CONTRIBUTED PAPERS

An impact evaluation of conservation investments targeting long-distance migratory species

C. Josh Donlan[1,2] | Diana Eusse-González[3] | Gloria M. Luque[1] | Matthew E. Reiter[4] |
Viviana Ruiz-Gutierrez[2] | Michael C. Allen[5] | Richard Johnston-González[3] |
Orin J. Robinson[2] | Guillermo Fernández[6] | Eduardo Palacios[7] | Jorge Valenzuela[8]

[1]Advanced Conservation Strategies, Midway, Utah, USA

[2]Cornell Laboratory of Ornithology, Ithaca, New York, USA

[3]Asociación para el Estudio y Conservación de las Aves Acuáticas en Colombia (Calidris), Cali, Colombia

[4]Point Blue Conservation Science, Petaluma, California, USA

[5]Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, New Jersey, USA

[6]Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México, Mazatlán, México

[7]Centro de Investigación Científica y de Educación Superior de Ensenada – Unidad La Paz, La Paz, México

[8]Centro de Estudio y Conservación del Patrimonio Natural (CECPAN), Ancud, Chile

Correspondence
C. Josh Donlan, Advanced Conservation Strategies, PO Box 413 Midway, 84049, Utah, USA. Email: jdonlan@advancedconservation.org

Article Impact Statement: An impact evaluation reveals the complexity of protecting migratory species and the importance of counterfactuals to assess impact.

Abstract

We evaluated the impact of a philanthropic program investing in the conservation of sites along the Pacific Americas Flyway, which spans >16,000 km of coastline and is used by millions of shorebirds. Using a quasi-experimental, mixed methods approach, we estimated what would have happened to shorebird populations at 17 wintering sites without the sustained and additional investment they received. We modeled shorebird populations across the entire flyway and at sites with and without investment. Combining shorebird abundance estimates with a land-cover classification model, we used the synthetic control method to create counterfactuals for shorebird trends at the treatment sites. We found no evidence of an overall effect across three outcome variables. Species- and site-level treatment effects were heterogeneous, with a few cases showing evidence of a positive effect, including a site with a high level of overall investment. Results suggest six shorebirds declined across the entire flyway, including at many Latin American sites. However, the percentage of flyway populations present at the sites remained stable, and the percentage at the treatment sites was higher (i.e., investment sites) than at control sites. Multiple mechanisms behind our results are possible, including that investments have yet to mitigate impacts and negative impacts at other sites are driving declines at the treatment sites. A limitation of our evaluation is the sole focus on shorebird abundance and the lack of data that prohibits the inclusion of other outcome variables. Monitoring infrastructure is now in place to design a more robust and a priori shorebird evaluation framework across the entire flyway. With this framework, it will prove easier to prioritize limited dollars to result in the most positive conservation outcomes.

KEYWORDS
counterfactual, impact evaluation, shorebirds, synthetic control method, Pacific Americas Flyway

Evaluación del impacto de la inversión para la conservación enfocada en especies migratorias de largo recorrido
Resumen: Evaluamos el impacto de un programa filantrópico que invierte en la conservación de sitios a lo largo de la Ruta Migratoria Pacífico-Américas, la cual abarca >16,000 km de la línea costera y millones de aves playeras la usan. Estimamos con una estrategia cuasiexperimental y de métodos mixtos lo que habría pasado con las poblaciones de estas aves en 17 sitios invernales sin la inversión adicional y continua que recibieron. Modelamos estas poblaciones en toda la ruta y en sitios con y sin inversión. Combinamos

las estimaciones de aves playeras con el modelo de clasificación de la cobertura del suelo y usamos el método de control sintético para crear contrafactuales para las tendencias de las aves playeras en sitios de tratamiento. No encontramos evidencia alguna de un efecto generalizado en las tres variables de los resultados. Los efectos del tratamiento de especies y de sitio fueron heterogéneos, con unos cuantos casos que mostraron evidencia de un efecto positivo, incluido un sitio con un nivel elevado de inversión general. Los resultados sugieren que seis especies de aves playeras declinaron a lo largo de toda la ruta, incluyendo en varios sitios de América Latina. Sin embargo, el porcentaje de poblaciones de la ruta presentes en los sitios permaneció estable y el porcentaje en los sitios de tratamiento (sitios de inversión) fue más elevado que en los sitios control. Muchos mecanismos son posibles detrás de nuestros resultados, incluidas las inversiones que todavía no han mitigado impactos y los impactos negativos en otros sitios que están causando las declinaciones en los sitios de tratamiento. Una limitación en nuestra evaluación es el enfoque único en la abundancia de aves playeras y la falta de datos que impiden la inclusión de otras variables de los resultados. El monitoreo de la infraestructura ahora está en una posición en la que puede diseñar un marco de evaluación más robusto y a priori de las aves playeras a lo largo de toda la ruta. Con este marco, será más fácil priorizar los dólares limitados para que los resultados de conservación sean lo más positivos posible.

INTRODUCTION

For over a decade, there have been calls for a greater focus on the rigorous evaluation of investments for environmental conservation (Ferraro & Pattanayak, 2006). Many have advocated using impact evaluations and counterfactual frameworks (Adams et al., 2019; Ferraro & Pattanayak, 2006). Impact evaluation goes beyond program monitoring; instead, it attempts to measure the causal effect of an intervention using a credible counterfactual scenario and seeks to understand the conditions under which any effect arises (Ferraro & Hanauer, 2014). The environmental sector has been slower to widely adopt impact evaluation approaches compared with others. There are many reasons for this, including that conservation interventions often target multiple outcomes, are susceptible to spatial spillovers or leakage, and are often not amenable to randomized research designs (Baylis et al., 2016). However, these challenges are not unique to environmental conservation and have not stopped researchers and practitioners from conducting impact evaluations in other sectors. Many opportunities exist to apply counterfactual thinking and impact evaluation methodologies to environmental policies and programs (Ferraro & Pattanayak, 2006; Wauchope et al., 2022).

We evaluated the impact of a philanthropic program investing in the conservation of shorebirds and their wintering habitat along the Pacific Americas Flyway (hereafter, flyway). The flyway spans >16,000 km of coastline between Alaska and Chile. Millions of shorebirds use it biannually, moving between breeding and nonbreeding grounds and back again. Across 13 countries, 170 priority sites along the flyway have been identified (Senner et al., 2016). Shorebirds have life-history traits that make them vulnerable to population declines. First, they

rely on coastal and interior wetland habitats susceptible to loss, degradation, and the effects of climate change (Donnelly et al., 2020). Second, they have high site fidelity, often depending on a few breeding, stopover, and wintering sites (Gibson et al., 2018). Third, shorebirds tend to have low reproductive rates and high adult survivorship and thus are sensitive to factors that increase adult mortality (Myers et al., 1987). As a group, shorebirds are showing evidence of declines in the Western Hemisphere (Rosenberg et al., 2019; Smith et al., 2023). Threats to shorebirds along the flyway have been identified and include climate change, development, invasive species, human disturbance, water use and management, aquaculture, and habitat modification (Senner et al., 2016).

The David and Lucile Packard Foundation (2019) (hereafter, foundation) had been investing in shorebird conservation along the flyway for over a decade. Investments supporting nongovernmental organizations (NGOs) working on shorebird conservation have focused on 17 sites in Latin America that provide important wintering habitat (hereafter, treatment sites) (Figure 1). Although the foci of investments have evolved since the program's inception, the foundation has been investing in NGOs working at these sites since 2006 (Figure 1). The theory of change related to these investments is based on three main assumptions (Figure 2). Although their investment is approaching US$10 million, the foundation recognizes that building capacity requires long-term investments and that pressures from multiple threats are significant—both factors that challenge the theory of change.

