The Biodiversity Loss from Data Centers

Valentina Bosetti* Michele Fioretti† Christoph Oberthür‡

March 26, 2026

Abstract

Data centers have tripled in number since 2010, yet their local ecological footprint has never been assessed. Using staggered difference-in-differences applied to the US Breeding Bird Survey and eBird citizen-science data, we estimate the causal effect of data center construction on local biodiversity across 3,063 US facilities. Openings reduce bird abundance by 2.4 percent within 25 km—comparable in magnitude to the effect of wind turbines on raptors—while leaving species richness unchanged, indicating community restructuring rather than uniform decline. Migratory specialists are replaced by resident generalists; nocturnal communities shift toward visually oriented predators. Water consumption, light pollution, and habitat disruption operate simultaneously: mitigating one channel leaves the others fully active. Effects concentrate in water-scarce regions, yet current siting practices steer facilities toward precisely those locations. These findings reveal a multi-channel environmental externality that no existing regulation addresses and that single-channel policy proposals cannot resolve.

JEL Codes: Q51, Q57, L86, H23

Keywords: artificial intelligence, biodiversity, data centers

*Bocconi University and RFF-CMCC EIEE. Email: valentina.bosetti@unibocconi.it

†Bocconi University, IGIER, and CEPR. Email: fioretti.m@unibocconi.it

‡Paris School of Economics. Email: christoph.oberthur@psemail.eu

1 Introduction

The digital economy runs on physical infrastructure. Data centers, the purpose-built facilities housing the servers, storage, and networking equipment behind cloud computing, e-commerce, and artificial intelligence, are the fastest-growing class of industrial facility in the United States. They now consume over one percent of world electricity, more than many mid-sized countries, and the International Energy Agency projects this share to double by 2030 (Masanet et al., 2020; International Energy Agency, 2024). A single hyperscale campus draws 50–200 megawatts of continuous power—enough to supply 80,000 to 160,000 homes—consumes 3–5 million gallons of water per day for evaporative cooling, roughly the daily water supply of a town of 30,000 people, and occupies 50–100 acres of land, an area equivalent to 40–75 football fields (Mytton, 2021). More than 3,000 commercial facilities operate in the United States, and state and local governments actively compete to attract them. Virginia's Loudoun County alone hosts the largest concentration of data center capacity in the United States, with minimal environmental review of new approvals.

Yet data centers occupy a remarkable regulatory gap. Power plants emit criteria pollutants subject to the Clean Air Act. Factories discharge effluent regulated under the Clean Water Act (Keiser and Shapiro, 2019). Wind turbines require environmental impact assessments under the National Environmental Policy Act. Data centers face none of these requirements. Their primary environmental channels (water consumption, continuous light and noise emissions, land conversion, and induced electricity demand) fall largely outside existing regulatory frameworks. Unlike a coal plant or a chemical factory, a data center produces no visible smoke, no odor, and no effluent discharge; from the outside, it is an anonymous concrete box surrounded by security fencing and parking lots. This apparent benignity has allowed the industry to expand with almost no scrutiny of its local environmental footprint. Whether this gap has ecological consequences is the question we address.

This paper asks a question at the intersection of environmental and infrastructure economics: when a new class of industrial facility expands rapidly without environmental assessment, what are the local ecological costs? We study this in the context of data centers, whose staggered construction across the United States between 2000 and 2024 provides a natural experiment. The industry began with small colocation facilities embedded in commercial zones during the early 2000s and evolved, after 2010, into hyperscale campuses owned by Amazon, Google, Microsoft, and Meta, facilities that individually consume more water than a small city and more electricity than a steel mill. This evolution means that facilities built in different locations at different times vary dramatically in their resource intensity, generating the staggered variation in treatment timing that our empirical strategy exploits.

We use the US Geological Survey's Breeding Bird Survey (BBS)—a standardized roadside protocol in which trained observers count every bird species detected along roughly 40-kilometer routes each June—as our primary outcome data. Standardized bird-survey routes whose full geometry falls within a given distance of a new facility at different points in time serve as each other's counterfactual, allowing us to estimate the causal effect of facility openings on local biodiversity. We complement the BBS with eBird, the world's largest citizen-science bird observation platform, which we use to study nocturnal species that the daytime BBS protocol does not cover. Birds are among the most reliable indicators of broader ecosystem health, correlating strongly with changes in other taxa, habitat quality, and environmental conditions (Gregory and van Strien, 2010; Sekercioglu, 2006). Identifying what drives their decline is essential for conservation policy (IPBES, 2019), especially as recent evidence shows that public conservation investments can measurably increase bird populations (Hogan et al., 2025)—but only if the sources of ecological damage are understood. North America has already lost nearly three billion birds since 1970 (Rosenberg et al., 2019).

We assemble a novel facility-level database of 3,063 data centers from commercial real estate records, recovering missing construction dates for 30 percent of facilities through a satellite detection pipeline that combines optical change detection with multi-spectral index ensembles, validated against Google Earth historical imagery. We combine this with the BBS and eBird in a staggered difference-in-differences framework using the Sun and Abraham (2021) interaction-weighted estimator, which is robust to heterogeneous treatment effects across cohorts, a first-order concern given the dramatic evolution of the industry over our sample period. To trace the pathways from facility operations to ecological outcomes, we draw on Landsat-derived vegetation indices, VIIRS nighttime radiance imagery, species-level acoustic data, power plant records from EIA Form 860, and groundwater monitoring from the USGS.

The main finding is that data center construction reduces local bird abundance while leaving species richness unchanged. This combination is the signature of community restructuring rather than ecological destruction: some species lose individuals while others are unaffected or even increase, leaving the species count roughly stable but reducing total numbers. The abundance effect is economically meaningful, statistically significant, and robust across specifications. It decays monotonically with distance from the facility, strongest within 10 kilometers and detectable out to 100 kilometers, and shows no positive spillover at any range, ruling out spatial displacement and indicating true population loss rather than redistribution. The effects concentrate in water-stressed regions, where the ecological carrying capacity is lowest and the marginal cost of water extraction is highest.

Beneath the aggregate, community composition shifts toward homogenization (McKinney and Lockwood, 1999): dominant species increase their share, migratory

specialists decline relative to resident generalists, and nocturnal communities restructure along the light-pollution gradient. Multiple environmental channels (water consumption, light emission, habitat conversion, and noise) operate simultaneously, and mitigating one leaves the others active. This multi-channel architecture distinguishes data centers from infrastructure externalities that operate through a single pathway, and it implies that single-channel regulation will capture only a fraction of the total ecological cost.

The policy implication extends further. Current siting decisions optimize over a narrow set of private costs: land prices, electricity rates, and fiber connectivity. Ecosystem services (pollination, pest control, water regulation, the existence value of biodiversity) are absent from this calculus. The ecological costs we document are real, spatially concentrated, and disproportionately borne by water-stressed communities. A data center in Phoenix or Dallas imposes substantially higher ecological costs than the same facility in Portland or Charlotte, yet nothing in the current framework reflects this geographic variation in environmental carrying capacity. Reforming siting policy to internalize these costs, whether through environmental impact assessments, water-stress-conditional guidelines, or closed-loop cooling mandates, would address the most consequential channels at modest cost relative to total facility budgets.

Our work connects to several literatures. Most directly, it extends the growing body of work on infrastructure externalities and biodiversity. Katovich (2024) estimates that wind turbines reduce raptor abundance by 5–10 percent using a staggered difference-in-differences design, a single-channel externality (collision mortality) with a well-understood physical mechanism. Duflo and Pande (2007) study the downstream effects of dams on agricultural productivity. Loss et al. (2014) documents bird–building collisions. Each of these settings involves one dominant pathway from infrastructure to ecological harm. Data centers differ fundamentally: they affect ecosystems through multiple simultaneous channels, none of which is individually regulated, and no prior work has estimated these effects.

A related strand of work studies the environmental costs of industrial activity subject to regulation. Greenstone (2002) asks what the Clean Air Act costs industry; Currie et al. (2015) trace the local health effects of toxic plant openings; Keiser and Shapiro (2019) estimate the benefits of the Clean Water Act. Shapiro and Walker (2018) and Fowlie et al. (2012) study pollution in sectors where firms face binding environmental constraints. We ask the mirror-image question: what are the ecological costs when a resource-intensive industry expands outside of any environmental regulatory framework? Data centers occupy precisely this gap: facilities whose resource footprint is substantial but falls outside any existing constraint.

The water economics literature establishes that unpriced extraction generates classic externalities (Olmstead, 2010; Hagerty, 2021). Data centers are a new, large-scale en-

trant into local water markets whose consumption is invisible because they draw from municipal supply rather than private wells. The engineering literature documents the energy and water footprint of data centers through lifecycle analysis (Obringer et al., 2021; Masanet et al., 2020), while Faber and Krause (2023) studies the environmental costs of e-commerce logistics. We move beyond accounting to causal estimates of how resource consumption translates into ecological outcomes, showing that the water channel matters most where water is scarce and that the ecological cost varies dramatically with geography in ways that lifecycle analysis cannot capture.

Finally, our finding of community restructuring beneath stable aggregate metrics connects to the ecology of biotic homogenization, the systematic replacement of specialist species by widespread generalists under human disturbance (McKinney and Lockwood, 1999; Newbold et al., 2015). This phenomenon is well documented in the natural sciences but largely absent from environmental economics, where species richness remains the default metric of ecological health despite its well-known insensitivity to the functional changes that drive ecosystem service provision. Our results provide a direct empirical manifestation: standard biodiversity indicators give false reassurance while the composition of local communities shifts in ways that degrade specific ecosystem services.

We make three contributions relative to these literatures. First, we bring quasi-experimental evidence to the question of how digital infrastructure affects ecosystems, filling the gap between engineering lifecycle analysis and ecological outcomes. The staggered rollout of 3,063 facilities across diverse geographies, combined with a standardized monitoring program that has operated under the same protocol for decades, provides a credible setting for causal inference on a question that has so far been addressed only through engineering models and correlational studies. Second, we identify multiple active mechanisms operating simultaneously, distinguishing our setting from single-channel infrastructure externalities and demonstrating that mitigating one channel (for example, water recycling) leaves others fully active, a finding with direct implications for policy proposals that target individual channels in isolation. The multi-channel architecture of the externality means that a comprehensive assessment is necessary: the ecological cost is not the sum of independently measurable parts but emerges from the interaction among water stress, light pollution, habitat conversion, and noise in ways that vary with local environmental conditions. Third, we show that aggregate biodiversity metrics mask community restructuring: standard indicators such as species richness and Shannon diversity give false reassurance while beneath them the composition of local communities shifts in ways that degrade specific ecosystem services. This is a caution for environmental impact assessment broadly, not only in the context of data centers but for any infrastructure whose ecological effects operate through compositional change rather than outright species loss.

The remainder of the paper is organized as follows. Section 2 presents the data. Section 3 describes the empirical approach. Section 4 reports the main results and heterogeneity. Section 5 investigates mechanisms. Section 6 discusses interpretation and policy implications. Section 7 concludes.

2 Data

We combine four categories of data: a facility-level database of US data centers, two independent bird monitoring programs, satellite-derived environmental measures, and ancillary controls.

2.1 Data Center Facilities

Our facility data come from Datacentermap, a commercial database of 9,600 entries from 3,335 operators worldwide with geocoded locations, facility type classifications, and operational characteristics. We merge this with CoStar commercial real estate records to obtain building characteristics. After removing non-operational sites and facilities outside the continental United States, we retain 3,063 facilities.

Three facility types are relevant for our analysis, and they differ substantially in their resource consumption profiles and, as we show, in their ecological footprints. *Hyperscaler* data centers (574 facilities) are purpose-built campuses owned and operated by the major cloud providers (Amazon Web Services, Google, Microsoft, Meta) to serve their own computing workloads. They are the largest facilities by power draw (a median of 46 MW, roughly seven times the median non-hyperscaler) and tend to locate on greenfield land in less urbanized areas, close to cheap power and fiber routes rather than to end users. Hyperscalers concentrate geographically: Oregon hosts 52 percent of hyperscaler facilities in our data, followed by Virginia (47 percent) and Ohio (40 percent). They locate in greener areas (median pre-treatment LAI of 0.476 versus 0.306 for non-hyperscalers) and darker areas (median pre-treatment VIIRS radiance of 21.3 versus 40.1 $\text{nW}/\text{cm}^2/\text{sr}$), consistent with greenfield development on previously undeveloped land.

Colocation facilities (896) are shared data centers operated by third-party providers (Equinix, Digital Realty, CyrusOne) that lease rack space, power, and cooling to multiple tenants. They tend to locate in metropolitan areas near the businesses they serve and are typically embedded in existing commercial or light-industrial zones. Their individual water and power consumption is smaller than hyperscaler campuses, but their collective footprint can be substantial: a metropolitan area may host dozens of colocation facilities drawing on the same municipal water system.

The remaining 1,593 *other/enterprise* facilities comprise single-firm data centers (a bank's private server room, a hospital's on-premises computing) and smaller facilities that do not fit neatly into either category. These are the smallest facilities by power draw but the most numerous, and as we show in Section 4.3, their cumulative effect is at least as large as the hyperscaler effect.

Appendix A provides a detailed comparison of facility types across size (Figure A1), geography and water stress (Figure A2), and pre-treatment environmental conditions (Figure A3).

Start-year recovery. Operational start years are available in the database for 2,152 facilities (70 percent). We recover dates for the remaining 30 percent using a satellite detection pipeline that combines Continuous Change Detection and Classification (CCDC) optical breakpoints (Zhu and Woodcock, 2014) with a Sentinel-2 multi-spectral index ensemble, assigning each facility its highest-confidence estimate through a priority cascade. Validation against database-dated facilities shows a mean absolute error of 1.3 years for high-confidence inferences (88 percent within ± 2 years). We visually verified 196 facilities using Google Earth Pro historical imagery, excluding 19 false positives and correcting 46 wrong-year assignments.

Figure 1 shows the spatial distribution of all 3,063 facilities alongside 2,931 BBS survey routes. Data centers cluster in a few metropolitan corridors (northern Virginia, Dallas–Fort Worth, the San Francisco Bay Area, and Chicago) but a long tail extends into rural areas across the South, Mountain West, and Pacific Northwest. BBS coverage is dense and broadly national, ensuring the data center expansion intersects with monitoring infrastructure across a wide range of habitats. Of the 2,931 active routes, 441 (15 percent) fall within 25 km of at least one operational data center, forming our treated group.

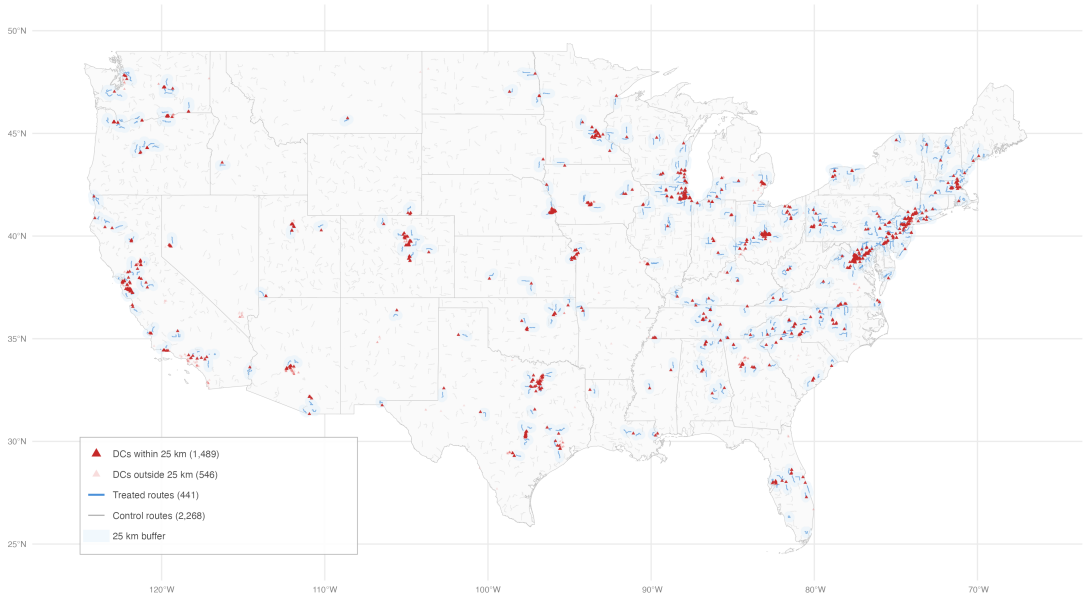
Figure 2 reveals two eras in data center construction. Before 2010, facilities entered at a pace of roughly 50–100 per year, predominantly small colocation and enterprise operations. After 2010—coinciding with the rise of cloud computing and the shift from air-cooled chillers to evaporative cooling systems that consume substantially more water—hyperscaler campuses account for a growing share. This temporal structure provides the staggered variation that identifies our treatment effects.

