Infrastructure, Institutions, and the Conservation of Biodiversity in India

JOB MARKET PAPER

Click Here For Latest Version

Raahil Madhok*

University of British Columbia

January 15, 2023

Abstract

Biodiversity in the tropics is severely threatened by land use change. This paper studies how infrastructure expansion degrades biodiversity in India and the role of local institutions in mitigating the tradeoff. Combining new data on infrastructure-driven deforestation with one million birdwatching diaries, and using within-observer variation for identification, I document a sizeable infrastructure-biodiversity tradeoff. Transport, irrigation, resettlement camps, and mining projects account for 20% of total species loss. Publicly owned projects are especially harmful, and species diversity does not recover in the medium run. Lastly, I find that species loss is more than halved when local institutions enable marginalized communities, who are excluded from project planning, to mobilize around their interests. Informed consent by indigenous tribes is a key mechanism, underscoring the importance of grassroots institutions for balancing development and conservation.

Keywords: sustainable development, economic development, infrastructure, biodiversity, conservation, institutions, political economy.

JEL Codes: Q01, Q56, Q57, Q20, O13.

---

*University of British Columbia, Food and Resource Economics Group, Faculty of Land and Food Systems. Email: madhokr@mail.ubc.ca. This paper has been previously circulated as "The Development-Biodiversity Tradeoff in India's Forests". For helpful comments, I thank Sumeet Gulati, Patrick Baylis, Frederik Noack, Siwan Anderson, Ryan Abman, Teevrat Garg, Katherine Wagner, Eyal Frank, Meredith Fowlie, Sonja Kolstoe, Subhrendu Pattanayak, Bianca Cecato, Tatiana Zarate, and the Wildlife and Conservation Economics Lab at UBC. I also thank conference/seminar audiences at the NBER Summer Institute (EEE), Vancouver School of Economics, Columbia IPWSD, MIT/Sloan Energy Camp, CREEA, AERE, WEAI, PacDev, and Camp Resources. Veena Vinod provided excellent research assistance for data digitization. I am grateful for the SSHRC CGS Doctoral Fellowship for generous funding. All errors are my own.

1 Introduction

Global infrastructure spending totalled $US 2.3 trillion in 2015 (Oxford Economics, 2017). Although crucial for economic growth, infrastructure expansion narrows the frontier between human activity and fragile ecosystems. The ecological threat from encroachment is especially acute in the tropics, home to two-thirds of Earth's biodiversity yet where over 60% of global infrastructure spending occurs (FAO and UNEP, 2020). This is exacerbated by the fact that millions of indigenous people—who have supported biodiversity for millennia—are displaced by, disaffected by, or otherwise excluded from project planning.

Economists have long sought how to reduce environmental costs of development (Grossman and Krueger, 1995; Dasgupta et al., 2002; Copeland and Taylor, 2004). Biodiversity receives little attention in this literature (Frank and Schlenker, 2016), let alone grassroots solutions for balancing development and conservation. Filling this gap thus requires not only estimates of the ecological threat from infrastructure, but also the role of local institutions for neutralizing it.

The first goal of this paper is to provide a deeper understanding of the extent to which infrastructure expansion drives biodiversity loss. I call this the infrastructure-biodiversity tradeoff. The second goal is to investigate the role of decentralized forest governance in mitigating the tradeoff. Better understanding these socio-ecological and institutional processes can assist countries to meet the dual objectives of development and conservation.

The broad setting is the tropics, where over half of global deforestation occurs (Pacheco et al., 2021). India notably avoided widescale forest loss despite recording rapid economic growth (Forest Survey of India, 2019). It is unclear whether this is due to concerted tree-planting or changing definitions of forest cover. Even if development did leave forests unscathed, important inhabiting species may still become threatened and require policy attention. Elusive measurement of such species has led to biodiversity being overlooked in studies of anthropogenic deforestation (Foster and Rosenzweig, 2003; Burgess et al., 2012).

The first part of this paper estimates the infrastructure-biodiversity tradeoff in India's forests between 2015-2020. This constitutes a valuable setting for three reasons. First, India is among the planet's most biodiverse countries, home to 8% of global biodiversity and 12% of bird diversity (Venkataraman and Sivaperuman, 2018; Jayadevan et al., 2016). Second, India's biodiversity is documented by active "citizen scientists" who upload sightings on species-specific (e.g. eBird) or general (e.g. iNaturalist) platforms. India boasts the highest eBird membership of any developing country. Their geocoded uploads serve as a new, high-resolution biodiversity repository unmatched in the literature. Third, India publicly reports forest encroachments by infrastructure projects. Deforestation for building roads, mines etc. now account for 17% of yearly forest loss (authors calculation). The Forest Act (1980) mandates environmental review of such projects before construction. The review process underwent a transparency initiative in 2014, unlocking new administrative data for estimating threats to biodiversity.

To measure infrastructure development, I digitize the universe of deforestation permits awarded to firms that passed environmental review. This includes 7000 scraped from a public portal and 2000 digitized by hand. Each permit describes a forest patch diverted for construction and uniquely bundles infrastructure development and deforestation into a single variable. For analysis, they are aggregated into a cumulative measure of district-monthly forest area diverted for development. This new data improves on conventional satellite measures because the latter overlooks the source of deforestation. Pixel values are also annual aggregates, which masks deforestation throughout the year. In contrast, my data directly measures development-driven deforestation and features sharp monthly landscape changes as projects roll out. I do, however, use satellite data to verify that *approved* projects trigger *actual* deforestation.

To measure biodiversity, I obtain one million geocoded birdwatching diaries from eBird, the world's largest crowd-sourced platform for wildlife sightings (Sullivan et al., 2009). Birds are a credible indicator species for ecosystem health, sensitive to environmental change, and documented with high spatial precision (Morrison, 1986; Fraixedas et al., 2020). Each diary reflects a unique birdwatching session—which I call a "trip"—and lists the date, start/end time, GPS coordinates, and a taxonomy of species sightings. I count the number of species listed on diaries that reflect the local species pool, and drop ones reporting only highlights. This yields a representative biodiversity dataset with unparalleled spatiotemporal resolution, spanning 95% of districts from the Himalayas to the Western Ghats.

The matched panel permits a two-way fixed effects (TWFE) design, which I employ to estimate the impact of infrastructure development on bird species diversity (hereafter, species diversity) in a typical Indian district. The data allows me to characterize heterogeneity in a way that has eluded previous studies. I decompose estimates by project category, ownership (public/private), and shape (linear/nonlinear) to show which types of infrastructure are the least and most harmful. I also stratify districts by initial fragmentation level to reveal whether projects have differential effects in pristine or already-fragmented habitats. These results can help policy makers optimize the allocation of conservation budgets by choosing the location and infrastructure mix with minimal harm to local ecosystems.

Despite the power of citizen science, its opportunistic nature yields more sample selection than typical administrative sources. For example, users tend to visit more biodiverse locations, especially in the Western Ghats. There is also a Siberian bird migration to India in winter, and a lull in birdwatching activity during monsoons, which induces stark seasonality. Lastly, users possess a range of abilities, complicating inference from cross-user comparisons. I employ district fixed effects to address site choice, state-month fixed effects to address seasonality, and individual fixed effects to compare trips *within* the same user. Even after removing the ability bias, within-user residuals still trend upward due to learning. User-by-year fixed effects address this, and I show that this design outperforms others in the literature.

Endogenous sorting of both birds and users is the main threat to identification, even with the fixed effects. If construction pushes birds into less-fragmented districts, then the control group is contaminated. Similarly, if users sort towards biodiverse districts, then estimates are upward biased. I address both issues with spatial lags. First, I show that species immigration into a district is uncorrelated with nearby development. Second, I show that the number of district users does not increase when nearby districts become more developed. These tests are robust to different distance cutoffs, ruling out most concerns of endogenous mobility.

My analysis yields three key findings. The main result is that infrastructure development triggers substantial species loss. Ten squared kilometres of infrastructure encroachments reduces species diversity by 4%, as observed by the average eBird user. In contrast, the portion of these projects falling on non-forest land has no impact on species diversity, suggesting that habitat loss is a key mechanism. In aggregate, approximately 20% of the observed decline in species diversity over the study period can be attributed to development in India's forests.

Second, nearly all project categories drive the infrastructure-biodiversity tradeoff. The top three most harmful are resettlement, transport, and irrigation projects. Resettlements are akin to camps for relocating displaced communities. The negative impact of mining is surprisingly small, which I show is due to sparse eBird activity in remote mining districts. The mining impact doubles when the sample is restricted to high-activity districts. When grouping projects by ownership, I find that public projects drive the infrastructure-biodiversity tradeoff, mainly because the most harmful categories are almost all publicly owned.

Third, species are more resilient to infrastructure development in intact forests. Heterogeneity by baseline forest cover shows that the infrastructure-biodiversity tradeoff is halved in districts with one standard deviation higher initial forest cover. This suggests that degraded landscapes should be prioritized for protection, advancing a long-standing debate about how to target conservation budgets (Betts et al., 2017).

The results are robust to a variety of sensitivity tests. Estimates are stable under alternative fixed effects, dropping outliers, and alternative infrastructure measures. More intricate diversity indices, which account for species abundance, adds noise to the estimates due to low-quality counts. Estimates are also robust to dropping districts with below-median eBird activity, suggesting that findings are not driven by selected samples in low-activity districts. They are also robust to the sample of users only ever active in a single district, which rules out sorting by construction. Lastly, estimates are stable when dropping the year 2020, indicating that they are unbiased by the switch to "balcony birdwatching" during the COVID-19 pandemic.

The second part of the paper investigates which institutions minimize biodiversity loss. India is home to nearly 200 million members of forest-dependent indigenous tribes, who have been custodians of biodiversity on traditional forestland for millennia. Today, they are among the country's most economically vulnerable, politically excluded, and face livelihood loss as forests are handed over to commercial interests. I study whether inclusive local institutions that emphasize decentralized decision-making can mitigate the infrastructure-biodiversity tradeoff.

Data are from Banerjee and Iyer (2005) and indicate whether district institutions favour elites (extractive) or are more inclusive of the masses. The measure is based on whether historic tax collection was through a middleman or not. They find that non-middleman districts feature higher income equality today and better ability of the disenfranchised to mobilize around their needs. If tribal groups are better able to protect their livelihoods—which hinges on protecting forests—in inclusive districts, then better conservation outcomes are expected there.

The infrastructure-biodiversity tradeoff estimated in the first section is significantly smaller in inclusive districts. The implied magnitudes are large; the tradeoff is 70% smaller in these districts, where disaffected groups are better able to engage in the development process. Results are independent of tribal population share, suggesting that heterogeneity reflects institutional differences and not population pressures. These results underscore the importance of inclusive forest governance in achieving sustainable development.

The paper concludes by probing the mechanisms through which inclusive institutions mitigate the infrastructure-biodiversity tradeoff. This helps answer *why* inclusive development fosters conservation. I extract unique data from project permits reporting whether indigenous tribes were consulted during the review process and whether a supplemental cost-benefit analysis was commissioned during project review. I find that projects approved in inclusive districts are associated with significantly higher rates of informed consent and environmental scrutiny compared to those approved in extractive districts in the same state.

The results indicate that direct, grassroots participation in project planning, along with higher environmental standards, are key features of inclusive institutions that balance development and conservation. These findings are even more important given India's plans to prioritize Northeastern states—which have the higher forest cover and the largest tribal populations in the country—for industrialization (Nayak et al., 2020).

Literature Contributions. This paper contributes to three literatures. My main contribution is to provide the first country-wide evidence that infrastructure expansion triggers local species loss. Most economics studies that quantify infrastructure externalities estimate pollution costs (Currie et al., 2015; Hanna and Oliva, 2015). A handful have estimated forest loss: Asher et al. (2020) and Garg and Shenoy (2021) find surprisingly little effect of infrastructure on forest cover in India, and Baehr et al. (2021) also find muted effects in Cambodia. While this suggests that ecosystems are resilient to infrastructure, my results indicate otherwise. Using detailed species-level data, I document the extent to which infrastructure development erodes biodiversity.

The most similar paper is Liang et al. (2021), who study GDP and biodiversity in the United States[1]. Biodiversity is compiled from hundreds of ecological studies covering many animals. Development is measured by state-level GDP, which subsumes many underlying mechanisms driving biodiversity loss. In contrast, my data captures infrastructure development at the forest frontier. Despite the differences, our results are consistent: development drives species loss.

The second contribution of this paper is to extend the ecology literature by expanding the spatiotemporal scope of data and integrating empirical techniques from economics. The ecology literature has long documented anthropogenic pressures on ecosystems. In these studies,

[1]Related studies include Liang et al. (2020), Noack et al. (2021) and Noack et al. (2019). The first studies pollution effects on bird abundance, whereas I focus on effects of habitat loss. The second studies the impact of farm size on bird species diversity. The third studies the impact of plant species diversity on income fluctuations in the tropics.

field workers often count species in transects with different levels of human activity. This yields accurate data but limits analysis to cross-sectional comparisons (Reis et al., 2012; Stephens et al., 2004). Although citizen science dramatically improves data coverage, most interest in the ecology literature has been in identifying endogeneity—i.e. from seasonality, site choice, and detection ability—rather than conducting quasi-experiments (Callaghan et al., 2019; Kelling et al., 2019). I advance this literature by accounting for documented as well as undocumented biases, especially within-user learning, to arrive at quasi-experimental estimates of the infrastructure-biodiversity tradeoff. These estimates can be generalized beyond existing cross-sectional estimates due to the national-scale panel nature of my data.

The third contribution of this paper is to extend research at the intersection of political economy and conservation. I do this by showing empirically that inclusive institutions matter for natural resource conservation. A seminal literature shows how historic institutions shape modern economic development (Nunn, 2009), yet few have considered conservation outcomes[2]. In contrast, the conservation literature acknowledges the importance of institutions in moderating economy-environment tradeoffs, yet few have tested the claim credibly (Börner et al., 2020). I advance this literature by credibly estimating of the role of institutions in reducing species loss.

I am also able to fill a gap in this literature by pinning down mechanisms. Duflo and Pande (2007) use the same institutional data to show that the poverty impact of dams is muted in inclusive districts, and argue that the poor are better able to obtain compensation in these districts. Lee (2019) show that inclusive districts indeed have better contemporary state capacity. My results point to a mechanism with "teeth", namely, higher rates of informed consent in inclusive districts. My paper therefore ties together this literature and shows that grassroots institutions are crucial for designing successful sustainable development strategies.

The rest of this paper is organized as follows. The next section provides background on infrastructure-driven deforestation in India. Section 3 introduces the construction permit and citizen science data. Section 4 presents stylized facts to motivate the research design, described in Section 5. Section 6 presents estimates of the impact of infrastructure expansion on biodiversity. Section 7 explores the role of institutions for mitigating the tradeoff. Section 8 concludes.

2 Background

India's Forest Act (1980) Regulates Construction in Forests. India's Forest (Conservation) Act (1980) protects its forests from "conversion to non-forest uses" (MoEFCC, 1980). Infrastructure is among the main regulated activities because it fragments important habitats. For example, roads split contiguous forests into smaller patches, restricting wildlife movement and gene flow since many fauna avoid clearings as narrow as 30 metres (Riley et al., 2006; Benítez-López et

[2]Prior work has studied institutions and water conservation (Libecap, 2011; Hagerty, 2021), as well as institutions and forest conservation (Börner et al., 2020; Lal et al., 2021). Neither consider biodiversity per se. Most related is Noack et al. (2021), who show that different land institutions in historic East and West Germany led to differences in modern farm size which, in turn, impacts bird diversity.

al., 2010). Infrastructure also generates noise pollution, changes soil properties, and facilitates invasive species dispersal (Laurance et al., 2009).

When non-forest sites are unfeasible, the Act permits infrastructure encroachments pending a rigorous environmental review process. It also sets up a forest advisory committee (FAC), comprised of government officials and forestry experts, to rule on construction proposals. Projects involving any amount of deforestation, on any land recorded as forest in government records irrespective of ownership (i.e. public/private forest), undergo this review process.

Despite the Act's intent, huge swathes of India's forests have been transferred to public and private firms. Between 1985-2014, approximately 4000 km^2 of forest was clearcut for the construction of 23,000 infrastructure projects. Total deforestation during this period was 24,223 km^2 (Meiyappan et al., 2017)[3], implying that infrastructure intrusions accounted for 17% of India's deforestation during the three decades preceding this study.

Informed Consent Required Since 2006. Infrastructure-driven deforestation sanctioned by the Act is often at the expense of India's 200 million indigenous Scheduled Tribes (STs). STs are mainly forest-dependent and conserve biodiversity through traditional knowledge. Yet, they have been excluded from development decisions because, until recently, the Act historically gave powers only to state institutions.

In 2006, the landmark Forest Rights Act (FRA) democratized forest governance by granting STs formal rights to inhabit and manage village forests. Importantly, these rights imply that firms cannot divert forests without due process i.e. informed consent from the concerned Gram Sabha (village council)[4]. While examples of successful implementation exist, the FRA has become diluted over the years and is often flouted (see Menon (2016) for a timeline). Empirically, however, this provides variation to study the merits of inclusive institutions (Section 7).

Project Approval is Granted via Forest Clearances. The step-by-step journey of a project proposal is known as the Forest Clearance (FC) Process (MoEFCC, 2003). There are two stages: stage-I approval is granted after environmental review. Stage-II is granted after FRA compliance and payment into an offsetting tree-planting fund.

To receive stage-I approval, the firm first submits an application to the District Forest Office (DFO). The DFO may commission a site inspection report, which typically includes stipulations to change project size or location. The report is then forwarded to the State Forest Department, which can add more stipulations. At this point, approval is granted to small projects (0-5 ha., except mining). Medium projects (5-40 ha.) are forwarded to the Regional Office (RO) and large projects ($>$ 40 ha.) are forwarded to the Ministry of Environment Forests and Climate Change (MoEFCC). The FAC[5] rules on stage-I, indicating that environmental review is complete.

[3]Forest loss was 18,000km^2 from 1985-2005 (Meiyappan et al., 2017), and 6223km^2 from 2006-14 (Global Forest Watch)

[4]FRA guildelines: http://forestsclearance.nic.in/writereaddata/public_display/schemes/981969732$3rdAugust2009.pdf

[5]The regional FAC consists of senior RO and DFO officers as well as non-government forestry experts.

To receive stage-II approval, firms pay into a compensatory afforestation fund. The money funds tree planting on contiguous non-forest land or, if unavailable, land elsewhere in the state. Rates are fixed by the state and planting begins within 1-2 years (MoEFCC, 2013). In parallel, the DFO verifies ST rights over the tract in question and submits evidence of Gram Sabha consent to the MoEFCC. After fundraising and FRA compliance, the central FAC makes their decision and, if approved, the firm begins deforestation. My data consists of firms with stage-II approval.

Two features of the approval process are particularly relevant for my research design. First, afforestation may offset negative impacts of deforestation. However, the 1-2 year lag between approval and tree planting alleviates this concern since my data are monthly. Furthermore, India's compensatory afforestation program is fraught with issues. A recent audit found that just 7% of land secured for afforestation between 2006-12 had been planted in 2013 (MoEFCC, 2013). Second, *authorized* deforestation is not the same as *actual* deforestation. In the absence of official monitoring data, I verify that stage-II project approvals trigger actual deforestation using remotely sensed forest cover (Section 4.1). Based on this equivalence, references to approved project area in this paper can be interpreted as actual area of forest diversion.

The Forest Clearance Process was fast-tracked and digitized in 2014. In 2014, then-prime minister candidate Modi promised to speed up the FC process, which the Environment Minister referred to as a "roadblock to growth". Delivering on this after election, the annual rate of stage-I and II approvals jumped 60% compared to the previous 40 years (authors calculation). The share of rejected proposals also considerably declined post-2014. Faster approvals became possible through easing norms (e.g. diluting no development zones), exempting certain projects from FRA compliance, and changing thresholds for MoEFCC review. Pre-2014, projects larger than 15 ha. were reviewed by the MoEFCC. This threshold nearly tripled in 2014, implying that many projects are currently approved without Central involvement.