Using a counterfactual framework, we sought insights into what may have happened to shorebird populations at the 17 treatment sites without the intervention (i.e., without sustained and additional conservation investment). We focused on a single
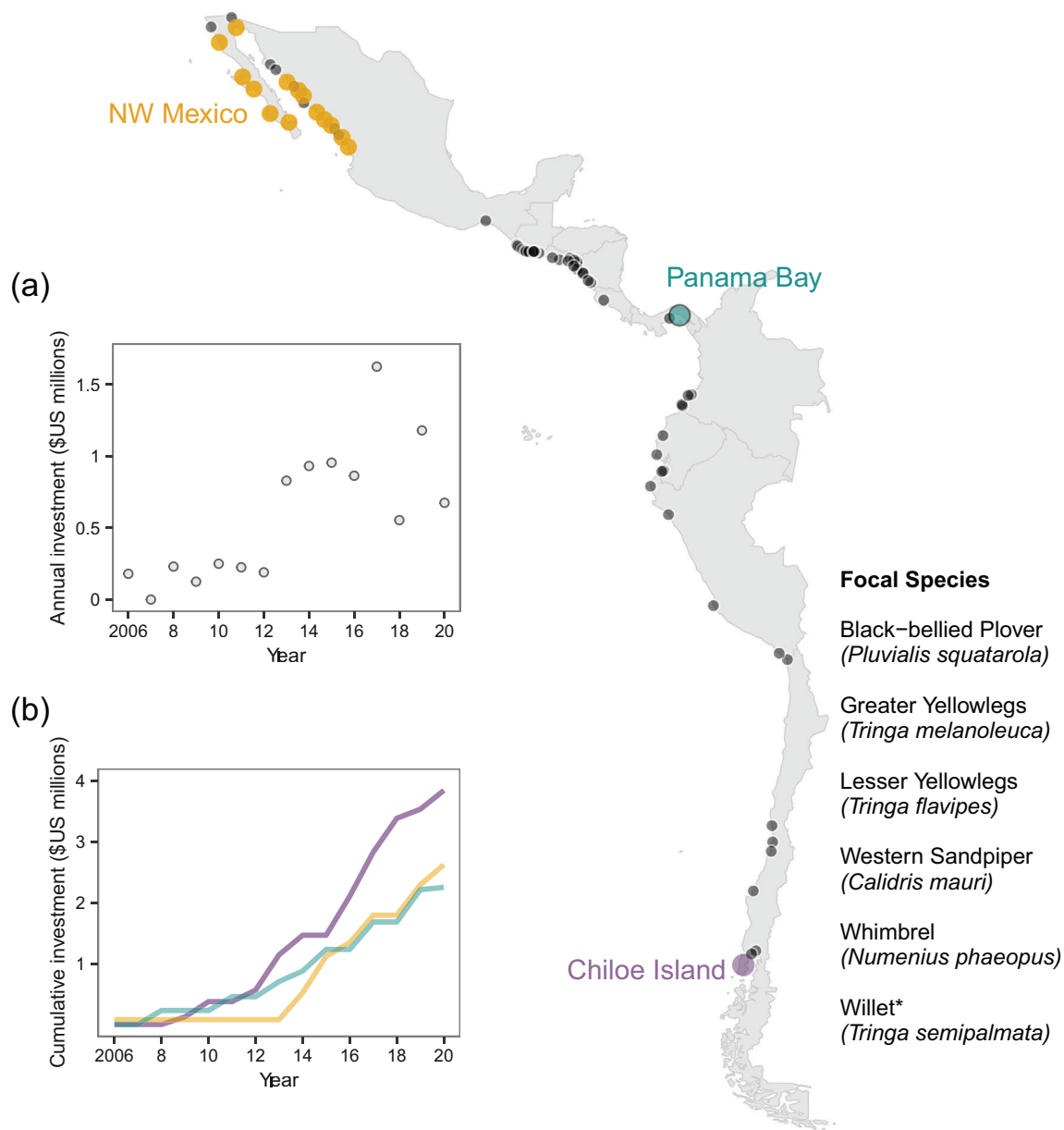
FIGURE 1 (Map) Sites along the Pacific Americas Flyway in Latin America included in an evaluation of shorebird populations (colored circles, 17 treatment [i.e., foundation funding] sites; black circles, 59 donor pool sites [i.e., no foundation funding]): (a) annual foundation investment in the treatment sites over time; (b) cumulative foundation investment (2006–2020) in each region. With one exception (Willet* absent on Chiloe Island), all focal species occur at all treatment sites.

impact from 2010 to 2021: shorebird abundance. The evaluation is summative and formative because we also sought to provide insights to improve the program. We used a quasi-experimental, mixed methods approach in which we combined multiple data sets and modeling approaches. Combining shorebird population models with a land-cover classification model, we used the synthetic control method (SCM) to create counterfactuals for shorebird trends at the treatment sites (Abadie, 2021). Although common in the social and medical sciences, the method has been applied less in the environmental sector (Adhikari, 2022; Sills et al., 2015).

METHODS

Evaluation framework

The SCM has several advantages over similar methods (Abadie, 2021). First, a counterfactual is selected objectively as opposed to being based on researchers' subjective judgment (Abadie et al., 2010). Second, unlike traditional regression models, the SCM safeguards against extrapolation (i.e., resulting weights are nonnegative and sum to one). Third, although other methods are based on the assumption that the effects of unobserved con-
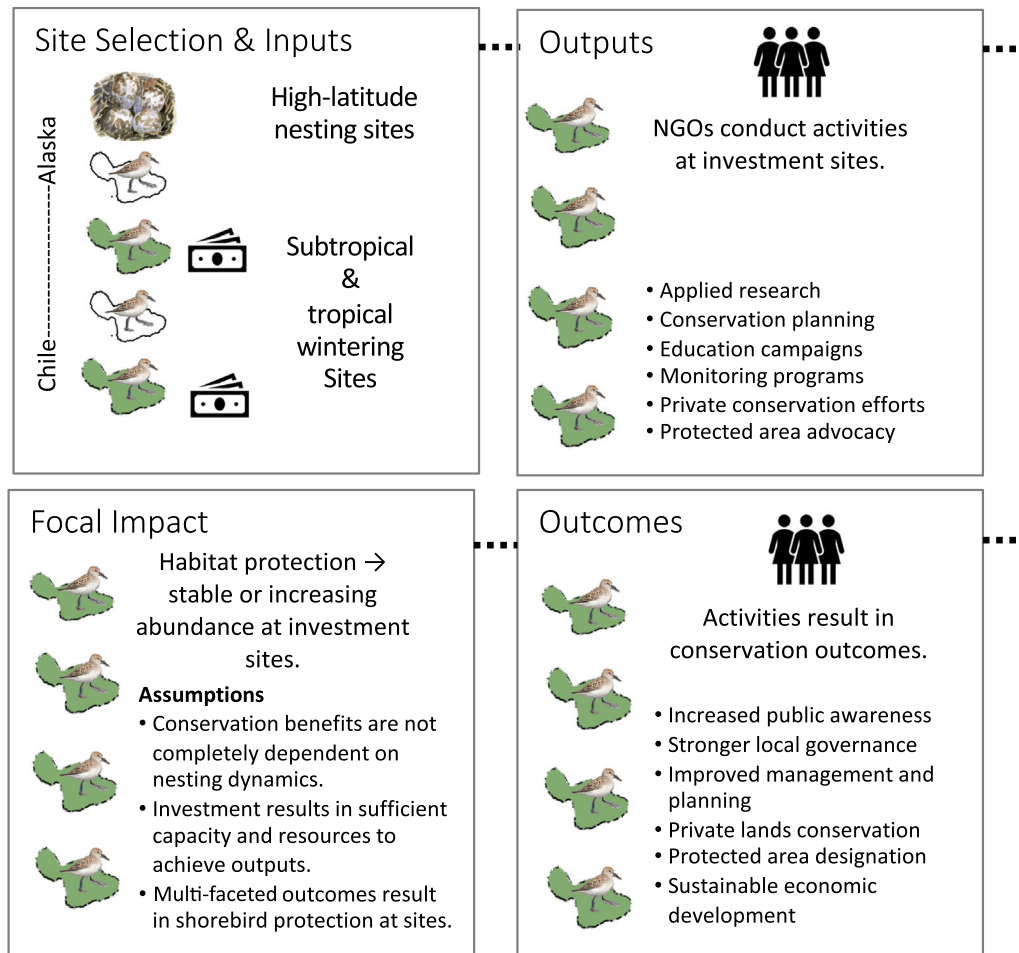
FIGURE 2 A theory of change for site-specific investments for shorebird conservation along the Pacific Americas Flyway (NGO, nongovernmental organization). The theory of change flows from site selection and inputs to outputs to outcomes to the focal impact of the evaluation. Other impacts, such as human capacity, were beyond the scope of the evaluation.

founders are constant over time, the SCM allows these effects to vary (i.e., parallel trends assumption) (Abadie et al., 2015). Fourth, the SCM provides a systematic way of constructing a control from a weighted average of units from a donor pool (i.e., group of comparison units), such that the constructed synthetic unit matches the path of an outcome variable for a pretreatment period. Finally, the method is premised on the idea that a combination of untreated sites provides a more appropriate comparison than any single untreated site (Abadie, 2021). A valid synthetic control analysis requires that the pretreatment outcome of the synthetic control closely matches the outcome of the treated site. If so, comparing the outcome paths posttreatment provides insight into a treatment effect. If not, the analysis is discarded as invalid.

Site selection

Sites in the donor pool were considered priority sites by the Pacific Shorebird Conservation Initiative (PSCI) (Senner et al., 2016), were located in Latin America and did not receive foun-dation investment, and had adequate shorebird population and land-cover data (see below). Treatment sites were in northwest-ern Mexico (15), Panama (1), and Chile (1) and were also PSCI priority sites (Figure 1). Site boundaries were delineated in the construction of the land-cover classification model (see below). We screened donor pool sites to avoid those that received sub-stantial investments that were not from the foundation, which could confound our results. Because some treatment sites also received investment from other sources, we viewed all sites as having some level of annual baseline (nonfoundation) shorebird conservation funding ($x_{baseline}$) from US$0 to some unknown maximum amount. Treatment sites received additional funding from the foundation ($x_{foundation}$), so the total annual funding equaled $x_{baseline} + x_{foundation}$. If a donor pool site's baseline funding was equal to or greater than an investment site's fund-ing ($x_{baseline} + x_{foundation}$), we deemed the site invalid as a potential control. Using this logic, we first documented annual foundation investment from 2006 to 2020 for the 17 treat-ment sites from internal grant documents ($x_{foundation}$) (Figure 1). We then screened all sites for nonfoundation shorebird fund-ing ($x_{baseline}$) from publicly available sources and personal

communications. This allowed us to compare estimates of total funding ($x_{baseline} + x_{foundation}$) between treatment and donor pool sites. Any sites that received significant shorebird funding relative to treatment sites were removed from the donor pool. Although imperfect, the screening captured any significant non-foundation investments and provided insights into the validity of the donor pool and additional investments made at treatment sites. We also compared the size of sites in the synthetic controls to treatment sites to assess other potential confounding factors.

