additionally, will the increased collection of opportunistic data, citizen science data have enormous potential to advance scientific knowledge, influence policy, and guide resource management decisions (Kosmala and others 2016).

The results from running the Baywide survey data for select species through the GLMM created to address Study Question 2 are summarized in Table 3-6.

Table 3-6. Model Results Using the Baywide Survey Data from 2006-2007, 2009-2010, and 2016-2017 Survey Years.

Species Assemblage	Species	Variable	Estimate (Trend)	Std. Error	Z-Value	P-Value	Model Interpretation
seabirds	brown pelican	*Year*	-0.07	0.03	-2.36	**0.02**	decreased predicted observations over time
		PC1	0.26	0.09	2.81	**0.01**	increased predicted observations as wetland/shallow water availability increases
		PC2	0.12	0.12	0.99	0.32	--
	Brandt's cormorant	*Year*	NA	NA	NA	NA	not enough data
		PC1	NA	NA	NA	NA	not enough data
		PC2	NA	NA	NA	NA	not enough data
	CA least tern	*Year*	-0.01	0.07	-0.11	0.92	--
		PC1	0.14	0.10	1.35	0.18	--
		PC2	-0.48	0.23	-2.06	**0.04**	decreased predicted observations as wetland/shallow water availability increases
	elegant tern	*Year*	0.04	0.04	0.98	0.33	--
		PC1	1.11	0.28	3.91	**<0.001**	increased predicted observations as wetland/shallow water availability increases
		PC2	0.14	0.60	0.24	0.81	--
	royal tern	*Year*	0.07	0.06	1.23	0.22	--
		PC1	-0.48	0.25	-1.90	**0.06**	decreased predicted observations as wetland/shallow water availability increases
		PC2	1.58	0.34	4.67	**<0.001**	
	osprey	*Year*	-0.01	0.02	-0.86	0.39	--
		PC1	0.07	0.06	1.16	0.25	--
		PC2	0.00	0.07	-0.03	0.97	--
	western gull	*Year*	-0.06	0.03	-2.52	**0.01**	decreased predicted observations over time
		PC1	0.05	0.13	0.39	0.70	--
		PC2	-0.06	0.18	-0.34	0.74	--
shorebirds	black-necked stilt	*Year*	0.02	0.02	1.09	0.28	--
		PC1	0.28	0.09	3.11	**0.002**	increased predicted observations as wetland/shallow water availability increases

Species Assemblage	Species	Variable	Estimate (Trend)	Std. Error	Z-Value	P-Value	Model Interpretation
		PC2	NA	NA	NA	NA	not enough data
shorebirds (cont'd)	dowitcher	*Year*	-0.02	0.04	-0.49	0.63	--
		PC1	-0.11	0.20	-0.54	0.59	--
		PC2	-0.19	0.16	-1.24	0.22	--
	killdeer	*Year*	-0.03	0.03	-1.04	0.30	--
		PC1	-0.29	0.29	-0.98	0.33	--
		PC2	0.07	0.14	0.51	0.61	--
	marbled godwit	*Year*	-0.08	0.02	-4.04	**<0.001**	decreased predicted observations over time
		PC1	0.33	0.09	3.53	**<0.001**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.34	0.10	-3.36	**<0.001**	decrease in predicted observation as sandy beach/open water increases
	western sandpiper	*Year*	-0.02	0.04	-0.48	0.63	--
		PC1	0.90	0.64	1.41	0.16	--
		PC2	-0.58	0.51	-1.15	0.25	--
	western snowy plover	*Year*	-0.01	0.08	-0.10	0.92	--
		PC1	-0.61	0.32	-1.90	**0.06**	decreased predicted observations as wetland/shallow water availability increases
		PC2	-0.31	0.04	-8.97	**<0.001**	decreased predicted observation as sandy beach/open water increases
	willet	*Year*	-0.14	0.04	-3.43	**<0.001**	decreased predicted observations over time
		PC1	0.30	0.13	2.29	**0.02**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.59	0.16	-3.70	**<0.001**	decreased predicted observation as sandy beach/open water increases
marshbirds	snowy egret	*Year*	-0.04	0.01	-2.89	**<0.001**	
		PC1	0.10	0.06	1.61	0.11	--
		PC2	-0.11	0.08	-1.29	0.20	--
waterfowl	American coot	*Year*	-0.08	0.06	-1.20	0.23	--
		PC1	0.73	0.30	2.48	**0.01**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.11	0.29	-0.39	0.70	--
	brant	*Year*	0.04	0.05	0.66	0.51	--
		PC1	0.69	0.20	3.47	**<0.001**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.28	0.11	-2.53	**0.01**	decreased predicted observation as sandy beach/open water increases
	western/ clark's grebe	*Year*	0.02	0.03	0.60	0.55	--
		PC1	-0.06	0.10	-0.56	0.58	--
		PC2	-0.43	0.17	-2.57	**0.01**	decreased predicted observation as sandy beach/open water increases
	eared grebe	*Year*	0.00	0.04	0.01	0.99	--
		PC1	0.60	0.23	2.59	**0.01**	increased predicted observations as wetland/shallow water availability increases
		PC2	-0.15	0.39	-0.40	0.69	--