Another cornerstone of the fast-track initiative is the digitization of the FC process. An online portal automates each stage of the decision process and enables a reduction in turnaround time. For research purposes, an added benefit is process standardization, which reduces variation from state-level bureaucratic differences.

3 Data

I estimate the infrastructure-biodiversity tradeoff by drawing on several new datasets. I use newly digitized FC permits to measure development in India's forests. Species diversity is from eBird, a popular e-notebook for birdwatchers. I combine these with multiple high-resolution satellite datasets to control for weather, forest cover, and economic trends. The final panel covers all of India from 2015 to 2020. This section describes the data and provides summary statistics.

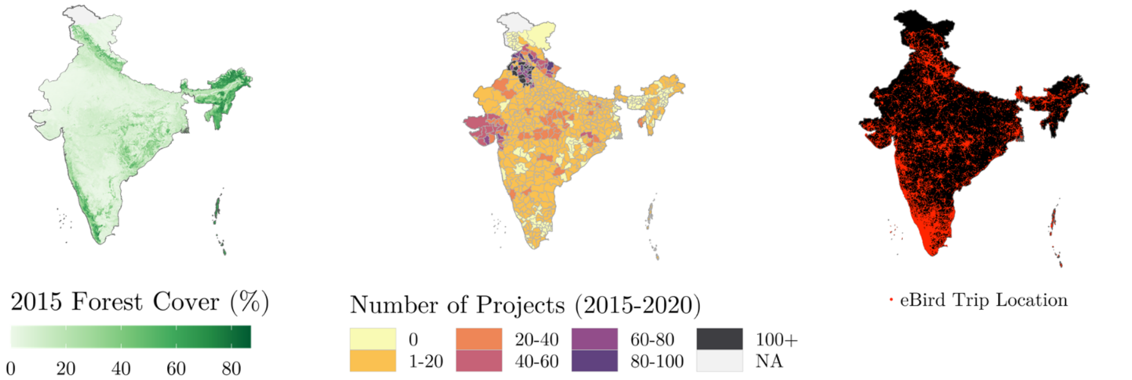


Figure 1: Infrastructure Encroachments and eBird Activity

Note: Panel A shows a heat map of 2015 forest cover from Townshend et al. (2017). Pixels are shaded by percent forest cover. Panel B maps the number of forest infrastructure projects that were approved for construction between 2015-2020. Panel C shows GPS coordinates of all birdwatching trips recorded through eBird during the study period.

3.1 Forest Clearances

Forests cover 22% of India (Figure 1A)(Forest Survey of India, 2019). This paper studies how encroachment of these forests by infrastructure impacts biodiversity. Administrative data on development-driven deforestation rarely exist, and previous work mainly relies on remote sensing. However, satellites have difficulty distinguishing anthropogenic intrusions from natural sources (e.g. forest fires). Moreover, the best satellite data report annual aggregates, which mask within-year encroachments that may have important local ecosystem impacts.

I construct a dataset of monthly infrastructure encroachments using newly digitized FC proposals approved between 2015-2020. All proposals submitted post-digitization and approved during the study period were scraped from the web portal ($N = 6,597$)[6]. I call this the digital subsample. Another 1,732 proposals submitted pre-digitization, but approved during the study period, were digitized by hand. I call these the manual subsample. These 8,329 projects, called the full sample, comprise the universe of industrial forest encroachments in India. Figure 1B shows the spatial distribution of projects. Figure A1 shows an example approval letter authorizing 185 ha. of deforestation for an irrigation project in Rajasthan.

Both the digital and manual subsamples report approved deforestation (in ha.) and project category (road, mine, etc), the most important variables for the analysis. District-wise deforestation is provided for multi-district projects (e.g. transmission lines). Digital applications additionally report: ownership type (public, private, neither[7]), non-forest land diversion, whether a cost-benefit analysis was commissioned, and whether informed consent was obtained from tribal land claimants. The latter enables analysis of how local institutions mediate the ecosystem impacts of infrastructure expansion (Section 7), a subject overlooked in prior studies. Ap-

[6] Data are publicly available at `www.parivesh.nic.in/`
[7] "Neither" are mainly joint public-private partnerships.

pendix B1 provides additional data details.

Project data are aggregated to the district and year-month level, both overall and category-wise (e.g. total deforestation in January 2018 for electricity projects in Delhi). I do this, firstly, because the district is the only consistent location identifier. Second, districts are the administrative unit immediately below the Indian state (similar to a United States county) and forms a natural unit for local policy implementation. The panel is balanced by zero-filling project approvals in districts not in the full sample (see cream color in Figure 1B). This is reasonable since all projects undergo the FC process, and the full sample contains the universe of approvals.

3.2 eBird

eBird entered the Indian market in 2014 and only requires a smartphone. Each birdwatching session (hereafter, "trip") is GPS-tracked, and users enter a taxonomy of sightings called a checklist. These are vetted by ornithologists on each upload (Sullivan et al., 2009). eBird is a revolutionary data product for research because it not only documents species observations, but also the observation process. This includes: trip date and time, duration, protocol (e.g. stationary or travelling), and whether all observed species were recorded, called a complete checklist, or only highlights. These data help identify checklists reflecting actual species diversity.

My sample frame is the eBird Basic Dataset for India (eBird Basic Dataset, 2019), which includes all trips between 2015-2020. To identify representative checklists, I start with stationary and traveling protocols, which characterize 99% of trips. The remainder often involve targeted searches (see Table A1 for a full breakup). Next, I keep complete checklists collected in < 5 hours and with group size ≤ 10. This produces the "gold standard of checklists for analysis", according to the eBird manual (Strimas-Mackey et al., 2020). Lastly, I link trip coordinates to 2011 district boundaries, which provides a matching key and also reveals off-coast boating trips, which are dropped. This leaves 1,049,930 trips by 17,634 users across 628 districts (out of 640). Figure 1C plots all locations from which species lists were uploaded.

Several studies use a similar selection procedure to ground-truth eBird. Horns et al. (2018) and Munson et al. (2010) find that population trends of North American species documented through eBird, and the more structured Breeding Bird Survey, are statistically similar. Callaghan et al. (2018) conduct their own bird census in Australia and find statistically similar species diversity compared with nearby eBird checklists. This ground-truthing builds confidence that eBird can be used to make reliable statistical inferences about factors affecting biodiversity.

My outcome measure is species richness, the number of unique species observed on a trip. Richness indicates conservation value, proxies the number and stability of ecosystem services, and is a widely used biodiversity metric (Fleishman et al., 2006). Alternative metrics integrate abundance weights, however eBird counts involve considerable measurement error; about 90% are approximated to the nearest tenth, and 10% of checklists have missing counts. Nevertheless, I estimate results with abundance-weighted diversity in the robustness checks.

The final panel aggregates mean species richness across users' trips in each district and time period, which keeps relevant variation and reduces noise. Otherwise, there would be no variation in development across trips in the same month and location since deforestation is district-monthly. I keep count of the number of trips over which the mean is computed and truncate at the 99th percentile to exclude outliers[8].

Who uses eBird? To estimate the infrastructure-biodiversity tradeoff, it is more important that eBird users collect representative data rather than themselves be representative of the population. The latter is implausible since users have smartphones and the privilege to engage in nature-based recreation. Nevertheless, citizen science is becoming the chosen data source in recent studies on the economic drivers of biodiversity loss, and will likely continue to be. Thus, it is important to frame a deeper discussion about who these citizen scientists are.

Since eBird does not record demographics, I characterize users by approximating home locations and then matching to the nationally representative Demographic and Health (DHS) survey. While this method, first proposed by Blanchard et al. (2021), does not directly impute demographics, it reveals how residents of users' home locations compare to the typical Indian. Perhaps unsurprisingly, I find that users are from more urban and better-off places. Yet despite statistical differences, they are not wildly atypical along either of these dimensions. Appendix B2 elaborates and provides supporting results (Figures B1-B2, Table B2).

3.3 Covariates

Environmental Covariates. The first set of covariates are environmental and include temperature and rainfall. Controlling for weather is important because it affects species detection. Monthly temperature (°C) is from the ERA5 reanalysis product on a $0.125^\circ \times 0.125^\circ$ grid (Hoffmann et al., 2019). Monthly rainfall (mm) is from the NASA GPM Level 3 product on a $0.1^\circ \times 0.1^\circ$ grid (Huffman et al., 2019). To compute the covariates, I extract means over all cells within a district, weighted by cell overlap fraction.

Observer Effort. The second set of covariates captures effort and includes: trip distance and duration, hour-of-day, protocol, group size, and spatial coverage. Duration (minutes), distance (km), and hour (0-23) are automatically recorded by eBird. Protocol equals one if the user is moving and zero if stationary. Group size is the birdwatching party size[9]. Spatial coverage accounts for projects opening up inaccessible forest patches (e.g. through supply roads), which may draw birders to new sites and upward bias the estimates. I measure it by overlaying trip coordinates onto a 10km grid and computing the fraction of district cells traversed by eBirders every month. All covariates are aggregated to their means during panel construction.

[8] As an example of one outlier, the maximum is 3779 trips in a district-month by a single user. This amounts to approximately 126 birdwatching trips *per day*.

[9] If multiple eBird users birdwatching together, only one needs to record a checklist. This is shared at the end, and the others can edit it to include additional species they observed.

Table 1: Summary Statistics of Forest Infrastructure Projects (2015-2020)

	Num. Projects	Mean Size (ha.)	SD (ha.)	Total Area (ha.)
Electricity	882	27.52	228.37	24,274.51
Irrigation	430	57.51	252.51	24,731.13
Mining	229	148.16	253.72	33,927.60
Other	4,448	2.36	34.51	10,486.43
Resettlement	44	71.53	92.92	3,147.23
Transportation	2,296	10.88	32.49	24,985.95
Total	8,329	14.59	110.63	121,552.87

Note: Data are arranged at the project level for 8,329 approved projects, prior to aggregating to the district level. Total area (ha.) is the summed deforestation area of all projects in the sample for the particular category.

Economic Trends. Consider a transport corridor being built to connect two fast-growing districts in a forested area. Biodiversity is threatened by habitat degradation from road construction and by polluting industry in each district (Liang et al., 2020). My goal is to disentangle the two and isolate the impact of infrastructure per se. In the absence of district GDP data, I control for satellite-detected nightlight radiance, a useful proxy for short-term GDP fluctuations (Henderson et al., 2012). Data are obtained from the VIIRS satellite at 500m resolution and aggregated to a district-month (Elvidge et al., 2017). Note that projects themselves should not be captured by satellites since construction is largely incomplete during the study period.

3.4 Summary Statistics

Figure 1 visualizes the main variables. Both sparse and dense forests (Panel A) have been fragmented by infrastructure (Panel B). The dense forests of Northern India suffered the most encroachment. These regions are also popular eBird destinations (Panel C), providing needed variation for the analysis. Users are also particularly active in Southern and Central India.

Table 1 shows that infrastructure triggered 122,000 ha. of deforestation between 2015-2020. The average encroachment is 14 ha., roughly 20 soccer pitches. Mines and resettlement projects are few in number but massive in size. Mines account for 3% of projects but 30% of total deforestation. Resettlements are least common and second-to-mines in size. Contrastingly, "other" projects are most common, but mainly reflect small patches. Transportation is the only category that is both numerous and accounts for a large (20%) share of total deforestation.

Table A2 shows summary statistics by ownership and shape. Over 70% of projects are publicly owned (Panel A). These are about five times larger than non-public projects. When grouped by shape, only 10% of projects are non-linear, but these are five times larger than linear projects (Panel B). Table A3 shows the distribution of projects by ownership, shape, and category. Public projects are largest because resettlement and irrigation projects—the second and third largest category—are almost all publicly owned (column 1). Mining and "other" feature a more even public-private split than any other category. Non-linear projects are largest because

Table 2: Correlation between approved and actual deforestation

	Linear	Weighted	Log
Outcome: Forest Cover (km^2)	(1)	(2)	(3)
Infrastructure (km^2)	-1.048	-3.369*	-0.022**
	(0.941)	(1.748)	(0.009)
Controls	Yes	Yes	Yes
District FEs	✓	✓	✓
Year FEs	✓	✓	✓
N	4480	4480	4480
R^2	0.973	0.977	0.978

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are aggregated district-yearly. The outcome is km^2 of forest cover in a district. Approved deforestation is cumulative up to the year end. Column 1 is a linear specification, with both outcome and explanatory variable in km^2. Column 2 adds regression weights equal to the inverse number of projects underlying the district aggregation. Column 3 is a log-log specification with no weights. All specifications include controls for nightlights as well as district and year fixed effects. Standard errors clustered by district.

nearly all mining and resettlement projects are non-linear (column 5). Note that statistics in both tables are from the digital subsample of permits report ownership and shape.

Table A4 summarizes eBird activity. Over 1600 trips by 100 users are recorded in the average district during the study period. Users themselves are quite active: the typical person records in four districts, two states, and six different periods. This within-user variation is the cornerstone of my identification strategy (Section 5). About 23 species are observed during the average trip.

4 Empirical Patterns

This section presents four stylized facts that make the data ideal for studying the infrastructure-biodiversity tradeoff in India. The first fact verifies that FC approvals actually trigger deforestation. The second and third illustrate shortcomings of citizen science for causal inference as well as remedies. The fourth fact is that users are highly mobile, providing a source of spatial variation for identification. These facts motivate the empirical strategy in section 5.

4.1 Fact 1: Approved deforestation triggers actual deforestation

This paper uses newly digitized infrastructure permits to study development-driven deforestation. Using this data to estimate biodiversity impacts requires a crucial assumption: that authorized deforestation equates to actual deforestation. I validate this using forest cover observed from outer space. Annual forest cover (% of a pixel) is from the VCF satellite product and converted to km^2 (Townshend et al., 2017)[10]. Since validation data is annual, I estimate the

[10] I convert to km^2 by multiplying cell values (% forest cover) by pixel area and then summing over district boundaries.

following equation on aggregated data:

$$ForestCover_{dst} = \alpha + \beta_1[Infrastructure]_{dst} + \beta_2[X]_{dst} + \gamma_d + \theta_t + \epsilon_{dsy} \quad (1)$$

$ForestCover_{dst}$ is *actual* forest cover in district d of state s in year t. $Infrastructure_{dst}$ is *approved* deforestation, the cumulative area of approved encroachments in year t in the same district. X_{dst} is a control for nightlight intensity, which helps disentangle infrastructure encroachment from other anthropogenic drivers of forest loss. γ_d and θ_t are district and time fixed effects, respectively. $\beta_1 < 0$ tests whether approved deforestation translates into actual deforestation.

Forest cover declines as districts approve more projects (Table 2). The linear specification in column 1 suggests that one km^2 of deforestation permits awarded to firms leads to one km^2 of forest actually cut down. Low precision may arise from the fact that $Infrastructure_{dst}$ is a district aggregate of differently sized projects. Small project patches in particular may not be captured by the satellite. Column 2 weights equation (1) by inverse number of underlying projects to give more weight to observations more likely to be captured by the satellite. Precision improves and the coefficient magnitude increases, suggesting that forest cover declines beyond the permitted amount[11]. The inverse association is robust to a log-log specification (column 3), which changes coefficient interpretation to percentages. A 1% increase in deforestation for infrastructure results in a much smaller percentage reduction in *total* forest cover since infrastructure makes up only a fraction of total deforestation.

4.2 Fact 2: eBird activity surges in winter and in "prettier" districts

Citizen science is revolutionizing biodiversity monitoring through crowd-sourced data. However, loose restrictions on when, where, and by whom data are collected yields more endogeneity than typical administrative datasets. eBird records details about the observation process which can help mitigate these data quality concerns.

Seasonality. The seasonality bias arises from the ability to record trips at any time[12]. Figure 2A demonstrates stark seasonality in collective species richness (left axis), with a peak in winter when Siberian birds migrate to India, and a trough during lulls in birdwatching activity (right axis) during monsoons. I address seasonality by exploiting within-month variation so that all time-invariant differences across months, such as seasonal species fluctuations, are eliminated. I do this separately by state since migratory patterns vary regionally.

Site Selection. The site selection bias arises from the ability to record species from anywhere[13]. This triggers a convergence of activity in more "attractive" districts. Figure 2B shows that eBird

[11] The 1-to-3 relationship in Table 2 Column 2 does not necessarily mean that developers illegally deforest beyond the permitted amount. It could be due to spillover activity alongside the infrastructure site that is not captured by the nightlights.

[12] In contrast, other surveys like the North American Breeding Bird Survey only operate in specific months

[13] In contrast, the North American Breeding Bird Survey mandates specific routes

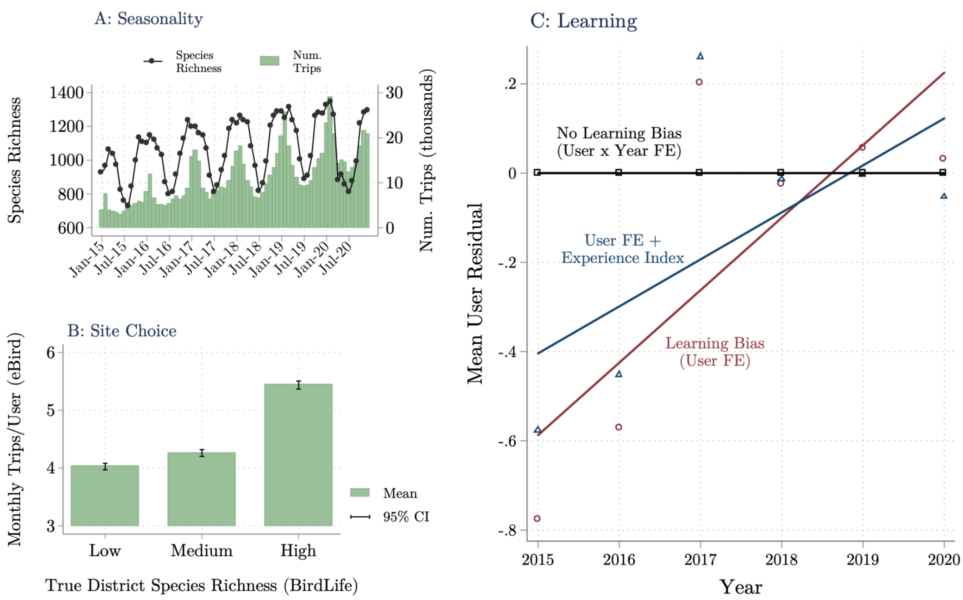


Figure 2: Biases in Citizen Science

Note: The left y-axis of Panel A shows total species richness across all users. The right y-axis shows the total number of their trips. Panel B shows mean number of trips per user-month in three quantiles of *true* species richness, obtained from historic range maps. In Panel C, red circles plot mean residuals per user from regressing species richness on user, district, state-month, and year fixed effects. Blue triangles show the same with a control for experience. Black squares partial out user-year, district, and state-month fixed effects.

users record more trips in districts with higher true species diversity (not eBird-reported) compared with less diverse ones. True species richness is computed by intersecting historic range maps of all known species in India (BirdLife International, 2018). Users populate longer checklists in more biodiverse districts since the species pool is larger. Previous studies account for this with a variety of fixed landscape covariates (Kelling et al., 2015), whereas I take a stricter approach with district fixed effects that rules out site selection more confidently.