Intervention period

The intervention did not occur at a single point; rather, investment was continuous and variable. Foundation investment at the treatment sites began in 2006, and there was a marked increase in annual investment starting in 2013 (mean$_{2006-2012}$ = US\$171,436, mean$_{2013-2020}$ = US\$951,788) (Figure 1). Of the five investments made in 2013, all were awarded in June or July, and four were 2-year grants. Thus, we used 2015 as the intervention period. This is also the year that ~US\$1 million was invested at the treatment sites in the three countries (Figure 1). We argue it is reasonable to assume a certain threshold of investment is needed before sufficient built capacity and resources are available to translate into positive shorebird outcomes. Although US\$1 million could be a premature threshold, a higher level of investment continued after the intervention year. Thus, 2015 represented an inflection point after which investment was sustained at a high rate. Investments were in the form of 1- to 2-year grants to 12 NGOs conducting shorebird conservation activities (Figure 2). Investments in Panama and Chile supported activities at a single site. Investments in northwest Mexico supported activities at 15 sites, and funds were not equally distributed. Data were not available to discern site-level investments. Because one NGO that worked on all Mexican treatment sites received ~70% of the foundation's investment in Mexico, we grouped the treatment sites in Mexico.

Outcome variables

We selected six focal shorebird species, all occurring widely across the study area and present at all treated sites (with one exception, willet [*Tringa semipalmata*]) (Figure 1). All had a global conservation status of least concern but decreasing populations (International Union for Conservation of Nature [IUCN], 2021). We selected species that differed in life histories to explore abundance patterns common to shorebirds spending the nonbreeding season in Latin America (e.g., breeding range, migration distance, body size [Appendix S1]). Using multiple data sets, we estimated trends of three outcome variables related to shorebird abundance: abundance relative to the entire flyway (relative abundance), proportion of the entire flyway population at a site (percent flyway), and shorebird density per 10 ha at a site (density).

Outcome variables were estimated from statistical models. Because we were interested in evaluating counterfactuals across different sites and species, we standardized relative abundance and density based on the rate of change from the first year of data. Percent flyway was already standardized. We used remote sensing data to build a land-cover classification model and used the results as covariates in the SCM analyses.

We took two approaches to estimate the average treatment effect on the treated (ATT), which contrasts the observed outcomes of the treated group with their counterfactual outcomes post-treatment. Using the geometric mean of the outcome variables, we first created aggregate shorebird indices that included all 6 species for Panama Bay, Chiloe Island, and the 15 treatment sites in Mexico. The geometric mean is an appropriate measure of central tendency for rates of decrease or increase because of the skewed distribution (i.e., population increases can be infinite, but population decreases cannot be >100% [Sheehan et al., 2010]). For each outcome variable, we ran SCM analyses for the indices. If an analysis was deemed valid (see below), we calculated the ATTs for each index-outcome combination. Second, we repeated the same analyses for each focal species for each treatment site separately. Doing so allowed us to explore any heterogeneity across species and sites. Finally, we tested for significance with permutation methods and assessed robustness with recommended methods (see below).

Land-cover model

We used the Google Earth Engine for land-cover classification modeling. Using digital elevation models and bathymetric data, we delineated the study area (i.e., the Pacific coast from Mexico to Chile) (Gorelick et al., 2017). To improve continuity, we joined the mask with a 2-km buffer around the coastline with country-specific data layers (GADM, 2022). We selected 105 sites (4.3 million ha), which included treatment and donor pool sites (Reiter et al., 2020). We obtained site boundaries from multiple data sets. When boundaries were not available, we created polygons encompassing the site. The same site boundaries were also used for shorebird modeling. We used four supervised classification techniques to identify cover types at the sites. We developed separate classification models for three provinces (modified from Spalding et al., [2007]). Using the wetland classification of Ramsar (2013) and Migratory Shorebird Project (MSP) data, we defined seven land-cover classes for the evaluation: bare, beach, mangrove, marsh, other, tidal, and water. The bare class consisted of built or bare land; the other class consisted of unflooded grassland, shrubland, forest, and cropland.

We compiled 100,000 Landsat 7 and 8 images (30 m^2 resolution) across the study area for 2001–2020. We removed images with >80% cloud cover, applied a cloud mask, conducted a spectral harmonization, and applied a gap filling procedure for Landsat 7 (Roy et al., 2016). We then built an image database for five 4-year periods, which allowed us to compile enough usable images in cloudy geographies and capture any variability in land-cover classes. We used 6 bands to calculate 10 spectral indices to identify and differentiate cover types, which were the predictors for the machine learning models. For each band

and spectral index, we created image mosaic stacks representing the temporal variation every 4 years. We calculated the median value (50th percentile) and the 10th and 90th percentiles to represent extreme values. For each land-cover type, we classified polygons based on ground sampling and the interpretation of high-resolution images. Within these land-cover polygons, we created random points and assigned the spectral responses of satellite image band combinations from 2017 to 2020. We used 80% of these points to train the model. Using the trained model, we classified each site during the pretreatment period. Using five metrics, we evaluated model accuracy and performance with 20% of the points for each cover type. We used confusion matrices to summarize the actual and predicted values and misclassifications. We further examined cover types at each site for potential problems, correcting classifications in some cases and excluding sites in others. For the pretreatment period, we calculated the percentage of each cover type for 76 donor pools and treatment sites. A complete land-cover classification model analysis will be presented elsewhere (details in Appendix S2).

Relative abundance

We used the eBird data set to model shorebird population trends (Johnston et al., 2015; Sullivan et al., 2014). First, we obtained complete checklist data for the focal species from the eBird Reference Dataset (Fink et al., 2020). We used only checklists submitted within the nonmigratory, nonbreeding season for each species as defined by the eBird Status and Trends Program and filtered the data accordingly (Fink et al., 2020). Second, we used eBird checklist data to model the spatial and temporal relative abundance patterns at 27-km^2 resolution during the nonbreeding seasons of 2010–2019 for the focal species along the entire flyway. Third, we used these estimates during species-specific wintering periods to model changes in relative abundance for the entire flyway and treatment and donor pool sites.

We modeled relative abundance for each species with a two-stage, locally fitted hurdle model (e.g., Johnston et al., 2015). To ensure that local habitat associations were captured across the study area, we first partitioned the flyway into 20 overlapping modeling regions. We fitted the model 27 times for each species and region, slightly varying the latitude and longitude boundaries. Each repeated run included a different random subsample of checklists chosen to ensure that no 2 checklists in the same year were <1 km apart (Strimas-Mackey et al., 2020). In each model run, we fitted a two-stage random forest model. The first stage predicted the probability of occurrence (i.e., ≥1 individual) of the species based on the checklist detection or nondetection data (i.e., 1 or 0) and 67 covariates. The second stage of the model predicted the expected count of each species present. This stage used the same set of covariates but included only those checklists for which the species was predicted to be present in the first stage (or was not predicted to be present at a given location but was observed on a checklist).

We then predicted the relative abundance for each focal species across the 27-km^2 grid in the flyway for each year (2010–2019). These predictions represent the expected average count of a skilled observer birding for 1 h and traveling 2.5 km on 1 January at the optimal time of day to observe each species. We used an ensemble approach to map predictions across the flyway for each year by averaging predicted values at each grid cell among model runs. Next, we estimated the relative abundance for the treatment and donor pool sites for all the focal species by summing relevant grid cell values on the resulting final maps (i.e., within a 25-km area around sites). Finally, we estimated uncertainty with the 2.5% and 97.5% quantiles of bootstrapped estimates calculated from 1000 relative abundance maps generated by resampling the pixel values from the repeated model runs (details in Appendix S2).

Percent flyway

To estimate the proportion of a population occurring in any given site, the relative abundance estimates for each site were divided by the total relative abundance for the entire flyway (Johnston et al., 2020). We did this for each year, species, and site to estimate trends.

Density

We used structured survey data from the MSP, conducted annually in January–February (Reiter et al., 2020). To control for variability in effort, we included sites that only had ≥5 years of data from before 2013 to 2021. We removed survey units where the focal species were never observed. The Chiloe treatment site is not part of the MSP network. Thus, we integrated a similar data set that included annual surveys at 12 areas on Chiloe Island (Centro de Estudio y Conservación del Patrimonio Natural, Ancud, Chile). Those surveys included only 1 focal species (i.e., whimbrel [*Numenius phaeopus*]). Like the MSP data, surveys were spatially defined into distinct units.

We used Bayesian hierarchical models to estimate shorebird abundance yearly at each site (Kéry, 2010). The total number of birds of a species counted at a site in a year was the response variable and was assumed to follow a negative binomial distribution. We included the total area surveyed in each site as a covariate to account for variable survey effort at sites across years. Not all sites and sampling units were sampled in all years. First, we modeled count as a function of year for each site and included individual site and year effects. Second, we modeled count with a year and site effect but no trend effect. Third, we used the deviance information criterion to compare the two models and its inverse to generate weighted average estimates of bird density from the predictions of each of the models (i.e., model with lower DIC received more weight) (Hooten & Hobbs, 2015). We fitted the models in JAGS with noninformative priors for all parameters. We considered model convergence successful if R-hat was ~1 and used residual and trace plots to assess model fit (Gelman & Hill, 2006). Finally, we sampled from the posterior distributions of the parameter estimates to generate estimates of abundance (birds 10/ha) and

uncertainty (95% credible interval) for each site and year (details in Appendix S2).