2.2 Bird Biodiversity Data

2.2.1 Breeding Bird Survey (BBS)

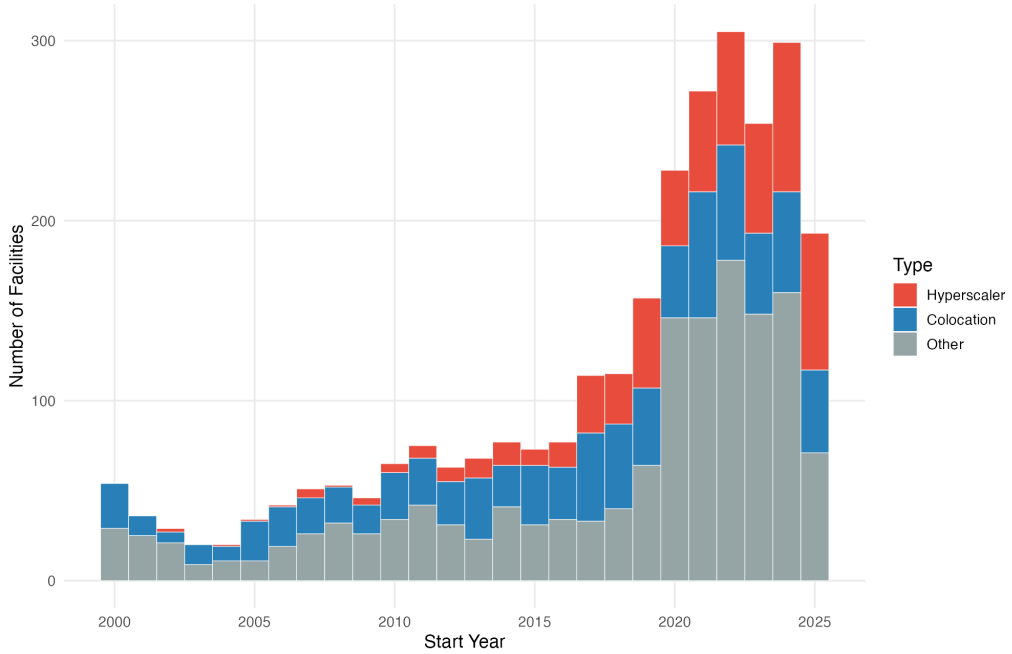
The BBS is a standardized avian monitoring program administered by the US Geological Survey. Each June, trained observers drive a 39.4 km route, stopping at 50 evenly spaced

Figure 1: Data Center Locations and BBS Survey Routes



Notes: Map shows 3,063 data center facilities (red points) and 2,931 BBS survey routes (blue lines) across the continental United States. Virginia's "Data Center Alley" in Loudoun County is the most concentrated cluster. Of the 2,931 active routes, 441 have their full route linestring geometry within 25 km of at least one operational data center (treated); 2,353 routes more than 50 km from any facility serve as never-treated controls (after removing 137 routes near excluded-but-real facilities). Source: Datacentermap; USGS BBS.

Figure 2: Data Center Construction Timeline



Notes: Annual count of new data center facilities becoming operational, 2000–2024. Colors distinguish hyperscaler (Amazon, Google, Microsoft, Meta), colocation, and other/enterprise facilities. The acceleration after 2015 reflects the cloud computing and AI boom. The median hyperscaler facility opened in 2021, compared to 2018 for colocation facilities. Source: Datacentermap; CoStar.

points to conduct three-minute point counts of all birds seen or heard within a 400-meter radius. The protocol has been essentially unchanged since 1966.

Our sample covers 2,931 active routes observed at least five times during 2000–2024. We construct two primary outcomes at the route-year level: *species richness* (the count of distinct species detected) and *total abundance* (the sum of individuals across all species), both transformed using the inverse hyperbolic sine.[1]

We also construct guild-specific outcomes (forest songbirds, aerial insectivores, urban-adapted species, raptors), community-structure measures (Shannon diversity, Pielou's evenness, Berger-Parker dominance), and migratory composition metrics (share of migratory species, neotropical migrant abundance).

Treatment assignment. A route is treated when a data center begins operating within X km of its full route linestring geometry (the driven path, not the starting point). We use 10 km and 25 km buffers in the main analysis. Never-treated routes, those more than 50 km from any data center throughout the sample, serve as controls.

2.2.2 eBird Nocturnal Species

The BBS surveys routes during early morning hours and does not adequately sample nocturnal species—owls, nightjars, and other night-active birds that are most directly exposed to the 24-hour operational profile of data centers. To test the light-pollution channel, we use eBird, the world's largest citizen-science bird observation platform, with over 59 million checklists submitted by 539,000 observers during 2000–2025. We apply standard eBird quality filters: checklist duration 5–300 minutes, effort distance $\leq$10 km, group size $\leq$10. We select 19 nocturnal and crepuscular species based on three criteria: (i) primarily active at night or during twilight, maximizing exposure to the continuous lighting and noise profile of data centers; (ii) sufficiently common in eBird records across the continental United States to support statistical inference at the 25 km grid-cell level; and (iii) spanning distinct ecological niches—visual predators (owls), acoustic insectivores (nightjars), and ground-foraging or ground-nesting crepuscular species—to permit within-analysis tests of whether light pollution differentially affects species by foraging strategy. The 19 species comprise 9 owls (Great Horned, Barred, Eastern Screech, Western Screech, Barn, Burrowing, Short-eared, Long-eared, Northern Saw-whet), 5 nightjars (Common Nighthawk, Chuck-will's-widow, Eastern Whip-poor-will, Common Poorwill, Lesser Nighthawk), and 5 crepuscular species (Killdeer, American Woodcock, Black-crowned Night Heron, Yellow-crowned Night Heron, Mexican Whip-poor-will).

We deliberately include checklists from all hours of the day, not only nighttime

[1]The IHS transformation, $\text{IHS}(x) = \ln(x + \sqrt{x^2 + 1})$, approximates the natural log for large values while being defined at zero.

surveys. Many nocturnal species are detectable during daytime—owls are located at roost sites, nightjars are flushed from ground cover, and crepuscular species are active at dawn and dusk—and restricting to nighttime checklists alone would discard the majority of eBird effort and drastically reduce statistical power. Instead, we control for the share of nocturnal-hour checklists (7 PM–5 AM) in each cell-year, allowing the regression to absorb differences in survey timing across cells while retaining all available detections.

A critical methodological step is *proper zero construction*. Raw eBird species downloads contain only positive detections, meaning checklists where the species was observed. This creates a selection problem: a cell-year with no detections could reflect either genuine absence of nocturnal birds or simply no birders visiting at night. We address this by loading the complete eBird sampling file (all 59 million checklists, regardless of whether nocturnal species were detected) and left-joining nocturnal detections onto it. Any complete checklist without a nocturnal detection is coded as a zero. This procedure substantially increases sample size and, as we show, produces larger treatment-effect estimates than a detection-only analysis, consistent with proper zeros reducing attenuation bias from non-random effort allocation.

The resulting dataset comprises 11.1 million records across 207,000 observers and 15,748 hexagonal grid cells (25 km H3 cells). Following Madhok (2025), we include observer$\times$year fixed effects to absorb citizen-science learning bias, the tendency for birders to detect more species as they gain experience, a confound that is particularly important for nocturnal species, which require specialized knowledge to identify by call.

2.3 Environmental and Mechanism Data

Vegetation health. To test whether data centers degrade local habitat, we measure vegetation density using Leaf Area Index (LAI), a dimensionless ratio of total one-sided leaf area to ground area, from the Landsat satellite archive at 30-meter resolution (2000–2025), extracted via Google Earth Engine. LAI captures the greenness and density of vegetation canopy and is a standard indicator of ecosystem productivity used in remote sensing and ecology. We compute annual mean LAI in concentric donut buffers around each data center facility at 5, 15, 25, and 50 km. The donut design excludes the inner 1 km ring around each facility, separating the mechanical effect of land clearing at the construction footprint (which is trivially expected) from the broader watershed-level vegetation response that would indicate water stress or hydrological disruption extending well beyond the facility boundary.

Nighttime light. To measure the light-pollution channel, we use VIIRS (Visible Infrared Imaging Radiometer Suite) Day/Night Band radiance at 750-meter resolution

(2012–2024). VIIRS captures artificial light emitted at night, measured in nanowatts per square centimeter per steradian ($nW/cm^2/sr$). We compute mean annual radiance in full circles of 1, 3, 5, and 15 km around each data center facility. Unlike LAI, we use full circles rather than donuts because the facility is the light source: radiance radiates outward from the campus, and the full circle captures the total nocturnal light environment that wildlife experiences. VIIRS data begin in 2012, limiting the pre-treatment window for facilities that opened before approximately 2014.

Water stress. Baseline water stress scores from the WRI Aqueduct 4.0 dataset measure the ratio of total water withdrawals to available renewable surface and groundwater supply at the watershed level. We spatially join each data center to its containing HydroBASINS polygon and classify facilities as high stress (above the 40th percentile, corresponding to WRI categories "High" and "Extremely High") or low stress. Water stress serves as a cross-sectional heterogeneity moderator; it does not vary over time within a facility, allowing us to test whether the ecological effects of data centers are amplified where water resources are already constrained. Of the 3,063 facilities, 1,131 are classified as high stress and 1,929 as low stress (3 facilities in arid regions with negligible water use are unclassified).

Grid capacity. To test whether data centers induce local electricity grid expansion, we draw on EIA Form 860, which provides the nameplate generating capacity (MW), fuel type, and location of all US electricity generators. The 2024 snapshot contains 31,283 generators at 14,184 power plants across the continental United States. We compute cumulative generation capacity within 10, 25, and 50 km of each BBS route for each year, and separately compute the clean energy share (solar, wind, hydro, nuclear, battery, and geothermal as a fraction of total capacity) to assess whether data center-induced grid expansion brings clean or fossil fuel generation.

Controls. Time-varying controls include BBS-recorded weather conditions (mean temperature, wind speed, sky condition, and survey start time), which account for the well-documented sensitivity of bird detection rates to survey-day conditions. A first-year observer indicator controls for the tendency of new surveyors to undercount species in their initial year on a route. NLCD land-cover shares (percent developed, forest, and agriculture within the route's buffer) capture ongoing changes in the built environment around each route, addressing the concern that data center areas may be on different development trajectories regardless of the facilities. For LAI regressions, we include the Palmer Drought Severity Index (PDSI) from GRIDMET as a climate control, since vegetation density responds to drought independently of data center water consumption.

Table 1 reports summary statistics across the four analysis samples. Panel A describes the 3,063 data center facilities: 19 percent are hyperscaler, 29 percent colocation,

and 52 percent other/enterprise, with a mean start year of 2012. Panel B covers the BBS route-year panel: the average route detects 53 species and 368 individual birds per survey, and 16 percent of routes are treated at the 25 km buffer. Panel C describes the facility-level LAI panel used for the vegetation mechanism analysis: the mean LAI is 1.76 (m^2 leaf area per m^2 ground), with substantial variation reflecting the range of climates in which data centers operate. Panel D covers the eBird nocturnal sample: the average cell detects fewer than 2 nocturnal species and 4 individuals, reflecting the inherent rarity of nocturnal bird observations, but the large number of cells (15,748) and checklists (mean 705 per cell) provides sufficient statistical power.

Table 2 compares pre-treatment characteristics of treated and control routes at 25 km. Two features are noteworthy. First, treated routes are substantially more developed at baseline (16.2 percent versus 5.8 percent), consistent with data centers co-locating near existing infrastructure: highways, fiber routes, and electrical substations. They also have slightly more agricultural land (30.6 versus 25.4 percent) and comparable forest cover. This selection on observables is expected and does not threaten identification: route fixed effects absorb all level differences, and time-varying NLCD shares control for differential trends. Second, pre-treatment biodiversity levels are similar across the two groups: treated routes have slightly higher richness (54.7 versus 51.4) and abundance (390.2 versus 350.3), but these differences are small relative to the within-route variation that identifies the treatment effect. The parallel-trends tests in Section 3 provide direct evidence that treated and control routes were on indistinguishable biodiversity trajectories before data center construction.[2]

3 Empirical Approach

3.1 Specification

We estimate the effect of data center openings on biodiversity using a staggered difference-in-differences design. The baseline specification is:

$$y_{it} = \alpha_i + \gamma_t + \sum_{k \neq -1} \beta_k D_{it}^k + X_{it}'\delta + \varepsilon_{it} \tag{1}$$

where i indexes BBS routes, t indexes years, α_i are route fixed effects, γ_t are year fixed effects, D_{it}^k are relative-time indicators equal to one when route i is k periods from first treatment, and X_{it} is a vector of time-varying controls. Standard errors are clustered

[2]As a further check, we re-estimate the main specification on a propensity-score-matched sample, using nearest-neighbor 1:1 matching with replacement on baseline NLCD land-cover shares, state-level electricity prices, and fiber infrastructure proximity. The matched sample eliminates the observable differences in Table 2 and produces qualitatively identical results, with the abundance ATT at 10 km strengthening relative to the full-sample estimate.

Table 1: Summary Statistics

	Mean	SD	Min	Max
Panel A: Data Center Facilities (N = 2,060)				
Hyperscaler	0.29	0.45	0	1
Colocation	0.45	0.50	0	1
Other/Enterprise	0.26	0.44	0	1
Start year	2016.6	6.2	2001	2024
Panel B: BBS Routes (N = 48,121 route-years, 2,794 routes, 2000–2024)				
Species richness	53.0	14.1	1	127
Total abundance	367.8	134.3	1	1,842
Ever-treated (25 km)	0.16		0	1
Ever-treated (10 km)	0.05		0	1
Panel C: LAI Facility Buffers (N = 2,739 facilities, 2000–2025)				
LAI (15 km donut)	1.76	1.12	0.01	5.89
Treated	0.69	0.46	0	1
Panel D: eBird Nocturnal Cells (N = 15,748 cells, 2000–2025)				
Nocturnal richness	1.82	1.41	0	15
Nocturnal abundance	3.94	6.32	0	412
N checklists/cell	705	2,148	1	89,312

Notes: Panel A: 2,060 data center facilities with operational start years 2001–2024, from DCM-CoStar after cleaning (false positives excluded, visually verified start years prioritized). Panel B: BBS route-year observations with $\geq$5 survey years during 2000–2024. Panel C: Landsat LAI extracted via Google Earth Engine in donut buffers. Panel D: 19 nocturnal/crepuscular species from eBird; proper zero construction from 59M complete checklists. Source: Datacentermap; CoStar; USGS BBS; eBird; Google Earth Engine.

at the route level. The outcome y_{it} is the IHS-transformed species richness or total abundance at the route-year level.

Route fixed effects absorb all time-invariant route characteristics: geographic location, elevation, surrounding land use, baseline habitat quality, proximity to highways and urban areas. These are important because data centers tend to locate near existing infrastructure (highways, fiber routes, electrical substations), which correlates with pre-existing development pressure. Year fixed effects absorb common temporal shocks that affect all routes simultaneously: national bird population trends driven by continent-wide habitat loss, broad weather patterns, and any changes to the BBS survey protocol.

Time-varying controls address remaining confounders. BBS-recorded weather conditions (mean temperature, wind speed, sky condition, survey start time) account for the well-documented sensitivity of bird detection to survey-day conditions. A first-year observer indicator controls for the "rookie effect," the tendency for new observers to detect fewer species in their first year on a route. NLCD land-cover shares (percent developed, forest, agriculture within the route's buffer) capture ongoing land-

Table 2: Pre-Treatment Balance: Treated vs. Control Routes (25 km)

	Treated	Control	Diff.	t-stat
Routes	441	2,353		
Route-years	7,702	40,419		
Biodiversity outcomes				
Mean richness	54.7	51.4	3.3	5.89
Mean abundance	390.2	350.3	39.9	7.66
NLCD land cover (2001 baseline, %)				
% Developed	16.2	5.8	10.4	11.33
% Forest	32.9	30.1	2.8	1.65
% Agriculture	30.6	25.4	5.2	3.40
Geography				
Latitude (°N)	39.0	40.1	−1.0	−3.62

Notes: Route-level pre-treatment means (2000–2005) for BBS routes within 25 km of a data center (treated, based on full route linestring geometry) vs. routes >50 km from any data center (control). t-statistics from two-sample t-tests of equality of means. NLCD land-cover shares are baseline (2001) values within the route's 25 km buffer. Treated routes are significantly more developed (16.2% vs. 5.8%), consistent with data centers co-locating near existing infrastructure. These level differences are absorbed by route fixed effects; time-varying NLCD shares control for differential land-use trends. Source: USGS BBS; National Land Cover Database via Google Earth Engine.

use change, addressing the concern that areas near data centers may be on different development trajectories independent of the facilities themselves.

We estimate treatment effects using the Sun and Abraham (2021) interaction-weighted estimator, which recovers cohort-specific effects and aggregates them using appropriate weights. Conventional two-way fixed effects estimation produces biased estimates in staggered adoption settings when treatment effects vary across cohorts (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Borusyak et al., 2024), a first-order concern given the dramatic evolution of the data center industry over our sample period, from small colocation facilities to hyperscale campuses consuming orders of magnitude more resources. The reference period is $k = -1$, the year immediately before treatment.

3.2 Treatment and Control Group Construction

Treatment definition. Route i is treated in year t if at least one data center is operational within X km of the route's full linestring geometry—the driven path recorded by USGS, not merely the starting point. This distinction matters: BBS routes are approximately

40 km long, and defining treatment from the start point alone would misclassify routes whose midpoint or endpoint passes near a facility. We use 10 km and 25 km buffers in the main analysis, with additional distances (5, 50, 75, 100 km) in the spatial gradient analysis (Section 4.2). Treatment is absorbing: once a route enters the treatment group, it remains treated for all subsequent years, even if additional facilities open nearby. At 25 km, 441 routes are treated; at 10 km, 137 routes.

Control group. Never-treated routes, those whose full geometry remains more than 50 km from any data center throughout the sample period, serve as the control group. The 50 km exclusion buffer is chosen to ensure a clean separation between treated and control routes: if we used all untreated routes regardless of distance, some "control" routes would be within 30 or 40 km of a data center, potentially experiencing partial treatment effects that would attenuate our estimates. The spatial decay results in Section 4.2 validate this choice by showing that treatment effects extend to approximately 50 km.