4.3 Fact 3: Learning is a crucial source of bias in citizen science

Besides seasonality and site selection, another bias arises from pooling users with wide-ranging abilities (Farmer et al., 2014; Fitzpatrick et al., 2009). I decompose this into a fixed and variable component, called innate ability and the learning curve, respectively. Low-ability users might misidentify species or overlook them, whereas the opposite is true for high-ability users. Volunteer training or ability scores are typically used to reduce inter-observer variability. Kelling et al. (2015) construct a fixed ability score for eBird users based on predictions from a random effects model. For causal inference this requires a strong orthogonality assumption between the score and other unobserved user attributes. I relax this assumption by comparing species

Table 3: Variation in Species Richness Under Various Fixed Effects

	$1 - R^2$	σ_ϵ
	(1)	(2)
District FE	0.825	16.998
District + State-Month + Year FE	0.806	16.798
User + District + State-Month + Year FE	0.515	13.418
User-Year + District + State-Month FE	0.441	12.401

Note: This table summarizes regressions of species richness on various sets of fixed effects (rows). Data is at the user-district-month level. Column 1 reports $1 - R^2$ of the regression, indicating the fraction of variation not explained by the fixed effects. Column 2 is the standard deviation of the residuals (units = number of species).

richness across time and space *within the same user*, making the ability score superfluous.

Red circles in Figure 2C show species richness residualized on user, state-month, and district fixed effects. A steep upward trend remains, which is illustrative evidence of learning. Blue triangles show the same with a control for experience, in line with Kelling et al. (2015). The experience index increments with each additional trip and assumes constant returns. The learning curve flattens, but is not fully absorbed. This is evidence that learning is driven not only by short term experience, but also by longer-term unobservables (see gradual upward trend in Figure 2A). For example, a novice may detect the same common species month-to-month, gradually listing rarer species over longer time spans after learning their songs.

My solution hinges on restricting variation from within-user to within-user-by-year. This has three advantages. First, user-specific annual trends are removed, including accumulated trips, number of months per year of activity, and other longer-term learning indicators. Second, it is agnostic about the shape of the learning curve. A line of best fit (Figure 2C) is one possibility, but the true shape is unknown. Lastly, it allows for differential learning rates between users. Partialling out user-by-year fixed effects mechanically flattens the trend (black line), implying that residual variation has been stripped of the learning bias. This is the variation that I exploit to estimate the impact of infrastructure encroachment on biodiversity in Section 5.

4.4 Fact 4: Users are highly mobile across space and time

One concern with high-dimensional fixed effects is that they might absorb too much variation in species diversity. Removing district and state-month fixed effects leaves monthly within-state deviations from district means, e.g. the amount by which a district in Kerala is more species diverse than normal in a given month compared to its neighbour. User fixed effects subtracts additional variation by restricting district comparisons to those traversed by individual users. However, this is less of a problem if users are sufficiently mobile.

Summary statistics showed that the average user visits multiple districts and states (section 3.4). Figure A2 shows distributions of within-user mobility. About 90% of users visit between 1 to 4 states. Over 40% of users are active in multiple months over the study period.

Table 3 presents the identifying variation more formally. It summarizes regressions of species richness on different fixed effects and reports the amount of residual variation (column 1) and the standard deviation of residual variation (column 2). One-fifth of the variation in species richness is explained by seasonality and site choice (second row). About half is explained when user heterogeneity and learning is also accounted for (third and fourth row). Overall, substantial identifying variation remains——driven by users traveling across space and time—even after removing important biases in citizen science data. The residual standard deviation is 12-13 species in the most saturated specifications, providing a wide margin for identification. These findings underscore the richness of crowd-sourced data.

5 Empirical Strategy

My analysis leverages panel fixed effects to quantify the infrastructure-biodiversity tradeoff. Development projects fragment district forests throughout the study period. eBird users venture to these districts to record birds. My specifications compare species diversity *within* a user's trips as they travel for birdwatching. This identification strategy exploits rich cross-sectional and time-series variation, giving rise to plausible control groups within users: their recorded species diversity in districts and time periods with different levels of encroachment.

5.1 Main Specification

I estimate the following equation to reveal the infrastructure-biodiversity tradeoff:

$$SR_{idsym} = \alpha + \beta_1 [Infrastructure]_{dsym} + \beta_2 [X]_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \tag{2}$$

where SR_{idsym} is mean species richness observed by user i across their trips in district d of state s during year y and month m. $Infrastructure_{dsym}$ is the cumulative area of development encroachments in the same district and time period. X_{idsym} is a vector of weather and effort covariates described in section 3. Importantly, it includes a measure of spatial coverage that accounts for projects opening up previously inaccessible parts of the forest. It also includes a measure of general economic trends. User-by-year fixed effects, ϕ_{iy}, absorb cross-user differences in ability and account for individual-specific learning. District fixed effects, γ_d, absorb mean checklist length and other other fixed district attributes. State-by-month fixed effects, θ_{sm}, control for state-specific seasonality.

Conditional on covariates and fixed effects, β_1 identifies the impact of infrastructure development on species diversity by leveraging remaining variation across users' trips to different locations within the year. A negative sign means user i observes one fewer species *across trips in a district-month* following the encroachment. If species relocate elsewhere in the district, then it would likely be spotted by user i or others in the district, leaving species diversity unchanged. The fact that 10 users are active in the average district-month, each covering 10% of district area (Table A4), helps ensure this. A negative β_1 is even more alarming, then, as it implies the species is fully displaced from the district, along with the ecosystem services it provides.

Estimator and Counterfactuals. Since Equation (2) includes group (district) and time (state-month) fixed effects, estimation of β_1 falls under the umbrella of TWFE estimators. An emerging literature characterizes robust TWFE estimators (Roth et al., 2022) with a binary treatment (i.e. district receives a project) and control (i.e. no project). However, my setting features variation in treatment *intensity* whereby "no project" is just one of many counterfactuals.

Callaway et al. (2021) provide the only theoretical decomposition of TWFE estimators with continuous treatment. They show that, in my setting, β_1 represents the weighted average change in outcomes from incremental changes in infrastructure development across and within periods. Identification thus requires that the parallel trends assumption holds at every level of infrastructure development. I expand on this and other identification assumptions in Section 5.3. I defer to Callaway et al. (2021) for mathematical details of the estimator.

Clustering. Standard error clustering is an experimental design issue in this analysis, which leaves the choice of cluster somewhat subjective (Abadie et al., 2017). I cluster at the biome level, which forms the most appropriate cluster because biomes delineate biological communities with shared eco-climatic characteristics that are unobserved in my model. These characteristics may generate arbitrary correlation of ϵ_{idsym} across time and space within a biome. Digital maps of India's 12 biomes are obtained from the Nature Conservancy[14]. For districts spanned by multiple biomes, I select the one with the largest overlapping area as the cluster. An alternative choice of cluster is the district, since $Infrastructure_{dsym}$ varies at the district level. Although unobserved ecological components of biodiversity are unlikely to adhere to political boundaries, I report results with district-level clustering in a robustness check.

5.2 Additional Specifications

Decomposed Specification. I decompose $Infrastructure_{dysm}$ in Equation (2) into six separate categories: electricity, transportation, mining, resettlement, irrigation, and other. This not only reveals particularly harmful categories, but also those with negligible, or even positive, impacts on species diversity. I estimate the following specification:

$$SR_{idsym} = \alpha + \sum_{k=1}^{6} \beta_{1k}[Infrastructure]_{kdsym} + \beta_2[X]_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \tag{3}$$

where the term under summation is cumulative forest area diverted for projects of category k in district d. Remaining terms and subscripts are defined as in Equation (2). β_{1k} represents the conditional impact of infrastructure category k on species richness.

[14] I use the "Terrestrial Ecoregion" files accessed from https://worldmap.maps.arcgis.com

I use the same approach to estimate species diversity impacts by project ownership (public, private) and shape (linear, nonlinear). Together, these decomposed estimates reveal useful information for policymakers tasked with allocating the right mix of projects that balance development and conservation.

Treatment Heterogeneity. There is debate among conservation practitioners about whether biodiversity is better conserved by protecting intact or already-fragmented landscapes. While both approaches are valuable, some species may be surprisingly resilient to landscape modifications in certain habitats. I help resolve this debate by investigating whether the same infrastructure intrusion has differential effects by baseline ecosystem quality. I estimate heterogeneous treatment effects with the following specification:

$$\begin{aligned} SR_{idsym} = \alpha + \beta_1 [Infrastructure]_{dsym} + \beta_2 [Infrastructure_{dsym} \times E_d] \\ + \sum_{k=1}^{6} \beta_{2k} [Share]_{kdysm} + \beta_2 [X]_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \end{aligned} \tag{4}$$

where E_d is a fixed measure of ecosystem quality in district d. It is measured in two ways: 1) with forest cover in 2014, the year before my study period, and 2) with actual bird diversity from historic range maps (see section 4.2). I also estimate specifications controlling for the share of approved projects in category k, which disentangles area effects from category effects. This important since some project categories may dominate certain landscapes (e.g. mines are usually sited in remote, intact forests). β_2 reveals whether the infrastructure-biodiversity tradeoff is accentuated or muted in more pristine landscapes, independent of project type.

5.3 Identifying Assumptions

5.3.1 Parallel Trends

Identification with continuous treatment requires the parallel trends assumption: species diversity in districts that received an additional km^2 of projects, had they not received it, must be on the same outcome path as districts that never received the increment. I use a standard event study design to assess parallel trends. To clearly define before-after groups, the event dummy switches on at the date of the first forest diversion and persists thereafter. I estimate:

$$SR_{idsym} = \sum_{k=a}^{k=b} \beta_{1k} \times \mathbb{1}[t - e_d = k] + \beta_2 [X]_{idsym} + \phi_{iy} + \gamma_d + \theta_{sym} + \epsilon_{idsym} \tag{5}$$

where e_d is the event date in district d and $\mathbb{1}[t - e_d = k]$ is a dummy for k periods before or after the event. $\mathbb{1}[t - e_d = -1]$ is normalized to zero and omitted. All fixed effects and covariates are the same as Equation (2). Each β_{1k} captures mean species richness k months relative to the month before the event date[15]. Non-existence of pre-trends is indicated by $\beta_{1k} = 0 \; \forall \, k < 0$.

The results in Figure A3 provide evidence supporting the parallel trends assumption. The lack of pre-trends suggests that project placement is generally unpredictable. For example, if projects were selectively approved in high-growth districts, then species diversity would trend downwards even prior to project approval. Thus, while the FAC may approve projects based on certain district characteristics, these characteristics appear unrelated to changes in local species diversity around the time of approval.

In addition to parallel trends, identification also requires that districts be otherwise similar without the marginal encroachment at every base level, which boils down to a no-selection-bias condition (Callaway et al., 2021). While the event study results rule out many types of selection, I investigate two cases in more detail.

5.3.2 No Spatial Spillovers

The coefficient β_1 in Equation (2) is unbiased assuming no interference between units, known as the stable-unit treatment value assumption (SUTVA) (Imbens and Rubin, 2015). This requires that potential outcomes in district d depend on development in district d and nowhere else. SUTVA is violated in my context since habitat loss triggers species dislocation to other districts, introducing spatial dependencies unmodeled in Equation (2).

The severity of the SUTVA violation is ex-ante unclear. The bias is zero if species relocation is random, since spillovers would be orthogonal to infrastructure development. It becomes positive (negative) if development causes species dislocation to less (more) fragmented districts. Bias magnitude is determined by marginal species; positive bias converges to zero if dislocated species are already found in destination districts, even though the spillover is non-random.

I address SUTVA by explicitly modelling spatial spillovers. Spatial dependencies are specified with a spatial matrix, $\mathbf{W}$, which transforms the explanatory variable into a "spatial lag of X" (SLX) (Elhorst and Vega, 2015). This measures infrastructure in "other" districts, where other districts are spatially related through $\mathbf{W}$. I compute the below term and add it to equation (2):

$$SLX_{dsym} = (I_T \otimes \mathbf{W}_D)[Infrastructure]_{dsym} \tag{6}$$

where $\mathbf{W}_D$ is a symmetric $D \times D$ spatial weight matrix, where D is the number of districts in India. I_T is a $T \times T$ identity matrix where T is the number of year-months in the study period. The kronecker product signifies that $\mathbf{W}_D$ is applied to the infrastructure variable in each period and then stacked into a time series of "spatially lagged" development.

I use contextual knowledge to populate elements $w_{dj} \in \mathbf{W}_D$. In the preferred specification, I assume displaced birds relocate to other districts *within the same biome*, but less so to further

[15] I use the window [-6, 12] because a wider one creates variable sample composition across estimates. Since I study the first event only, the pre-post period differs depending on subsequent event timing. For example, with a 24-month post-period, $\beta_{1,24}$ from equation (5) would be noisily estimated off the few districts where no subsequent projects were approved for two years after the first.

away districts. This is modelled by setting $w_{dj} = \frac{1}{distance_{dj}}$, where $distance_{dj}$ is distance between centroids of district d and j, if they are in the same biome, and zero otherwise. The coefficient on SLX_{dsym} captures spillover effects of infrastructure development in one district on species diversity elsewhere. Conditional on this term, β_2 is purged of the spillover bias and less likely to violate SUTVA. Alternative spatial dependencies are implemented as robustness checks.

5.3.3 Exogenous Mobility

Another threat to identification in Equations (2)-(4) is endogenous user sorting. Since mobility is a key source of variation, estimates of the infrastructure-biodiversity tradeoff are unbiased only if unobserved determinants of species diversity are conditionally uncorrelated with site choice. To this end, my research design allows selection on time-varying observables and fixed unobservables. Suppose experience increases species detection and the probability of visiting pristine locations. This will not bias my estimates since I observe user experience. Similarly, citizen scientists may exhibit fixed, unobserved, heterogeneous preferences over nature (e.g. enjoying hiking). These preferences do not bias my estimates insofar as they affect site choice because they are absorbed by user fixed effects.

The orthogonality assumption is therefore violated if mobility reflects systematic *changes* in unobserved preferences. Suppose deforestation causes sorting towards more pristine districts. District fixed effects account for overall biodiversity, user fixed effects account for static preferences, but neither accounts for evolving deviations from within-user species diversity as they sort in and out of districts. These deviations are non-random if sorting correlates with project development. I test for sorting with the following spatial lag specification:

$$\begin{aligned} Users_{dsym} &= \alpha + \beta_1[Infrastructure]_{dsym} + \beta_2[X]_{dsym} \\ &+ \beta_3[SLX]_{dsym} + \gamma_d + \theta_{sm} + \mu_y + \epsilon_{dsym} \end{aligned} \tag{7}$$

where $Users_{dsym}$ are the number of users active in district d during time ym. The fourth term is defined as in equation (6), except $w_{dj} = \frac{1}{distance_{dj}}$ varies across the full support of $\mathbf{W}_D$ (instead of only within-biome). Remaining terms are as in Equation (2). A positive β_3 implies that development elsewhere attracts users into district d, but less so with distance. A negative β_1 implies that d's own infrastructure development pushes out users, conditional on those who sort in from elsewhere. I find no evidence of cross-district sorting, corroborating the orthogonality assumption and improving confidence in my research design. Within-district sorting is accounted for by the spatial coverage covariate. These results are discussed further in Section 6.1.

5.3.4 Cross-Cutting Programs

My research design guards against potential biases from cross-cutting development programs. These programs are unlikely to unfold with the same sharp timing as project approvals. For example, the Pradhan Mantri Gram Sadak Yojana program rolled out during my study period

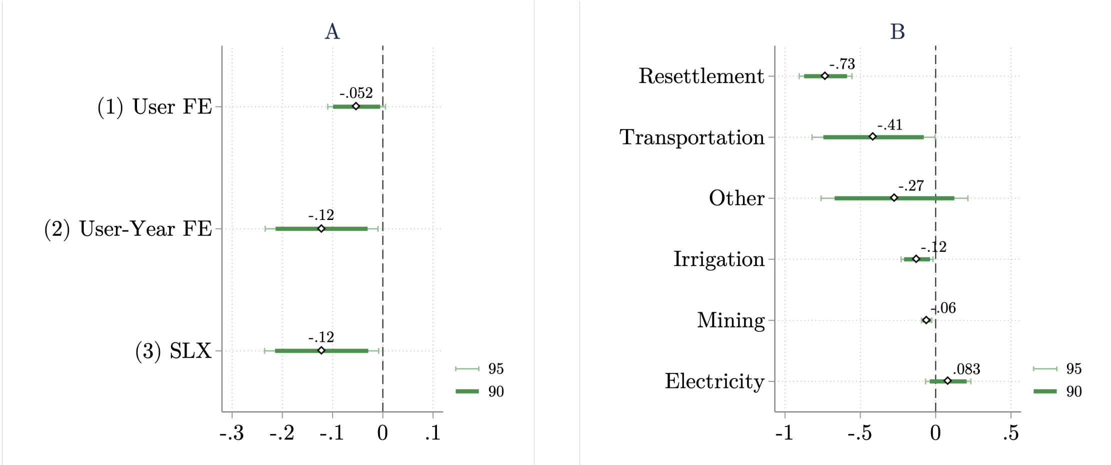


Figure 3: Estimates of the Infrastructure-Biodiversity Tradeoff in India

Note: The outcome in all specifications is mean species richness across users' trips in a district-month. Panel A shows coefficients on cumulative area of infrastructure encroachments in a district-month. Specification (1) includes fixed effects for user, district, state-month, and year. Specification (2) includes user-by-year, district, and state-month fixed effects. Specification (3) adds a term to (2) that captures spatial spillovers within the biome (see section 5.3.2). Panel B is a single regression with deforestation decomposed into project categories. Shaded bars denote confidence intervals. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Standard errors clustered by biome.

to provide roads to unconnected villages. These roads affect districts with and without infrastructure projects in my sample (or those yet to be treated). The threat is thus limited to factors differentially affecting districts with forest infrastructure and with the same timing as project approvals. This threat is minimal given that the targeted nature of most infrastructure programs are based on rules unrelated to forest cover (Asher et al., 2020; Burlig and Preonas, 2016).

6 Results

This section presents several pieces of evidence on the impact of infrastructure expansion on biodiversity. Species diversity is significantly threatened by infrastructure development in India's forests, across a variety of specifications. Resettlement, transport, irrigation, and mining projects are particularly harmful. The tradeoff is smaller in intact forests compared to fragmented ones. Lastly, species diversity does not rebound in the medium run.

6.1 Estimates of the Infrastructure-Biodiversity Tradeoff

Main Estimates. Figure 3A illustrates the main infrastructure-biodiversity tradeoff and Table A5 provides additional tabular estimates. Specifications (1) and (2) of Figure 3A estimate equation (2) with and without the learning curve, respectively. The main coefficient (β_1) is negative in both specifications, indicating that infrastructure intrusions reduce local species diversity. The difference represents a simple test of the learning bias in citizen science, which has rarely been quantified (Kelling et al., 2015). The upward learning curve counteracts species declines in specification (1), yielding a small coefficient with low precision. Removing this counterbalancing pressure in specification (2) yields a steep decline in species richness. Comparing estimates, the learning curve erases 60% of the species diversity impact if unaccounted for.

An additional km^2 of forest infrastructure in a district causes users to observe 0.12 fewer species, equivalent to 0.5% of the average checklist. To put this in perspective, eBird users observed 0.8 fewer species at the end of the study period compared to the start. The average district had 1.14 km^2 of forest occupied by infrastructure during this period, implying a loss of $1.14 \times -0.12 = 0.14$ species. Thus, infrastructure development accounted for $0.14/0.80 \approx 17.5\%$ of species diversity loss across India between 2015-2020.

Biases from spatial spillovers are minimal. Specification (3) in Figure 3A shows that species loss is unchanged when accounting for species displacement within the biome. Column 3 of Table A5 shows that, conditional on the direct effect, the spillover coefficient is positive but insignificant. Lack of spillovers does not mean species do not relocate following habitat disturbance. It means they do so in a way that is seemingly uncorrelated with local development.

Sensitivity: Controlling for Observables. Columns 4-6 of Table A5 probe sensitivity of the estimates by successively adding controls. When controls for observer behaviour and nightlights are removed (column 4), the coefficient remains negative but loses precision. The tradeoff appears when behaviour is included (column 5), suggesting that observer behaviour the key source of bias. This coefficient is equivalent to the main specification, which also includes a control for economic trends (column 3). Equal point estimates with or without nightlights suggests that species loss is driven by habitat loss, not general economic activity.