Synthetic control method

The treatment (and donor pool) sites consisted of 17 (59) sites for relative abundance and percent flyway and 16 (7) for density. In addition to the seven land-cover types as covariates, we included two lagged variables (Abadie, 2021): outcome variable of interest in 2012 and 2014 (pretreatment period). The covariates were observed characteristics that affect outcomes at all sites pre- and post-treatment. The geometric mean of the covariates was used for the aggregate analyses.

We used a pretreatment fit index for the SCM analyses to decide which results to include in evaluating treatment effects (Adhikari & Alm, 2016). We excluded analyses if the index was >0.10, which means there was >10% difference between the treated and synthetic unit in the pretreatment period (i.e., 0 is a perfect fit, and 1 represents a synthetic unit that is twice as big [or half as small] as the treated unit). For included analyses, we computed the ATT by taking the mean of the annual differences between the treatment site and synthetic control for the outcome variables in the post-treatment period.

To assess whether estimated effects were statistically significant, we conducted placebo tests and calculated pseudo *p* values. A placebo test applies the SCM to donor pool sites and calculates the associated estimated effect (Abadie et al., 2010). The permutational procedure produces a distribution of estimated gaps for the sites where no intervention occurred. We excluded placebo runs where the pretreatment mean squared prediction error (MSPE) was two times greater than the treated site. We then estimated the probability of the observed magnitude of the treatment gap. We evaluated significance by assessing the ratios of pre- and post-treatment MSPEs for all donor pools and treatment sites and calculated the probability of obtaining a ratio as large as the treatment ratio, adopting an a of 0.05. We used leave-one-out tests to check for sensitivity to including specific donor pool sites (Abadie et al., 2010). We used *synth* in R for the analyses (Abadie et al., 2011).

RESULTS

Site selection

The combined total area of the treatment and donor pool sites was 1.8 million and 2.1 million ha, respectively. Site size ranged from 27 to 804,087 ha (mean [SD] = 51,089 [120,107], median 3136 ha). The mean size of treatment sites (105,792 ha [113,595]) was larger than donor pool sites (35,327 ha [181,171]). However, there were no differences in size between the treatment sites and those selected from the donor pool for synthetic controls (see the "Discussion" section). We documented US$48.1 million of nonfoundation funding (i.e., baseline) at treatment and donor sites from 2006 to 2021. Forty-eight (81%) donor pool sites received no baseline
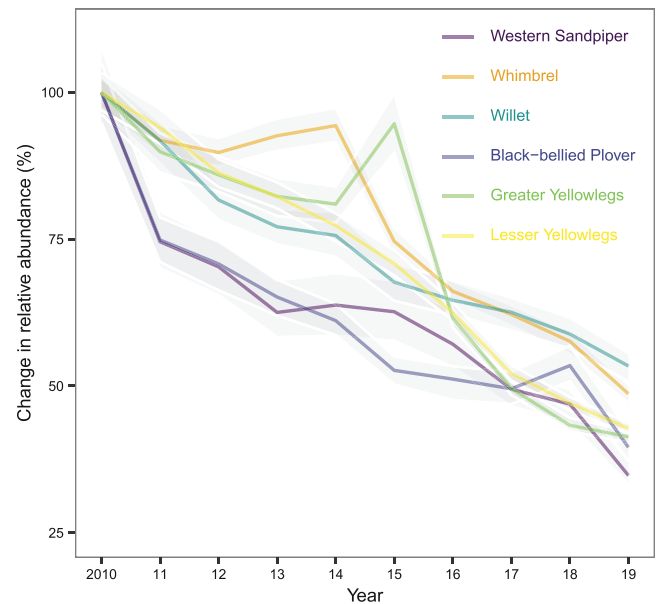

FIGURE 3 Rate of change of relative abundance (2010–2019) for six shorebird species across the entire Pacific Americas Flyway during the nonbreeding season (gray shading, uncertainty of estimates based on 2.5% and 97.5% quantiles of bootstrapped estimates).

funding. For the 11 donor pool sites that received funding, the median total funding was US$159,270. In contrast, 13 (76%) treatment sites received baseline funding, with a median of US$2.6 million. Overall, treatment sites received an order magnitude more baseline funding than donor sites (US$44.8 million versus US$3.3 million). Based on this analysis, no donor pool sites were removed from the analyses (details in Appendix S1).

Land-cover model

Overall, the random forest model predicted land-cover classes the best, with an accuracy of >0.90 for all three provinces. The mean F1 score across the seven classes and three provinces was 0.92 (SD 0.04). Percent cover varied widely across class and site. However, the distributions of percent cover types for all classes overlapped between treatment and donor pool sites (details in Appendix S2).

Relative abundance

Annual estimates had declining trajectories across the entire flyway for all species (Figure 3). For treatment sites, all of the site-species combinations declined (*n* = 101) (Figure 4). For donor pool sites, 98% of the combinations declined (*n* = 352). Although both site types showed similar average declines, dynamics differed between donor and treatment sites (Figure 4). Changes in relative abundance were heterogenous at both site types. At treatment sites, the magnitude of change for site-species combinations ranged from 91% of 2010 levels for whimbrel at the Colorado River Delta (Mexico) to 13% for black-bellied plover (*Pluvialis squatarola*) at Panama Bay. At the
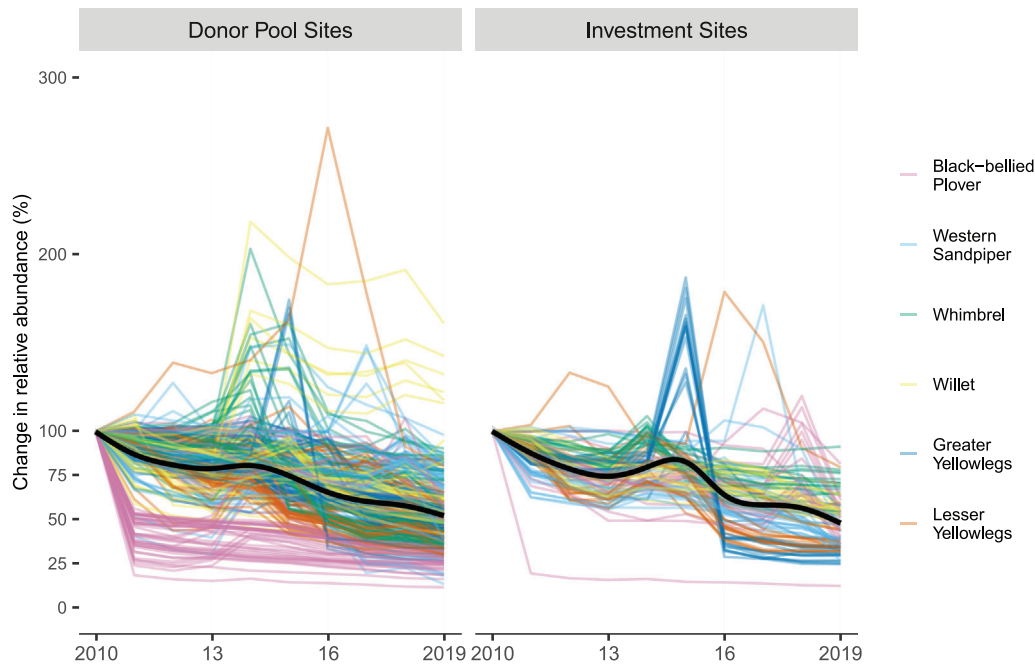
FIGURE 4 Rate of change of relative abundance (2010–2019) for six shorebird species at treatment (received foundation funding) ($n = 17$) and donor pool sites (no foundation funding) ($n = 59$, sites that are potentially selected to create counterfactuals; black lines, regression with a general additive model).

donor pool sites, the change ranged from 161% of 2010 levels for willet at Tumbes Mangroves (Peru) to 12% for black-bellied plover at Chame Bay (Panama) (details in Appendix S2).

Percent flyway

Percent flyway at a site remained relatively constant across species and sites (Figure 5). On average, treatment sites had higher values of percent flyway than donor pool sites ($\bar{x}_{2019} = 1.57\%$ and $\bar{x}_{2019} = 0.41\%$, respectively). In addition, species-site combinations with the highest percent flyway were dominated by treatment sites: Six of the seven combinations >4% in 2019 were treatment sites (Figure 5). For all species, the average percent flyway in 2019 for treatment sites was more than twice that of donor pool sites.

Density

Although density trends varied by treatment and donor pool site, on average, both site types showed little change (Figure 6). For treatment sites, 57% of the site-species combinations showed a decline ($n = 91$). For donor pool sites, 64% of the site-species combinations showed a decline ($n = 42$). Changes were heterogenous at both site types. At the treatment sites, the rate of change for site-species combinations ranged from 313% of 2011 levels for willet at San Ignacio Lagoon (Mexico) to 28% for willet at Ceuta Bay (Mexico). At the donor pool sites, changes ranged from 392% for willet at Paracas National Reserve (Peru) to 24% for western sandpiper (*Calidris mauri*) at the Iscuande River (Colombia) (details in Appendix S2).

There were differences in trends between the three outcome variables. Although relative abundance decreased for almost all sites and species for donor and treatment sites, density (i.e., a more localized scale) was more variable, with ∼60% of site-species combinations showing a decline. At treatment sites, site-level density trends were variable, whereas relative abundance trends decreased (Appendix S2). Percent flyway trends were relatively stable at all sites (Figure 5).