A subtlety in control-group construction deserves attention. Our facility database contains 3,063 data centers with usable start-year estimates, but the broader database includes additional facilities that were excluded from the analysis sample, for example because they lacked any start-year information (even from satellite detection) or fell outside the sample window. These excluded facilities are real data centers that affect local ecosystems; routes near them are not clean controls. We identify 974 such excluded-but-real facilities and drop 137 control routes that fall within 25 km of them. This "contaminated control" removal ensures that the control group consists only of routes that are genuinely unexposed to data center effects, not routes near facilities that happen to be missing from our analysis sample. The treated group is unaffected: a route near both an in-sample and an excluded facility remains treated based on the in-sample facility.

After these adjustments, the final sample at 25 km comprises 441 treated routes and 2,353 control routes, yielding 48,121 route-year observations over the 2000–2024 panel.

3.3 Facility-Level Specifications

We apply the same staggered DiD framework to facility-level environmental outcomes to test specific mechanisms. For vegetation health:

$$\text{LAI}_{jt} = \alpha_j + \gamma_t + \sum_{k \neq -1} \beta_k \, D_{jt}^k + \text{PDSI}_{jt} \, \delta + \varepsilon_{jt} \tag{2}$$

where j indexes data center facilities, LAI is the Landsat-derived Leaf Area Index measured in concentric donut buffers around each facility, and PDSI is the Palmer Drought Severity Index from GRIDMET as a time-varying climate control. The donut

design excludes the inner 1 km around each facility to separate the mechanical effect of land clearing at the construction footprint from the broader watershed-level vegetation response. Facility fixed effects and year fixed effects, with standard errors clustered at the facility level. The treated group comprises facilities that opened during 2001–2022 (requiring at least one pre-treatment year in the panel, which begins in 2000). The control group consists of not-yet-treated facilities—those with later start dates that serve as counterfactuals for earlier cohorts before they themselves become treated—plus a small number of planned or under-construction facilities that had not yet become operational by the end of the sample period. This is the standard control-group structure for the Sun-Abraham estimator in staggered adoption settings.

VIIRS nighttime radiance uses the same panel structure with raw radiance ($\mathrm{nW/cm^2/sr}$) as the outcome, measured at 1, 3, 5, and 15 km full circles around each facility. We use full circles rather than donuts because the facility is the light source: radiance radiates outward, and the full circle captures the total light environment that nocturnal species experience.

For eBird nocturnal species, we estimate cell-level Sun-Abraham regressions on 25 km hexagonal grid cells. The specification includes cell fixed effects, year fixed effects, effort controls (number of checklists, number of observers, mean checklist duration, mean effort distance), and observer$\times$year fixed effects following Madhok (2025). The observer$\times$year interaction is critical for citizen-science data: it absorbs the tendency for individual birders to detect more species as they gain experience, a confound that could otherwise bias treatment-effect estimates if more experienced observers happen to survey cells near data centers.

3.4 Identification

The identifying assumption is that, absent data center construction, treated and control units would have followed parallel biodiversity trajectories. Several pieces of evidence support this assumption.

Pre-treatment placebo tests are the most direct assessment. For the primary BBS specification (abundance, 25 km), event-study coefficients at $t = -4$ and $t = -5$ are small and statistically insignificant, but the coefficients at $t = -2$ and $t = -3$ are negative and individually significant. We interpret this pattern as construction-phase anticipation rather than a parallel-trends violation: data center construction typically begins 18–36 months before the operational start date, and the negative coefficients are confined to the construction window rather than extending to earlier periods. The pattern holds separately for richness (which is precisely null at all pre-treatment periods), and pre-treatment coefficients are flat in the facility-level LAI and VIIRS analyses.

Our treatment is defined at the facility's operational start date—when servers begin running and the continuous environmental channels (evaporative cooling, security lighting, HVAC noise) activate. Construction-phase disruption (site clearing, earthwork, heavy equipment) begins earlier and may itself affect local bird populations, particularly at shorter distances. The pre-treatment dip at $t = -2$ and $t = -3$ is consistent with this timeline. Importantly, the post-treatment effect grows substantially beyond the pre-treatment level, reaching -0.05 by $t = +5$, indicating that operational impacts compound the construction-phase effect rather than the entire result being driven by construction alone.

Section 4.2 provides a direct test of the stable unit treatment value assumption (SUTVA) through the spatial decay analysis. Effects decline monotonically with distance from the nearest facility and show no positive coefficient at any distance band. If data centers were merely displacing birds to nearby areas rather than reducing populations, we would observe positive treatment effects at intermediate distances as displaced birds concentrate in the surrounding landscape. The absence of any such pattern rules out spatial displacement and supports the interpretation of true population loss.

4 Results

4.1 Aggregate Effects

Table 3 reports the Sun-Abraham average treatment effects on the treated (ATT) for BBS routes within 10 and 25 km of data centers. The headline result is at 25 km: bird abundance declines by 2.4 percent (ATT $= -0.024$, $p = 0.029$), while species richness is statistically indistinguishable from zero (ATT $= +0.001$, $p = 0.89$). This combination, significant abundance decline with null richness, is informative about the nature of the ecological impact. If data centers caused uniform habitat destruction, we would expect both richness and abundance to decline: fewer species and fewer individuals. Instead, the pattern implies selective population loss. Some species lose individuals, perhaps many individuals, while others are unaffected or even increase, leaving the species count roughly unchanged but reducing total numbers. The aggregate abundance decline is the *net* effect of heterogeneous species-level responses, a point we develop in Section 5.1.

At 10 km, the point estimate is roughly twice as large (ATT $= -0.051$, $p < 0.001$) with 137 treated routes. The larger effect at closer distance is consistent with localized environmental stressors whose intensity diminishes with distance from the facility.

To put the magnitude in concrete terms, the mean BBS route records approximately 364 individual birds across 53 species per survey. A 2.4 percent decline at 25 km translates to roughly 9 fewer birds per route per year; at 10 km the 5.1 percent effect

Table 3: Main Results: Bird Abundance and Richness

	10 km buffer		25 km buffer	
	Richness	Abundance	Richness	Abundance
SA ATT	−0.008	−0.051***	+0.001	−0.024**
	(0.007)	(0.015)	(0.008)	(0.011)
Treated routes	137	137	441	441
Control routes	2,751	2,751	2,353	2,353
Route-years	48,121	48,121	48,121	48,121

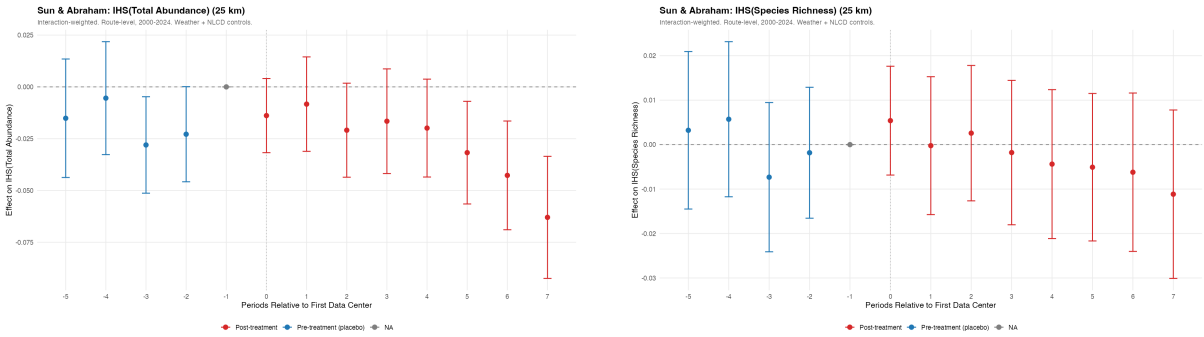
Notes: Sun-Abraham interaction-weighted ATT estimates. Dependent variables: IHS(richness) and IHS(abundance). All specifications include route fixed effects, year fixed effects, weather controls (temperature, wind speed, sky condition, survey start time), first-year observer indicator, and NLCD land-cover shares (percent developed, forest, agriculture). No anticipation imposed. Standard errors clustered at the route level in parentheses. ***p<0.01, **p<0.05, *p<0.1. Source: USGS BBS; Datacentermap.

implies approximately 19 fewer birds. These are net losses after accounting for route and year fixed effects; that is, relative to what the same route would have recorded absent data center construction.

For broader context, Rosenberg et al. (2019) document a 29 percent decline in North American bird abundance over 50 years, equivalent to roughly 0.7 percent per year. A single data center causing a 2.4 percent decline thus compresses approximately three years of background population loss into the footprint of one facility. Compared to other infrastructure, Katovich's (2024) wind turbine estimate of 5–10 percent raptor decline operates through direct mortality at very short range (<1 km), while the data center effect operates through habitat degradation at much larger spatial scales (10–25 km) and across the entire avian community rather than a single guild.

Figure 3 presents the event-study coefficients at 25 km. Three features are noteworthy. First, pre-treatment coefficients at $t = -4$ and $t = -5$ are centered on zero, but the coefficients at $t = -2$ and $t = -3$ are negative and individually significant. We attribute this to construction-phase disruption: data center construction typically begins 18–36 months before the operational start date that defines our treatment, involving site clearing, earthwork, and heavy equipment activity that directly disturbs local habitat. The fact that the pre-trend is confined to the two years immediately before treatment—rather than extending further back—is consistent with a construction anticipation effect rather than a violation of parallel trends. Second, the abundance effect grows substantially larger after treatment, reaching approximately -0.05 by five years post-opening. This post-treatment acceleration beyond the pre-trend level indicates that operational impacts (water consumption, continuous lighting, noise) compound the construction-phase effect. Third, the richness event study is flat both before and after treatment, confirming that species richness is a null result rather than a noisy

Figure 3: Event Studies: Abundance and Richness (25 km)



(a) Total Abundance **(b)** Species Richness

Notes: Sun-Abraham event-study coefficients for IHS(abundance) and IHS(richness) at 25 km. 95% confidence intervals shown. Reference period: $t = -1$. Abundance coefficients at $t = -2$ and $t = -3$ are negative, consistent with construction-phase disruption beginning before the operational start date. The effect grows substantially after treatment, reaching -0.05 by five years post-opening. Richness is precisely null both before and after treatment. Same specification as Table 3. Source: USGS BBS; Datacentermap.

negative: the confidence intervals are tight enough to rule out declines of the same magnitude as the abundance effect.

4.2 Spatial Decay and No Displacement

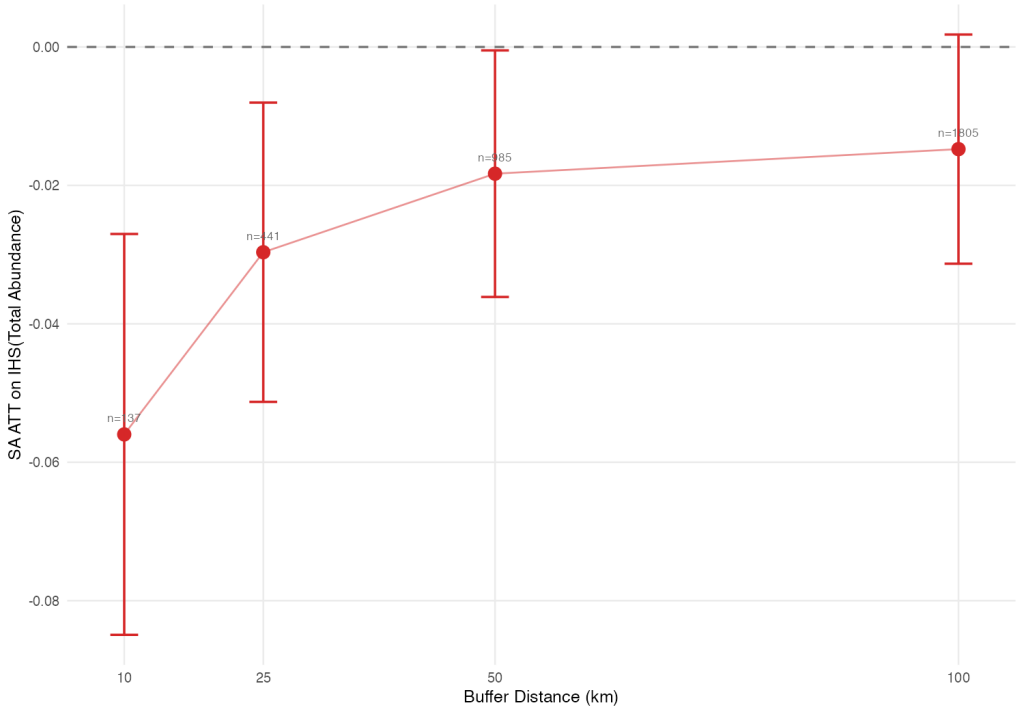
A key question for interpreting the abundance decline is whether it reflects true population loss or spatial displacement—birds relocating away from data centers to nearby, untreated areas. If displacement were the explanation, the aggregate decline would overstate the ecological cost: birds would be redistributed across space but not lost. We test this by estimating treatment effects at progressively larger buffer distances and examining the shape of the resulting spatial gradient.

Figure 4 plots the Sun-Abraham ATT for IHS(abundance) as a function of treatment buffer distance. The effect is strongest at 10 km (-5.1 percent) and decays monotonically: -2.4 percent at 25 km, -1.9 percent at 50 km, and -1.8 percent at 100 km. Two features of this gradient carry important implications.

First, the smooth, monotonic decay is consistent with localized environmental stressors (noise, light, water consumption, habitat conversion) whose intensity diminishes with distance from the facility. A sharp boundary effect (for example, a large coefficient at 10 km but zero at 25 km) would suggest a single dominant channel operating at a fixed radius. Instead, the gradual attenuation implies multiple overlapping channels with different spatial footprints: noise attenuates within a few kilometers, light pollution extends somewhat further, and water-mediated vegetation effects may operate at the watershed scale.

Second, and critically for identification, there is no positive coefficient at any dis-

Figure 4: Spatial Gradient: Distance Decay of Treatment Effects



Notes: Sun-Abraham ATT estimates for IHS(abundance) plotted against treatment buffer distance (10, 25, 50, 75, 100 km). Effects decay monotonically with no positive coefficient at any distance, ruling out spatial displacement and indicating true population loss. Each point represents a separate regression with the corresponding buffer distance defining treatment. Same controls and clustering as Table 3. Nested buffer detail in Appendix B. Source: USGS BBS; Datacentermap.

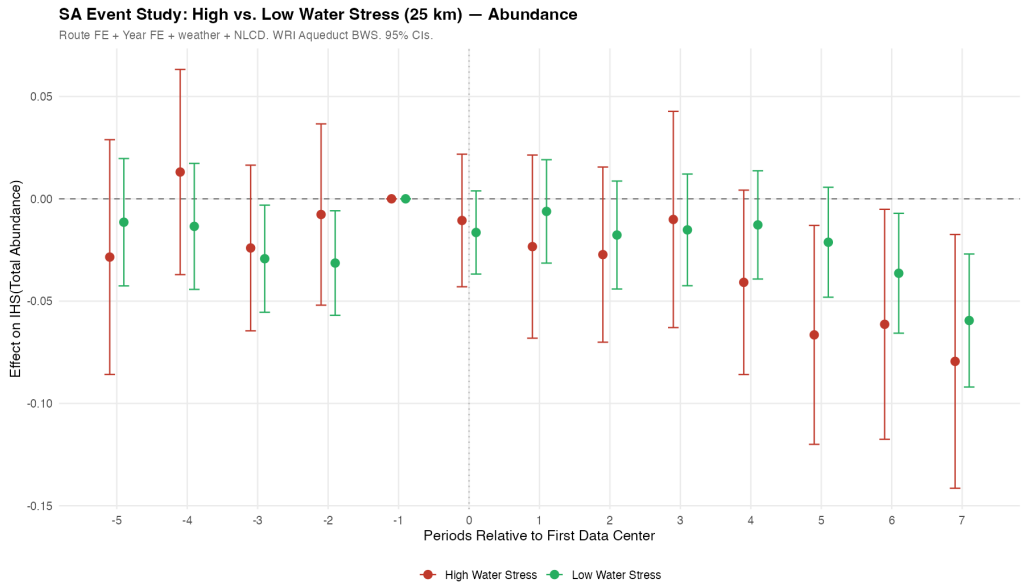
tance. If birds were simply relocating from the vicinity of data centers to nearby areas, we would observe a positive "pile-up" at some distance band, with displaced birds concentrating in the next ring outward. The monotonic decline to zero instead indicates true population loss: fewer birds overall, not the same birds in different places. This result is important because it rules out the main alternative interpretation of the aggregate decline. The nested buffer results are reported in Appendix B (Table C1 and Figure C1).

4.3 Heterogeneity

The aggregate treatment effect masks substantial variation across environmental and facility characteristics. Table 4 reports heterogeneity along four dimensions that are informative about both the mechanisms at work and the policy implications.