Column 6 adds diversion of *non-forest* land for the same projects as a covariate. It has no impact on species diversity, underscoring habitat loss as the key mechanism driving species loss as opposed to other infrastructure-driven disturbances such as pollution. It also implies that reorganizing economic activity away from forests can blunt biodiversity decline. Note that non-forest diversion is reported only in digital project proposals (80% of the full sample).

Ruling out Sorting. These results are not driven by sorting. Table A6 shows estimates from Equation (7). The outcome is log number of users, and infrastructure measures are standardized to ease interpretation. Infrastructure development in other districts j does not draw users into district d (column 1, second row). Neither does development in d push users elsewhere (first row). Other-district development is an inverse-distance weighted average of encroachments in districts within 100km of d. This result is also visible under alternative distance cutoffs (columns 2-3). Overall, eBird users are highly mobile (Fact 4 in Section 4.4), but not because of infrastructure development. This finding supports causal interpretation of the main estimates.

6.2 Estimates by Project Category, Ownership, and Shape

Estimates by Category. In a novel addition to the literature, Figure 3B provides estimates of the infrastructure-biodiversity tradeoff by project category (Equation 3). Each coefficient describes the impact of a marginal encroachment by projects of that category, conditional on that by all other categories. Five out of six categories negatively affect species diversity. Four of them—resettlement, transportation, irrigation, and mining—do so with statistical precision. Resettlements threaten species the most. An example is the diversion of 2.85 km^2 of forest in Betul district, Madhya Pradesh for relocating a village previously located in a nearby Tiger Reserve. The project was approved in April 2017 and includes housing, playgrounds, and roads[16].

The negative impact of "other" projects is imprecise. These are the smallest projects on average, but feature a standard deviation 17 times greater than the mean, the largest ratio of any category (Table 1). When aggregated to the district, a marginal encroachment thus comprises a variety of underlying patch sizes. The coefficient magnitude is likely driven by the few large projects, where marginal encroachments comprise a single patch, and the noise by the numerous smaller projects, each too small to affect species diversity with statistical precision. The same logic may explain the noisy impact of electricity projects, which has the second highest noise-to-signal ratio of project size. The largest ones are dams, which may explain the positive coefficient. Dams create reservoirs, which may attract previously unseen waterbirds.

Mining threatens species diversity minimally, which is surprising given the sector's notoriety for disrupting local ecology. The coefficient, however, is likely attenuated since mines are often sited in remote areas where few eBird users go. Those who do may be a selected sample that miscount the species pool, despite attempts to prevent this (section 3.2). Half of sample mines are in Odisha, Madhya Pradesh, and Chhattisgarh, with 27% in Odisha alone. The median number of users and trips in Odishi mining districts is under half of the national median.

Table A7 probes sensitivity of the estimates. Similar to the overall tradeoff, category-wise tradeoffs materialize when observer behaviour is accounted for (column 2) and remain stable when controlling for economic activity (column 3). To test the conjecture about the small mining effect, I estimate equation (3) in districts with above-median numbers of users recording above-median trips per user. If the bias is mining-specific, then the sample restriction should only accentuate the mining coefficient. Indeed, mining projects are twice as harmful in the high-activity sample whereas other coefficients are stable (column 4). This implies that non-mining projects are sited in districts with sufficient eBird activity to begin with.

Estimates by Ownership and Shape. Table A8 groups estimates by project ownership and shape. Each column successively adds controls and each coefficient describes the impact of a marginal encroachment by projects in that group. Public projects are the main threat to biodiversity (Panel A). This is unsurprising because over 88% of resettlement, transport, and irrigation

[16] The site inspection report for this project can be found at http://forestsclearance.nic.in/writereaddata/SIR/06022017561SBScan_02-06-2017_1501.pdf

Table 4: Treatment Effects by Baseline Forest Intactness

	(1)	(2)	(3)	(4)
Infrastructure (km^2)	-0.131***	-0.140***	-0.134***	-0.143***
	(0.019)	(0.016)	(0.024)	(0.021)
Infrastructure (km^2) × Baseline Forest Cover	0.063*	0.065**		
	(0.030)	(0.029)		
Infrastructure (km^2) × Baseline Species Richness			0.049**	0.051**
			(0.021)	(0.021)
Controls	Yes	Yes	Yes	Yes
Category Shares	No	Yes	No	Yes
User x Year FEs	✓	✓	✓	✓
District FEs	✓	✓	✓	✓
State × Month FEs	✓	✓	✓	✓
N	161782	161782	161782	161782
R^2	0.694	0.694	0.694	0.694

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is mean species richness across users' trips in a district-month. Forest infrastructure is cumulative area of development encroachments in a district-month. Columns 1 and 2 include an interaction with baseline forest cover in 2015 (% of district area). Columns 3 and 4 show interactions with baseline species richness, obtained from overlapping historic species range maps. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Columns 2 and 4 additionally control for the share of approved projects in each category. Standard errors clustered by biome.

projects—the three most harmful categories—are publicly owned (Table A2). Similarly, mining and "other" projects have the highest private ownership but smallest biodiversity impacts (Figure 3B), which explains the noisy estimates for private projects. Overall, these results suggest that deforestation for large-scale public infrastructure is a major source of ecological concern.

Similar logic helps interpret the estimates by project shape (Panel B). Non-linear projects have a robust negative impact on species diversity. The magnitude is a combination of the small mining and large resettlement coefficients in Figure 3B, which are almost entirely non-linear (Table A2). The tradeoff for linear projects is weaker but the magnitude is twice as large. The magnitude is a combination of the transportation, electricity, and irrigation coefficients, which are predominantly linear (Table A2). The large standard errors likely come from "other" projects, which are also mostly linear.

6.3 Species are More Resilient in Intact Forests

Given scarce budgets, should conservation focus more on intact or fragmented landscapes? This question is especially important in India, home to some of the most biodiverse, but also most degraded, places in the world. My estimates of Equation (4) suggest greater returns from conserving fragmented landscapes. Table 4 shows heterogeneous treatment effects using two measures of baseline ecosystem quality. Both are standardized so that a one-unit change can be interpreted on the same scale despite raw units differing.

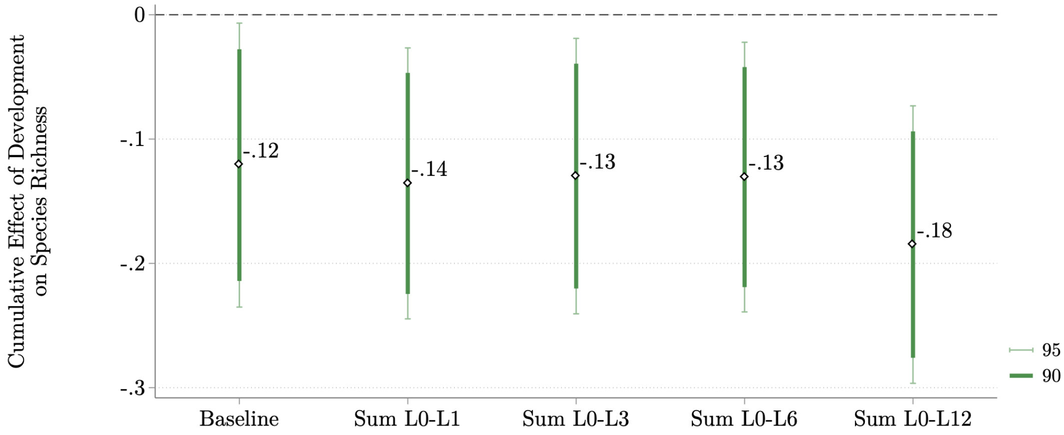


Figure 4: Cumulative Dynamic Lag Results

Note: "Baseline" repeats the main result with user-year, district, and state-month fixed effects. "Sum L0-L1" adds the first lag of forest diversion to the main specification and reports the sum of coefficients on the first lag and baseline effect. "Sum L0-L3" sums up to the third lag, and so on. Shaded bars are confidence intervals. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Standard errors clustered by biome.

Species are more resilient to infrastructure development in pristine districts. Column 1 shows that the adverse impact of infrastructure on species diversity is halved in districts with one standard deviation higher baseline forest cover[17]. This mitigating effect is robust to controlling for category shares (column 2), which ensures that the interaction reveals heterogeneous impacts of 1 km^2 of habitat loss independent of underlying project category. Remaining columns explore sensitivity to using baseline species richness from BirdLife range maps as the dimension of heterogeneity. The tradeoff reduces by a similar amount without (column 3) and with (column 4) controls for category shares. Overall, these results suggest that conservation should target degraded landscapes where species are more threatened.

This finding corroborates existing theory from landscape ecology (Hanski, 1998) and is among the first empirical tests. An exception is Betts et al. (2017), who find the opposite result: that species are more threatened in intact landscapes. They interact time-varying forest loss with an indicator for baseline forest cover above 90%, whereas I use a continuous measure of baseline forest cover that exploits substantially more variation.

6.4 Species Diversity Loss is Persistent

My estimates reflect *instantaneous* responses of species to anthropogenic habitat loss. Scientists increasingly view biodiversity as only partially determined by current habitat, and the rest by legacies of landscape change (Odum, 1969). The infrastructure-biodiversity tradeoff may therefore exhibit a lag as species diversity equilibrates. Whether current biodiversity exceeds

[17] Forest cover (% of a pixel) is from the VCF satellite product on a 250×250m grid (Townshend et al., 2017)

the carrying capacity of a recently modified landscape, generating "extinction debts", can be revealed through analyses of cumulative effects. Such tests are rare in the literature, risking conservation strategies becoming quickly outdated (Haddou et al., 2022).

I estimate a cumulative dynamic lag model to formally study treatment effect dynamics. This reveals the evolution of species diversity changes in the medium run. I estimate Equation (2) with lags of $Infrastructure_{dsym}$ and report the sum of the baseline and lagged coefficients, which is the net impact of habitat loss several periods later. Sharp increases in cumulative effects are evidence of the extinction debt being paid[18].

Species declines are triggered soon after habitat disturbance and largely persist through the medium run (Figure 4). The cumulative impact three months later ("Sum L0-L3") is nearly equivalent to the baseline impact, with stable point estimates under a narrower (1 month lag) and wider (6 month lag) window. Evidence of a small extinction debt appears one year after the disturbance ("Sum L0-L12"), when species diversity declines beyond the initial impact. It is unlikely that these results are contaminated by offsetting afforestation since tree-planting typically occurs 1-2 years after project approval (see section 2).

Overall, I find weak evidence of the extinction debt (i.e. lagged effects). While this is an important finding, I acknowledge that species richness may imperfectly capture cumulative responses to habitat change. Functional traits of species also affect long-term resilience. Furthermore, my study period is likely too short to reveal the full legacy of habitat disturbance.

6.5 Additional Sensitivity Checks

This section presents sensitivity checks on the estimated infrastructure-biodiversity tradeoff. These include: alternative specifications of spatial spillovers (Table A9), as well as alternative specifications for seasonality, extreme values, different diversity metrics and several sample restrictions (Table A10). Robustness tests by project category are shown in Table A11.

Spatial Spillovers. The SLX specification in Figure 3A accounts for spillover effects of habitat disturbance on species diversity elsewhere in the biome. That birds relocate within the biome is a contextual assumption modelled by a spatial dependency matrix (section 5.3.2). Table A9 shows results from allowing spillovers to instead materialize over different distances. Spatial dependencies are inverse distance weighted until a threshold and zero thereafter. Columns 1, 2 and 3 allow spillovers within 100, 200 and 500km, respectively. Column 4 allows nationwide spillovers. Direct and spillover variables are standardized for comparability.

Estimates of species loss remains stable and significant across all specifications. Spillovers also remain positive but noisy, similar to the baseline result. This increases confidence that the lack of spillovers is pervasive, not a data artifact from the within-biome assumption. Again, these results do not imply that spillovers are non-existent, but rather that species relocate in

[18] Increases in cumulative effects could also be evidence of lagged construction. However, Fact 1 (section 4.1) along with the fact that species diversity declines in the month of project approval suggests that this mechanism is unlikely.

way that is unrelated to local habitat loss.

User-specific Seasonality. The main estimating equation accounts for state-specific cycles of eBird activity. However, if more advanced users can better identify migratory species, then seasonality may exhibit an individual component (i.e. winter observations mainly reported by experts) (Johnston et al., 2018). My estimates would then become biased by not accounting for changing distributions of user types across months. I specify this scenario with user-by-month fixed effects, a demanding specification that relies on comparisons across districts and years within a user-month. To account for learning, I control for the number of months per year of birdwatching (this was swept away by the user-year fixed effect in Equation 2). Species richness declines by 0.15 per km^2 of development (Table A10 column 1), similar to the baseline estimate. This suggests that my estimates are not biased by individual-specific seasonality.

Regression Weights. Species richness is a mean over users' trips in a district-month. Part of the error variance in Equation (2) may be explained by differences in the number of underlying trips. Figure A4 shows the cumulative distribution: 10% of observations are a mean over more than 10 trips (precisely measured) and 90% are over fewer than 10 trips (imprecisely measured). I estimate the main equation with weighted least squares, with weights equal to the number of trips underlying each observation. This ensures that observations influence the coefficient in proportion to their measurement precision rather than be treated equally. The coefficient is virtually unchanged (column 2) and remains significant at the 5-percent level.

Removing Outliers. I transform the sample in two ways to test robustness to removing outliers. First, before aggregating, I drop India's top three "mega-projects": 1) the world's largest lift irrigation[19] project, located in Telangana and requiring 3168 ha. of deforestation, 2) a 4000 MW coal plant also in Telangana and requiring 4334 ha. of deforestation, and 3) the world's largest concrete dam, located in Arunachal Pradesh and requiring 5056 ha. of deforestation. The coefficient of interest from the truncated sample remains negative and significant, doubles in magnitude, and aligns with the lower bound of the baseline estimate (column 3). The larger magnitude is likely from the two irrigation projects, which I claimed create water habitat that attract new species. Dropping these reduces counterbalancing pressure on the coefficient.

Second, instead of dropping mega-projects, which involves some arbitrariness, I apply the inverse hyperbolic sine (arcsinh) transformation[20] to $Infrastructure_{dsym}$ in the full sample. Small changes in $arcsinh(x)$ reflect proportional changes in x and can be interpreted similar to a log-transform (Bellemare and Wichman, 2020). There are two advantages in my context. First, $arcsinh(x)$ is defined at $x = 0$, which is common in districts with no forest or no projects (see Figure 1B). Second, since it mimics the natural log, $arcsinh(x)$ reduces the influence of outliers. The coefficient on the transformed infrastructure variable shows a statistically significant

[19] Lift irrigation is a method whereby water is transported by pumps rather than by exploiting natural flow.
[20] This uses the function $arcsinh(x) = ln(x + (x^2 + 1)^{1/2})$.

infrastructure-biodiversity tradeoff with better precision (column 4). Overall, these two robustness checks suggest that my main estimates are relatively unbiased by outliers.

Alternative Diversity Measures. Species richness has been criticized for its simplicity. A location with one pigeon and 99 crows, and another with fifty of each, both have a species richness of two despite the latter being more "even". I compute two alternative diversity measures that account for evenness. The Shannon Index is measured as $SH_j = -\sum_{s=1}^{S} p_{sj} ln(p_{sj})$, where p_{sj} is the proportion of all observations on checklist j belonging to species s, and increases in diversity. The Simpson Index is $SI_j = 1/\sum_{s=1}^{S} p_{sj}^2$ and reflects the probability that two randomly drawn individuals belong to the same species (Magurran, 2013). I use $1 - SI_j$ so that the index increases in diversity. The disadvantage of both is that eBird abundance data are notoriously imprecise given difficulties with recording accurate counts of quickly moving flocks.

Columns 5 and 6 show that an infrastructure-biodiversity tradeoff is still observed using these alternative measures, but coefficients are imprecise as expected. In terms of magnitude, effects on Shannon and Simpson diversity are 1.7% and 4.0% of their means, respectively.

Alternative Infrastructure Measures. Forest intrusions are measured in km^2. Absolute units are easy to interpret, but overlook proportionality. For example, a 1 km^2 infrastructure encroachment in a district with 1 km^2 of forest signifies complete habitat loss, whereas the same intrusion in a densely forested district is a negligible disturbance. My forest cover covariate accounts for this by ensuring comparisons are made between equally-forested districts. As an extra check, I estimate the main equation with infrastructure as a percentage of baseline forest cover. The coefficient of interest remains negative and significant, although at the 10% level (column 7).

Sample Restrictions. I test the sensitivity to three relevant sample restrictions. First, as an added safeguard against user sorting, I restrict the sample to users active in a single district throughout the study period. This fully rules out endogenous sorting by construction. However, it is highly restrictive since only 5% of the sample is selected. Estimates remain statistically significant at the 10% level (column 8).

Second, I drop districts with low eBird activity, measured as districts with below-median number of users. Remaining users recording below-median trips per user are also dropped. This allows me to spotlight areas with high data collection activity. The resulting coefficient is remarkably similar to the baseline effect (column 9), suggesting that my estimates are not driven by peculiarities in districts with sparse eBird usage.

Third, I drop observations from 2020, the year COVID-19 swept the globe. India faced one of the world's toughest lockdowns between March and May. In the wake of this tragedy, "balcony birdwatching" was popularized and eBird sign-ups quadrupled (Madhok and Gulati, 2022). Project approvals also accelerated as FAC meetings moved online, with some projects receiving "just 10 minutes for consideration" (Gokhale, 2020). Estimates from the non-COVID sample are virtually unchanged (column 10), implying that my estimates are robust to the shock. This is unsurprising since year fixed effects absorb macro-shocks, state-month fixed effects absorb state-level project approval rates, and the protocol covariate controls for the shift indoors.

Clustering. Whereas the unobservable determinants of species diversity are assumed to be arbitrarily dependent within biomes with similar biophysical conditions, the treatment nevertheless varies at the district level. When standard errors are clustered by district, precision remains approximately the same (column 11).

Alternative Source of Variation. My research design exploits within-user mobility across districts for identification, which removes biases from cross-user heterogeneity, but requires a no-sorting assumption. The reverse is to track species diversity in a fixed location as it develops, which obviates this assumption, but pools checklists from heterogeneous users. I implement this alternative design with location, state-month, and year fixed effects. I use 10km $\times$ 10km grid-cell fixed effects as an extra safeguard against endogenous sorting, even within districts. Raw eBird data are aggregated to the user-cell-month level for this robustness check.

Species loss is still observed with cell fixed effects, but estimates are noisy (column 12). This is likely from pooling observations by heterogenous users in a cell. Although effort covariates remove some heterogeneity, differences in fixed unobservables (e.g. ability) inflate the error. My chosen design, with user fixed effects, solves this issue and, coupled with evidence of minimal sorting (see section 6.1), delivers credible estimates of the infrastructure-biodiversity tradeoff.

Robustness by Project Category. Table A11 reports robustness checks with infrastructure decomposed by project category. Coefficients on all categories remain negative across all twelve stress tests. Given that the decomposition already reduces variation mechanically, the restrictive fixed effects and sample restrictions inflate standard errors even more. Nevertheless, irrigation, transportation, mining, and resettlement projects continue to drive species loss in many specifications. Electricity and Other projects continue to have no impact on species diversity.

7 The Political Economy of Conservation

Having established that infrastructure expansion degrades biodiversity, this section turns to an exploration of which institutions can minimize the tradeoff. I estimate the tradeoff from the previous section as a function of whether districts have inclusive or extractive institutions. The evidence suggests that inclusive institutions promote conservation, particularly through higher rates of involvement by local communities in the development process.