Synthetic control method

In the aggregate, the SCM analyses performed well for all outcome variables, with a mean prefit index of 0.02 ($n = 8$, maximum = 0.07). Synthetic control solutions were sparse (median = 5.5 sites, range = 4–8, $n = 8$). The median number of predictor variables included was 8.5 (range = 5–9, $n = 8$). The mean weight was 0.11 to 0.13, with the lag covariates contributing more weight than land-cover covariates (0.17–0.40 vs. 0.03–0.09, respectively). Based on placebo tests, no post-treatment effects were significant (Appendix S3). Aggregate ATTs revealed no evidence of a treatment effect. The mean ATT across the three indices was 0.62% for density, −0.37% for relative abundance, and 0.01% for percent flyway (Figure 7). We conducted two additional SCM analyses to explore the robustness of our results. First, we used a donor pool restricted to sites that received no funding prior to the intervention period ($n = 48$). Second, we used conservation funding prior to the intervention period as an additional covariate. The results were similar with no evidence of a treatment effect (mean ATT = −0.18 and 0.44) (details in Appendix S3).
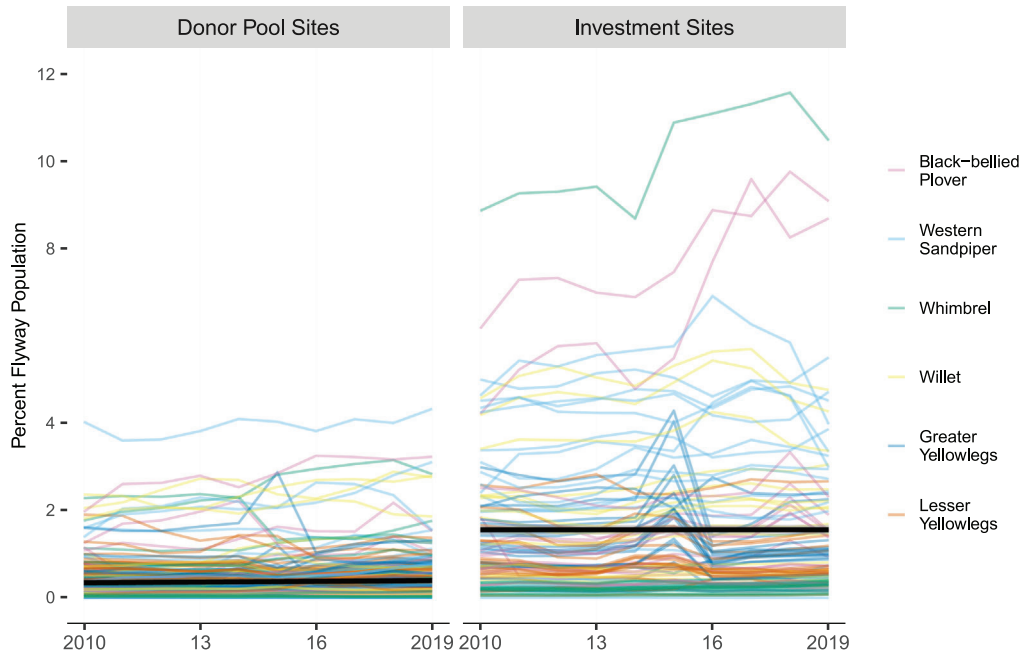
FIGURE 5 Changes in percentage of the flyway population (2010–2019) for six shorebird species at treatment (received foundation funding) ($n = 17$) and donor pool sites (no foundation funding) ($n = 59$, sites that are potentially selected to create counterfactuals; black lines, regression with a general additive model).

FIGURE 6 Change in shorebird density (2011–2021) for six shorebird species at treatment (foundation funding) ($n = 16$) and donor pool sites (no foundation funding) ($n = 7$, sites that are potentially selected to create counterfactuals; black lines, regression with a general additive model).
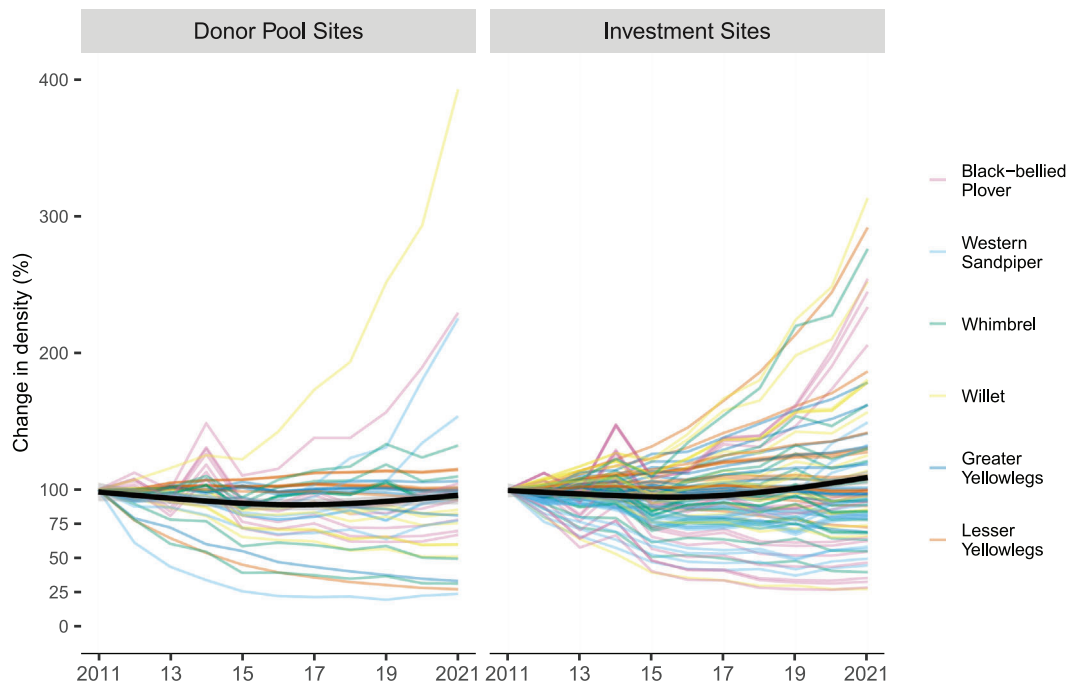
At the species level, the SCM analyses also performed well. For relative abundance and percent flyway, 92% (93) and 71% (72) of the analyses had a prefit index of <0.10 ($\bar{x}_{\text{relative abundance}} = 0.02$ and $\bar{x}_{\text{percent flyway}} = 0.03$). For density, 92% (84) of the analyses had a prefit index <0.10

($\bar{x}_{\text{density}} = 0.01$). For all outcome variables, all donor sites were selected at least once. Synthetic control solutions were, however, relatively sparse. The median number of sites selected for synthetic controls was 6.5 for relative abundance, 4 for density, and 5 for percent flyway. All nine covariates contributed weight
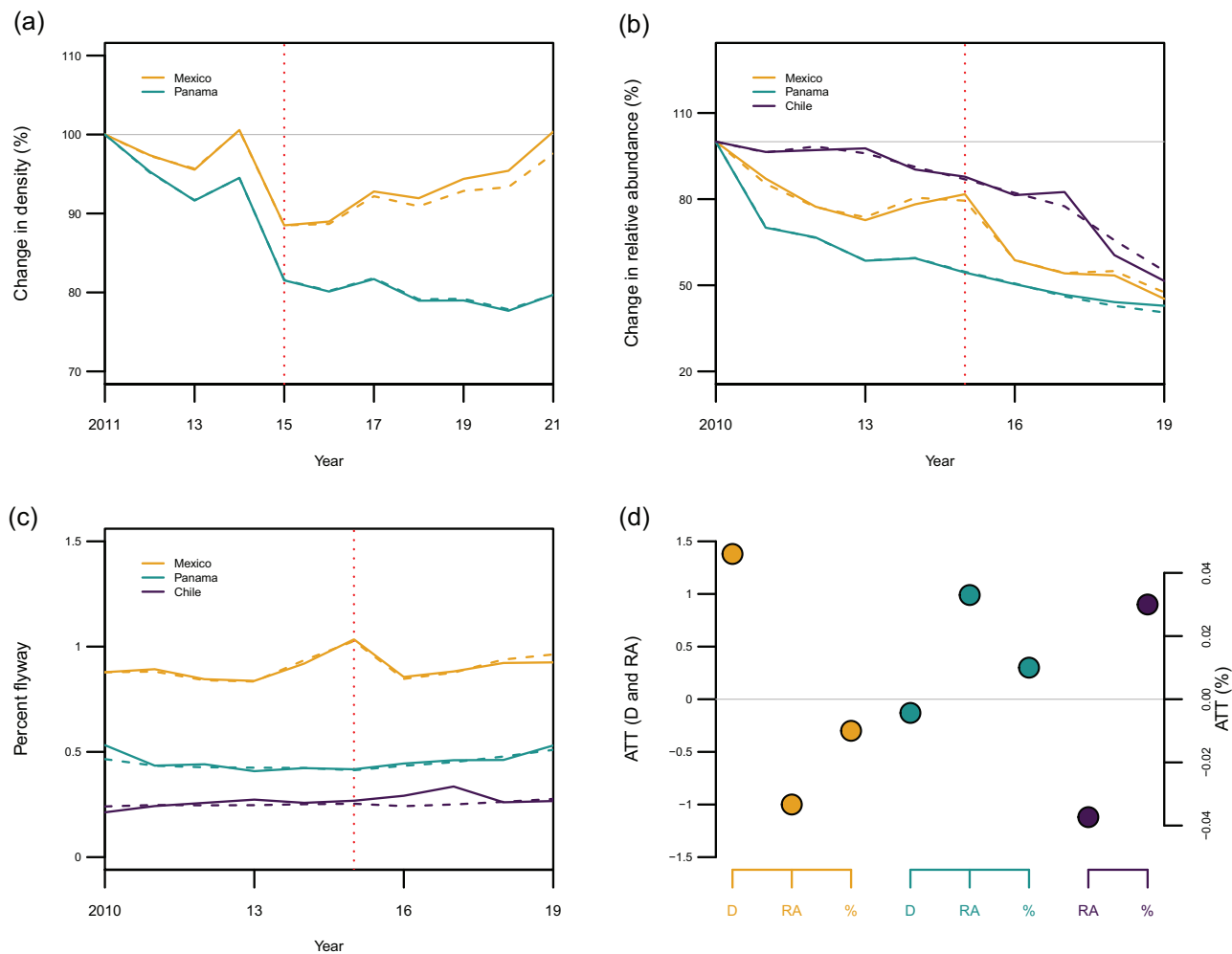
FIGURE 7 Synthetic control analyses with three outcome variables: (a) shorebird density (D), (b) relative shorebird abundance (RA), and (c) percentage of the flyway population (%) (solid line, treatment, geometric mean of six shorebird species aggregated across treatment sites in Mexico [15], Panama [1], and Chile [1]; dotted line, synthetic control). For density and relative abundance, the *y*-axis is rate of change since the initial year (=100%). Data for density at the Chile site was not available. In all cases, the treatment year is 2015. (d) The average treatment effect on the treated (ATT) for the eight synthetic control analyses.