Water stress. The ecological effects of data centers are amplified where water is scarce. Splitting routes by baseline water stress of the nearest data center (WRI Aqueduct index), species richness declines only in high water-stress areas (ATT $= -0.029$, $p = 0.055$ at 25 km), while low-stress areas show no richness effect. The abundance ATTs are

Figure 5: Water Stress Heterogeneity: Abundance Event Studies (25 km)



Notes: Sun-Abraham event-study coefficients for IHS(abundance) at 25 km, estimated separately for routes near data centers in high and low baseline water stress areas (WRI Aqueduct). Pre-treatment trajectories are indistinguishable; the divergence in biodiversity effects emerges after facility activation. Same specification as Table 3. Source: USGS BBS; WRI Aqueduct; Datacentermap.

negative in both groups but individually insignificant when averaged across all post-treatment periods (-3.6 percent, $p = 0.16$ for high stress; -2.0 percent, $p = 0.11$ for low stress). However, the event studies in Figure 5 reveal that the effect builds progressively over time: in high-stress areas, abundance declines reach -6.3 percent by five years and -7.9 percent by seven years post-treatment (both $p < 0.02$), while low-stress areas show a smaller and later trajectory. The ATT understates the mature effect because it averages the large late-period declines with near-zero early-period coefficients. The wide standard errors for the high-stress subsample reflect the smaller number of treated routes in that group (120 versus 321 for low stress), combined with higher outcome variance in arid environments where bird populations are naturally more volatile. This gradual intensification is consistent with the vegetation mechanism explored in Section 5.2: evaporative cooling demand degrades vegetation progressively, and the cascading effects on bird populations take years to materialize fully.

The water stress result also has the sharpest policy implication of any finding in the paper. It identifies a specific, measurable geographic characteristic, baseline water stress (publicly available from WRI Aqueduct), that predicts where the ecological cost of a new data center will be highest. Water-stress-conditional siting guidelines would be straightforward to implement and would target the most consequential margin.

Dose-response. Routes exposed to more facilities show larger effects, consistent with cumulative environmental pressure. Splitting treated routes by the number of data

centers within 25 km, routes with 5 or more facilities, typically in dense corridors like northern Virginia or the Dallas–Fort Worth metroplex, show the strongest decline (-3.4 percent, $p = 0.052$). Routes with a single facility show a similar effect (-2.7 percent, $p = 0.12$). The medium-intensity bin (2–4 facilities) is insignificant, possibly reflecting the small sample of routes with exactly 2–4 facilities, but the stronger effect at high intensity relative to low supports the interpretation that the abundance decline is caused by cumulative data center operations rather than correlated shocks that happen to coincide with facility openings. Event-study overlays by intensity bin appear in Figure D1 in Appendix C.

Facility type. Both hyperscaler and non-hyperscaler data centers produce significant abundance declines of similar magnitude (-2.7 percent for hyperscaler, $p = 0.044$; -2.2 percent for non-hyperscaler, $p = 0.083$). This finding demonstrates that the biodiversity externality is not driven exclusively by the largest, most resource-intensive facilities. The cumulative effect of many smaller colocation and enterprise operations—which individually consume far less water and electricity than a hyperscale campus—is at least as large in aggregate. It also carries a direct regulatory implication: a policy framework that targets only hyperscaler facilities (above, say, a 50 MW threshold) would miss approximately half the total ecological impact.

As we show in Section 5.2, the two facility types operate through different mechanisms at different spatial scales: non-hyperscalers through cumulative watershed-level water stress from the aggregation of many smaller facilities; hyperscalers through direct land conversion and light pollution from individual large campuses.

Rural versus urban. Splitting routes by baseline urbanization (median split on NLCD 2001 percent developed within the route buffer), rural data centers produce a larger and more significant abundance decline (-5.6 percent, $p = 0.004$) than urban facilities (-1.8 percent, $p > 0.10$). The pattern reflects the higher baseline habitat quality in rural areas—there is simply more ecological value to displace. A data center built on former cropland adjacent to a forest fragment disrupts a functioning ecosystem; the same facility in an already-developed suburban corridor adds one more building to an environment that birds have largely already adapted to or abandoned.

This result carries a stark policy implication. The current siting calculus favors rural locations precisely because land is cheap and power is available. But the ecological cost per facility is highest in these same locations—a spatial mismatch between private siting incentives and social costs that policy reform could address.

Table 4: Heterogeneity: Water Stress, Intensity, Facility Type, and Urbanization

	Richness	Abundance
Panel A: Water Stress (25 km)		
High BWS	−0.029*	−0.036
	(0.015)	(0.026)
Low BWS	+0.010	−0.020
	(0.009)	(0.012)
Panel B: Treatment Intensity (25 km)		
Low (1 DC)	+0.002	−0.028
	(0.012)	(0.017)
Medium (2–4 DCs)	−0.008	−0.010
	(0.010)	(0.018)
High (5+ DCs)	+0.008	−0.034**
	(0.014)	(0.017)
Panel C: Facility Type (25 km)		
Hyperscaler routes	+0.005	−0.027**
	(0.010)	(0.013)
Non-hyperscaler routes	+0.001	−0.022*
	(0.008)	(0.013)
Panel D: Urbanization (25 km)		
Rural routes		−0.056***
		(0.020)
Urban routes		−0.018
		(0.012)

Notes: Sun-Abraham ATT estimates split by subgroup. Panel A: routes classified by WRI Aqueduct baseline water stress of nearest data center (median split). Panel B: routes binned by cumulative DC count within 25 km (Low = 1 facility, Medium = 2–4, High = 5+). Panel C: routes classified by whether nearest DC is hyperscaler-owned (Amazon, Google, Microsoft, Meta). Panel D: routes classified by median split on baseline NLCD 2001 percent developed within the route buffer. All specifications include route fixed effects, year fixed effects, weather controls, first-year observer indicator, and NLCD land-cover shares. Standard errors in parentheses, clustered at the route level. Event-study overlays for each dimension in Appendix C. ***p<0.01, **p<0.05, *p<0.1. Source: USGS BBS; WRI Aqueduct; Datacentermap.

5 Mechanisms

The aggregate abundance decline with null richness points to compositional change within the avian community. This section documents those shifts and traces the environmental pathways through which data centers produce them. We organize the analysis

around four channels: community-level restructuring (what is happening to the composition of local bird communities), water and vegetation (the habitat degradation pathway), light pollution (the nocturnal pathway), and noise (the acoustic pathway). For each, we test a specific prediction, present the evidence, and assess whether the pattern is consistent with the aggregate results in Section 4.

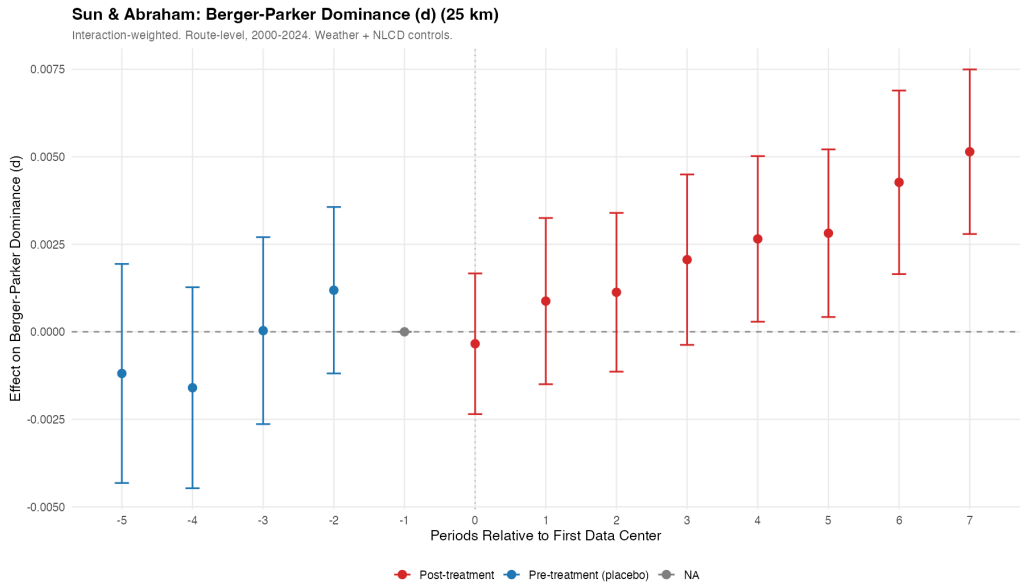
5.1 Community Restructuring

The combination of declining abundance and stable richness documented in Section 4.1 implies that the net effect averages over winners and losers. To characterize who wins and who loses, we estimate treatment effects on diversity indices, migratory composition, and guild-level outcomes.

Table 5 reports the results at 25 km. Panel A presents diversity indices. The Berger-Parker dominance index, the share of total individuals belonging to the single most abundant species, increases significantly ($+0.003$, $p = 0.001$). On a baseline mean of 0.093 for treated routes (i.e., the most common species accounts for 9.3 percent of individuals on a typical route), this represents a 3.2 percent relative increase in dominance. The most common species becomes *more* dominant relative to others after data center construction, the hallmark of biotic homogenization (McKinney and Lockwood, 1999). Figure 6 shows the event-study path: pre-treatment coefficients are flat and centered on zero, and dominance rises gradually after data center activation, reaching $+0.005$ by seven years post-treatment. The progressive trajectory is consistent with a cumulative process, gradual community restructuring as sensitive species decline and dominant generalists expand, rather than a one-time compositional shock from construction. Pielou's evenness (J') is null, which is consistent: this composite index can remain stable when one species increases its dominance while rare species decline symmetrically. The Shannon diversity index (H') is also null (Appendix D), confirming that standard diversity metrics fail to capture the compositional shift.

Panel B documents changes in migratory composition. The share of migratory species in route-level richness declines by 0.4 percentage points ($p = 0.011$); residents are replacing migrants in local assemblages. In absolute terms, the typical route records 42 migratory and 11 resident species; a 0.4 pp shift in the migratory share corresponds to losing approximately one migratory species from every two to three routes. Neotropical migrants, long-distance species that breed in North America and winter south of the Tropic of Cancer (including Wood Thrush, Scarlet Tanager, and Baltimore Oriole), are disproportionately affected, with IHS-transformed abundance declining by 4.2 percent ($p = 0.035$). On a baseline of approximately 74 neotropical migrant individuals per route, this implies roughly 3 fewer neotropical migrants per survey, a modest but ecologically meaningful decline given that these species are already under continent-

Figure 6: Berger-Parker Dominance: Event Study (25 km)



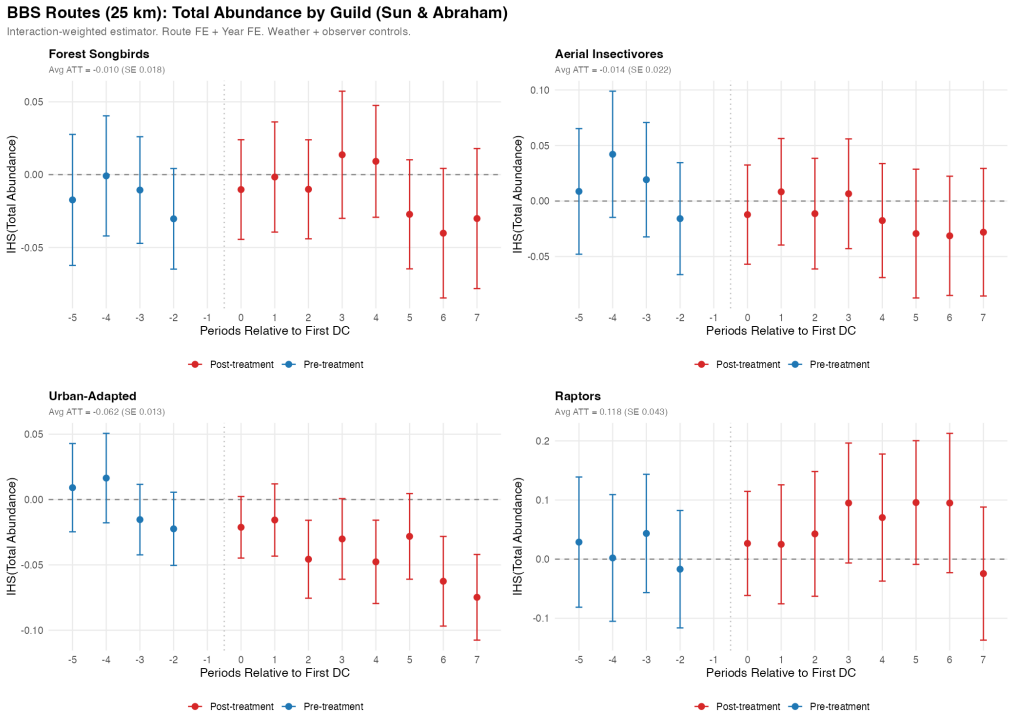
Notes: Sun-Abraham event-study coefficients for the Berger-Parker dominance index ($d = N_{\max}/N$) at 25 km. Higher values indicate greater dominance by the single most abundant species. Pre-treatment coefficients are flat; dominance rises progressively after data center activation, consistent with biotic homogenization. Same specification as Table 3. Source: USGS BBS; Datacentermap.

wide pressure from habitat loss on both breeding and wintering grounds. These 153 species, spanning seven families (warblers, vireos, swallows, flycatchers, nightjars, hummingbirds, cuckoos), provide ecosystem services including pest control and seed dispersal that operate at continental scales (Noack et al., 2025). Their decline may carry economic consequences beyond the ecological impact: a community that loses its Wood Thrushes loses not only a bird species but the pest-regulation services that species provides to surrounding agricultural land.

The migratory decline is particularly informative about mechanism because migratory species are more sensitive to local breeding-habitat quality than residents. A resident species can adapt to gradual environmental change by shifting territories within its home range. A migratory species returning from a 5,000-kilometer journey to Central America has no such flexibility: if its breeding territory has degraded, it either attempts to breed in suboptimal conditions (reducing reproductive success) or moves to a different site (reducing local abundance). The disproportionate decline of migrants is therefore consistent with habitat degradation as a proximate cause.

Panel C presents guild-level abundance effects. The pattern across guilds is striking. Raptors, including Red-tailed Hawks, American Kestrels, and Cooper's Hawks, *increase* significantly (+11.0 percent, $p = 0.012$). These predators exploit altered landscapes in several ways: data center campuses provide open sight lines for hunting, manicured grounds attract rodent prey, security lighting makes prey visible at dusk and dawn, and the displacement of competing species reduces interspecific competition. Urban-

Figure 7: Guild-Level Treatment Effects on Abundance (25 km)



Notes: Sun-Abraham ATT estimates and 95% confidence intervals for IHS(abundance) by species guild at 25 km. Raptors increase significantly (+11.0%, $p = 0.012$), consistent with predators exploiting altered landscapes. Urban-adapted species decline (−5.5%, $p < 0.001$) despite proximity to developed areas, reflecting the low habitat complexity of data center campuses. Forest songbirds and aerial insectivores show negative but insignificant point estimates. Same specification as Table 3. Source: USGS BBS; Datacentermap.

adapted species (House Sparrows, European Starlings, Rock Pigeons) decline (−5.5 percent, $p < 0.001$), a counterintuitive result that likely reflects the distinction between urban habitats (dense, diverse built environments with varied food sources and nesting sites) and data center campuses (large, uniform, low-complexity structures that offer little nesting or foraging diversity). Forest songbirds and aerial insectivores show negative but insignificant point estimates, consistent with these guilds being less uniformly exposed to data center effects across the heterogeneous landscapes in our sample.

Figure 7 summarizes the guild-level ATTs and confidence intervals. The divergence between raptors (strongly positive) and urban-adapted species (strongly negative) encapsulates the restructuring interpretation: data centers do not create "better" or "worse" habitat in any absolute sense—they create habitat that systematically favors certain ecological strategies over others. Event-study plots for individual guilds, showing the dynamic treatment paths, appear in Figure E1 in Appendix D. Community turnover coefficients, summarizing the shifts in migratory and neotropical richness and abundance proportions, are in Figure E2.

Table 5: Community Composition Effects (25 km)

Outcome	ATT	SE
Panel A: Diversity Indices		
Berger-Parker Dominance d	$+0.003^{***}$	(0.001)
Pielou Evenness J'	-0.000	(0.001)
Panel B: Migratory Status		
Migratory richness share	-0.004^{**}	(0.002)
IHS(Neotropical abundance)	-0.042^{**}	(0.020)
Panel C: Guild Abundance (IHS)		
Forest songbirds	-0.007	(0.018)
Aerial insectivores	-0.010	(0.022)
Urban-adapted	-0.055^{***}	(0.013)
Raptors	$+0.110^{**}$	(0.044)

Notes: Sun-Abraham ATT at 25 km buffer. Same specification as Table 3. Panel A: Berger-Parker dominance $d = N_{\max}/N$ is the share of individuals belonging to the most abundant species; higher values indicate greater dominance. Pielou evenness $J' = H'/\ln S$ normalizes Shannon diversity by maximum possible value. Panel B: Migratory richness share is the fraction of detected species classified as migratory (family-level classification). Neotropical migrants: 153 species in 7 families (Parulidae, Vireonidae, Hirundinidae, Tyrannidae, Caprimulgidae, Trochilidae, Cuculidae). Panel C: guild definitions based on BBS family-level classification. Diversity index event studies, neotropical migrant breakdowns, and guild-level event studies in Appendix D. ***p<0.01, **p<0.05, *p<0.1. Source: USGS BBS; Datacentermap.

5.2 Water and Vegetation

Data centers consume water primarily through evaporative cooling: warm air from server rooms passes through cooling towers where water evaporates to dissipate heat. A typical hyperscale facility consumes 1.2–1.8 liters per kilowatt-hour of IT load (Mytton, 2021). Most facilities connect to municipal water systems rather than operating private wells, meaning their consumption enters the broader watershed supply. If this water demand degrades local vegetation, we should observe declining Leaf Area Index (LAI), a satellite measure of vegetation density, in facility buffers, with stronger effects where water is already scarce.

Table 6 confirms both predictions. At the 15 km scale, LAI declines by 3.3 percent (ATT $= -0.033$, $p < 0.001$). The effect is present at all buffer distances: -2.3 percent at 5 km, -3.2 percent at 25 km, and -2.3 percent at 50 km. The persistence of the effect out to 50 km indicates that the vegetation response extends well beyond the facility footprint, consistent with watershed-level water stress rather than localized land clearing alone. The effect intensifies with data center density, as expected if

cumulative water demand from multiple facilities depletes shared water resources.