7.1 Measuring Institutional Quality

I start by categorizing districts as having inclusive or extractive institutions, broadly defined, and then estimating the size of the infrastructure-biodiversity tradeoff under each type. Data on institutional quality is obtained from Banerjee and Iyer (2005) for 163 districts. They distinguish between two colonial institutions. In *zamindari* districts, landlords set land taxes, could dispossess peasants for nonpayment, and kept residuals after paying the British. In *raiyatwari* districts, cultivators paid taxes without a middleman. The authors argue that institutional choice was driven by "historic accidents" and provide various tests of exogeneity. Perhaps unsurprisingly, *zamindari* districts perform worse today on several equality and development measures. Persistence of class-based inequality and lower ability of the disenfranchised to mobilize around their interests in *zamindari* districts are key mechanisms explaining the lack of convergence. In a follow up paper, Lee (2019) provide additional evidence that state capacity is indeed the most plausible mechanism driving the results in Banerjee and Iyer (2005).

Leaning on these these two papers, I re-conceptualize *raiyatwari* and *zamindari* districts as inclusive and extractive, respectively. If disaffected groups are better able to engage in the development process and protect their livelihoods in inclusive districts, then the adverse environmental impacts of infrastructure development should be smaller in these districts.

7.2 Inclusive Institutions Minimize Species Loss

Estimation. To investigate the role of institutions in mediating the infrastructure-biodiversity tradeoff, I first estimate the tradeoff in inclusive and extractive districts separately, before turning to a more formal analysis of heterogenous treatment effects with the following equation:

$$\begin{aligned} SR_{idsym} = \alpha &+ \beta_1[Infrastructure]_{dsym} + \beta_2([Infrastructure]_{dsym} \times [Inclusive]_d) \\ &+ \beta_3[X]_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \end{aligned} \tag{8}$$

where $Inclusive_d$ is a dummy for whether district d has a history of inclusive institutions and all other terms are as in Equation (2). Data are aggregated to 1991 census boundaries to match data provided by Banerjee and Iyer (2005). The coefficient of interest is β_1 and β_2, which captures the main infrastructure-biodiversity tradeoff, and any moderation of the tradeoff depending on the type of local institution. I focus on the hypothesis that $\beta_2 > 0$ i.e. biodiversity is conserved in districts with better institutions.

The main identification concern is endogenous institutions (Aghion et al., 2004). This is less of an issue in my context because *zamindar* status was based on British politics and not local characteristics (Banerjee and Iyer, 2005). Moreover, time-invariant differences in the ecology of inclusive and extractive districts are absorbed by district fixed effects. Yet, infrastructure may exhibit heterogeneous effects along dimensions correlated with institutional type, in which case

Table 5: The Impact of Infrastructure on Biodiversity as a Function of Institutional Type

	(1)	(2)	(3)	(4)	(5)	(6)
Infrastructure (km^2)	-0.359***	-0.571***	-0.476***	-0.492***	-0.414***	-0.571***
	(0.057)	(0.058)	(0.031)	(0.060)	(0.071)	(0.204)
Infrastructure (km^2) × Inclusive (=1)	0.448**	0.433**	0.347*	0.367*	0.419**	0.433***
	(0.143)	(0.126)	(0.159)	(0.148)	(0.136)	(0.137)
Infrastructure (km^2) × Tribal Pop. Share	-0.045	0.037	0.054	0.131	-0.122	0.037
	(0.213)	(0.272)	(0.286)	(0.181)	(0.296)	(0.411)
Baseline Forest Cover and Interactions	No	Yes	Yes	Yes	Yes	Yes
High-Activity eBird District and Interactions	No	No	No	No	Yes	No
User × Year FEs	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓
State x Month FEs	✓	✓	✓	✓	✓	✓
Spillovers			✓			
Weighted				✓		
Clustering	Biome	Biome	Biome	Biome	Biome	District
N	58760	58760	58760	58760	58760	58760
R^2	0.709	0.709	0.709	0.788	0.709	0.709

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all specifications is mean species richness across users' trips in a district-month. Inclusive (=1) means the district has historically inclusive institutions. Sample restricted to 163 districts in Banerjee and Iyer (2005) and aggregated to 1991 boundaries. ST share is the fraction of district population belonging to a tribal group as measured in 2011. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Sensitivity checks are described in the footer.

β_2 is biased. Although the first point protects against this, I test sensitivity to interactions between $Infrastructure_{dsym}$ and various district characteristics as a safeguard.

Results. Table A12 shows estimates of the infrastructure-biodiversity tradeoff in subsamples of inclusive and extractive districts. The coefficient in inclusive districts (column 1) mirrors the main effect (Figure 3A). In contrast, development projects threaten species diversity three times more in extractive districts (column 5). Estimates are stable when controlling for spillovers, regression weights, and district clustering (columns 2-4, 6-8). However, precision is sensitive across specifications, likely because user mobility becomes restricted to the subset of districts under each institution[21]. Stacking the subsamples and estimating heterogeneity with an interaction term would spotlight the moderating role of institutions while boosting statistical power.

Estimates of Equation 8 are in Table 5. All specifications control for interactions between infrastructure and baseline tribal population share (from the 2011 Census) to separate heterogeneity through population effects from that through institutions. Column 1 shows that development-driven species loss is more than offset in inclusive districts. The offset is potentially upward biased if inclusive districts are more densely forested today, in which case β_2 also picks up higher species resilience in these districts (see Table 4). Column 2 reports estimates controlling for the interaction of infrastructure with baseline forest cover. Although the

[21] In the full sample (640 districts), Table 3 showed that 44% of variation in species richness remains after partialling out user-year, district, and state-month fixed. In Table A12, there are 99 inclusive districts and 66 extractive districts.

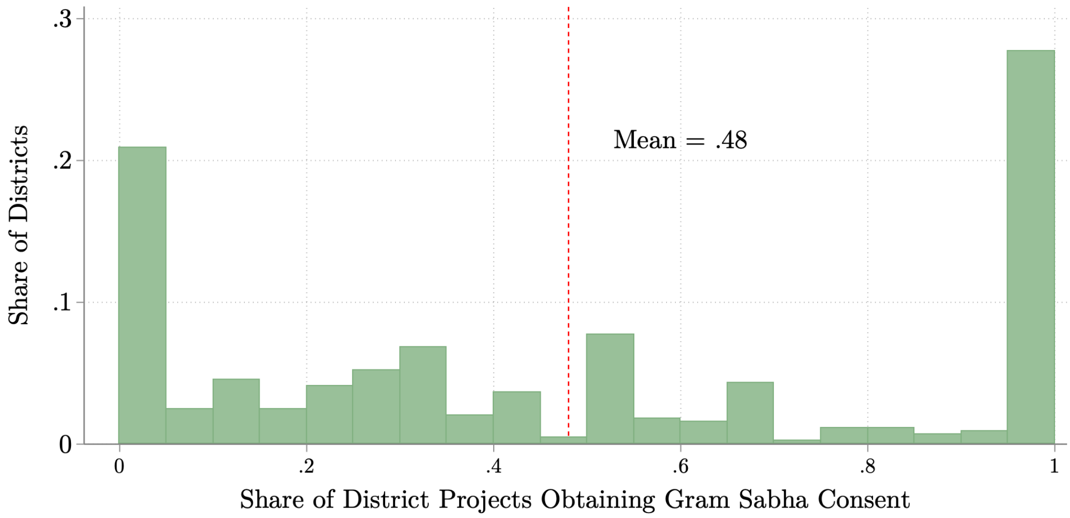


Figure 5: Enforcement of FRA (2006)

Note: Data are the share of district projects approved with informed consent during the study period. Data represent an 80% subsample of total projects that reported informed consent (the digital subsample).

counterbalancing force weakens, 70% of species loss is still erased in inclusive districts.

The mitigating effect of inclusive institutions is robust to a range of sensitivity checks. Estimates are very similar when controlling for spatial spillovers within the biome (column 3), weighting by number of eBird trips underlying SR_{idsym} (column 4), and controlling for the interaction between infrastructure and a district dummy for high eBird activity (column 5). The latter is defined in Section 6.5 and accounts for β_2 potentially confounding differences in eBird usage across institution types. Lastly, the mitigating effect remains highly significant under district-level clustering (column 6). The moderating role of institutions is independent of tribal population in all specifications. This suggests that institutions empowering disaffected people, not their population per se, determine the extent of sustainable development.

These results highlight the importance of inclusive institutions in mitigating anthropogenic pressures on ecosystems. However, it is difficult to glean specific policy lessons since the muted tradeoff in inclusive institutions may operate through numerous channels. I turn to an investigation of mechanisms in the next section.

7.3 Policy Mechanisms

Local Governance and FRA Compliance Why are development projects more sustainable in districts with historically inclusive institutions? I claim it is because these districts emphasize grassroots rather than top-down governance. Banerjee and Iyer (2005) argue that the absence of a landed gentry in inclusive districts left a legacy enabling “elites and the masses to act together in the collective interest" even today. Lee (2019) additionally show that more "contact" between the state and cultivators in inclusive districts created a legacy of better state capacity compared to extractive districts where the state was absent. This suggests that STs have higher capacity to mobilize around their interests in inclusive districts. Participatory development and stringent environmental criteria on project proposals should thus be more common in these districts.

The simplest test is whether projects in inclusive districts are more likely to follow the FRA, which requires inclusion of STs in the project approval process (Section 2). If the policy is binding, however, then there would be no variation. Recent reports have indicated weak FRA implementation, with many exemptions made, non-recognition of land titles, and even bypassing consent altogether (Dubey et al., 2017). Fortunately, the permit sample reports whether consent was obtained (see Figure A5 for an example), enabling me to characterize variation.

Figure 5 shows the distribution of projects obtaining Gram Sabha consent. The lack of right-tail bunching is evidence of imperfect compliance. There are districts where inclusive development is always, sometimes, and never observed. I exploit this weak implementation to study if inclusive institutions, as defined by Banerjee and Iyer (2005), are actually more inclusive.

Besides FRA compliance, two other variables from project permits highlight mechanisms. The first is whether a supplemental cost-benefit report was commissioned, beyond the site reports required during project review. This reflects the rigour of environmental review since commissioning is based on value judgement[22]. The second variable is whether the project is sited in a protected area buffer.

Estimation. I match project permits with the district level inclusive-extractive dummies and use pooled OLS to compare outcomes across projects in the two types of districts. Since institution type is time-invariant, I make cross-district comparisons within the same state and time-period:

$$Y_{pdsym} = \alpha + \beta_1 [Inclusive]_d + \beta_2 [X]_{pdsym} + \theta_{sm} + \epsilon_{pdsym} \tag{9}$$

where Y_{pdsym} is a dummy for whether project p approved in district d of state s in year y and month m received informed consent, completed a supplemental cost-benefit report, or was sited near a protected area. $Inclusive_d$ is the same institutional dummy as equation (8). X_{pdsym} is a set of covariates including project size, tribal population share, baseline forest cover, and district size. θ_{sm} are state-month fixed effects. β_1 reveals the proportion of projects with each feature in inclusive compared to extractive districts.

Results. Figure 6 reports the results. Projects in districts with inclusive institutions are more conservation friendly. They are 8% more likely to obtain informed consent from Gram Sabhas and follow FRA provisions compared to projects approved in extractive districts in the same

[22] Value judgment is used for projects $>$ 20 ha., which is more than 90% of my projects. Official guidelines here: http://forestsclearance.nic.in/writereaddata/Addinfo/0_0_7111512571261CostBenefitAnalysisGuidelinesforforestlanddiversion-2017.pdf

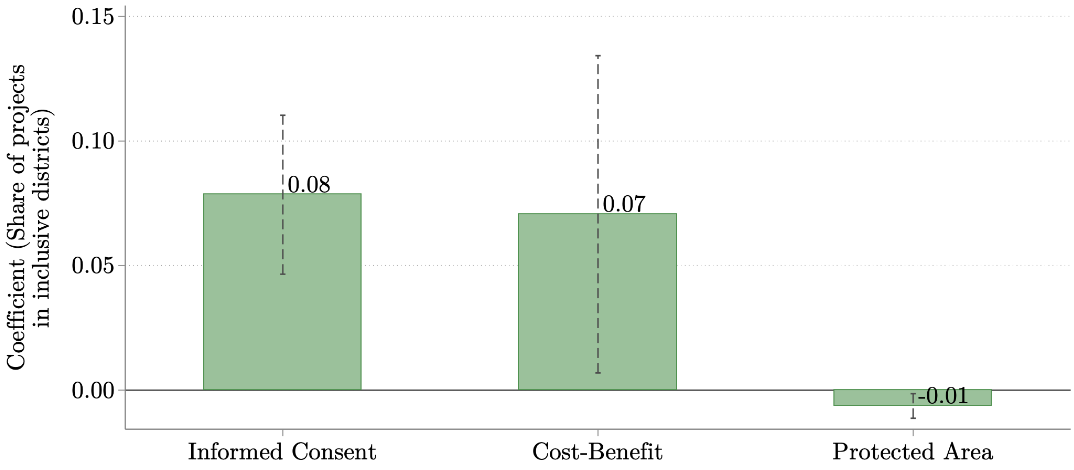


Figure 6: Mechanisms by which Institutions Mitigate the Infrastructure-Biodiversity Tradeoff
Note: Data are at the project level for the digital subsample. Bars represent coefficients from Equation (9) and describe the share of projects in inclusive districts compared to extractive ones. Sample is restricted to 163 districts in Banerjee and Iyer (2005) under British Rule. Informed consent indicates whether the Gram Sabha was consulted and the FRA followed. Cost-Benefit Analyses are reports commissioned during project review. Protected Area equals one if the project is sited in or near one. All specifications include controls for: project size, district population share belonging to a tribal group as measured in 2011, baseline forest cover, and district area. Grey bars represent 95% confidence intervals.

state (column 1). Column 2 shows that forest officers in inclusive districts are 7% more likely to commission a supplemental cost-benefit report during project review. Lastly, projects in inclusive districts are 1% less likely to be sited near a protected area (column 3). Tabular estimates are provided in Table A13.

These correlations corroborate the logic of Banerjee and Iyer (2005) and other studies. Duflo and Pande (2007) use the same institutions classification to claim that populations affected by dams are more effective at demanding compensation in inclusive districts. Lal et al. (2021) show that inclusive governance in Indian districts increased tree cover. My results thus represent a test of mechanisms through which institutions drive conservation. They suggest that engaging forest-dependent communities in the development process, especially through informed consent and more stringent checks-and-balances during project approval, are vital for protecting biodiversity. Besides providing direct policy guidance, these results also add a procedural justice lens to the environment-development literature.

8 Conclusion

Economic development and associated infrastructure-building in the tropics raises concern about harmful impacts that may percolate through our fragile ecosystems. This paper provides rigorous evidence on the impact of infrastructure expansion on biodiversity in a biodiverse developing nation. It also quantifies the role of inclusive institutions in mitigating the tradeoff. Between 2015-2020, development in India's forests accounted for nearly 20% of the decline in bird diversity, an important proxy for overall biodiversity. Species loss does not rebound in the medium-run, and is accentuated in already-fragmented areas. This, however, is not a foregone result. The tradeoff is more than halved when local institutions emphasize community forest governance and amplify the voices of indigenous groups in the development planning process.

My results are especially relevant as emerging economies prioritize the types of infrastructure studied here. China's Belt and Road Initiative is the textbook example, envisioning enhanced regional connectivity across Asia and Africa. Surprisingly, studies from these regions find limited ecological costs of infrastructure development projects (Asher et al., 2020; Garg and Shenoy, 2021; Baehr et al., 2021). In the absence of biodiversity data, these studies use tree cover to measure ecosystem health, whereas I leverage several million verified species sightings. After accounting for observer biases and spatial spillovers, this novel data yields robust evidence of anthropogenic species decline, and can be used to inform infrastructure planning as economies expand throughout the 21st century.

Resettlement, transportation, irrigation, and mining projects are the main drivers of species loss. Village resettlements are particularly damaging. Several thousand communities await relocation from India's 600 protected areas, primarily due to human-wildlife conflict and forced evictions (Lasgorceix and Kothari, 2009). For example, Kerala's proposed Silver Line railway is set to divert forests and displace enroute communities. My results imply that species loss from such projects can be compounded if displaced villages are resettled by fragmenting forests elsewhere. Relocation to non-forested or less fragmented areas can achieve a better net outcome.

My results are policy relevant at a broad and grassroots level. In places where institutions favour the economically advantaged, development is associated with more biodiversity loss. This highlights the need for people-centred conservation policy. India has made strides with the FRA (2006), which promises forest rights to indigenous people and their inclusion in development decisions. Yet nearly two decades later, half of forest rights claims remain legally unrecognized and face other forms of weak enforcement (Ministry of Tribal Affairs, 2022). I find that upholding the FRA, and requiring informed consent during the development process, helps neutralize the infrastructure-biodiversity tradeoff. In sum, inclusive institutions and procedural justice are critical for meeting the dual objectives of development and conservation.

This paper is not without limitations. First, species richness is a simplified diversity measure and abstracts from notions of functional diversity, genetic diversity, and other dissimilarity indices (Weitzman, 1992, 1993). Second, with a six year study period, I am unable to study whether species diversity rebounds or collapses in the long-run. Lastly, without reliable species values, I am unable to benchmark the economic cost of development-driven species loss. A comprehensive cost-benefit analysis is left for future work. Despite these limitations, this study provides powerful insights into the dynamics of biodiversity in human-modified landscapes and is relevant for decision-makers tasked with conserving local and global biodiversity.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge, "When should you adjust standard errors for clustering?," Technical Report, National Bureau of Economic Research 2017.

Aghion, Philippe, Alberto Alesina, and Francesco Trebbi, "Endogenous political institutions," *The Quarterly Journal of Economics*, 2004, *119* (2), 565–611.

Asher, Sam, Teevrat Garg, and Paul Novosad, "The ecological impact of transportation infrastructure," *The Economic Journal*, 2020, *130* (629), 1173–1199.

Baehr, Christian, Ariel BenYishay, and Bradley Parks, "Linking Local Infrastructure Development and Deforestation: Evidence from Satellite and Administrative Data," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (2), 375–409.

Banerjee, Abhijit and Lakshmi Iyer, "History, institutions, and economic performance: The legacy of colonial land tenure systems in India," *American economic review*, 2005, *95* (4), 1190–1213.

Bellemare, Marc F and Casey J Wichman, "Elasticities and the inverse hyperbolic sine transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Benítez-López, Ana, Rob Alkemade, and Pita A Verweij, "The impacts of roads and other infrastructure on mammal and bird populations: a meta-analysis," *Biological conservation*, 2010, *143* (6), 1307–1316.

Betts, Matthew G, Christopher Wolf, William J Ripple, Ben Phalan, Kimberley A Millers, Adam Duarte, Stuart HM Butchart, and Taal Levi, "Global forest loss disproportionately erodes biodiversity in intact landscapes," *Nature*, 2017, *547* (7664), 441–444.

BirdLife International, "Bird species distribution maps of the world. Version 2018.1," 2018.

Blanchard, Paul, Douglas Gollin, and Martina Kirchberger, "Perpetual Motion: Human Mobility and Spatial Frictions in Three African Countries," *CEPR Discussion Paper No. DP16661*, 2021.

Börner, Jan, Dario Schulz, Sven Wunder, Alexander Pfaff et al., "The effectiveness of forest conservation policies and programs," *Annu. Rev. Resour. Econ*, 2020, *12* (1), 45–64.

Burgess, Robin, Matthew Hansen, Benjamin A Olken, Peter Potapov, and Stefanie Sieber, "The political economy of deforestation in the tropics," *The Quarterly journal of economics*, 2012, *127* (4), 1707–1754.

Burlig, Fiona and Louis Preonas, "Out of the darkness and into the light? development effects of rural electrification," *Energy Institute at Haas WP*, 2016, *268*, 26.

Callaghan, Corey T, Jodi JL Rowley, William K Cornwell, Alistair GB Poore, and Richard E Major, "Improving big citizen science data: Moving beyond haphazard sampling," *PLoS Biology*, 2019, *17* (6), e3000357.