Species Assemblage	Species	Variable	Estimate (Trend)	Std. Error	Z-Value	P-Value	Model Interpretation
waterfowl (cont'd)	surf scoter	*Year*	0.01	0.03	0.37	0.71	--
		PC1	0.47	0.26	1.79	**0.07**	increased predicted observations as wetland/shallow water availability increases
		PC2	0.05	0.37	0.15	0.88	--
	red-necked phalarope	*Year*	-0.08	0.00	-18.14	**<ins><0.001</ins>**	decreased predicted observations over time
		PC1	1.64	0.20	8.12	**<ins><0.001</ins>**	increased predicted observations as wetland/shallow water availability increases
		PC2	NA	NA	NA	NA	not enough data
terrestrial	horned lark	*Year*	-0.11	0.02	-6.22	**<ins><0.001</ins>**	decreased predicted observations over time
		PC1	-0.63	0.21	-2.99	**<ins>0.003</ins>**	increased predicted observations as development/overwater structures increase
		PC2	-0.23	0.09	-2.46	**<ins>0.01</ins>**	decreased predicted observation as sandy beach/open water increases

Notes: NA = not available; p-values **<ins>underlined in bold</ins>** indicate a significance level of <0.05. P-values in **bold** indicate a lesser significance level (<0.1, but greater than 0.05).

To compare the two datasets, the resultant coefficients for each model run (i.e., each bird species) were compared. The results of these comparisons are provided in Figure 3-7; for each coefficient and each species, if both the models were in agreement (i.e., both observed a significant effect, or neither observed any effect) are noted by a green box; if the model results for both datasets were in disagreement (i.e., either the eBird data observed an effect and the Baywide survey did not, or the other way around, or both datasets observed a significant effect with contradictory trends), the cell was noted in yellow. Overall, the datasets were in agreement for about half (46 percent) of the species analyzed. For the year coefficient, the eBird data and Baywide survey model results were in agreement for 35 percent of species analyzed; for the PC1 coefficient, the datasets were in agreement for about half (45 percent) of the species analyzed; and lastly, PC2 showed about 58 percent agreement for species analyzed.

Figure 3-7. Model Output Comparison for Opportunistic (eBird) and Systematic (Baywide) Datasets

Coeff	Comparison	black-necked stilt	brant	brown pelican	CA least tern	American coot	dowitcher	eared grebe	elegant tern	horned lark	killdeer	marbled godwit	osprey	red-necked phalarope	royal tern	snowy egret	snowy plover	surf scoter	western gull	western/ clark's grebe	western sandpiper	willet
Year	agreement (effect present)																					
	one dataset shows effect or trend disagreement	x		x	x	x		x		x	x	x		x	x	x			x			x
	agreement (no effect)		x				x						x				x	x		x	x	
PC1	agreement (effect present)								x	x		x		x								x
	one dataset shows effect or trend disagreement	x	x	x	x	x	x	x			x		x		x	x					x	
	agreement (no effect)																x	x	x	x		
PC2	agreement (effect present)											x			x							x
	one dataset shows effect or trend disagreement		x		x					x	x					x	x			x		
	agreement (no effect)			x		x	x	x	x				x					x	x		x	

Note: Each coefficient and each species are provided in the summary table above. Models that were in agreement (i.e., both eBird and baywide data observed a significant effect, or neither observed any effect) are noted by a green box; models that were in disagreement (i.e., either the eBird data observed an effect and the Baywide survey did not, or the other way around, or both datasets observed a significant effect with contradictory trends) are noted in yellow.

4. Discussion

The results of each study question, as well as study limitations are discussed herein.

Project Question 1: Are the trends of birds using San Diego Bay increasing or decreasing over time?

Based on the model outputs presented in Figure 3-1, our results reveal that the trends over time of bird observations of the eBird dataset are species dependent. Looking at the dataset as a whole (i.e., Figure 3-1), we see that 20 percent of all species showed a significant positive effect (i.e., its estimate was positive and the confidence interval did not overlap zero). Conversely, about 43 percent of all species showed a significant negative significant effect. Approximately 37 percent of species showed no significant effect. We also evaluated the eBird data season-by-season. Overall, we saw a decrease in species richness during the spring season, an increase in richness during the winter and spring seasons, and the greatest amount of richness in the fall months, which is expected given that San Diego Bay is a refuge for migratory species.