The heterogeneity across facility types is informative about the spatial scale at which the water channel operates. Non-hyperscaler facilities drive the watershed-scale LAI effect (-4.2 percent, $p < 0.001$), while hyperscalers show no significant vegetation loss ($+2.0$ percent, $p > 0.10$; Panel B). This pattern is consistent with cumulative stress from many smaller facilities drawing on shared municipal water systems. A cluster of twenty colocation facilities in a metropolitan area may collectively withdraw enough water to affect the regional supply, even though each individual facility's consumption is modest. Hyperscaler campuses, by contrast, tend to be newer facilities that invest in on-site water management (recycling, air-cooled backup systems) and are often large enough to negotiate dedicated water supply agreements that reduce their draw on the shared municipal system.

The baseline vegetation heterogeneity is even more striking (Panel C). The aggregate LAI decline is entirely driven by facilities in arid and semi-arid environments: low-baseline-LAI sites experience a -6.5 percent decline ($p < 0.001$) at 15 km, while high-LAI (temperate) sites show a slight, insignificant positive effect ($+0.6$ percent). This asymmetry has a clear ecological interpretation. In arid environments, vegetation is already near a degradation threshold—even modest additional water extraction can push plant communities past a tipping point from which recovery is slow. In well-watered environments, the hydrological system can absorb the additional demand without measurable stress on vegetation.

Figure 8 presents the event-study overlay by water stress at 15 km. The divergence between high- and low-stress facilities emerges after facility activation: pre-treatment LAI trajectories are indistinguishable, and the gap widens progressively over the post-treatment period. High-stress areas experience 1.7 times larger declines than low-stress areas.

This evidence is consistent with a causal pathway linking water scarcity to the biodiversity results in Section 4.3. The sequence runs: data center water consumption $\rightarrow$ vegetation degradation (measured via LAI) $\rightarrow$ habitat loss $\rightarrow$ selective species decline (measured via BBS). The fact that both the bird abundance effect and the vegetation effect are amplified by water stress, and that both emerge with similar timing after facility activation, supports the interpretation that the water–vegetation pathway is a primary mechanism through which data centers affect local ecosystems.

5.3 Light Pollution and Nocturnal Species

Data centers operate continuously, 24 hours a day, 365 days a year. Their campuses include extensive security lighting, illuminated parking areas, and the glow of equipment through ventilation systems. Artificial light at night is a well-documented disruptor of

Table 6: Vegetation Health (LAI) Results

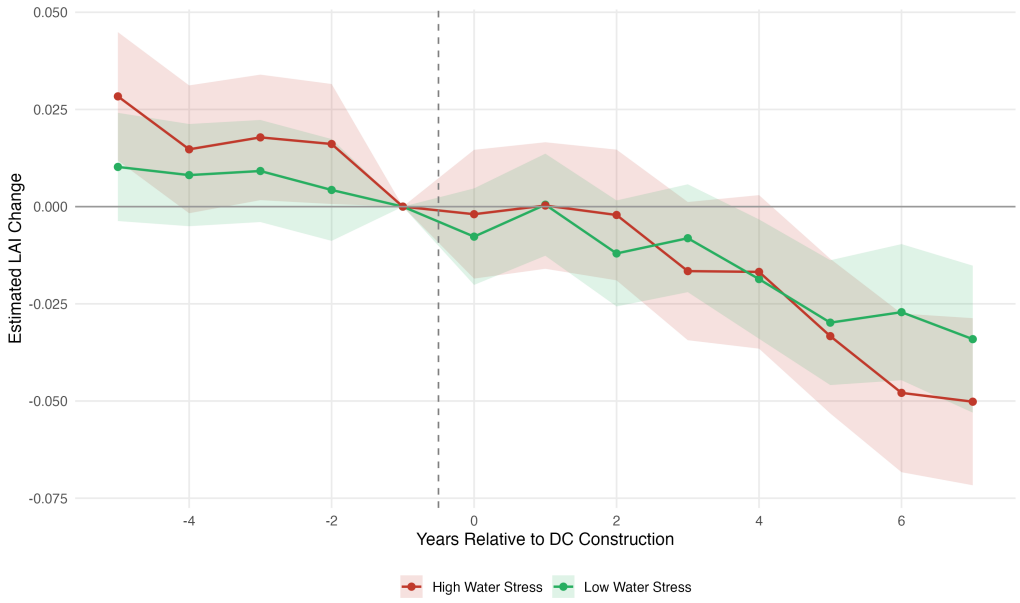
Specification	ATT	SE
Panel A: Buffer Distance		
5 km	-0.023^{***}	(0.006)
15 km	-0.033^{***}	(0.005)
25 km	-0.032^{***}	(0.005)
50 km	-0.023^{***}	(0.004)
Panel B: Hyperscaler Heterogeneity (15 km)		
Hyperscaler	$+0.020$	(0.012)
Non-hyperscaler	-0.042^{***}	(0.006)
Panel C: Baseline LAI Heterogeneity (15 km)		
Low LAI (arid)	-0.065^{***}	(0.009)
High LAI (temperate)	$+0.006$	(0.009)

Notes: Sun-Abraham ATT on facility×year panels. 1,796 treated facilities + 806 never-treated controls. Donut buffers exclude inner 1 km ring to separate land-clearing footprint from watershed effects. Facility fixed effects, year fixed effects, PDSI drought control. No anticipation imposed. Clustered at facility level. Panel B: non-hyperscalers drive the watershed-scale decline through cumulative water demand from many facilities on shared municipal systems. Panel C: median split on pre-treatment baseline LAI at 15 km; the 6.5% decline in arid environments represents vegetation near a degradation threshold. Full multi-buffer event studies (Figure G1) and water stress heterogeneity detail (Figure G2) in Appendix F. ***p<0.01, **p<0.05, *p<0.1. Source: Landsat via Google Earth Engine; GRIDMET PDSI; WRI Aqueduct; Datacentermap.

nocturnal wildlife, affecting foraging behavior, predator-prey dynamics, and navigation (Gaston et al., 2015). We test the light-pollution channel using VIIRS nighttime radiance as a first-stage measure and eBird nocturnal species data as an independent ecological outcome.

First stage: VIIRS radiance. VIIRS nighttime radiance increases sharply within 1 km of data centers ($+1.38$ nW/cm^2/sr, $p < 0.001$). On a pre-treatment mean of approximately 45 nW/cm^2/sr, this represents a 3 percent increase in ambient light, roughly comparable to adding a small commercial parking lot to the area. Figure 9 presents event-study coefficients at multiple buffer distances: radiance rises upon facility activation and persists throughout the post-treatment period, with the strongest effects at close range. The immediate jump at $t = 0$ and stable post-treatment level are consistent with the continuous nature of data center lighting; unlike seasonal or cyclical light sources, data center illumination is constant from the day of activation.

Figure 8: Vegetation Health (LAI): Event Studies by Water Stress (15 km)



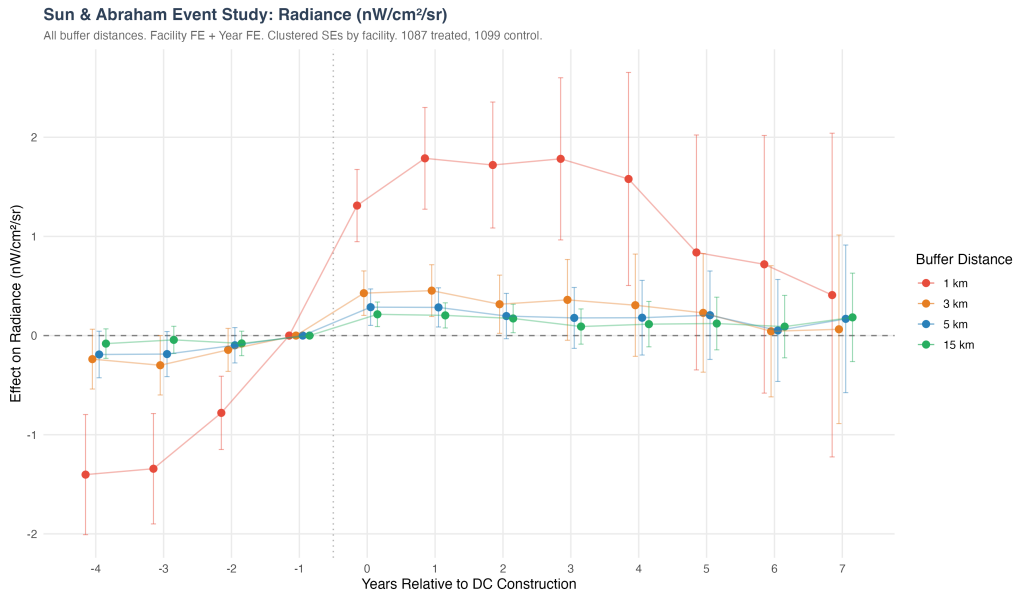
Notes: Sun-Abraham event-study coefficients for LAI at 15 km donut buffer, estimated separately for data centers in high and low baseline water stress areas (WRI Aqueduct). Pre-treatment trajectories are indistinguishable. The divergence in vegetation loss emerges after facility activation, with high-stress areas experiencing 1.7 times larger declines. Facility fixed effects, year fixed effects, PDSI drought control. Clustered at facility level. Source: Landsat via Google Earth Engine; WRI Aqueduct; Datacentermap.

Hyperscaler facilities are dramatically brighter than non-hyperscalers (+5.77 nW at 1 km versus −0.12, $p > 0.10$), consistent with their larger physical footprints and greenfield locations in previously dark areas. A hyperscaler campus built on former agricultural land in rural Oregon introduces a substantial new light source into an environment that was previously near-dark, while a colocation facility in a suburban office park adds incrementally to an already illuminated environment. The light treatment is therefore strongest precisely where the ecological sensitivity to light is also highest—dark-adapted rural ecosystems.

Ecological consequences: nocturnal species. To test whether this light increase has ecological consequences, we turn to an entirely independent dataset, eBird citizen-science observations of 19 nocturnal and crepuscular species, using different methodology (observer×year fixed effects on hexagonal grid cells) and an entirely different set of species from the daytime BBS. Table 7 presents the results. At 25 km, IHS-transformed nocturnal richness declines by 10.9 percent ($p < 0.001$) and abundance by 19.1 percent ($p < 0.001$). The effects are substantially larger than the daytime BBS effects, consistent with nocturnal species being more sensitive to the 24-hour operational profile of data centers. The distance gradient mirrors the BBS pattern: effects are strongest at 10 km (richness −11.2 percent, abundance −23.2 percent) and decay at 50 km.

The species-level results (Figure 10) are where the light-pollution mechanism be-

Figure 9: VIIRS Nighttime Radiance: Event Studies (All Distances)



Notes: Sun-Abraham event-study coefficients for VIIRS nighttime radiance (mean nW/cm^2/sr) at multiple buffer distances (1, 3, 5, 15 km). Light pollution increases sharply upon data center activation and persists throughout the post-treatment period. The immediate jump at $t = 0$ is consistent with continuous facility illumination. Facility fixed effects and year fixed effects. Clustered at facility level. Source: VIIRS DNB via Google Earth Engine; Datacentermap.

Table 7: Nocturnal Species Results (eBird)

	IHS(Richness)	IHS(Abundance)
10 km	-0.112^{***}	-0.232^{***}
	(0.032)	(0.066)
25 km	-0.109^{***}	-0.191^{***}
	(0.031)	(0.055)
50 km	-0.047^{***}	-0.167^{***}
	(0.013)	(0.048)

Notes: Sun-Abraham ATT on cell×year panels (25 km hexagonal grid). 19 nocturnal and crepuscular species (9 owls, 5 nightjars, 3 crepuscular shorebirds, 2 herons), 11.1M records, 207K observers, 15,748 cells. Proper zero construction from full eBird sampling file (59M checklists): all complete checklists are loaded regardless of whether nocturnal species were detected, distinguishing genuine non-detection from absence of survey effort. Controls: effort variables (number of checklists, observers, mean duration, distance), cell fixed effects, year fixed effects, observer×year fixed effects. Clustered at cell level. Species-level heterogeneity in Figure 10. Event studies (Figure F1) and species-level heterogeneity (Figure F2) in Appendix E. Source: eBird; Datacentermap.

comes most visible. The 19 species do not respond uniformly—they respond in a pattern that is precisely what the ecological literature on artificial light at night would predict.

Owls, nocturnal predators that rely primarily on vision to hunt, increase dramatically. Great Horned Owl abundance rises by 51 percent ($p < 0.001$) and Barred Owl by 29 percent ($p < 0.001$). These are large, adaptable raptors at the top of the nocturnal food chain. Artificial light makes their prey (rodents, small birds, insects) more visible, directly enhancing hunting efficiency. The magnitude of the increase is substantial: a 51 percent rise in Great Horned Owl abundance is among the largest species-level effects in our entire analysis, and it is consistent with the daytime BBS finding that raptors increase near data centers.

In contrast, nightjars, nocturnal insectivores that depend on darkness for their sit-and-wait foraging strategy, decline. Chuck-will's-widow, a cryptic species that relies on camouflage to avoid predators while sitting motionless on the ground waiting for insects to pass within range, declines by 8 percent ($p = 0.02$). Artificial light disrupts this species on two fronts: it draws moths and other flying insects toward the light source and away from the bird's foraging territory, and it compromises the bird's camouflage-based predator defense by making it visible to the very owls whose populations are increasing nearby. Common Nighthawk, however, *increases* by 19 percent ($p = 0.01$), an exception that proves the rule. Unlike other nightjars, the Common Nighthawk is the most urban-adapted member of its family, already accustomed to foraging on insects attracted to streetlights in cities. It exploits the insect aggregations created by data center lighting rather than being harmed by them.

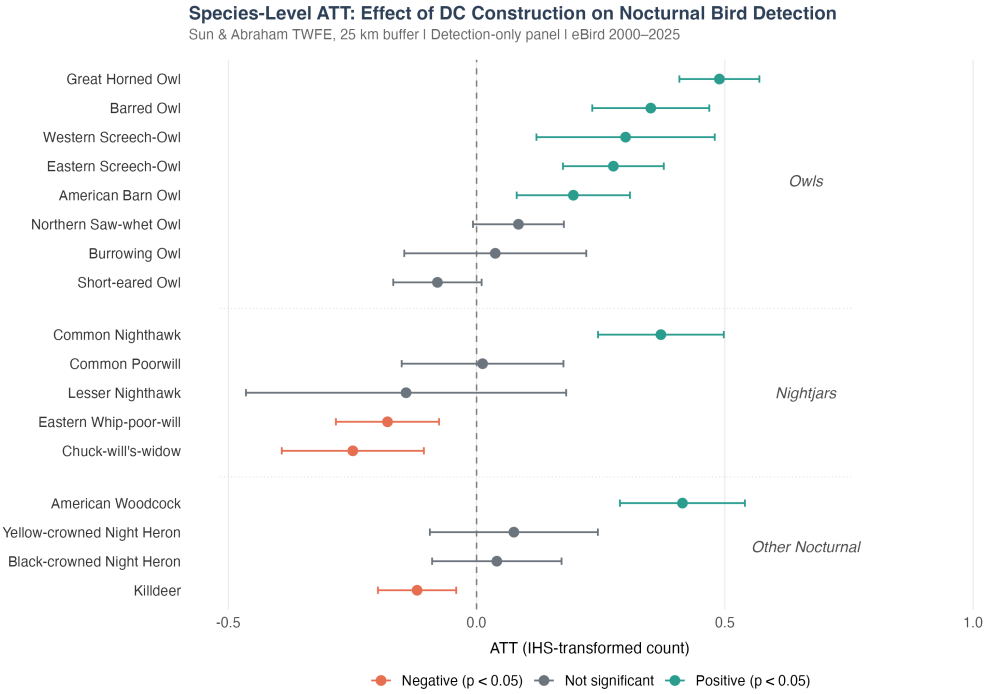
Killdeer, a ground-nesting shorebird, declines by 11 percent ($p = 0.02$). Continuous artificial illumination can disrupt incubation behavior in ground-nesting species and expose nests to visually oriented predators—the same owls whose populations are increasing nearby.

The consistency of the eBird nocturnal results with the daytime BBS finding of community restructuring, obtained from an entirely independent dataset using different methodology on different species, substantially strengthens the causal interpretation. Both analyses point to the same conclusion: data centers reorganize local ecological communities by shifting the competitive balance among species, with predators and urban-adapted generalists gaining at the expense of cryptic, darkness-dependent, and ground-nesting specialists.

5.4 Acoustic Environment

Data center HVAC systems, backup diesel generators (tested regularly), and electrical transformers generate continuous low-frequency noise in the 50–500 Hz range (Buxton

Figure 10: Nocturnal Species: Treatment Effects by Species (eBird, 25 km)



Notes: Sun-Abraham ATT estimates for individual nocturnal species at 25 km. Green = significant positive, red = significant negative, grey = not significant. Owls (Great Horned +51%, Barred +29%) increase, consistent with visually oriented predators exploiting illuminated environments. Killdeer (−11%) and Chuck-will's-widow (−8%) decline. Common Nighthawk (+19%) is the most urban-adapted nightjar and exploits insect aggregations near lights. 95% confidence intervals shown. Same specification as Table 7. Source: eBird; Datacentermap.

et al., 2017). This overlaps with the vocal frequencies of many songbird species. The acoustic masking hypothesis (Cardoso et al., 2020; Barber et al., 2010) predicts that species whose vocalizations overlap most with anthropogenic noise should be most affected, either because their songs are masked (reducing mating success) or because they avoid noisy areas entirely.