—, **John M Martin, Richard E Major, and Richard T Kingsford**, "Avian monitoring–comparing structured and unstructured citizen science," *Wildlife Research*, 2018, *45* (2), 176–184.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna, "Difference-in-differences with a continuous treatment," *arXiv preprint arXiv:2107.02637*, 2021.

Copeland, Brian R and M Scott Taylor, "Trade, growth, and the environment," *Journal of Economic literature*, 2004, *42* (1), 7–71.

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*, February 2015, *105* (2), 678–709.

Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler, "Confronting the environmental Kuznets curve," *Journal of economic perspectives*, 2002, *16* (1), 147–168.

Dubey, Sanghamitra, Tushar Dash, and Geetanjoy Sahu, "Political economy of Community forest rights," *Economic and Political Weekly*, 06 2017, *52.*

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

eBird Basic Dataset, "Version ebd_IN_201401_201904_relApr-2019," 2019.

Elhorst, JP and SH Vega, "The SLX model," *Journal of Regional Science*, 2015, *55* (3), 339–363.

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh, "VIIRS night-time lights," *International Journal of Remote Sensing*, 2017, *38* (21), 5860–5879.

FAO and UNEP, "State of the World's Forests: Forests, Biodiversity and People," Technical Report, FAO 2020.

Farmer, Robert G, Marty L Leonard, Joanna E Mills Flemming, and Sean C Anderson, "Observer aging and long-term avian survey data quality," *Ecology and Evolution*, 2014, *4* (12), 2563–2576.

Fitzpatrick, Matthew C, Evan L Preisser, Aaron M Ellison, and Joseph S Elkinton, "Observer bias and the detection of low-density populations," *Ecological applications*, 2009, *19* (7), 1673–1679.

Fleishman, Erica, Reed F Noss, and Barry R Noon, "Utility and limitations of species richness metrics for conservation planning," *Ecological indicators*, 2006, *6* (3), 543–553.

Forest Survey of India, "India State of Forest Report," Technical Report, Ministry of Environment, Forests and Climate Change. Government of India 01 2019.

Foster, Andrew D and Mark R Rosenzweig, "Economic growth and the rise of forests," *The Quarterly Journal of Economics*, 2003, *118* (2), 601–637.

Fraixedas, Sara, Andreas Lindén, Markus Piha, Mar Cabeza, Richard Gregory, and Aleksi Lehikoinen, "A state-of-the-art review on birds as indicators of biodiversity: Advances, challenges, and future directions," *Ecological Indicators*, 2020, *118*, 106728.

Frank, Eyal G and Wolfram Schlenker, "Balancing economic and ecological goals," *Science*, 2016, *353* (6300), 651–652.

Garg, Teevrat and Ajay Shenoy, "The Ecological Impact of Place-Based Economic Policies," *American Journal of Agricultural Economics*, 2021.

Gokhale, Nihar, "To kickstart the economy, India?s environment ministry is clearing projects in 10 minutes," 2020.

Grossman, Gene M and Alan B Krueger, "Economic growth and the environment," *The Quarterly Journal of Economics*, 1995, *110* (2), 353–377.

Haddou, Yacob, Rebecca Mancy, Jason Matthiopoulos, Sofie Spatharis, and Davide M Dominoni, "Widespread extinction debts and colonization credits in United States breeding bird communities," *Nature ecology & evolution*, 2022, *6* (3), 324–331.

Hagerty, Nick, "Adaptation to surface water scarcity in irrigated agriculture," *Working paper*, 2021.

Hanna, Rema and Paulina Oliva, "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City," *Journal of Public Economics*, 2015, *122*, 68–79.

Hanski, Ilkka, "Metapopulation dynamics," *Nature*, 1998, *396* (6706), 41–49.

Henderson, J Vernon, Adam Storeygard, and David N Weil, "Measuring economic growth from outer space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hoffmann, Lars, Gebhard Günther, Dan Li, Olaf Stein, Xue Wu, Sabine Griessbach, Yi Heng, Paul Konopka, Rolf Müller, Bärbel Vogel et al., "From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations," *Atmospheric Chemistry and Physics*, 2019, *19* (5), 3097–3124.

Horns, Joshua J, Frederick R Adler, and Çağan H Şekercioğlu, "Using opportunistic citizen science data to estimate avian population trends," *Biological conservation*, 2018, *221*, 151–159.

Huffman, George J, David T Bolvin, Dan Braithwaite, Kuolin Hsu, Robert Joyce, Pingping Xie, and Soo-Hyun Yoo, "NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG)," *Algorithm Theoretical Basis Document (ATBD) Version*, 2019, *4*, 26.

Imbens, Guido W and Donald B Rubin, *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press, 2015.

Jayadevan, Praveen, Rajah Jayapal, and Aasheesh Pittie, "A checklist of the birds of India," *Indian BIRDS*, 07 2016, *11*, 113–172.

Johnston, Alison, Daniel Fink, Wesley M Hochachka, and Steve Kelling, "Estimates of observer expertise improve species distributions from citizen science data," *Methods in Ecology and Evolution*, 2018, *9* (1), 88–97.

Kelling, Steve, Alison Johnston, Aletta Bonn, Daniel Fink, Viviana Ruiz-Gutierrez, Rick Bonney, Miguel Fernandez, Wesley M Hochachka, Romain Julliard, Roland Kraemer et al., "Using semistructured surveys to improve citizen science data for monitoring biodiversity," *BioScience*, 2019, *69* (3), 170–179.

_ , _ , Wesley M Hochachka, Marshall Iliff, Daniel Fink, Jeff Gerbracht, Carl Lagoze, Frank A La Sorte, Travis Moore, Andrea Wiggins et al., "Can observation skills of citizen scientists be estimated using species accumulation curves?," *PloS one*, 2015, *10* (10), e0139600.

Kolstoe, Sonja and Trudy Ann Cameron, "The non-marketo tht value of birding sites and the marginal value of additional species: biodiversity in a random utility model of site choice by eBird members," *Ecological economics*, 2017, *137*, 1–12.

Lal, Apoorva, Saad Gulzar, and Ben Pasquale, "Representation and Forest Conservation: Evidence from India?s Scheduled Areas," Feb 2021.

Lasgorceix, Antoine and Ashish Kothari, "Displacement and relocation of protected areas: a synthesis and analysis of case studies," *Economic and political weekly*, 2009, pp. 37–47.

Laurance, William F, Miriam Goosem, and Susan GW Laurance, "Impacts of roads and linear clearings on tropical forests," *Trends in ecology & evolution*, 2009, *24* (12), 659–669.

Lee, Alexander, "Land, state capacity, and colonialism: Evidence from India," *Comparative Political Studies*, 2019, *52* (3), 412–444.

Liang, Yuanning, Ivan J Rudik, and Eric Zou, "Economic Production and Biodiversity in the United States," Technical Report, National Bureau of Economic Research 2021.

_ , Ivan Rudik, Eric Yongchen Zou, Alison Johnston, Amanda D Rodewald, and Catherine L Kling, "Conservation cobenefits from air pollution regulation: Evidence from birds," *Proceedings of the National Academy of Sciences*, 2020, *117* (49), 30900–30906.

Libecap, Gary D, "Institutional path dependence in climate adaptation: Coman's" Some unsettled problems of irrigation"," *American Economic Review*, 2011, *101* (1), 64–80.

Madhok, Raahil and Sumeet Gulati, "Ruling the roost: Avian species reclaim urban habitat during India's COVID-19 lockdown," *Biological Conservation*, 2022, p. 109597.

Magurran, Anne E, *Measuring biological diversity*, John Wiley & Sons, 2013.

Meiyappan, Prasanth, Parth S Roy, Yeshu Sharma, Reshma M Ramachandran, Pawan K Joshi, Ruth S DeFries, and Atul K Jain, "Dynamics and determinants of land change in India: integrating satellite data with village socioeconomics," *Regional environmental change*, 2017, *17* (3), 753–766.

Menon, Meena, "The unmaking of the forest rights act," *Economic and Political Weekly*, 2016, pp. 15–18.

Ministry of Tribal Affairs, "Monthly Update on Status of Implementation of the Scheduled Tribes and Other Traditional Forest Dweller (Recognition of Forest Rights) Act, 2006," 2022.

MoEFCC, "Forest (Conservation) Act, 1980," 1980.

_ , "Forest (Conservation) Rules, 2003, Ministry of Environment and Forests Notification," 2003.

_ , "Report of the Comptroller and Auditor General of India on Compensatory Afforestation in India," 2013.

Morrison, Michael L, "Bird populations as indicators of environmental change," in "Current ornithology," Springer, 1986, pp. 429–451.

Munson, M Arthur, Rich Caruana, Daniel Fink, Wesley M Hochachka, Marshall Iliff, Kenneth V Rosenberg, Daniel Sheldon, Brian L Sullivan, Christopher Wood, and Steve Kelling, "A method for measuring the relative information content of data from different monitoring protocols," *Methods in Ecology and Evolution*, 2010, *1* (3), 263–273.

Nayak, Rajat, Krithi K Karanth, Trishna Dutta, Ruth Defries, K Ullas Karanth, and Srinivas Vaidyanathan, "Bits and pieces: Forest fragmentation by linear intrusions in India," *Land Use Policy*, 2020, *99*, 104619.

Noack, Frederik, Christian Levers, Johannes Kamp, and Ashley Larsen, "A bird's eye view on farm size and biodiversity," *Mimeo*, 2021.

_ , Marie-Catherine Riekhof, and Salvatore Di Falco, "Droughts, biodiversity, and rural incomes in the tropics," *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (4), 823–852.

Nunn, Nathan, "The Importance of History for Economic Development," *Annual Review of Economics*, 2009, *1* (1), 65–92.

Odum, Eugene P, "The Strategy of Ecosystem Development: An understanding of ecological succession provides a basis for resolving man's conflict with nature.," *science*, 1969, *164* (3877), 262–270.

Oxford Economics, "Global Infrastructure Outlook," Technical Report, Global Infrastructure Hub 2017.

Pacheco, Pablo, Karen Mo, Nigel Dudley, Aurelie Shapiro, Naikoa aguilar amuchastegui, Pui-Yu Ling, Christa Anderson, and Alexander Marx, "Deforestation fronts: Drivers and responses in a changing world," Technical Report, WWF 01 2021.

Reis, Elivânia, Germán Manuel López-Iborra, and Renato Torres Pinheiro, "Changes in bird species richness through different levels of urbanization: Implications for biodiversity conservation and garden design in Central Brazil," *Landscape and urban planning*, 2012, *107* (1), 31–42.

Riley, Seth, John Pollinger, Raymond Sauvajot, Eric York, Cassity Bromley, Todd Fuller, and Robert Wayne, "A southern California freeway is a physical and social barrier to gene flow in carnivores," *Molecular Ecology*, 07 2006, *15*, 1733–41.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *arXiv preprint arXiv:2201.01194*, 2022.

Stephens, Scott E, David N Koons, Jay J Rotella, and David W Willey, "Effects of habitat fragmentation on avian nesting success: a review of the evidence at multiple spatial scales," *Biological conservation*, 2004, *115* (1), 101–110.

Strimas-Mackey, M., W.M. Hochachka, V. Ruiz-Guiterrez, O.J. Robinson, E.T. Millet, T. Auer, S. Kelling, D. Fink, and A. Johnston, *Best Practices for Using eBird Data. Version 1.0*, Cornell Lab of Ornithology, Ithaca, NY, 2020.

Sullivan, Brian L, Christopher L Wood, Marshall J Iliff, Rick E Bonney, Daniel Fink, and Steve Kelling, "eBird: A citizen-based bird observation network in the biological sciences," *Biological conservation*, 2009, *142* (10), 2282–2292.

Townshend, John, Matthew Hansen, Mark Carroll, Charlene DiMiceli, Robert Sohlberg, and Chengquan Huang, "Annual global automated MODIS vegetation continuous fields (MOD44B) at 250 m spatial resolution for data years beginning day 65, 2000?2014, collection 5 percent tree cover, version 6.," Technical Report, University of Maryland, College Park 2017.

Venkataraman, K. and C. Sivaperuman, in "Biodiversity Hotspots in India" 2018.

Weitzman, Martin L, "On diversity," *The Quarterly Journal of Economics*, 1992, *107* (2), 363–405.

_ , "What to preserve? An application of diversity theory to crane conservation," *The Quarterly Journal of Economics*, 1993, *108* (1), 157–183.

A Supplementary Tables and Figures

A1 Tables

Table A1: Trip Protocols in Sample Frame

Protocol	Num. Trips	Pct.
Traveling	651579	61.29
Stationary	403452	37.95
Historical	6050	0.57
Area	1346	0.13
Nocturnal Flight Call Count	392	0.04
Banding	115	0.01
eBird Pelagic Protocol	122	0.01
CWC Point Count	14	0
Greater Gulf Refuge Waterbird Count	1	0
International Shorebird Survey (ISS)	8	0
Oiled Birds	1	0
Random	3	0
Tricolored Blackbird Winter Survey	1	0

Note: Traveling and stationary trips comprise the analysis sample. The traveling protocol is used if the observer walks or drives during the trip. The stationary protocol is used when the user remains in one place. Many of the remaining protocols describe community initiatives to document specific species. For details of these protocols see manual in eBird Basic Dataset (2019).

Table A2: Summary Statistics of Projects by Ownership and Shape

	Num. Projects	Mean Size (ha.)	SD (ha.)	Total Area (ha.)
Panel A: Ownership				
Public	4,666	9.61	89.88	44,861.30
Private	1,549	1.88	15.23	2,910.62
Neither	382	2.67	24.94	1,019.81
Panel B: Shape				
Linear	5,768	4.76	28.24	27,472.61
Nonlinear	829	25.72	200.97	21,319.13

Note: Data are arranged at the project level for 6,597 approved projects that reported ownership, prior to aggregating to the district level. Panel A splits projects by ownership type. Panel B splits projects by shape. Total area (ha.) is the summed deforestation area of all projects in the sample for the particular project group.

Table A3: Percent of Projects in Each Category by Ownership and Shape

	Ownership (%)			Shape (%)	
	Public (1)	Private (2)	Neither (3)	Linear (4)	Nonlinear (5)
Electricity	81.48	15.86	2.67	86.03	13.97
Irrigation	95.34	2.48	2.17	72.05	27.95
Mining	49.15	35.59	15.25	0.00	100.00
Other	58.07	33.53	8.40	85.04	14.96
Resettlement	100.00	0.00	0.00	5.00	95.00
Transportation	88.41	9.54	2.05	99.03	0.97

Note: Data are arranged at the project level for 6,597 approved projects that reported ownership and shape, prior to aggregating to the district level. Cell values denote row percentages *within each group* (i.e. % of projects in each category falling under different ownership types). Thus, the row sum of columns 1-3 equals 100, and same for columns 4-5.

Table A4: eBird Summary Statistics (2015-2020)

	Mean	Std. Dev.	Obs.
District			
Num. Users	109.01	193.91	628
Num. Trips	1671.86	5497.55	628
User			
Num. Districts	3.99	7.45	16899
Num. States	1.93	2.21	16899
Num. Year-months	6.41	11.32	16899
User-District-Time			
Species Richness	23.39	18.72	173813
Coverage (%)	9.58	16.98	173813
Duration (min)	85.51	70.70	173813
Distance (km)	3.06	6.02	173813
Hour of day	10.50	3.42	173813
District-Time			
Forest Cover (%)	16.09	14.90	21750
Rainfall (mm)	0.34	0.82	21750
Temperature ($^\circ$ C)	23.30	7.22	21750
Nightlights (radiance)	2.61	7.28	21750

Note: District variables reflect total eBird activity in a district during the study period. User variables describe number of locations and time-periods in which the user is active. Variables at the user-district-time level are means over users' trips in a district-month. Coverage is measured as the percentage of district cells traversed by a user on a 10km grid. Remaining covariate details are explained in section 3.

Table A5: Impact of Forest Infrastructure on Species Diversity

	Main Estimates			Sensitivity		
	(1)	(2)	(3)	(4)	(5)	(6)
Infrastructure (km^2)	-0.052*	-0.122**	-0.122**	-0.105	-0.120**	-0.110*
	(0.026)	(0.051)	(0.052)	(0.062)	(0.052)	(0.056)
Infrastructure (district j $\neq$ d)			0.801			
			(0.889)			
Non-forest Land Diversion (km^2)						-0.047
						(0.056)
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Behaviour Controls	Yes	Yes	Yes	No	Yes	Yes
General Economic Trends	Yes	Yes	Yes	No	No	Yes
Outcome Mean	23.672	23.748	23.748	23.748	23.748	23.748
User FEs	✓					
User $\times$ Year FEs		✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓
State $\times$ Month FEs	✓	✓	✓	✓	✓	✓
Year FEs	✓					
N	167256	161896	161896	161896	161896	161896
R^2	0.639	0.694	0.694	0.559	0.694	0.694

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Columns 1-3 are the same as in Figure 3A. Columns 4-5 successively add controls. Weather controls include temperature and rainfall. Behaviour controls include: traveling trips, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. General economic trends is measured by nightlights. Column 6 adds cumulative non-forest land diversion, which is available only the digital subsample of project proposals. Standard errors clustered by biome.

Table A6: Tests of Endogenous Sorting

	(1) Users	(2) Users	(3) Users
Infrastructure (district d)	0.009	0.008	0.007
	(0.023)	(0.022)	(0.027)
Infrastructure (district $j \neq d$)	-0.025	-0.015	-0.011
	(0.030)	(0.038)	(0.016)
Controls	Yes	Yes	Yes
Data Aggregation	District	District	District
Distance Cutoff	100km	200km	500km
District FEs	✓	✓	✓
State $\times$ Month FEs	✓	✓	✓
Year FEs	✓	✓	✓
N	21681	21681	21681
R^2	0.808	0.808	0.808

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is log number of users in a district. Forest infrastructure (district d) is cumulative area of forest occupied by infrastructure in district d during a year-month. In column 1, forest infrastructure (district $j \neq d$) is inverse-distance weighted infrastructure in districts j within 100km of d. In columns 2 and 3, the distance cutoff is extended to 200km and 500km, respectively. Both infrastructure variables are standardized in all specifications. Controls are the same as the main specification. Experience, duration, distance, hour of day, group size, and % traveling trips are aggregated to district means and logged. Standard errors clustered by biome.

Table A7: Impact of Forest Infrastructure on Species Diversity by Category

	(1)	(2)	(3)	(4)
Electricity	0.088	0.088	0.083	0.087
	(0.066)	(0.068)	(0.068)	(0.068)
Irrigation	-0.105*	-0.123**	-0.124**	-0.113**
	(0.052)	(0.049)	(0.048)	(0.043)
Mining	-0.061	-0.059***	-0.060***	-0.106***
	(0.035)	(0.015)	(0.016)	(0.025)
Other	-0.165	-0.276	-0.273	-0.284
	(0.229)	(0.221)	(0.221)	(0.217)
Resettlement	-1.100***	-0.732***	-0.731***	-0.714***
	(0.062)	(0.081)	(0.079)	(0.091)
Transportation	-0.265	-0.411**	-0.412**	-0.464**
	(0.305)	(0.183)	(0.186)	(0.193)
Weather Controls	Yes	Yes	Yes	Yes
Behaviour Controls	No	Yes	Yes	Yes
General Economic Trends	No	No	Yes	Yes
Outcome Mean	23.748	23.748	23.748	23.983
Sample	Full	Full	Full	High-Activity
User x Year FEs	✓	✓	✓	✓
District FEs	✓	✓	✓	✓
State × Month FEs	✓	✓	✓	✓
N	161896	161896	161896	150011
R^2	0.559	0.694	0.694	0.692

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all specifications is mean species richness across users' trips in a district-month. Rows denote cumulative area of infrastructure encroachments by a particular category in a district-month. Column 1-3 successively add controls. Weather controls include temperature and rainfall. Behaviour controls include: traveling trips, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. General economic trends is measured by nightlights. Column 3 is the same as Figure 3B. Column 4 restricts the sample to districts with high eBird usage, measured as districts with above-median numbers of users, recording above-median trips per user. Standard errors clustered by biome.