to the synthetic control for each outcome variable at least once. The median number of covariates included in an SCM analysis was 7 for all outcome variables. The mean weight was between 0.14 and 0.17, with the lag covariates contributing more weight on average than land-cover covariates (0.21–0.37 vs. 0.05–0.12, respectively).

Species-level ATTs showed heterogeneity in treatment effects. Across outcome variables, five cases had ATTs that did not overlap zero (Figure 8). For relative abundance, black-bellied plover (ATT = 7.4%) and whimbrel (ATT = 8.1%) had a positive effect, whereas lesser yellowlegs (*Tringa flavipes*) had a negative effect (ATT = −9.1%). For density, two species had a negative effect: western sandpiper (ATT = −6.4%) and willet (ATT = −31.2%). Of the two species with positive effects, two Mexican sites were consistently among the sites with the largest ATT: Colorado River Delta (ATT$_{\text{black-bellied plover}}$ = 30.3%, ATT$_{\text{whimbrel}}$ = 20.3%) and Ojo de Liebre Lagoon (ATT$_{\text{black-bellied plover}}$ = 13.9%, ATT$_{\text{whimbrel}}$ = 17.5%) (Figure 8). The Colorado River Delta was

also a positive outlier with the geometric mean of the six focal species for relative abundance and percent flyway (17.6% and 0.61%, respectively) (Figure 8). Based on placebo tests, no treatment effects were significant. The leave-one-out test did not reveal any site-specific sensitivities (details in Appendix S3).

DISCUSSION

Conserving migratory species is daunting, and shorebirds are no exception. Our results provide evidence that shorebirds may be declining across the entire flyway. Compared with counterfactuals, we failed to find evidence of an overall difference in shorebird population dynamics at 17 wintering sites where significant conservation investments had been made over the past decade. Depending on the data set, 61–99% of the populations at treatment and donor pool sites appeared to be declining for the six shorebird species evaluated. However, the percentage of flyway population at sites remained stable,
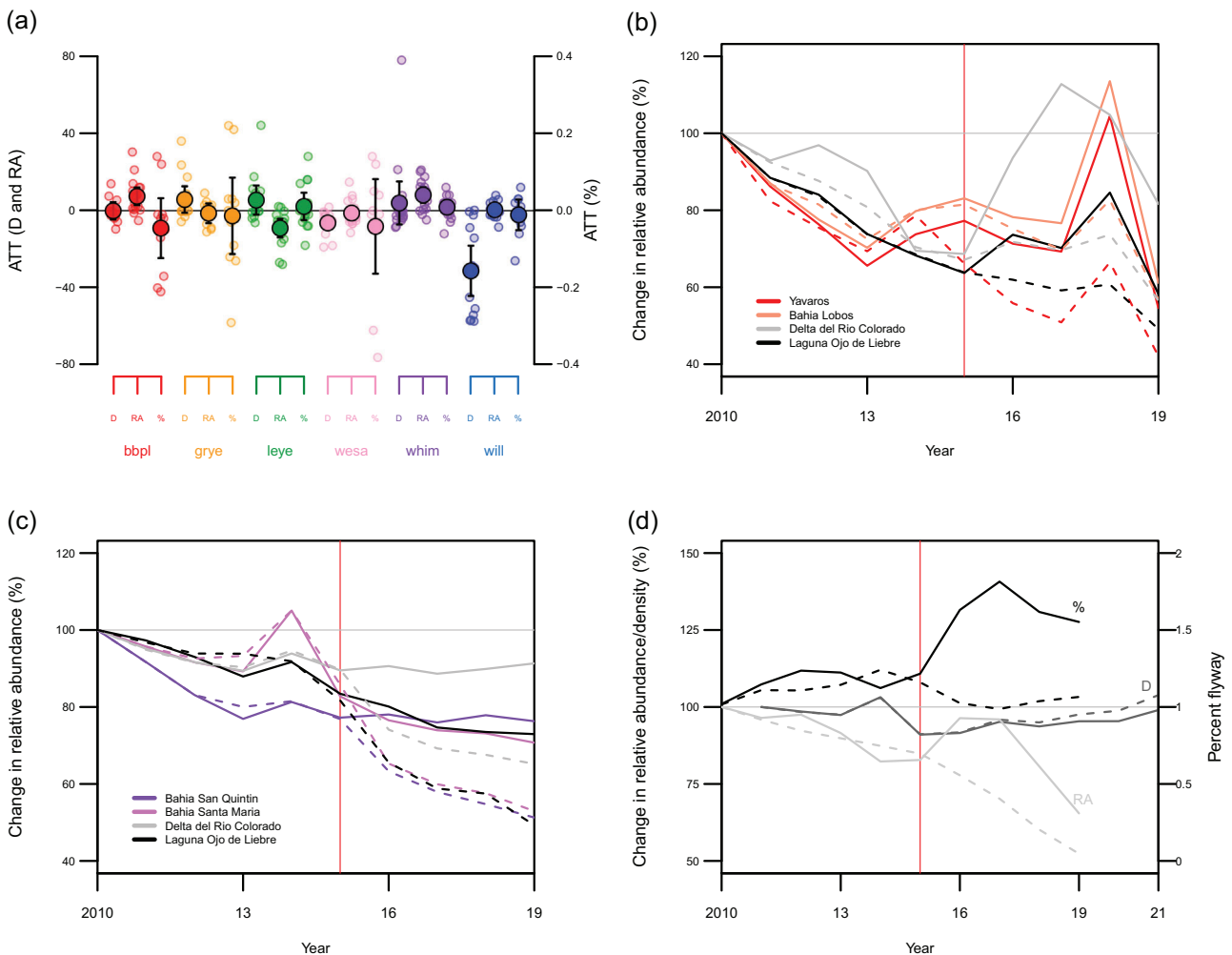
FIGURE 8   (a) Average treatment of the treated (ATT) and 95% CI for each of the six shorebird species across treatment sites for three outcome variables: density (*D*, *n* = 16), relative abundance (RA, *n* = 17), and percentage of the flyway population (%, *n* = 17). Five cases (*) do not overlap with zero. (b) Synthetic control analyses for black-bellied plover at the four treatment sites with the largest ATT. (c) Synthetic control analyses for whimbrel at the four treatment sites with the largest ATT. (d) Synthetic control analyses for the treatment site Colorado River Delta for the geometric mean of the three outcome variables (bbpl, black-bellied plover; grye, greater yellowlegs; leye, lesser yellowlegs; wesa, western sandpiper; whim, whimbrel; will, willet).

with a greater percentage of flyway populations present at treatment sites. Treatment effects were heterogeneous and positive for two species. Two treatment sites were also outliers with respect to positive effects. However, species- and site-level effects were not statistically significant. We discuss three important implications of our results: performance and limitations of our evaluation; potential explanations for our results through the lens of the funding program's theory of change; and recommendations for evaluating investments in shorebird conservation.

Evaluation framework

Overall, the SCM performed well, demonstrating its application and value in environmental conservation, which has been limited to deforestation reduction and water conservation programs to date (e.g., Sills et al., 2020; Zhang et al., 2022). Most

analyses were valid. Even when the donor pool was large, the number of sites selected tended to be sparse, suggesting no problems with overfitting. All covariates tended to contribute weight to the synthetic control selection across the analyses. Although the pretreatment period (i.e., 4 and 5 years) was shorter than the post-treatment (i.e., 7 and 5 years), the donor pool and predictor variables tended to perform well in creating synthetic controls.