We test this prediction by classifying BBS species into terciles by minimum vocal frequency using the Sagar et al. (2024) acoustic traits dataset (87 percent of BBS species matched directly, 100 percent with family-level fallback). Low-frequency species (bottom tercile, minimum frequency < 1.5 kHz) have the greatest spectral overlap with data center noise. We estimate separate Sun-Abraham regressions for each frequency group, allowing each to have its own treatment effect.

The results contradict simple masking. Low-frequency species, those predicted to be most vulnerable, show null effects on abundance (ATT $= -0.013$, $p = 0.35$). Mid-frequency species decline significantly (-3.3 percent, $p = 0.005$), and high-frequency species, those whose vocalizations are *above* the data center noise band and should be least affected by masking, decline even more (-3.8 percent, $p = 0.017$). The gradient runs in the opposite direction from the masking prediction.

This reversed gradient has a natural explanation rooted in the ecological characteristics of species in each frequency group. Low-frequency species are generally large-bodied birds (pigeons, doves, corvids, some raptors) that are well adapted to urban and peri-urban environments. These species face compensating positive effects from the built environment that data centers create: open areas for foraging, anthropogenic food sources (garbage, manicured lawns with earthworms and grubs), reduced predation pressure in simplified landscapes, and the manicured landscaping typical of facility campuses. The null effect on low-frequency species therefore reflects offsetting forces: they are likely harmed by noise but benefit from habitat modification, rather than absence of any impact.

Mid- and high-frequency species, by contrast, include many small-bodied, open-habitat and edge-dwelling birds (sparrows, warblers, finches) that are directly exposed to the landscape changes documented in Section 5.2. They inhabit the ecotones (habitat edges between forest and open land) most disrupted by data center construction and ongoing operations, and they do not benefit from the simplified built environment in the way that large-bodied urban generalists do.

Raptors, vision-driven predators whose foraging does not depend on acoustic communication, increase strongly in the frequency-stratified specification (+11.3 percent abundance, $p = 0.008$; consistent with the +10.5 percent in the main guild regression, Table 5), reinforcing the community restructuring interpretation. Their success near data centers is consistent with open sight lines for hunting, abundant rodent prey on manicured campuses, and reduced competition from displaced songbird species.

The overall pattern across frequency groups, body sizes, and foraging strategies points to habitat disruption as the operative mechanism rather than acoustic masking per se. The noise is real—data centers are loud—but its ecological effect is confounded by, and likely secondary to, the broader suite of environmental changes (vegetation loss, light pollution, land conversion) that accompany facility construction and operations.

5.5 Mechanism Summary

Table 8 summarizes the first-stage effects across all channels tested, providing a comprehensive view of how data centers alter local environmental conditions. Every channel we test shows a statistically significant first-stage effect: vegetation degrades, light increases, the electricity grid expands. The channels that do not conform to simple predictions—groundwater improving, clean energy share unchanged—turn out to reflect land-use substitution and compositional effects that we discuss in Section 6.

The multi-channel nature of the externality is itself a key finding. Unlike wind turbines (where the dominant channel is collision mortality) or dams (where the dominant channel is downstream flooding), data centers affect ecosystems through several path-

Table 8: Mechanism Tests Summary

Channel	ATT	SE	Interpretation
LAI (15 km)	-0.033^{***}	(0.005)	Vegetation degrades
VIIRS (1 km)	$+1.38^{***}$		Light increases
Grid MW (10 km)	$+0.34^{***}$	(0.08)	Grid expands (§6.2)
Clean share (50 km)	$+0.001$		Null; hyper. $+1$ pp (§6.2)
Groundwater (5 km)	-1.96^{**}	(0.79)	Water table rises (§6.3)

Notes: First-stage effects test whether data centers alter each environmental channel. No anticipation imposed. LAI: facility×year panel, donut buffer, PDSI control (Section 5.2). VIIRS: facility×year panel, 1 km circle (Section 5.3). Grid MW: BBS route×year panel, EIA Form 860 cumulative capacity (Section 6.2). Clean share: facility×year panel, 50 km buffer (Section 6.2). Groundwater: USGS NWIS well×year panel, 5 km buffer (Section 6.3). Grid expansion does not mediate the bird effect (1.2% attenuation). Groundwater "improvement" reflects agricultural displacement. Source: Landsat, VIIRS, EIA Form 860, USGS NWIS; Datacentermap.

ways simultaneously. This means that policy interventions targeting any single channel (water recycling, for example, or lighting standards) will address only a fraction of the total ecological cost. A comprehensive regulatory framework would need to consider all channels jointly, ideally through facility-level environmental impact assessments that account for local conditions (water stress, baseline habitat quality, existing light environment).

6 Discussion

6.1 The Reallocation Interpretation

The headline finding—a 2.4 percent decline in bird abundance—is a net effect that masks substantial heterogeneity across species. Data centers do not uniformly harm biodiversity. Owls increase dramatically, raptors thrive, and some nightjar species show positive responses. Neotropical migrants decline, community dominance rises, and the migratory share of local assemblages falls.

The unifying interpretation is *reallocation*: data centers reallocate ecological communities by shifting the competitive balance among species. Water is reallocated from vegetation to cooling towers. Light is reallocated from darkness to illumination. Land is reallocated from habitat to impervious surface. Each reallocation favors some species (raptors, owls, large-bodied generalists) at the expense of others (migratory specialists, nocturnal insectivores, small open-habitat birds).

This pattern matters for welfare analysis because the ecosystem services that birds provide (pest control, pollination, seed dispersal) depend on the functional composition of the community, not aggregate headcounts. When neotropical migrants that suppress

agricultural pest populations are replaced by generalists such as starlings and sparrows, the economic value of the avian community declines even if species richness remains unchanged. An environmental impact assessment based on species counts alone would conclude that data centers are ecologically benign.

6.2 Grid Expansion and Energy Composition

The ecological impact of data centers extends beyond their direct operational footprint through induced changes in the electricity grid. A hyperscale campus consuming 100–200 MW requires dedicated electrical infrastructure (substations, transmission lines, and often new generation capacity) that would not have been built absent the facility. Using EIA Form 860 data on all 31,283 US electricity generators at 14,184 power plants, we estimate that data center openings cause a significant increase in local generation capacity: $+0.34$ IHS megawatts within 10 km ($p < 0.001$; Figure H1 in Appendix G). The effect is marginally significant at 25 km ($+0.11$, $p = 0.052$) and turns negative at 50 km (-0.11), consistent with grid expansion being localized: new power plants are built near the data center, not distributed broadly.

Does this grid expansion mediate the bird-abundance effect? We test this by adding cumulative generation capacity as a control in the main BBS specification. The bird abundance ATT barely changes when controlling for grid capacity (-0.024 to -0.024), indicating that grid expansion does not operate through the same local pathway as the direct environmental channels. The induced power plants may have their own ecological footprints (emissions, cooling water demand, land use), but these effects are geographically dispersed and do not concentrate at the same locations as the data center's direct impact.

A natural follow-up question is whether data center-induced grid expansion brings clean or fossil fuel generation. The answer has implications for the net carbon footprint of the digital economy. Descriptively, data centers sit in gas-dominated grids: the mean clean energy share (solar, wind, hydro, nuclear, battery, geothermal as a fraction of total nameplate capacity) within 50 km of a data center facility is approximately 15 percent, compared to roughly 35 percent nationally. This gap reflects the industry's historical co-location with natural gas infrastructure for reliable baseload power.

The aggregate causal effect of data center openings on local clean energy share is null (ATT $= +0.001$, $p = 0.83$). However, the aggregate masks a striking divergence by facility type. To cleanly separate hyperscaler from non-hyperscaler effects, we use *type-matched controls*: hyperscaler regressions use only not-yet-treated hyperscalers as controls (215 facilities with start year after 2022), and non-hyperscaler regressions use only not-yet-treated non-hyperscalers (203 facilities). This addresses the concern that hyperscalers and colocation facilities site in fundamentally different areas (hyperscalers

in rural locations with cheap power and available land, colocations in urban centers near customers) and that a shared control group would conflate these location differences with the treatment effect.

With type-matched controls, hyperscaler facilities reduce nearby fossil fuel capacity by 3.8 percent (ATT $= -0.038$, $p < 0.001$) and natural gas capacity by 5.4 percent (ATT $= -0.054$, $p = 0.008$) within 50 km (Figure 11a–b). On a pre-treatment mean of approximately 2,400 MW fossil and 1,500 MW gas capacity within 50 km of the average hyperscaler, these effects translate to roughly 90 MW less fossil and 80 MW less gas generation, comparable to the retirement of one mid-sized gas peaker plant. The fossil MW event study shows flat pre-trends and a gradual post-treatment decline, consistent with a causal effect. Non-hyperscaler facilities show no effect on any energy outcome ($p > 0.4$ for all). Clean energy capacity increases by 14.5 percent for hyperscalers ($p = 0.15$), driven entirely by wind power (+18.8 percent, $p = 0.01$) rather than solar, consistent with the dominance of wind PPAs in hyperscaler procurement during our sample period. The clean MW effect materializes 5–7 years after facility activation, reflecting the lag between corporate PPAs and generation coming online. Table 9 reports the full set of ATT estimates across energy outcomes.

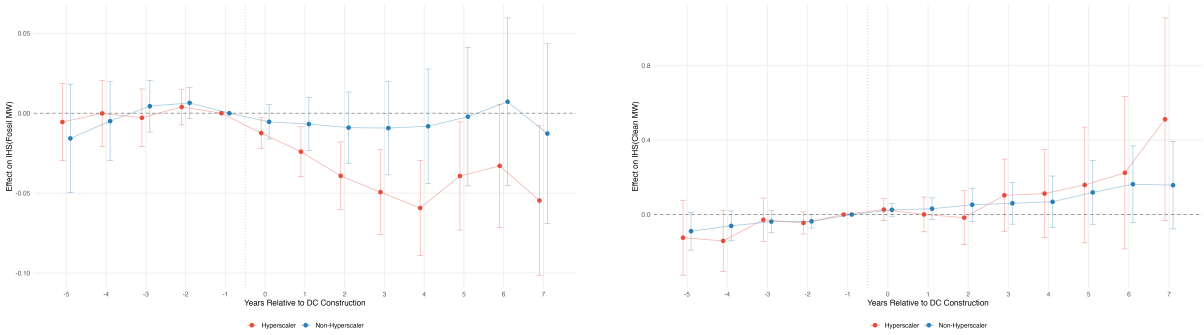
The policy implication is nuanced. Hyperscalers reduce nearby fossil generation—consistent with corporate renewable procurement displacing gas and coal plants. But this cannot overcome the fossil fuel dominance of the areas where data centers locate. The mean clean share near data centers (∼21 percent at 50 km) remains well below the national average (∼39 percent). The grid expansion channel therefore creates a multiplier: the facility's own operational footprint (water consumption, light pollution, habitat conversion) is compounded by the environmental costs of the predominantly gas-fired power plants that serve these areas, even as hyperscalers partially offset this through renewable procurement.

6.3 Land-Use Substitution Paradoxes

Two counterintuitive results in our mechanism tests illustrate a broader caution about interpreting facility-level environmental metrics. Both involve land-use substitution, the replacement of one activity by another, and both show how narrowly focused environmental monitoring can produce misleading conclusions.

The groundwater paradox. Using 13,790 USGS monitoring wells with 160,068 well-year observations (2000–2024), we find that local groundwater levels *improve* near data centers—depth to water decreases significantly at 5 km (-1.96 feet, $p = 0.013$) and the effect attenuates monotonically with distance, becoming insignificant at 25 km. At first glance, this might suggest an environmental benefit: the water table rises after a

Figure 11: Energy Composition Near Data Centers



(a) IHS(Fossil MW): Hyperscaler vs. Non-Hyp **(b)** IHS(Clean MW): Hyperscaler vs. Non-Hyp

Notes: Sun-Abraham event-study coefficients for IHS(fossil MW) and IHS(clean MW) at 50 km, estimated separately for hyperscaler and non-hyperscaler facilities with type-matched controls. Hyperscalers reduce fossil capacity ($p < 0.001$). Clean MW grows with a 5-year lag ($p = 0.15$), driven by wind procurement. Non-hyperscalers show no effect. Facility and year FE, NLCD control. Clustered by facility. Source: EIA Form 860; Datacentermap.

Table 9: Energy Composition: Hyperscaler vs. Non-Hyperscaler (50 km, Type-Matched Controls)

	Hyperscaler	Non-Hyperscaler
IHS(Clean MW)	+0.145	+0.095
	(0.102)	(0.074)
IHS(Wind MW)	+0.188**	+0.078*
	(0.073)	(0.041)
IHS(Solar MW)	−0.024	−0.006
	(0.082)	(0.065)
IHS(Fossil MW)	−0.038***	+0.001
	(0.010)	(0.019)
IHS(Gas MW)	−0.054***	+0.035
	(0.020)	(0.043)
IHS(Coal MW)	−0.211	+0.017
	(0.155)	(0.115)
Treated facilities	333	851
Control facilities	215	203

Notes: Sun-Abraham interaction-weighted ATT estimates at 50 km buffer. Type-matched controls: hyperscaler regression uses only not-yet-treated hyperscalers (start year > 2022) as controls; non-hyperscaler regression uses only not-yet-treated non-hyperscalers. All outcomes are IHS-transformed nameplate MW. All specifications include facility fixed effects, year fixed effects, and NLCD pct_developed (15 km). Standard errors clustered at the facility level in parentheses. ***p<0.01, **p<0.05, *p<0.1. Source: EIA Form 860; Datacentermap.

water-intensive facility opens. The reason is more nuanced.

Data centers frequently locate on former agricultural land; our NLCD data show

that treated areas have 30.6 percent agricultural land cover at baseline (Table 2). When a facility replaces irrigated farmland, the agricultural groundwater pumping that previously drew down the local aquifer ceases. Water levels recover not because the data center conserves water, but because the data center displaced the previous water-intensive use. Meanwhile, the data center itself consumes water through evaporative cooling, but because it connects to the municipal supply system rather than operating private wells, its consumption is invisible in local groundwater monitoring. The water enters the municipal system from wherever the utility sources its supply, which may be a distant reservoir, river intake, or deep aquifer well field many kilometers from the data center site.

The apparent aquifer "improvement" is therefore not an environmental benefit; it is a byproduct of agricultural displacement. This is analogous to the pollution haven effect in international trade: environmental degradation does not disappear when an activity relocates; it is spatially displaced. Data centers displace local groundwater extraction to wherever the municipal water system sources its supply, and the net effect on total water consumption is ambiguous without utility-level data, which is not available in a nationally standardized form.

The vegetation paradox. A parallel interpretive challenge arises from the LAI heterogeneity results (Table 6, Panel C). In areas with high baseline vegetation (temperate, well-watered environments), data center openings are associated with a slight, statistically insignificant *increase* in vegetation index (+0.6 percent). Does this mean data centers benefit vegetation in temperate areas?

No. The mechanism is again land-use substitution. In temperate areas, data centers replace cropland (which has periodically low LAI during harvest, plowing, and fallow periods) with developed land that may include maintained landscaping, tree buffers planted as visual screening, or simply reduced agricultural disturbance on adjacent parcels. The LAI signal reflects a shift from one land-use type (seasonal agriculture with fluctuating vegetation) to another (developed with stable, low-amplitude vegetation), not an ecological improvement. The satellite cannot distinguish between a forest canopy and a manicured corporate lawn—both register as green.

The policy-relevant finding is the 6.5 percent LAI decline in arid and semi-arid environments. In these water-stressed regions, evaporative cooling demand pushes already scarce water resources past a degradation threshold. Vegetation loss in arid landscapes is ecologically consequential: recovery is slow (decades, not years), habitat value is high relative to baseline, and the species that depend on these habitats, many of which are adapted to marginal conditions, are precisely those most vulnerable to further resource extraction.

Both paradoxes carry the same lesson for environmental regulation: facility-level en-

vironmental metrics can be actively misleading when the facility changes surrounding land use. A data center reporting zero direct groundwater impact is technically correct but substantively incomplete: it has merely displaced the impact to the municipal water system. A vegetation assessment showing "greening" near a temperate data center confuses land-use change with ecological improvement. A comprehensive accounting of the ecological costs of digital infrastructure must consider induced changes in surrounding human activities, not just the facility's direct footprint, a challenge for regulatory frameworks that traditionally target the facility itself.

6.4 Policy Implications

Data center siting decisions currently optimize over a narrow set of private costs: land prices, electricity rates, network latency, and fiber connectivity. Ecosystem services (pollination, pest control, water regulation, existence value of biodiversity) are absent from this calculus. Our results show that the ecological costs are real, spatially concentrated, and disproportionately borne by communities in water-stressed regions.

Several policy responses merit consideration. First, *environmental impact assessments* for large data center developments (above a power or land-area threshold) would force internalization of ecological costs at the project-approval stage. Second, *water-stress-conditional siting guidelines* would steer development away from regions where the marginal ecological cost of water extraction is highest. Third, *closed-loop cooling requirements* or water-recycling mandates would reduce the water consumption channel directly; the marginal cost of these technologies is modest relative to total facility construction costs. Fourth, *lighting design standards* (shielded, downward-facing, motion-activated exterior lighting) would reduce the light-spillage channel at minimal cost.

Current siting practices favor cheap rural land in arid regions—precisely where our results show the ecological damage per facility is greatest. Reforming siting policy to reflect geographic variation in ecological carrying capacity would align private incentives with social costs.