Table A8: Impact of Forest Infrastructure on Species Diversity by Project Ownership and Shape

	(1)	(2)	(3)
Panel A: Ownership			
Public	-0.197***	-0.204***	-0.204***
	(0.047)	(0.053)	(0.052)
Private	-0.806	-0.393	-0.391
	(0.925)	(0.827)	(0.825)
Neither	2.911	0.972	0.973
	(1.854)	(0.786)	(0.788)
Panel B: Shape			
Linear	-0.260	-0.340*	-0.339*
	(0.193)	(0.172)	(0.173)
Nonlinear	-0.177***	-0.172***	-0.173***
	(0.043)	(0.042)	(0.042)
Weather Controls	Yes	Yes	Yes
Behaviour Controls	No	Yes	Yes
General Economic Trends	No	No	Yes
Outcome Mean	23.748	23.748	23.748
User x Year FEs	✓	✓	✓
District FEs	✓	✓	✓
State × Month FEs	✓	✓	✓
N	161896	161896	161896
R^2	0.559	0.694	0.694

Note: The outcome in all specifications is mean species richness across users' trips in a district-month. In Panel A, rows denote cumulative area of infrastructure encroachments by projects of a particular ownership type in a district-month. Panel B reports the same by project shape. Column 1-3 successively add controls. Weather controls include temperature and rainfall. Behaviour controls include: traveling trips, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. General economic trends is measured by nightlights. Standard errors clustered by biome.

Table A9: Robustness—Spatial Spillovers

	(1)	(2)	(3)	(4)
Infrastructure	-0.400**	-0.405**	-0.413**	-0.414**
(district d)	(0.160)	(0.168)	(0.166)	(0.171)
Infrastructure	0.007	0.069	0.430	0.461
(district $j \neq d$)	(0.501)	(0.243)	(0.634)	(0.395)
Distance Cutoff	100km	200km	500km	None
User x Year FEs	✓	✓	✓	✓
District FEs	✓	✓	✓	✓
State x Month FEs	✓	✓	✓	✓
N	161782	161782	161782	161782
R^2	0.694	0.694	0.694	0.694

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is mean species richness across user's trips in a district-month. In column 1, forest infrastructure in other districts j describes cumulative encroachment area in other districts within 100km of the focal district d. In each time period, $Infrastructure_{dsym}$ is multiplied by a $N \times N$ (where N is the number of districts in India) dimensional weight matrix W with elements $w_{dj} = 1/distance_{dj}$ for districts j within 100km of d and zero otherwise. Columns 2 and 3 extend the distance cutoff to 200km and 500km, respectively. Column 4 applies the inverse distance weight to all districts. Section 5.3.2 elaborates the procedure. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Standard errors clustered by biome.

Table A10: Robustness Checks

	(1) SR	(2) SR	(3) SR	(4) SR	(5) Shannon	(6) Simpson	(7) SR	(8) SR	(9) SR	(10) SR	(11) SR	(12) SR
Infrastructure (km^2)	-0.146***	-0.123**	-0.248**	-0.627***	-0.035	-0.036	-0.165*	-0.960*	-0.125**	-0.111***	-0.123**	-0.052
	(0.034)	(0.053)	(0.081)	(0.124)	(0.277)	(0.062)	(0.083)	(0.483)	(0.051)	(0.012)	(0.060)	(0.056)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Infrastructure Unit	km^2	km^2	km^2	IHS	km^2	km^2	% Forest	km^2	km^2	km^2	km^2	km^2
Sample Restriction	None	None	Truncated	None	None	None	None	One District	High-Activity	2015-2019	None	None
User x Year FEs		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
User × Month FEs	✓											
District FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Cell FEs												✓
State x Month FEs		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State × Year FEs	✓											
Year FEs												✓
Weights	None	Num. Trips	None	None	None	None	None	None	None	None	None	None
Clustering	Biome	Biome	Biome	Biome	Biome	Biome	Biome	Biome	Biome	Biome	District	Biome
N	143269	161782	161782	161782	161782	157655	161782	6884	149905	123331	161782	282427
R^2	0.707	0.765	0.694	0.694	0.650	0.448	0.694	0.821	0.692	0.695	0.694	0.549

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all columns (except 5 and 6) is mean species richness (SR) across users' trips in a district-month. Column 1 includes user-month, district, and state-year fixed effects. Columns 2-10 include user-year, district, and state-month fixed effects. Column 2 is a weights the regression with number of trips underlying the mean outcome. Column 3 drops the three largest projects. Column 4 uses the inverse hyperbolic sine of the explanatory variable. Column 5 and 6 show results with two alternative species diversity metrics. Column 7 defines infrastructure encroachment as a percent of baseline forest cover. Column 8 restricts the sample to users active in a single district throughout the study period. Column 9 restricts the sample to districts with above-median numbers of users who record an above-median number of trips per user. Column 10 drops the year 2020. Column 11 clusters at the district level. Column 12 includes 10km $\times 10km$ cell fixed effects along with state-month and year fixed effects. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Column 1 includes an additional control for the number of months/year of birdwatching.

Table A11: Category-Wise Robustness Checks

	(1) SR	(2) SR	(3) SR	(4) SR	(5) Shannon	(6) Simpson	(7) SR	(8) SR	(9) SR	(10) SR	(11) SR	(12) SR
Electricity	-0.130	0.010	-0.224	0.120	-0.138	-0.099	0.176	-1.122	0.088	0.192	0.082	0.040
	(0.099)	(0.030)	(0.878)	(0.961)	(0.295)	(0.077)	(0.223)	(5.640)	(0.068)	(0.452)	(0.057)	(0.044)
Irrigation	-0.021	-0.176***	-0.150	-0.568**	0.000	-0.040	-0.535**	2.617***	-0.113**	-0.052*	-0.124**	-0.073
	(0.045)	(0.049)	(0.135)	(0.226)	(0.206)	(0.049)	(0.208)	(0.653)	(0.043)	(0.024)	(0.059)	(0.113)
Mining	-0.218***	0.064	-0.035	-0.422	-0.332**	0.019	-0.243	-5.899	-0.106***	-0.086*	-0.060	-0.158
	(0.069)	(0.117)	(0.039)	(0.434)	(0.128)	(0.080)	(0.400)	(3.292)	(0.025)	(0.041)	(0.098)	(0.261)
Other	-0.255	-0.340***	-0.274	-0.680	-0.080	-0.088	-0.373	-4.322***	-0.284	-0.223	-0.273	-0.262
	(0.173)	(0.074)	(0.214)	(0.462)	(0.827)	(0.172)	(0.305)	(0.833)	(0.216)	(0.228)	(0.183)	(0.251)
Resettlement	-0.774***	-0.560***	-0.735***	-1.525***	2.436***	0.366**	-8.620***	1.655	-0.714***	-0.566***	-0.731***	-0.174***
	(0.155)	(0.080)	(0.085)	(0.307)	(0.206)	(0.123)	(1.000)	(4.274)	(0.093)	(0.102)	(0.125)	(0.030)
Transportation	-0.464	-0.146	-0.400**	-0.900***	-0.877	0.057	-0.120	-0.179	-0.463**	-0.259*	-0.412	0.172
	(0.275)	(0.151)	(0.168)	(0.199)	(1.487)	(0.273)	(0.105)	(0.304)	(0.192)	(0.138)	(0.269)	(0.228)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Infrastructure Unit	km^2	km^2	km^2	IHS	km^2	km^2	% Forest	km^2	km^2	km^2	km^2	km^2
Sample Restriction	None	None	Truncated	None	None	None	None	One District	High-Activity	2015-2019	None	None
User x Year FEs		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
User × Month FEs	✓											
District FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Cell FEs												✓
State x Month FEs		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
State × Year FEs	✓											
Year FEs												✓
Weights	None	Num. Trips	None	None	None	None	None	None	None	None	None	None
Clustering	Biome	Biome	Biome	Biome	Biome	Biome	Biome	Biome	Biome	Biome	District	Biome
N	143269	161782	161782	161782	161782	157655	161782	6884	149905	123331	161782	282427
R^2	0.707	0.765	0.694	0.694	0.650	0.448	0.694	0.821	0.692	0.695	0.694	0.549

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all columns (except 5 and 6) is mean species richness (SR) across users' trips in a district-month. Coefficients describe the marginal impact of infrastructure encroachment by projects of a given category. Column 1 includes user-month, district, and state-year fixed effects. Columns 2-10 include user-year, district, and state-month fixed effects. Column 2 is a weights the regression with number of trips underlying the mean outcome. Column 3 drops the three largest projects. Column 4 uses the inverse hyperbolic sine of the explanatory variable. Column 5 and 6 show results with two alternative species diversity metrics. Column 7 defines infrastructure encroachment as a percent of baseline forest cover. Column 8 restricts the sample to users active in a single district throughout the study period. Column 9 restricts the sample to districts with above-median numbers of users who record an above-median number of trips per user. Column 10 drops the year 2020. Column 11 clusters at the district level. Column 12 includes 10km $\times 10km$ cell fixed effects along with state-month and year fixed effects. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Column 1 includes an additional control for the number of months/year of birdwatching.

Table A12: The Infrastructure-Biodiversity Tradeoff in Inclusive and Extractive Districts

	Inclusive Institutions				Extractive Institutions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infrastructure (km^2)	-0.124**	-0.089	-0.099**	-0.124	-0.380	-0.383	-0.298***	-0.380***
	(0.046)	(0.062)	(0.040)	(0.095)	(0.145)	(0.132)	(0.009)	(0.103)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User × Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓	✓	✓
State x Month FEs	✓	✓	✓	✓	✓	✓	✓	✓
Spillovers		✓				✓		
Weighted			✓				✓	
Clustering	Biome	Biome	Biome	District	Biome	Biome	Biome	District
N	62891	62891	62891	62891	11188	11188	11188	11188
R^2	0.695	0.696	0.774	0.695	0.722	0.722	0.804	0.722

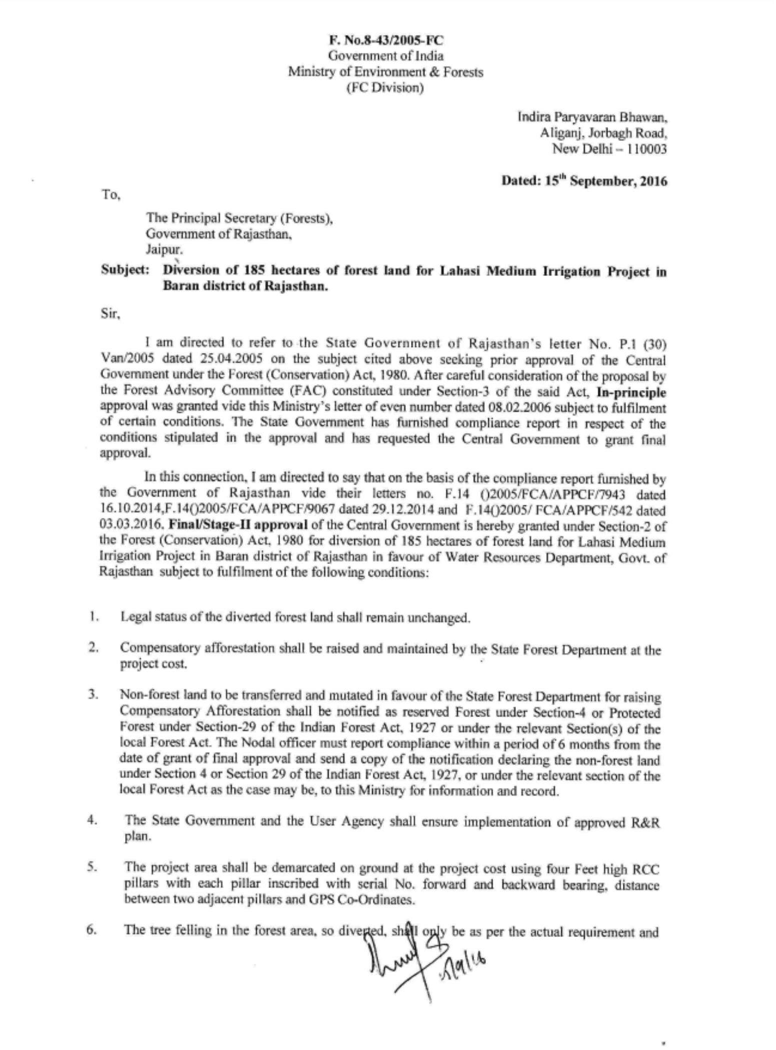
Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all specifications is mean species richness across users' trips in a district-month. Coefficients describe the impact of an additional km^2 of infrastructure encroachments in a district-month. Sample restricted to 99 inclusive districts (columns 1-4) and 66 extractive districts (columns 5-8) as defined by Banerjee and Iyer (2005) and aggregated to 1991 boundaries. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Sensitivity checks are described in the footer.

Table A13: Mechanisms by Which Institutions Mitigate the Infrastructure-Biodiversity Tradeoff

	(1) Informed Consent	(2) Cost-Benefit	(3) Protected Area
Inclusive (=1)	0.078***	0.071**	-0.006**
	(0.015)	(0.029)	(0.002)
Controls	Yes	Yes	Yes
Outcome Mean	0.234	0.156	0.007
State × Time FEs	✓	✓	✓
N	2275	2275	2270
R^2	0.541	0.510	0.237

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the project level for the digital subsample. Inclusive (=1) means the district has historically inclusive institutions. Sample restricted to 163 districts in Banerjee and Iyer (2005) under British Rule. Informed consent indicates whether the Gram Sabha was consulted and the FRA followed. Cost-Benefit Analyses are reports commissioned during project review. Protected Area equals one if the project is sited in or near one. All specifications include controls for: project size, district population share belonging to a tribal group as measured in 2011, baseline forest cover, and district area.

A2 Figures



F. No.8-43/2005-FC
Government of India
Ministry of Environment & Forests
(FC Division)

Indira Paryavaran Bhawan,
Aliganj, Jorbagh Road,
New Delhi – 110003

Dated: 15th September, 2016

To,

The Principal Secretary (Forests),
Government of Rajasthan,
Jaipur.

Subject: Diversion of 185 hectares of forest land for Lahasi Medium Irrigation Project in Baran district of Rajasthan.

Sir,

I am directed to refer to the State Government of Rajasthan's letter No. P.1 (30) Van/2005 dated 25.04.2005 on the subject cited above seeking prior approval of the Central Government under the Forest (Conservation) Act, 1980. After careful consideration of the proposal by the Forest Advisory Committee (FAC) constituted under Section-3 of the said Act, **In-principle** approval was granted vide this Ministry's letter of even number dated 08.02.2006 subject to fulfilment of certain conditions. The State Government has furnished compliance report in respect of the conditions stipulated in the approval and has requested the Central Government to grant final approval.

In this connection, I am directed to say that on the basis of the compliance report furnished by the Government of Rajasthan vide their letters no. F.14 ()2005/FCA/APPCF/7943 dated 16.10.2014,F.14()2005/FCA/APPCF/9067 dated 29.12.2014 and F.14()2005/ FCA/APPCF/542 dated 03.03.2016. **Final/Stage-II approval** of the Central Government is hereby granted under Section-2 of the Forest (Conservation) Act, 1980 for diversion of 185 hectares of forest land for Lahasi Medium Irrigation Project in Baran district of Rajasthan in favour of Water Resources Department, Govt. of Rajasthan subject to fulfilment of the following conditions:

1. Legal status of the diverted forest land shall remain unchanged.
2. Compensatory afforestation shall be raised and maintained by the State Forest Department at the project cost.
3. Non-forest land to be transferred and mutated in favour of the State Forest Department for raising Compensatory Afforestation shall be notified as reserved Forest under Section-4 or Protected Forest under Section-29 of the Indian Forest Act, 1927 or under the relevant Section(s) of the local Forest Act. The Nodal officer must report compliance within a period of 6 months from the date of grant of final approval and send a copy of the notification declaring the non-forest land under Section 4 or Section 29 of the Indian Forest Act, 1927, or under the relevant section of the local Forest Act as the case may be, to this Ministry for information and record.
4. The State Government and the User Agency shall ensure implementation of approved R&R plan.
5. The project area shall be demarcated on ground at the project cost using four Feet high RCC pillars with each pillar inscribed with serial No. forward and backward bearing, distance between two adjacent pillars and GPS Co-Ordinates.
6. The tree felling in the forest area, so diverted, shall only be as per the actual requirement and

Figure A1: Example Approval Letter

Note: Figure shows scanned letter from Principal Secretary of State Forest Ministry (Rajasthan) approving proposal for deforestation of 185 hectares for irrigation project. Encroachment area, district, and state are extracted for construction of manual sample. Additional fields from the actual proposal/application itself is only provided for the digital sample.

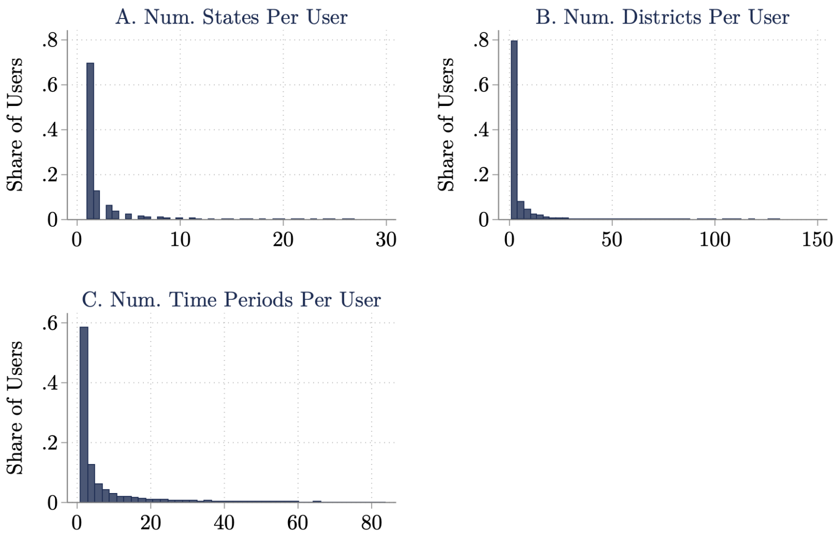


Figure A2: Within-User Distribution of Spatiotemporal Activity

Note: Distributions are based on aggregating eBird data (over all locations and time periods) to the user level (N=17,634 users). Panel A illustrates the distribution of total number of states traversed per user across all their trips during the study period. Panel B shows the same for total number of districts traversed per user. In Panel C, a time period is a year-month. There are 72 possible values (12 months*6 years). The distribution describes the number of periods in which a user recorded a trip.

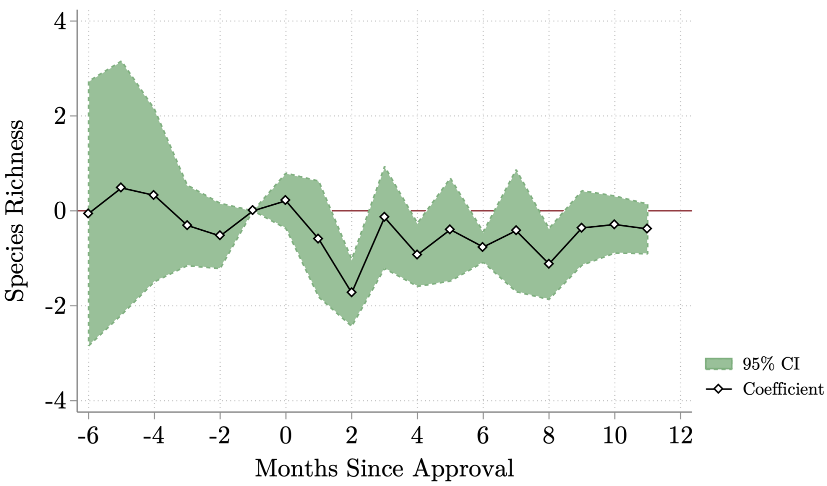


Figure A3: Event Study Results

Note: White diamonds are coefficients from a regression of species richness per user-district-month on the number of months before and after the first forest diversion event in a district. Time zero is the month before the event and all coefficients are normalized relative to this date. Dotted lines represent 95% confidence intervals. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log hour-of-day, log group size, and log spatial coverage. Standard errors clustered by biome. Standard errors are clustered by biome.