Over 400 species were captured in the eBird dataset, including nearly all of the known 300 species that utilize the San Diego Bay tidelands according to the Baywide survey results (TierraData 2018). However, opportunistically collected data tends to favor species that are more ubiquitous. This type of data collection does not target bird species with more secretive behavior, such as marsh birds like rails. Additionally, there are several locations that are difficult for the public to see, even though there are known roosting or high use areas. These include the entire extent of areas within the salt ponds (access is limited to the road and the Bayshore bikeway), nesting sites that are isolated from public access, and others.

Project Question 2: Does land/water use affect bird populations in San Diego Bay?

As with the first project question, results on whether habitat/land and/or water use affected bird observations in San Diego Bay varies by species. For the species analyzed, the GLMM revealed that the PC1 (positively correlated with presence of wetland and shallow waters, negatively correlated with development including riprap and overwater structures) had a significant effect on twelve of the 22 species that were analyzed. Of these species, over 83 percent showed a positive response to the increase of wetland and shallow water habitats. The GLM revealed that PC2 (positively correlated with deep water and sandy beach, negatively correlated with open space) nine of 22 species showed a significant effect. Of these nine species, six showed a negative response to the increase of open water and/or sandy beach.

Most of these results are not surprising – most of the shorebirds analyzed, such as the willet, marbled godwit, red-necked phalarope, and killdeer showed a positive response to PC1, which makes sense for wetland-dependent species. Similarly, species that forage in shallow waters, such as tern species (like the endangered California least tern, royal tern, and elegant tern) showed a positive response to PC1. Similarly, wetland-dependent shorebirds and marshbirds, such as the willet, marbled godwit, and snowy egret showed a negative response to PC2. This makes sense because areas with open (deep) water and sandy beaches typically do not coincide with marshy areas. The California least tern responded positively to PC2, which may indicate more access to prey, however, other tern species either showed no effect (elegant tern) or a negative effect (royal tern). Surprisingly, PC2 had no effect on species known to occupy sandy beaches, including western snowy plovers.

Project Question 3: What species and observation numbers should we expect to see upon the establishment of a 110-acre wetland mitigation bank?

The predicted number of observations is expected to increase as well based on the results of our GLMM. However, while there is a predicted increase in bird observations based on the model, the actual number may be better predicted using better more refined habitat models and dedicated bird surveys within the Pond 20 area. Additionally, based on the results from Study Question 2, the establishment of a wetland and shallow water is likely to bring in species that are positively associated with PC1.

Further studies may want to explore specific species that are expected to occupy the newly established wetland habitat at Pond 20. Such studies could look at monitoring data from nearby wetland establishment and/or restoration projects, such as the establishment of tidal influence to Ponds 10 and 11[10], or evaluate species of nearby salt marshes, such as the Tijuana River Estuary National Estuarine Research Reserve[11]. It may also be beneficial to evaluate how the establishment of a wetland mitigation at Pond 20 may impact the overall bird population trends of San Diego Bay, as our study only evaluated the predicted change within the Pond 20 area.

Project Question 4: Is using opportunistic data a valid substitute for systematic data collection to monitor bird trends in San Diego Bay? Are there detectable changes between both datasets? Are they consistent with one another?

Land managers and restoration project managers rely on monitoring programs to assess spatial and temporal biodiversity trends. However, the cost of ensuring sufficient spatial and temporal coverage is

[10] https://www.fws.gov/story/salt-pond-restoration-san-diego-bay-nwr
[11] https://trnerr.org/

often a limitation for those monitoring such trends (Callaghan and Gawlik 2015). Therefore, the incorporation of citizen science data, including eBird, which has broad spatial and temporal coverage, has important implications for avian conservation.

However, when it came to similarities in model outputs for the two datasets, the results were a mixed bag. The eBird data and Baywide survey data aligned approximately 46 percent of the time; about 12 percent were instances where both datasets showed a significant effect; 34 percent were instances where both datasets showed no effect at all to the variable.

At the Baywide level, opportunistic data may be more useful in determining overall trends for more ubiquitous species. For example, as shown in Figure 2-2, observer efforts are clustered throughout the Bay. On a smaller scale, this could introduce sampling bias for smaller scale studies, however, Baywide inferences can be drawn using multiple popular points. Furthermore, they cannot replace species-specific surveys, such as clapper rail surveys and surveys in areas that are inaccessible to the public (i.e., most of the south San Diego Bay salt ponds, the Chula Vista Wildlife Reserve, and others).