7 Conclusion

Data centers causally reshape local ecosystems. Using the staggered rollout of 3,063 facilities across the United States, we estimate a 2.4 percent decline in bird abundance within 25 km, driven by selective species loss and community homogenization rather than uniform ecological decline. Migratory and specialist species are disproportionately affected, while predators and resident generalists benefit from the altered landscape. An independent analysis of nocturnal species from eBird confirms community restructuring through the light-pollution channel.

The mechanisms are multiple and simultaneous. Water-mediated vegetation loss is the primary channel in arid environments. Light pollution restructures nocturnal communities. Habitat disruption shifts daytime composition. Induced grid expansion creates regional multiplier effects. These channels interact, and mitigating one leaves the others fully active, a finding that argues for comprehensive siting assessment rather than single-channel regulation.

The ecological costs concentrate where water is scarce and habitat baselines are high—precisely the rural, arid areas where current siting practices steer facilities. These facilities fall outside every existing environmental regulatory framework despite a substantial and measurable ecological footprint. Water-stress-conditional siting guidelines, environmental impact assessments, and closed-loop cooling requirements would address the most consequential channels at modest cost relative to facility budgets.

The staggered expansion of data center capacity continues to accelerate, driven by artificial intelligence workloads projected to double electricity and water demand within the decade. As new facilities are built, the natural experiment we exploit will intensify. Future work should pursue utility-level water consumption data to quantify the true water footprint, NLCD land-use conversion analysis to measure agricultural displacement directly, and migration phenology data to test whether data centers disrupt the timing of migratory movements, a dimension that annual breeding-season surveys cannot capture.

References

Barber, Jesse R., Kevin R. Crooks, and Kurt M. Fristrup, "The Costs of Chronic Noise Exposure for Terrestrial Organisms," *Trends in Ecology & Evolution*, 2010, *25* (3), 180–189.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Buxton, Rachel T., Megan F. McKenna, Daniel Mennitt, Kurt Fristrup, Kevin Crooks, Lisa Angeloni, and George Wittemyer, "Noise Pollution Is Pervasive in U.S. Protected Areas," *Science*, 2017, *356* (6337), 531–533.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cardoso, Gonçalo C., Yun Hu, and Clinton D. Francis, "Exposure to Noise Pollution across North American Passerines Supports the Noise Filter Hypothesis," *Global Ecology and Biogeography*, 2020, *29* (7), 1300–1308.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker, "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings," *American Economic Review*, 2015, *105* (2), 678–709.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Duflo, Esther and Rohini Pande, "Dams," *Quarterly Journal of Economics*, 2007, *122* (2), 601–646.

Faber, Benjamin and Monika Krause, "The Environmental Costs of E-Commerce," *American Economic Review: Insights*, 2023. Forthcoming.

Fowlie, Meredith, Stephen P. Holland, and Erin T. Mansur, "What Do Emissions Markets Deliver and to Whom? Evidence from Southern California's NO_x Trading Program," *American Economic Review*, 2012, *102* (2), 965–993.

Gaston, Kevin J., Marcel E. Visser, and Franz Hölker, "The Biological Impacts of Artificial Light at Night: The Research Challenge," *Philosophical Transactions of the Royal Society B*, 2015, *370* (1667), 20140133.

Greenstone, Michael, "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures," *Journal of Political Economy*, 2002, *110* (6), 1175–1219.

Gregory, Richard D. and Andrea van Strien, "Wild Bird Indicators: Using Composite Population Trends of Birds as Measures of Environmental Health," *Ornithological Science*, 2010, *9* (1), 3–22.

Hagerty, Nick, "Liquid Constrained in California: Estimating the Potential Gains from Water Markets," *Journal of Political Economy*, 2021. Forthcoming.

Hogan, Dylan, Eyal G. Frank, Joséphine Gantois, and Anouch Missirian, "The Effectiveness of Local Conservation Ballots," *American Economic Review: Papers & Proceedings*, 2025, *115*, 403–408.

International Energy Agency, "Electricity 2024: Analysis and Forecast to 2026," Technical Report, IEA, Paris 2024.

IPBES, "Global Assessment Report on Biodiversity and Ecosystem Services," *Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, 2019. Bonn, Germany.

Katovich, Erik, "Birds and Energy Infrastructure: Quasi-Experimental Evidence from the United States," 2024. Working paper.

Keiser, David A. and Joseph S. Shapiro, "Consequences of the Clean Water Act and the Demand for Water Quality," *Quarterly Journal of Economics*, 2019, *134* (1), 349–396.

Loss, Scott R., Tom Will, Sara S. Loss, and Peter P. Marra, "Bird–Building Collisions in the United States: Estimates of Annual Mortality and Species Vulnerability," *The Condor*, 2014, *116* (1), 8–23.

Madhok, Raahil, "Infrastructure, Institutions, and the Conservation of Biodiversity in India," 2025. Working paper.

Masanet, Eric, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey, "Recalibrating Global Data Center Energy-Use Estimates," *Science*, 2020, *367* (6481), 984–986.

McKinney, Michael L. and Julie L. Lockwood, "Biotic Homogenization: A Few Winners Replacing Many Losers in the Next Mass Extinction," *Trends in Ecology & Evolution*, 1999, *14* (11), 450–453.

Mytton, David, "Data Centre Water Consumption," *npj Clean Water*, 2021, *4*, 11.

Newbold, Tim, Lawrence N. Hudson, Samantha L. L. Hill, Sara Contu, Igor Lysenko, Rebecca A. Senior et al., "Global Effects of Land Use on Local Terrestrial Biodiversity," *Nature*, 2015, *520*, 45–50.

Noack, Frederik, Oliver Engist, and Ashley Larsen, "Migratory Birds and Ecosystem Services," 2025. Working paper.

Obringer, Renee, Benjamin Rachunok, Debora Maia-Silva, Maryam Arbabzadeh, Roshanak Nateghi, and Kaveh Madani, "The Overlooked Environmental Footprint of Increasing Internet Use," *Resources, Conservation and Recycling*, 2021, *167*, 105389.

Olmstead, Sheila M., "The Economics of Water Quality," *Review of Environmental Economics and Policy*, 2010, *4* (1), 44–62.

Rosenberg, Kenneth V., Adriaan M. Dokter, Peter J. Blancher, John R. Sauer, Adam C. Smith, Paul A. Smith et al., "Decline of the North American Avifauna," *Science*, 2019, *366* (6461), 120–124.

Sagar, Tushar, Dominic M. Ferraro, Jesse R. Barber, and Clinton D. Francis, "Acoustic Traits of Bird Species Predict Sensitivity to Anthropogenic Noise," *Dryad Digital Repository*, 2024.

Sekercioglu, Cagan H., "Increasing Awareness of Avian Ecological Function," *Trends in Ecology & Evolution*, 2006, *21* (8), 464–471.

Shapiro, Joseph S. and Reed Walker, "Why Is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade," *American Economic Review*, 2018, *108* (12), 3814–3854.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

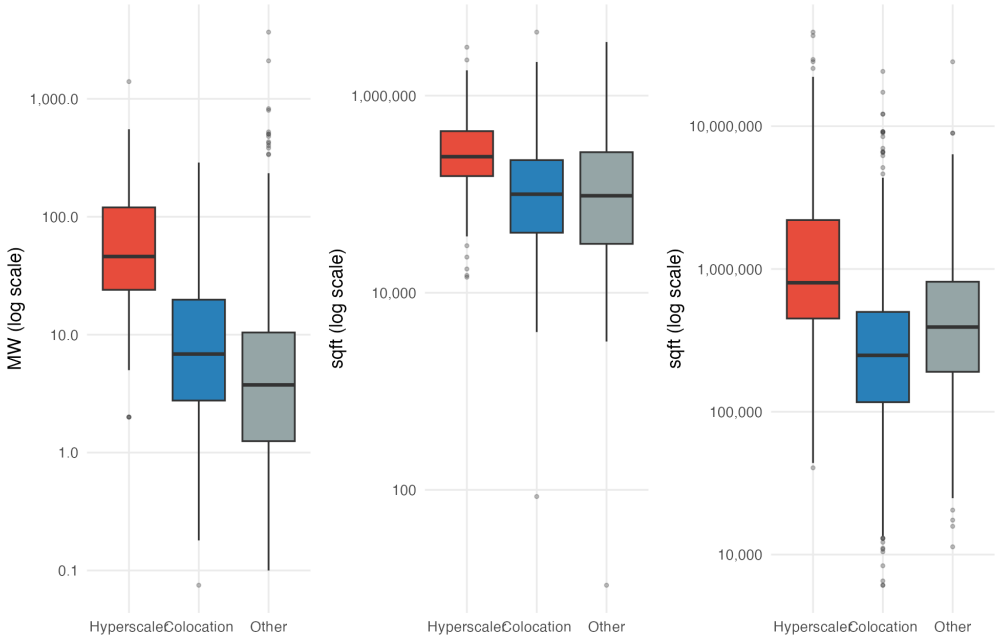
Zhu, Zhe and Curtis E. Woodcock, "Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data," *Remote Sensing of Environment*, 2014, *144*, 152–171.

Online Appendix

A Facility Characteristics and Placement

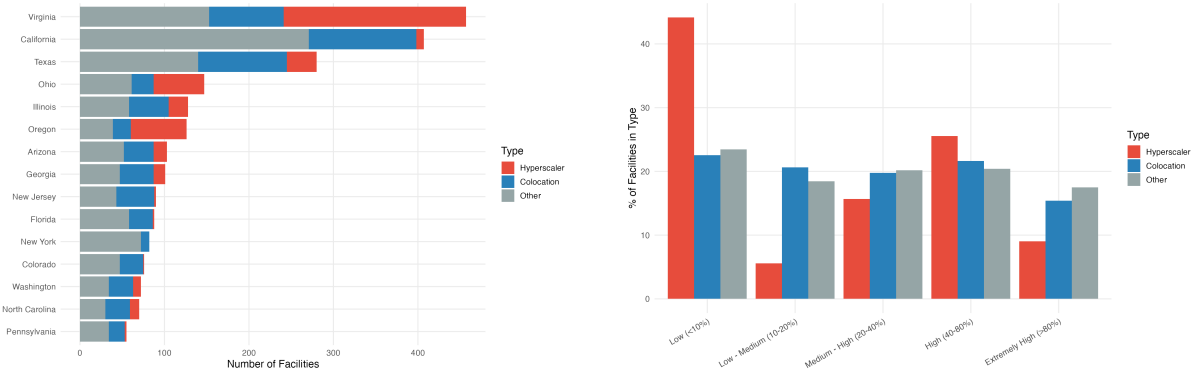
This section examines how hyperscaler data centers differ from non-hyperscaler facilities in physical size, geographic distribution, and pre-treatment environmental conditions.

Figure A1: Hyperscaler Placement: Facility Size



Notes: Distribution of facility size for hyperscaler vs. non-hyperscaler data centers. Hyperscaler campuses (Amazon, Google, Microsoft, Meta) are approximately seven times larger (median 46 MW vs. 5.6 MW). Source: Datacentermap; CoStar.

Figure A2: Hyperscaler Placement: Geography and Water Stress

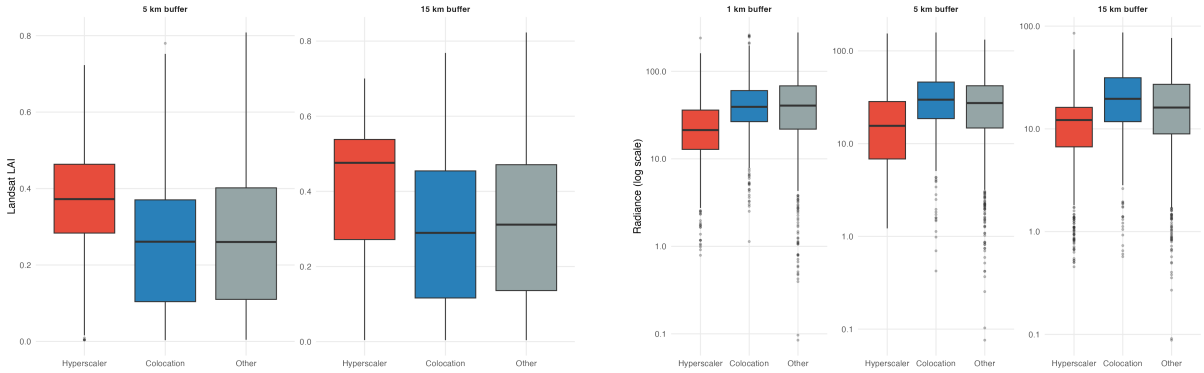


(a) Geographic Distribution

(b) Water Stress Distribution

Notes: (a) Geographic distribution of hyperscaler vs. non-hyperscaler data centers. Hyperscalers concentrate in Oregon (52%), Virginia (47%), and Ohio (40%). (b) Distribution of baseline water stress (WRI Aqueduct) by facility type. Source: Datacentermap; WRI Aqueduct.

Figure A3: Hyperscaler Placement: Pre-Treatment Environmental Conditions



(a) Pre-Treatment LAI **(b)** Pre-Treatment VIIRS

Notes: (a) Pre-treatment vegetation index (LAI) by facility type. Hyperscalers locate in greener areas (median LAI 0.476 vs. 0.306). (b) Pre-treatment nighttime radiance (VIIRS) by facility type. Hyperscalers locate in darker areas (median 21.3 vs. 40.1 $nW/cm^2/sr$), consistent with greenfield development. Source: Landsat; VIIRS; Datacentermap.

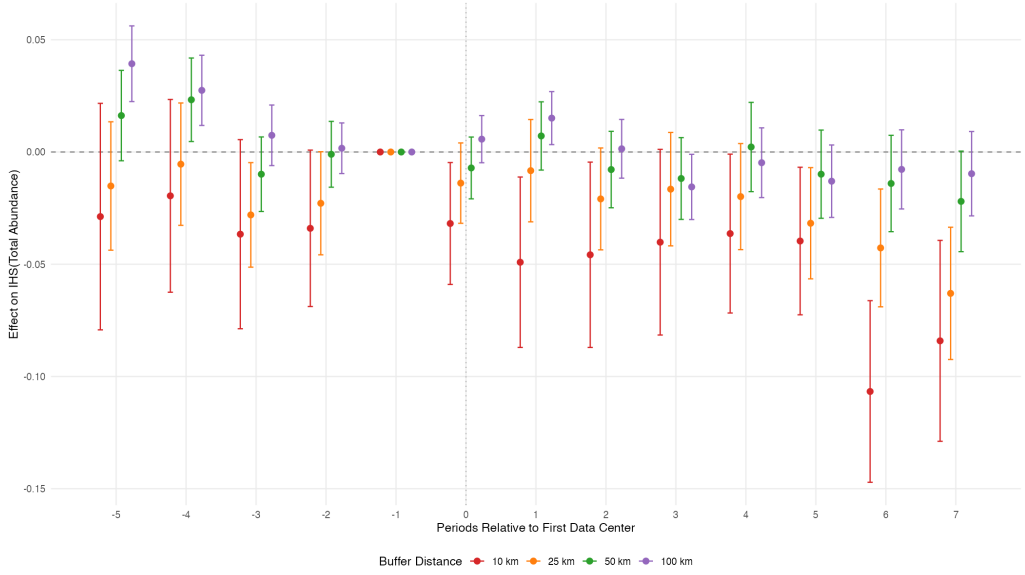
B Spatial Gradient Detail

Table C1: Spatial Gradient: Nested Buffer Results

Distance	ATT	Sig.
10 km	−0.055	***
25 km	−0.030	***
50 km	−0.019	**
75 km	−0.009	
100 km	−0.018	**

Notes: Sun-Abraham ATT for IHS(abundance) at nested buffers. The monotonic decay and absence of positive coefficients rule out spatial displacement. Same controls as Table 3. ***p<0.01, **p<0.05, *p<0.1. Source: USGS BBS; Datacentermap.

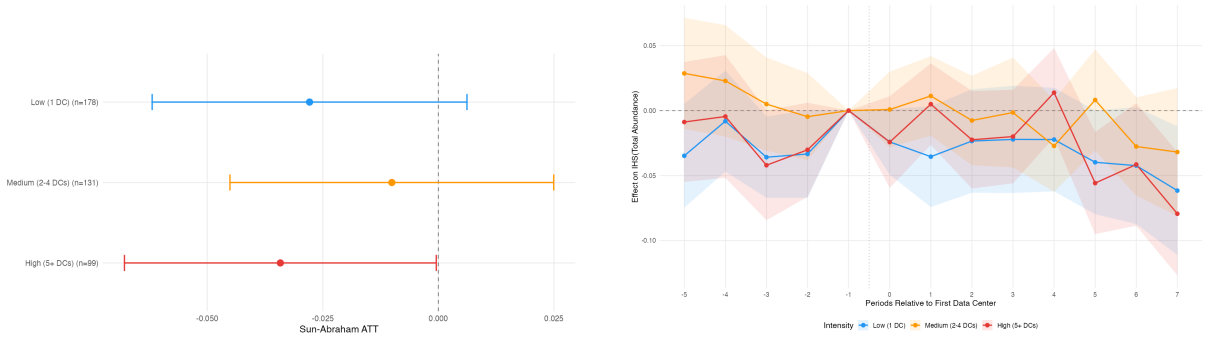
Figure C1: Spillover: Event Study Overlay by Buffer Distance (Abundance)



Notes: Sun-Abraham event-study coefficients for IHS(abundance) at multiple buffer distances (10, 25, 50, 75, 100 km) overlaid on the same axes. Monotonic attenuation with distance and absence of positive coefficients supports true population loss rather than spatial displacement. Source: USGS BBS; Datacentermap.