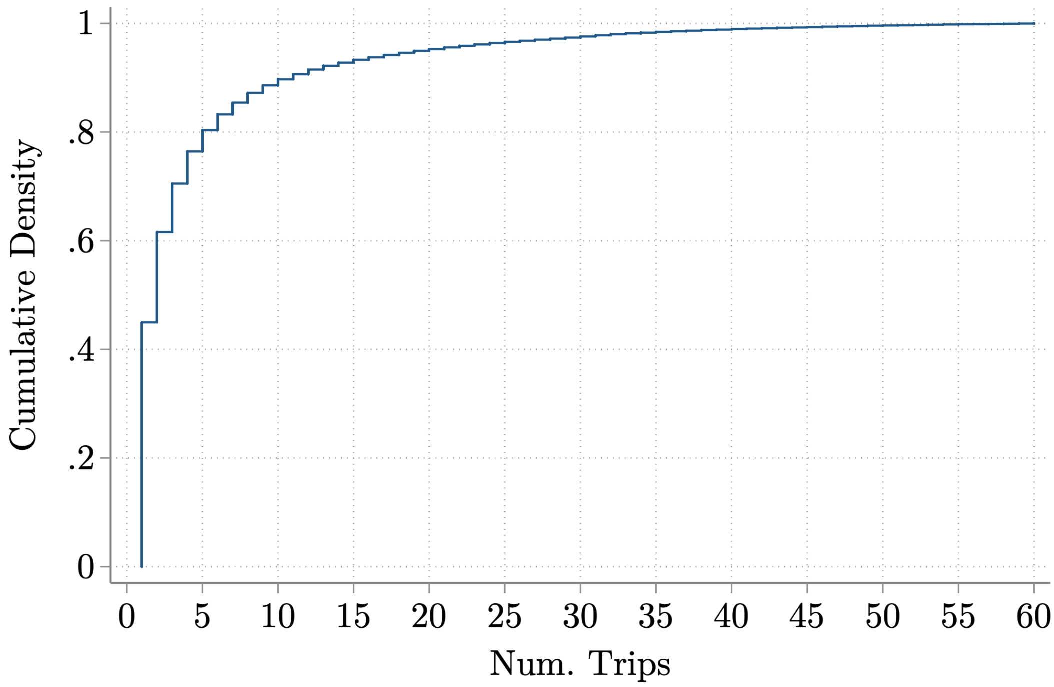
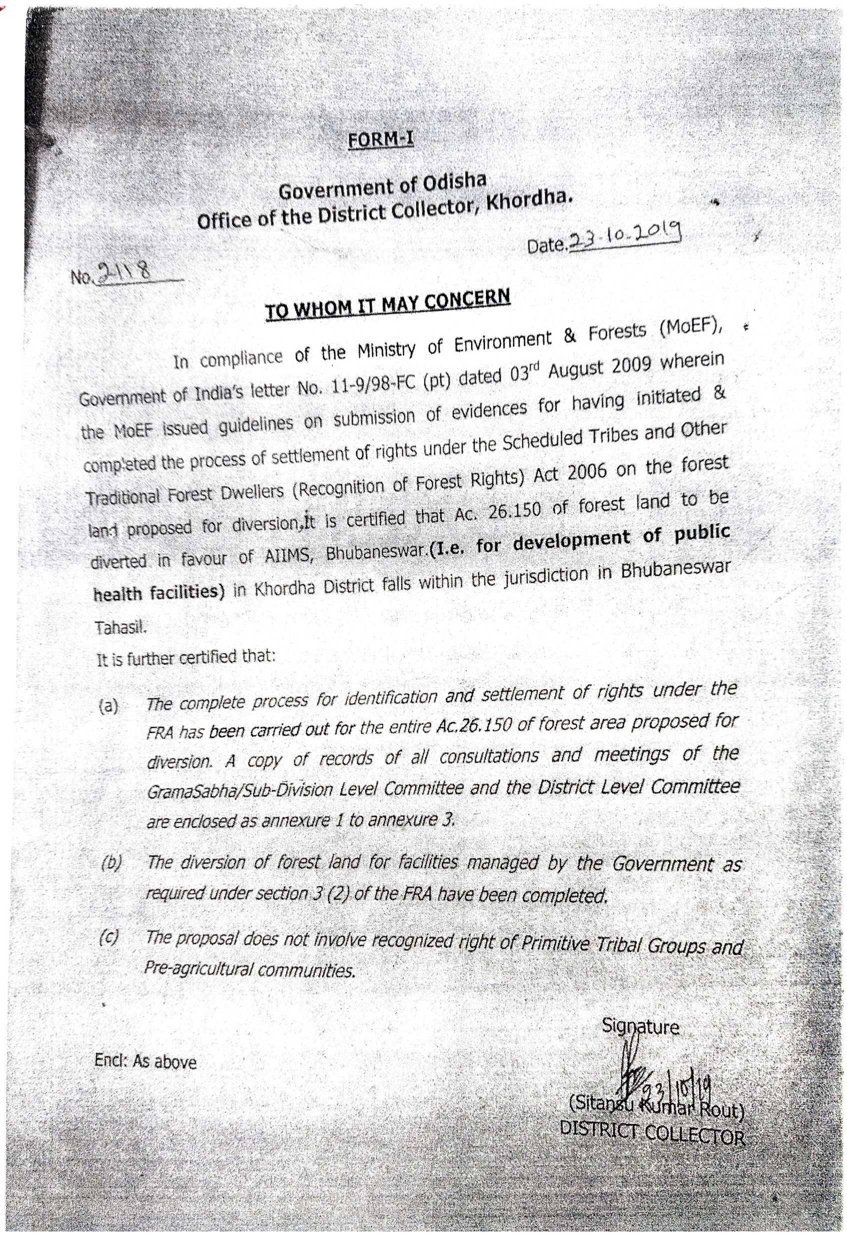


Figure A4: Cumulative Distribution of Weighting Variable

Note: The x-axis measures the total number of birdwatching trips taken by a user in a district-time period. Data are truncated at the 95th percentile to remove outliers. The y-axis is the percentage of observations with values less than the x-axis value.



FORM-I

Government of Odisha
Office of the District Collector, Khordha.

Date.23.10.2019

No.2118

TO WHOM IT MAY CONCERN

In compliance of the Ministry of Environment & Forests (MoEF), Government of India's letter No. 11-9/98-FC (pt) dated 03rd August 2009 wherein the MoEF issued guidelines on submission of evidences for having initiated & completed the process of settlement of rights under the Scheduled Tribes and Other Traditional Forest Dwellers (Recognition of Forest Rights) Act 2006 on the forest land proposed for diversion,It is certified that Ac. 26.150 of forest land to be diverted in favour of AIIMS, Bhubaneswar.**(I.e. for development of public health facilities)** in Khordha District falls within the jurisdiction in Bhubaneswar Tahasil.

It is further certified that:

(a) *The complete process for identification and settlement of rights under the FRA has been carried out for the entire Ac.26.150 of forest area proposed for diversion. A copy of records of all consultations and meetings of the GramaSabha/Sub-Division Level Committee and the District Level Committee are enclosed as annexure 1 to annexure 3.*

(b) *The diversion of forest land for facilities managed by the Government as required under section 3 (2) of the FRA have been completed.*

(c) *The proposal does not involve recognized right of Primitive Tribal Groups and Pre-agricultural communities.*

Signature
23/10/19
(Sitansu Kumar Rout)
DISTRICT COLLECTOR

Encl: As above

Figure A5: Example Letter of Informed Consent

Note: Figure shows scanned letter from Principal Secretary of State Forest Ministry (Rajasthan) approving proposal for deforestation of 185 hectares for irrigation project. Encroachment area, district, and state are extracted for construction of manual sample. Additional fields from the actual proposal/application itself is only provided for the digital sample.

B Data Appendix

B1 Infrastructure Sample Construction

Table B1: Procedure for Project Categorization

Raw Entry	Recategorized Entry
hydel, sub station, thermal, transmission line, village electricity, wind power, solar power	Electricity
road, railway	Transportation
canal, irrigation, drinking water	Irrigation
forest village conversion, rehabilitation	Resettlement
mining, quarrying, borehole prospecting	Mining
Everything else	Other

Note: The left column is the verbatim project categories as entered in application by the firm. The right column describes the synthesized categories for the purpose of simplifying project types.

Sample Construction and Digitization. The project sample consists of projects approved between 2015-2020. Applications submitted after 2014 (N=6,597) were scraped from the online portal (the digital subsample). Applications submitted before 2014 but approved afterwards (N=1,732) were manually digitized (the manual subsample). The application itself was not available for the manual subsample, but a PDF of the approval letter listing project size, location etc., along with other bureaucratic documents, was available in the portal.

The manual subsample was digitized as follows. First, a PDF of each approval letter was downloaded. The district of each project was extracted from the subject header (see Figure A1) and cross-checked online. If only the village was given, the district was identified on Google. Second, project size (hectares approved for forest diversion) was also extracted from the letter. For projects that span multiple districts (e.g. roads), a separate document called "Form A" (also available in the portal) was downloaded to identify hectares per district. 26 multi-district projects did not specify a district-wise breakdown, in which case total project size was divided equally across districts. Lastly, project category was extracted from the letter. In some cases it was taken from the Form A document which includes a detailed project description[23] .

Project Categorization. Verbatim project categories often refer to the same type of project. To simplify the analysis, I re-categorize projects according to Table B1[24]. Because of inconsistencies in raw data entry by firms, I manually examine and reassign projects mistakenly categorized as "other". "Other" project descriptions with the word "power", "substation", and "kv" are

[23]The category of each project in the manual subsample was available digitally, and scraped, but the majority were listed as "Other". I thus opted to categorize them based on the subject header text (see Figure A1)

[24]There are 89 projects categorized as "industry" which together make $< 0.1\%$ of total area cleared during the study period. For this reason, I include industrial projects in the "other" category.

placed in the Electricity category. "Other" projects with the word "resettle", "relocate", and "pattayam"[25] are placed in the Resettlement category.

District Splitting. I address the problem of district splitting and name changes. In the application, district name is an unstructured string with many spelling inconsistencies. To form a consistent matching key, I link the official district names and codes from the 2011 Census. To do this, I first manually rename district names to their original 2011 census name if there was a name change after 2011. Second, in the case of a district split after 2011, I rename the "child" districts to the name of the 2011 "parent" district. These two steps produce a single geographical unit that can be tracked consistently across years. Third, I perform a fuzzy match between district names in the application and the 2011 key using the Levensthein distance between the district strings. This algorithm identifies the official census code for 98% of districts in the application sample. The remaining 2% (8 districts) represent districts with more complex redrawing procedures and are dropped.

Panel Aggregation. I transform project-level data into a district-monthly panel in three steps. First, since each project can span multiple districts (e.g. roads), I reshape the data from project to project-district level. Each row contains the project component area falling into a specific district. The sum of rows for a single project equals total project size. Second, I aggregate to the total forest area diverted in each district and year-month, overall and category-wise (e.g. total deforestation in January 2018 for electricity projects in Delhi). Third, I balance the panel by assuming districts and time periods not in the portal had zero stage-II approvals.

[25] *Pattayam* means land deed and refers to a scheme for providing land to the landless, typically tribal families.

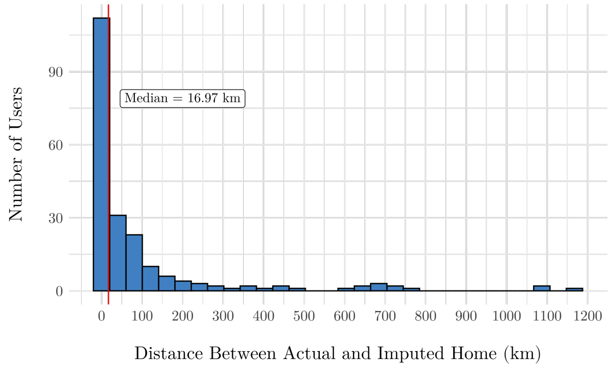


Figure B1: Distance between real and imputed home locations

Note: Data are from 210 eBird users who volunteered their actual home locations. Distance is the straight-line distance between their actual home and the gravitational centre of their trips, accounting for outliers (imputed home).

B2 User Demographics

eBird does not release data on user demographics. This appendix describes a method for inferring demographics when official data is unavailable. First, I impute user home locations as the gravitational centre of their trips. Second, I compare the distribution of user home locations to the general population to see whether they are rural or urban. Lastly, I characterize users more precisely by studying respondents from a large household survey who live near eBird users. The last two steps are inspired by Blanchard et al. (2021).

User Home Locations. Home is defined as the gravitational centre of users' trips. I start with the full sample frame (all protocols) and find the centroid of users' trips. Since trips far away from the main cluster (e.g. trips during vacation) warp the centroid, I drop outliers and then recompute the home. Outliers are identified by computing the straight-line distance from home to each trip destination, and then dropping those with distances below Q1-1.5*IQR or above Q3+1.5*IQR, where Q1 and Q3 are the first and third quartiles, respectively.

This method produces a fairly accurate approximation of home location. 210 users volunteered their real home locations, which I use for corroboration. I compute the straight-line distance (in km) between their real and imputed home. Figure B1 shows the distribution of offsets. The median difference is only 17km, which suggests considerable accuracy of the imputation.

I acknowledge this check is based on a selected sample. However, a similar imputation is applied for estimating eBird travel costs by Kolstoe and Cameron (2017), who received special access to eBird member profiles (including home address) for a much larger sample. Their results are robust to using imputed and real home locations.

Where Do Users Live? To visualize how representative users are in terms of where they live, I map the imputed homes of all 17,634 users in my analysis against gridded population den-

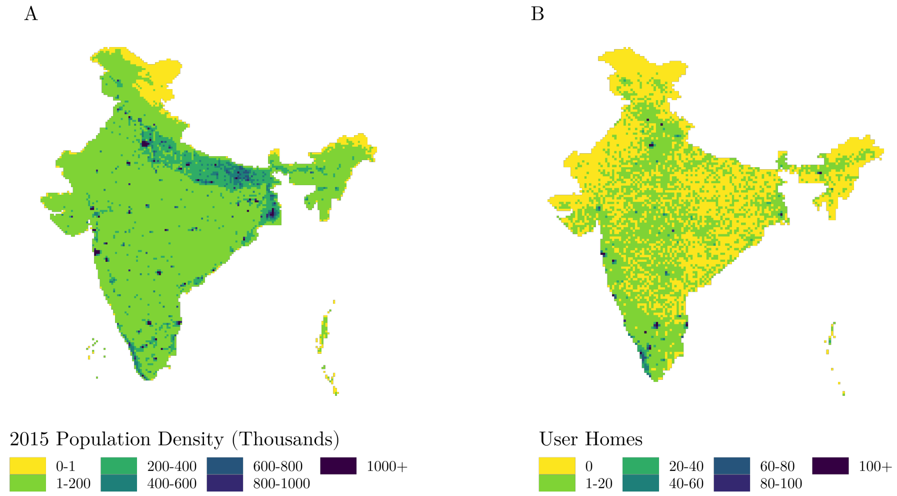


Figure B2: Population density of overall population and eBird users

Note: Both maps are at 20km resolution. Panel A) is the total population count in a cell from WorldPop. The method used for calculating population counts is described in the manual: https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained. Panel B) shows the count of eBird user home locations in each cell.

sity data for India. User density is mapped by constructing a 20×20km resolution grid and counting the number of user homes in each cell. Population density for 2015 is obtained from WorldPop[26]. Data are at 1km resolution and aggregated to 20km for consistency.

The dark hotspots in panel A of Figure B2 are India's largest cities. Many of these cities are also home to the highest density of eBird users (Panel B). While it may be unsurprising that eBird users live in big cities, Panel B also shows many remote eBird users (green).

To assess representativity, I compare the fraction of users living in "mega-cities" with more than 1 million population to that of the overall population. City polygons are obtained from the Global Rural-Urban Mapping Project (GRUMP), and I add a 3km buffer to include suburbs. Overlapping boundaries are dissolved into a single region. Extracting WorldPop counts over these polygons reveals that 27% of the Indian population live in megacities. The equivalent number for eBird users is 43%.

Location Profiles from the DHS 2015-16 Survey. As a last step to characterize eBird users, I draw on the DHS, a nationally representative household survey of 600,000 households. Households are grouped into georeferenced clusters, usually a village or town. There are 28,395 clusters with available coordinates. My goal is to identify clusters comparable to where eBird users live. The main challenge is that DHS geocodes are displaced to ensure confidentiality. Urban and rural clusters are displaced by up to 2 and 10 kilometres, respectively.

I start by defining eBird users as urban if they live in cities and rural if not. Cities are defined by GRUMP polygons (see above). Next, I identify DHS clusters within 5km of urban user homes

[26]Data accessed from: https://www.worldpop.org/. I use the 1km resolution unconstrained mosaic.

Table B2: T-test for equality of means between matched eBird and DHS samples

Variable	All	Urban	Rural
HH Size	-0.375***	-0.193***	-0.341***
Cellphone (=1)	0.053***	0.013***	0.040***
Fridge (=1)	0.253***	0.089***	0.181***
Car (=1)	0.089***	0.056***	0.067***
Sep. Kitchen (=1)	0.139***	0.035***	0.190***
Colour TV (=1)	0.212***	0.045***	0.190***
Internet (=1)	0.110***	0.062***	0.041***
Washing Machine (=1)	0.191***	0.109***	0.101***
Flush Toilet (=1)	0.271***	0.053***	0.216***

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Difference in characteristics between DHS respondents living in similar locations as eBird users compared to the overall DHS survey. The former dataset is weighted by the number of users to which each cluster is matched and the latter uses DHS survey weights. Robust standard errors are bootstrapped.

and 10km of rural user homes to account for displacements. This may generate mismatched pairs if, for example, a user living in a Delhi suburb is matched to a nearby rural cluster as well as urban clusters inside Delhi. Therefore, I only keep matches if the population density of the DHS cluster is within 25% of that in the user's home location, both calculated over a 5km buffer. This method matches 61% of users with at least one comparable DHS cluster. Note that the same cluster can match to several nearby users, resulting in duplicates. This is equivalent to a weighted dataset of unique DHS respondents with weights equal to the number of users to which the cluster is matched (Blanchard et al., 2021). I call this the "matched eBird" sample.

This procedure presents a new way to assess whether eBird users live in locations that are statistically similar to the average population. As such, the citizen scientists of India can be characterized in a data-poor context. I compare users along several wealth indicators reported in DHS, including: household size and ownership of various physical assets. T-tests for equality of means are conducted between the matched eBird sample and the overall DHS sample, with bootstrapped standard errors robust to heteroskedasticity. Survey weights are used for the overall sample and the number of matched users for the matched eBird sample.

Figure B2 shows the results. Overall, there are statistically significant differences in wealth between the matched and overall sample, indicating that eBird users live in non-representative locations. Compared to the overall population, eBird users live in places with smaller household sizes and better access to amenities such as a fridge, car, separate kitchen, and flush toilets. These differences persist even within rural and urban subsamples. Put differently, the urban locations where eBird users live are wealthier than the average urban location. Yet it should also be noted that these wealth differences are quantitatively small. Thus, while eBird users live in places that are not nationally representative, these places are not markedly atypical either.