The donor pool sites appeared to function as credible counterfactuals despite the potential for two confounding variables: size and level of investment. Because we used the rate of change for outcome variables, it is potentially problematic that donor pool sites were smaller, on average, than treatment sites. If smaller sites with smaller populations were selected in a synthetic control, the underlying mechanisms of population dynamics could be comparatively different due to factors related to density-dependent processes. However, this appears not to be the case. For example, 24 sites were selected in the synthetic

controls for the 3 aggregate analyses, of which 15 contributed weights >10%. The mean (103,931 ha) and median (55,208 ha) size of those 15 sites are comparable to treatment sites (105,782 and 47,688 ha, respectively). In addition, the investment screening demonstrated that the donor pool sites with conservation funding received substantially less than treatment sites, further supporting the credibility of the donor pool. Only 11 donor pool sites received any conservation investment, all of which were substantially less than the treatment sites. Further, leave-one-out analyses suggest that no 1 site was driving our results, and our additional analyses revealed similar results (details in Appendices S1 and S3).

Our overall results were consistent across three outcome variables estimated by different data sets and statistical models. Although relative abundance estimates suggested more of a general decline than density estimates, these differences do not influence our conclusions and underscore the value of a mixed methods approach. Disentangling the causes of these differences is important but was beyond the scope of and unnecessary for our evaluation focused on estimating counterfactuals. At the aggregate level, no treatment effect was detected for any of the three outcome variables.

There were three limitations to our evaluation. First, we focused solely on shorebird abundance. Other important outcomes were not evaluated, including human and other environmental dimensions. The human capacity built at the treatment sites has been significant, and foundation investments were a contributing factor. However, this outcome has yet to be tracked or documented in detail. Important environmental factors (e.g., habitat quality) are challenging and costly to track at the site level. Yet, the quality of tidal flats plays a critical role in maintaining energy and water balance in shorebirds (Piersma, 2012). Although we successfully incorporated the quantity of land cover, important changes or differences in habitat quality remain elusive. Relatedly, climate factors independent of the conservation status of a site can affect shorebird populations (Ward et al., 2009). However, the SCM should control for these potential effects (e.g., any unique population site-level effect would likely result in an invalid SCM analysis and would be excluded). Second, we use US dollars invested as the single input for the evaluation as opposed to attempting to conduct SCM analyses on outputs or outcomes. This was not possible due to the multi-faceted nature of the foundation's program (i.e., multiple conservation activities supported) and the lack of detailed information available for treatment and donor pool sites. Third, our data were limited for the preintervention period. Although our data sets were within the range of many other studies that have applied the SCM (i.e., 4 years for density and 5 years for relative abundance and percent flyway), small preintervention periods can result in spuriously obtained fit (Abadie, 2021). We mitigated for this risk by using multiple data sets and multiple levels of analyses (i.e., index- and species-based analyses), which resulted in consistent results and sparse solutions.

Theory of change

There are several potential and nonmutually exclusive explanations for our results. First, it is possible that investments were insufficient to date to translate into positive shorebird population outcomes at the treatment sites. Although conservation outcomes have been achieved, the relationship between them and shorebird populations is likely not linear. Building capacity and many conservation activities are long-term endeavors. Thus, one hypothesis is that investments have yet to mitigate the negative impacts at treatment sites. Second, conservation outcomes at treatment sites may be insufficient to safeguard shorebird populations. For example, human disturbance or declining habitat quality could occur at sites even in the face of conservation outcomes (e.g., protected area designation). If a critical driver of site-specific shorebird decline (e.g., human disturbance; Palacios et al., 2022) is not being mitigated with the activities being supported by investments, the lack of a difference in trends compared with sites with no conservation investment is not surprising.

Another explanation for limited impact is the possibility that factors at other sites along the flyway influence shorebird trends at treatment sites. For example, negative impacts at breeding or stopover sites along the flyway could drive the observed population declines, irrespective of treatment site quality (Runge et al., 2014). Although the migratory connectivity of shorebirds is poorly understood, our results are consistent with this hypothesis. Our flyway-wide estimates suggest that all six species are in decline. Density estimates suggest that >50% of the species-site combinations are declining. Relative abundance estimates suggest a decline at nearly all the 456 site-species combinations. Yet, the percentage of flyway populations present at sites remained relatively stable. These patterns are consistent with a flyway-wide decline that negatively impacts all focal species, perhaps driven by factors at breeding sites, critical stopover sites, or both. We selected the focal species to represent different breeding grounds and migratory routes. Given their wide distribution during the nonreproductive period, our results under this hypothesis suggest that all breeding areas and migratory routes are experiencing changes that are affecting shorebirds similarly. More work is needed to assess these hypotheses.

Although species- and site-level treatment effects were inconsistent, the observed heterogeneity provides insights into potential effects. The relative abundance ATT for black-bellied plover and whimbrel was positive, suggesting that individuals of populations that overwinter (or stopover) at treatment sites are declining less compared with counterfactuals. This pattern suggests that conservation investments might have somewhat alleviated threats at treatment sites. For example, there is evidence that black-bellied plover populations respond favorably to wetland restoration (Brawley et al., 1998). In contrast, ATTs for three species were negative, suggesting a greater decline compared with counterfactuals. For lesser yellowlegs and willet, the effect was not small (>9%). Potential mechanisms could

be related to the relative threats at sites. For example, if threats are more severe at treatment sites, one might expect a negative treatment effect. The Colorado River Delta and Ojo de Liebre Lagoon in Mexico were positive outliers for the two shorebird species with positive effects. Further, for the Colorado River Delta, on average, the six focal species declined by 18% less than the counterfactuals, and the percent flyway population increased by 0.5%. One potential mechanism contributing to this result is the total investment amount over the past decade. In addition to the foundation investment, we identified an additional US$8.6 million of conservation funding for the site, which included significant habitat restoration funds.

Evaluating investments

Evidence suggests that shorebirds are declining across the Western Hemisphere (Rosenberg et al., 2019; Smith et al., 2023). Shorebirds are relatively unique with their annual long-distance migrations that can include multiple sites across multiple continents. However, evaluations targeting shorebirds have been limited to assessing interventions at a single site (or several nearby sites) (e.g., Burger & Niles, 2013; Dai et al., 2021). Although valuable, evaluations at this scale are unlikely to capture overall population dynamics and are of limited utility to inform comprehensive strategies to reverse declines. We believe two factors are critical for evaluating shorebird investments: defining success and evaluating infrastructure.

Evaluating shorebird conservation successfully centers on defining a response variable for success. Most objectives focus on addressing factors that limit the abundance of individuals, assuming that doing so will lead to stable or increasing population growth rates. For migratory species, population dynamics must be captured at a sufficient spatial scale at a given site to differentiate between actual increases in individuals versus movement in and out of areas of interest. Thus, evaluations should strive to tease apart actual population increases versus temporary increases and fluctuations in individuals using a site. Population size is a challenging metric for success for long-distance migrants, given how unlikely a single program can address threats across the entire life cycle. Although we found limited evidence of positive and negative effects, our results suggest that the six shorebirds appear to be declining across the flyway. Estimating trends in the percentage of a population at a site provides an alternative metric for defining success. If a species is declining across an entire flyway, its proportion at a site can remain stable or increase. While modest, this positive outcome could have important future implications.

Establishing infrastructure to enable impact evaluations is critical. Investments in flyway conservation will likely continue. Importantly, it will occur in the face of (and in response to) population declines. With additional investment and coordination, the existing monitoring infrastructure is well poised to deliver impact evaluations regularly, which could help improve the effectiveness of investments. Our classification model provided a basis for tracking land-cover changes at sites over time. The eBird models provided site-level and flyway population information. The MSP network has developed the human capacity to collect site-level information to estimate trends. Ultimately, integrating these data may provide the most accurate estimates for evaluation (Robinson et al., 2020). This infrastructure presents an opportunity to design a more robust and a priori evaluation framework that targets the entire flyway and has clear connections among investments, activities, and outcomes.

While not yet mainstream, impact evaluations are becoming more common in the environmental sector. Perhaps not surprisingly, these improved approaches are having mixed results that are often context-dependent (e.g., Sills et al., 2020). They also demonstrate that simpler evaluation designs often estimate different or opposite results than more robust designs (Wauchope et al., 2022). Our results demonstrate that impact evaluations are possible even for investments that target species with complex life histories and migrate across continents. For shorebirds, a next step is to improve and expand our approach to cover the entire flyway and its associated investments and apply impact evaluations to other flyways. Without the ability to evaluate the impact of conservation across entire flyways, it will be challenging to prioritize limited conservation dollars toward the most compelling portfolio of actions that will result in positive conservation outcomes.