C Heterogeneity Detail

Figure D1: Intensity/Dose-Response Event Studies



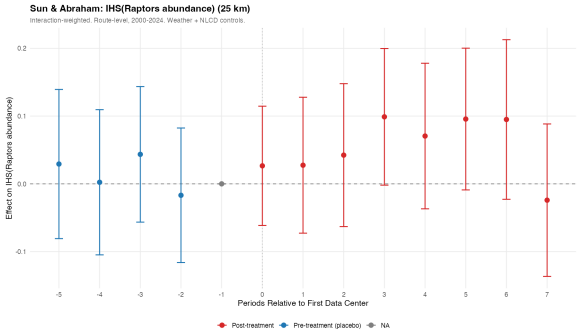
(a) ATT Comparison by Intensity

(b) Event Study Overlays

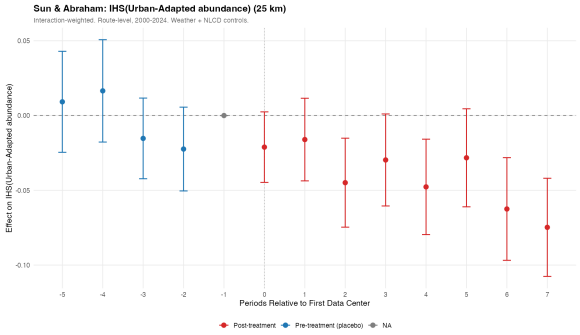
Notes: (a) ATT for IHS(abundance) by DC intensity bin. High intensity (5+ DCs) shows the largest and only significant decline. (b) Overlaid event studies by intensity tercile. Source: USGS BBS; Datacentermap.

D Community Composition and Ecological Restructuring

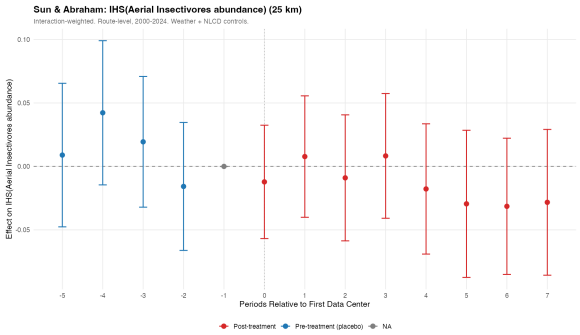
Figure E1: Guild-Level Event Studies: Abundance (25 km)



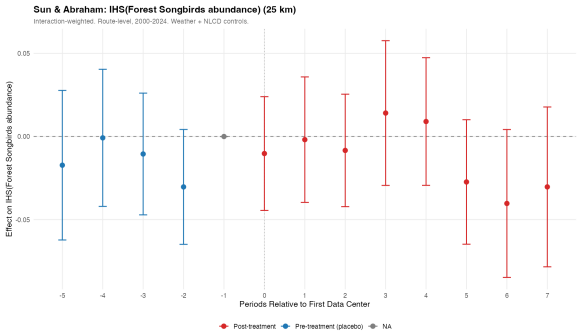
(a) Raptors



(b) Urban-Adapted



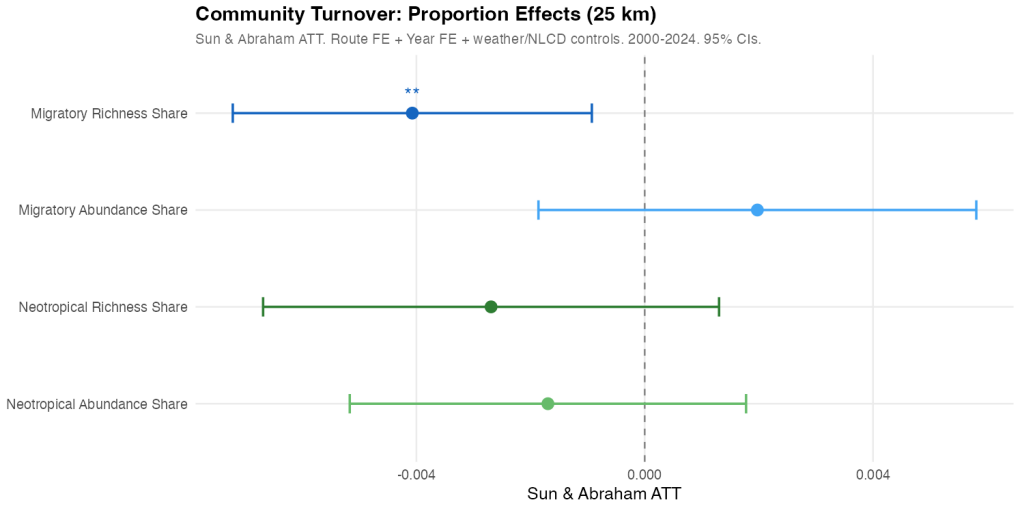
(c) Aerial Insectivores



(d) Forest Songbirds

Notes: Sun-Abraham event-study coefficients for IHS(abundance) by species guild at 25 km. Raptors show sustained post-treatment increase; urban-adapted species decline; aerial insectivores and forest songbirds show weak effects. Source: USGS BBS; Datacentermap.

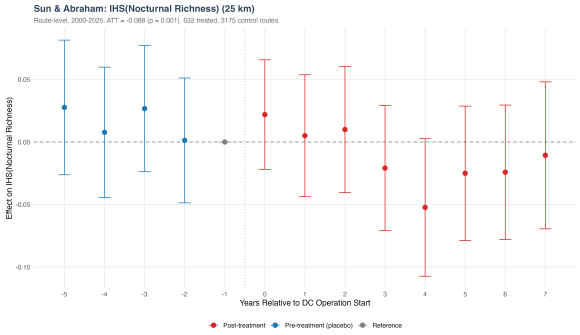
Figure E2: Community Turnover ATT Comparison (25 km)



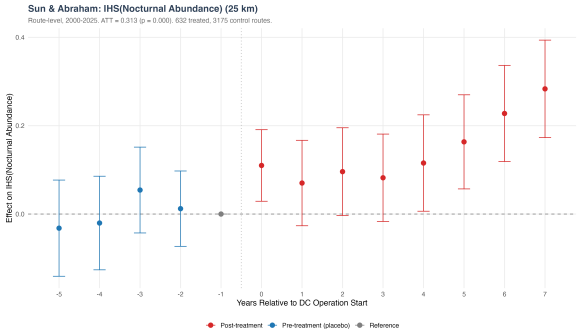
Notes: Sun-Abraham ATT and 95% CIs for community turnover proportions at 25 km: migratory richness share, migratory abundance share, Neotropical richness share, Neotropical abundance share. The significant decline in migratory richness share indicates selective loss of migratory species. Source: USGS BBS; Datacentermap.

E Nocturnal Species Detail (eBird)

Figure F1: eBird Nocturnal Species: Main Event Studies (25 km)



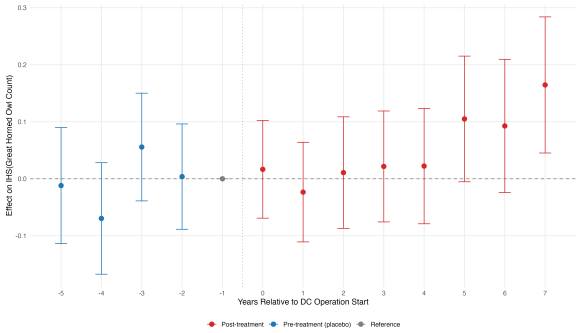
(a) Nocturnal Species Richness



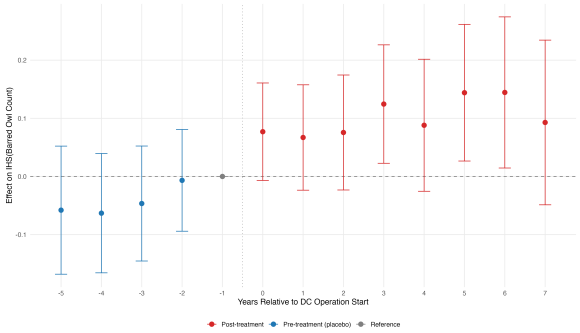
(b) Nocturnal Species Abundance

Notes: Sun-Abraham event-study coefficients for IHS(richness) and IHS(abundance) of nocturnal species from eBird at 25 km. Cell fixed effects, year fixed effects, effort controls. Source: eBird; Datacentermap.

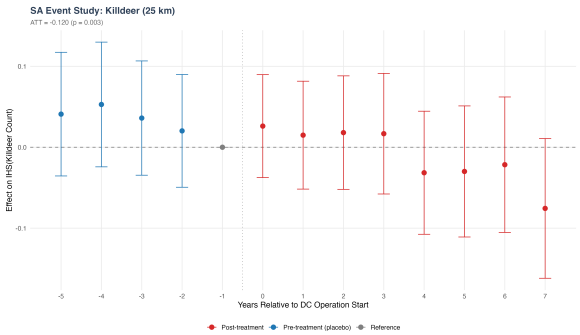
Figure F2: eBird Species-Level Event Studies (25 km)



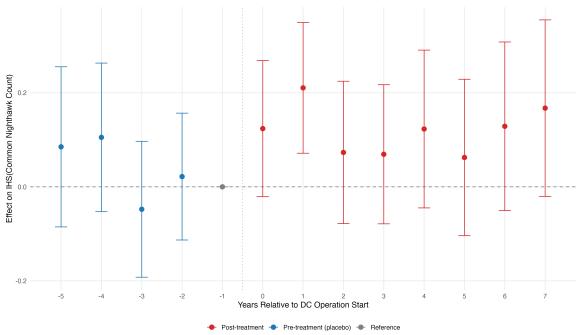
(a) Great Horned Owl



(b) Barred Owl



(c) Killdeer

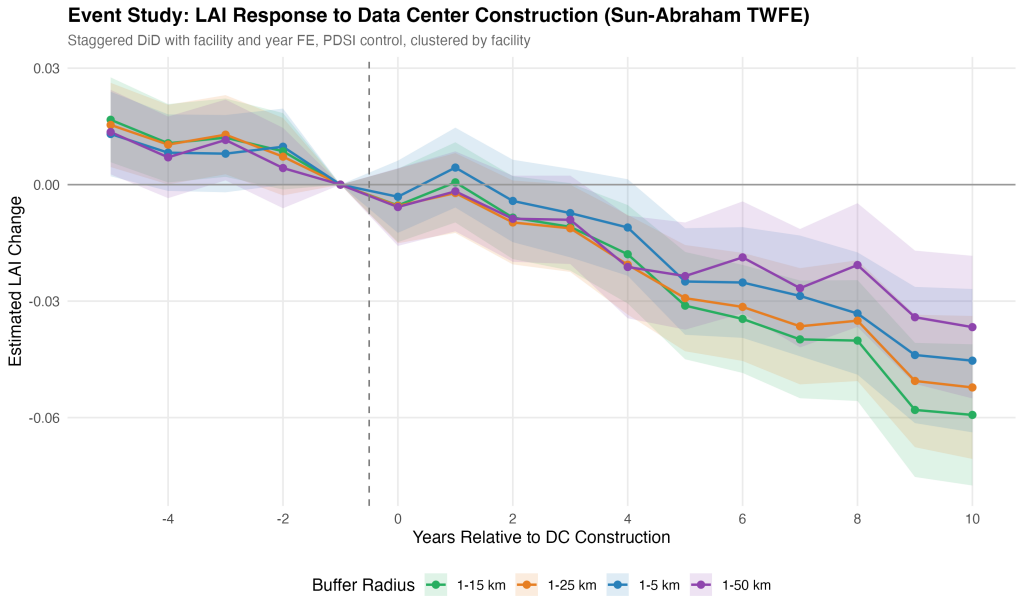


(d) Common Nighthawk

Notes: Sun-Abraham event-study coefficients for individual nocturnal species at 25 km. Great Horned Owl (+51%) and Barred Owl (+29%) are visually oriented predators that benefit from illumination. Killdeer (−11%) is a ground-nesting species disrupted by continuous light. Common Nighthawk (+19%) exploits insects attracted to artificial light. Source: eBird; Datacentermap.

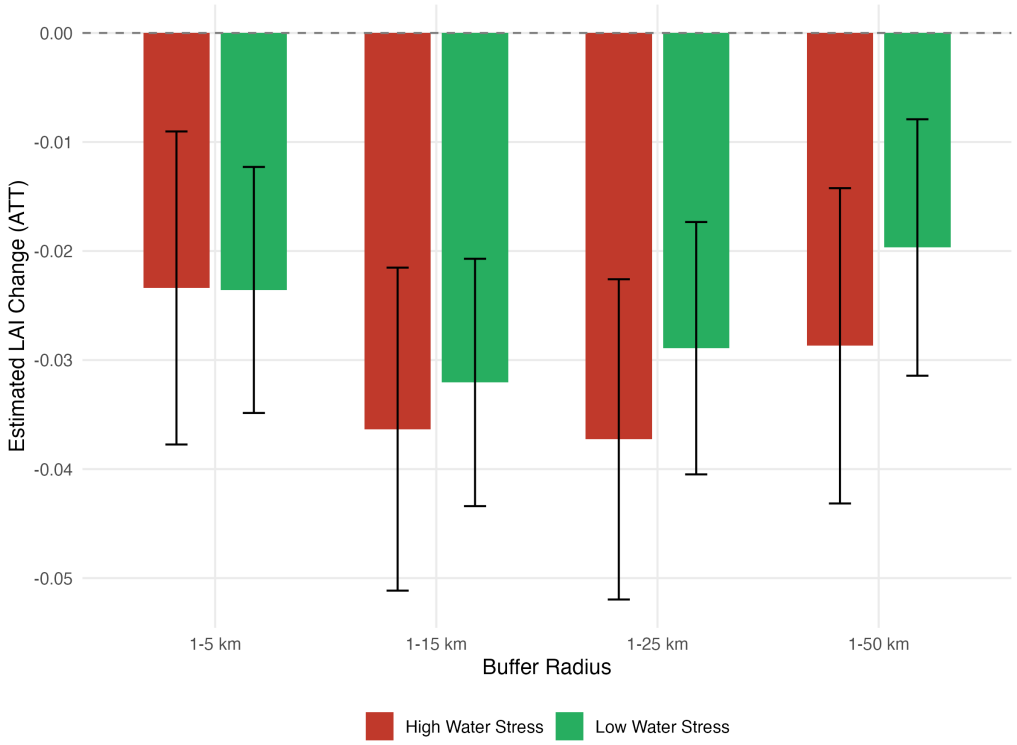
F Vegetation Health Detail (LAI)

Figure G1: LAI: Sun-Abraham Event Study (All Buffers)



Notes: Sun-Abraham event-study coefficients for LAI at multiple buffer distances. Vegetation loss is immediate at close distances (direct land conversion) and more gradual at farther distances (water table and microclimate effects). Source: Landsat via Google Earth Engine; Datacentermap.

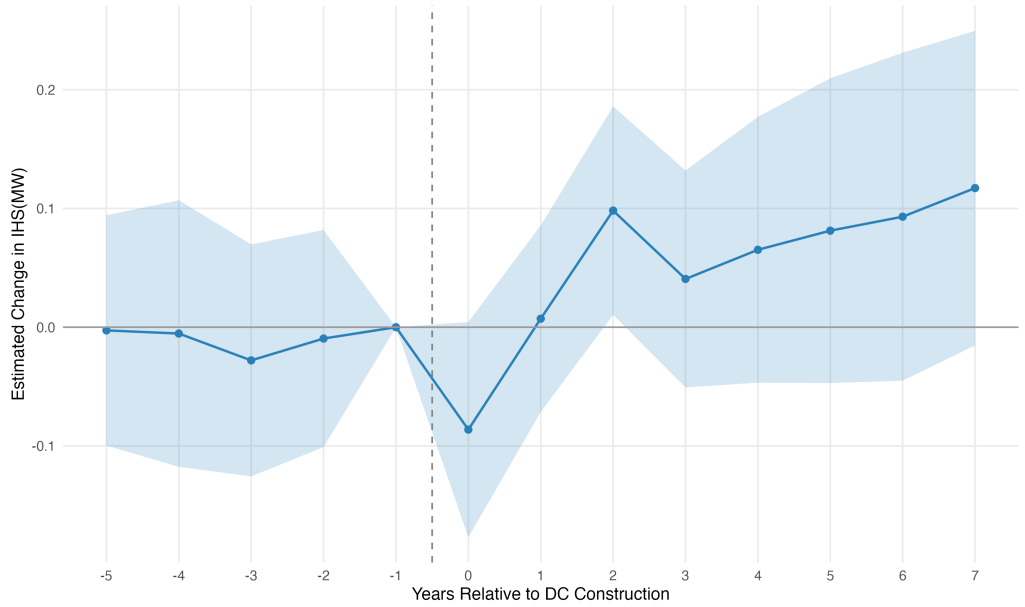
Figure G2: LAI: Heterogeneity by Water Stress



Notes: LAI ATT by WRI Aqueduct baseline water stress. High-stress facilities produce 1.7 times larger vegetation declines. Source: Landsat; WRI Aqueduct; Datacentermap.

G Grid Expansion and Groundwater

Figure H1: Grid Capacity First Stage: Data Centers and Local Generation (10 km)



Notes: Sun-Abraham event-study coefficients for IHS(cumulative MW) within 10 km of BBS routes. Data center openings cause significant grid expansion (ATT $= +0.33$, $p < 0.001$). EIA Form 860 data, excluding wind generation. Route fixed effects, year fixed effects. Source: EIA Form 860; USGS BBS; Datacentermap.