Another advantage to incorporating opportunistic data is to get a better understanding on the status of a particular species' population on a larger scale, not just as a function of the use of San Diego Bay as habitat. The ability to compare local data to larger, regional datasets allows the ability to better assess the successes of local efforts, such as habitat restoration. This also allows the ability to identify species to focus on for conservation efforts.

5. Conclusions

In this study, we show that large volumes of opportunistic eBird data can approximate bird population trend estimates, however, valid estimations may be limited to more widely spread species given that San Diego Bay has large numbers of birds that take refuge in areas that are mostly inaccessible to the public. In part because of this, eBird data may be used to fill data gaps for some, but not all bird species, and therefore systematic surveys should still be completed on the recommended three-to-five-year basis. We saw that there is indeed a connection between habitat and bird occurrence, and that the presence of wetland habitat and shallow waters generally has a positive impact on many seabird, waterfowl, and shorebird species. The presence of sandy beach, open water, and undeveloped upland also supports several seabird, shorebird, marshbird, and waterfowl species. And lastly, based on our analysis of habitat and eBird data, we expect that the establishment of 110-acres of wetland habitat in south San Diego Bay will result in about a 60 percent increase in total bird observations within the area.

6. References

Callaghan, C.T. and Gawlik, D.E. (2015). Efficacy of eBird data as an aid in conservation planning and monitoring. *J. Field Ornithol.* 2015; 86(4):298–304. doi: 10.1111/jofo.12121

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. New York: Cambridge Press.

Clark, C.J. (2017). eBird records show substantial growth of the Allen's Hummingbird (*Selasphorus sasin sedentarius*) population in urban Southern California. *American Ornithological Society*. 2017; 119:122-130. doi: 10.1650/CONDOR-16-153.1

Desrochers, D. W., Keagy, J. C., & Cristol, D. A. (2008). Created versus natural wetlands: Avian communities in Virginia salt marshes. *Ecoscience*, 15(1), 36–43. https://doi.org/10.2980/1195-6860(2008)15[36:CVNWAC]2.0.CO;2

Horns, J. J., Adler, F. R., & Şekercioğlu, Ç. H. (2018). Using opportunistic citizen science data to estimate avian population trends*. Biological Conservation,* 221, 151–159. https://doi.org/10.1016/j.biocon.2018.02.027

Hostetler, M., & Knowles-Yanez, K. (2003). Land use, scale, and bird distributions in the Phoenix metropolitan area. *Landscape and Urban Planning*, 62(2), 55–68. https://doi.org/10.1016/S0169-2046(02)00096-8

Howe, M.A., Geissler, P.H., & Harrington, B.A. (1989). Population trends of North American shorebirds based on the International Shorebird Survey. *Biological Conservation*, 49(3): 185-199. https://doi.org/10.1016/0006-3207(89)90035-9

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T., & Donald, P. F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, 22(10), 1024–1035. https://doi.org/10.1111/ddi.12463

Kosmala, M., Wiggins, A., Swanson, Simmons, B. (2016). Assessing data quality in citizen science. *Front Ecol Environ* 2016; 14(10): 551– 560. doi:10.1002/fee.1436

Strimas-Mackey, M., W.M. Hochachka, V. Ruiz-Gutierrez, O.J. Robinson, E.T. Miller, T. Auer, S. Kelling, D. Fink, & Johnston, A. (2020). *Best Practices for Using eBird Data*. Version 1.0. Cornell Lab of Ornithology, Ithaca, New York. https://doi.org/10.5281/zenodo.3620739

TierraData Inc. (2009). *2006-2007 San Diego Bay Avian Species Surveys*. Retrieved on October 18, 2021 from https://pantheonstorage.blob.core.windows.net/environment/2006-07_San-Diego-Bay-Avian-Species-Surveys_April2009.pdf

TierraData Inc. (2011). *2009-2010 San Diego Bay Avian Species Surveys*. Retrieved on October 18, 2021 from https://pantheonstorage.blob.core.windows.net/environment/2009-10_San-Diego-Bay-Avian-Species-Surveys_May2011.pdf

TierraData Inc. (2018). *2016-2017 San Diego Bay Avian Species Surveys*. Retrieved on October 18, 2021 from https://pantheonstorage.blob.core.windows.net/environment/2016-17-San-Diego-Bay-Avian-Species-Surveys-April-2018.pdf

U.S. Department of the Navy, Naval Facilities Engineering Command Southwest (Navy) and Port of San Diego (Port) (2013). San Diego Bay Integrated Natural Resources Management Plan [INRMP], Final September 2013. San Diego, California. Prepared by Tierra Data Inc., Escondido, California.

United States Fish and Wildlife Service [USFWS] (n.d.) *Otay River Estuary Restoration Project*. Retrieved November 10, 2022 from https://www.fws.gov/story/otay-river-estuary-restoration-project