ORCID
C. Josh Donlan https://orcid.org/0000-0002-1331-5576
Diana Eusse-González https://orcid.org/0000-0001-8845-5904
Viviana Ruiz-Gutierrez https://orcid.org/0000-0001-7116-1168
Michael C. Allen https://orcid.org/0000-0002-6632-4337
Orin J. Robinson https://orcid.org/0000-0001-8935-1242
Guillermo Fernández https://orcid.org/0000-0002-8478-5575
Eduardo Palacios https://orcid.org/0000-0003-4557-2427

REFERENCES
Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*, 391–425. https://doi.org/10.1257/jel.20191450

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, *105*, 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2011). Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, *42*, 1–17.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*, 495–510. https://doi.org/10.1111/ajps.12116

Adams, V. M., Barnes, M., & Pressey, R. L. (2019). Shortfalls in conservation evidence: Moving from ecological effects of interventions to policy evaluation. *One Earth*, *1*, 62–75. https://doi.org/10.1016/j.oneear.2019.08.017

Adhikari, B. (2022). A guide to using the synthetic control method to quantify the effects of shocks, policies, and shocking policies. *The American Economist*, *67*, 46–63. https://doi.org/10.1177/05694345211019714

Adhikari, B., & Alm, J. (2016). Evaluating the economic effects of flat tax reforms using synthetic control methods. *Southern Economic Journal*, *83*, 437–463. https://doi.org/10.1002/soej.12152

Baylis, K., Honey Rosés, J., Börner, J., Corbera, E., Ezzine de Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A., & Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, *9*, 58–64. https://doi.org/10.1111/conl.12180

Brawley, A. H., Warren, R. S., & Askins, R. A. (1998). Bird use of restoration and reference marshes within the Barn Island Wildlife Management Area, Stonington, Connecticut, USA. *Environmental Management*, *22*, 625–633. https://doi.org/10.1007/s002679900134

Burger, J., & Niles, L. (2013). Shorebirds and stakeholders: Effects of beach closure and human activities on shorebirds at a New Jersey coastal beach. *Urban Ecosystems*, *16*, 657–673. https://doi.org/10.1007/s11252-012-0269-9

Dai, W., Goodale, E., He, R., Mammides, C., Liu, S., Zhou, L., Tang, S., Su, B., Lao, X., & Jiang, A. (2021). An eco-compensation policy increases shorebird diversity during the non-farming period for aquaculture. *Wetlands*, *41*, 1–9. https://doi.org/10.1007/s13157-021-01397-7

David and Lucile Packard Foundation. (2019). *Marine birds strategy*. The David and Lucile Packard Foundation.

Donnelly, J. P., King, S. L., Silverman, N. L., Collins, D. P., Carrera Gonzalez, E. M., Lafón Terrazas, A., & Moore, J. N. (2020). Climate and human water use diminish wetland networks supporting continental waterbird migration. *Global Change Biology*, *26*, 2042–2059. https://doi.org/10.1111/gcb.15010

Ferraro, P. J., & Hanauer, M. M. (2014). Advances in measuring the environmental and social impacts of environmental programs. *Annual Review of Environment and Resources*, *39*, 495–517. https://doi.org/10.1146/annurev-environ-101813-013230

Ferraro, P. J., & Pattanayak, S. K. (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology*, *4*, e105. https://doi.org/10.1371/journal.pbio.0040105

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M.Ligocki, S., Robinson, O., Hochachka, W., Jaromczyk, L., Rodewald, A., Wood, C., Davies, I., & Spencer, A. (2020). eBird status and trends [Data Version: 2019; Released: 2020]. Cornell Lab of Ornithology. https://doi.org/10.2173/ebirdst.2019

GADM. (2022). *GADM data* [Version 4.0.4]. GADM. https://gadm.org

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gibson, D., Chaplin, M. K., Hunt, K. L., Friedrich, M. J., Weithman, C. E., Addison, L. M., Cavalieri, V., Coleman, S., Cuthbert, F. J., & Fraser, J. D. (2018). Impacts of anthropogenic disturbance on body condition, survival, and site fidelity of nonbreeding Piping Plovers. *The Condor: Ornithological Applications*, *120*, 566–580. https://doi.org/10.1650/CONDOR-17-148.1

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*, 3–28. https://doi.org/10.1890/14-0661.1

International Union for Conservation of Nature (IUCN). (2021). *The IUCN red list of threatened species* [Version 2021-3]. IUCN. https://www.iucnredlist.org

Johnston, A., Auer, T., Fink, D., Strimas-Mackey, M., Iliff, M., Rosenberg, K., Brown, S., Lanctot, R., Rodewald, A., & Kelling, S. (2020). Comparing abundance distributions and range maps in spatial conservation planning for migratory species. *Ecological Applications*, *30*, e02058. https://doi.org/10.1002/eap.2058

Johnston, A., Fink, D., Reynolds, M. D., Hochachka, W. M., Sullivan, B. L., Bruns, N. E., Hallstein, E., Merrifield, M. S., Matsumoto, S., & Kelling, S. (2015). Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications*, *25*, 1749–1756. https://doi.org/10.1890/14-1826.1

Kéry, M. (2010). *Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses*. Academic Press.

Myers, J. P., Morrison, R. I. G., Antas, P. Z., Harrington, B. A., Lovejoy, T. E., Sallaberry, M., Senner, S. E., & Tarak, A. (1987). Conservation strategy for migratory species. *American Scientist*, *75*, 19–26.

Palacios, E., Vargas, J., Fernández, G., & Reiter, M. E. (2022). Impact of human disturbance on the abundance of non-breeding shorebirds in a subtropical wetland. *Biotropica*, *54*, 1160–1169. https://doi.org/10.1111/btp.13139

Piersma, T. (2012). What is habitat quality? Dissecting a research portfolio on shorebirds. In *Birds and habitat: Relationships in changing landscapes* (pp. 383–407). Cambridge University Press.

Ramsar. (2013). *The Ramsar convention manual: A guide to the convention on wetlands* (6th ed.). Ramsar Convention Secretariat,.

Reiter, M., Palacios, E., Eusse-Gonzalez, D., Johnston González, R., Davidson, P., Bradley, D., Clay, R., Strum, K., Chu, J., & Barbaree, B. (2020). A monitoring framework for assessing threats to nonbreeding shorebirds on the Pacific Coast of the Americas. *Avian Conservation and Ecology*, *15*, 7. https://doi.org/10.5751/ACE-01620-150207

Robinson, O. J., Ruiz-Gutierrez, V., Reynolds, M. D., Golet, G. H., Strimas-Mackey, M., & Fink, D. (2020). Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions*, *26*, 976–986. https://doi.org/10.1111/ddi.13068

Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., & Parr, M. (2019). Decline of the North American avifauna. *Science*, *366*, 120–124. https://doi.org/10.1126/science.aaw1313

Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, *185*, 57–70. https://doi.org/10.1016/j.rse.2015.12.024

Runge, C. A., Martin, T. G., Possingham, H. P., Willis, S. G., & Fuller, R. A. (2014). Conserving mobile species. *Frontiers in Ecology and the Environment*, *12*, 395–402. https://doi.org/10.1890/130237

Senner, S. E., Andres, B. A., & Gates, H. R. (Eds.). (2016). *Pacific Americas shorebird conservation strategy*. National Audubon Society.

Sheehan, D. K., Gregory, R. D., Eaton, M. A., Bubb, P. J., & Chenery, A. M. (2010). *The wild bird index—Guidance for national and regional use*. UNEP-WCMC.

Sills, E., Pfaff, A., Andrade, L., Kirkpatrick, J., & Dickson, R. (2020). Investing in local capacity to respond to a federal environmental mandate: Forest & economic impacts of the Green Municipality Program in the Brazilian Amazon. *World Development*, *129*, 104891.

Sills, E. O., Herrera, D., Kirkpatrick, A. J., Brandão, A., Jr., Dickson, R., Hall, S., Pattanayak, S., Shoch, D., Vedoveto, M., & Young, L. (2015). Estimating the impacts of local policy innovation: The synthetic control method applied to tropical deforestation. *PLoS ONE*, *10*, e0132590. https://doi.org/10.1371/journal.pone.0132590

Smith, P. A., Smith, A. C., Andres, B., Francis, C. M., Harrington, B., Friis, C., Morrison, R. G., Paquet, J., Winn, B., & Brown, S. (2023). *Accelerating declines of North America's shorebirds signal the need for urgent conservation action*. Ornithological Applications.

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B. S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K. D., McManus, E., Molnar, J., Recchia, C. A., & Robertson, J. (2007). Marine Ecoregions of the World: A bioregionalization of coastal and shelf areas. *Bioscience*, *59*, 573–583. https://doi.org/10.1641/B570707

Strimas-Mackey, M., Hochachka, W. M., Ruiz-Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S., Fink, D., & Johnston, A. (2020). Best practices for using eBird data [Version 1.0]. Cornell Lab of Ornithology.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., & Farnsworth, A. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, *169*, 31–40. https://doi.org/10.1016/j.biocon.2013.11.003

Ward, D. H., Dau, C. P., Tibbitts, T. L., Sedinger, J. S., Anderson, B. A., & Hines, J. E. (2009). Change in abundance of Pacific brant wintering in Alaska: Evidence of a climate warming effect? *Arctic*, *62*, 301–311.

Wauchope, H. S., Jones, J. P., Geldmann, J., Simmons, B. I., Amano, T., Blanco, D. E., Fuller, R. A., Johnston, A., Langendoen, T., & Mundkur, T. (2022). Protected areas have a mixed impact on waterbirds, but management helps. *Nature*, *605*, 103–107. https://doi.org/10.1038/s41586-022-04617-0

Zhang, J., Zhu, J., Liu, Y., Lu, N., & Fang, W. (2022). The economic impact of payments for water-related ecosystem services on protected areas: A synthetic control analysis. *Water Resources Management*, *36*, 1535–1551. https://doi.org/10.1007/s11269-022-03099-z

SUPPORTING INFORMATION

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

How to cite this article: Donlan, C. J., Eusse-González, D., Luque, G. M., Reiter, M. E., Ruiz-Gutierrez, V., Allen, M. C., Johnston-González, R., Robinson, O. J., Fernández, G., Palacios, E., & Valenzuela, J. (2023). An impact evaluation of conservation investments targeting long-distance migratory species. *Conservation Biology*, e14194. https://doi.org/10.1111/cobi.14194