

# The Economic Value of Biodiversity in India

## Running Notes

Raahil Madhok      Matt Braaksma      Ryan McWay  
Jovin Lasway

### Table of contents

Project Description . . . . .	2
Notes . . . . .	2
<b>TODO:</b>	<b>2</b>
Agenda . . . . .	2
<b>TODO:</b>	<b>3</b>
Agenda . . . . .	3
Project Management . . . . .	5
Coding Notes . . . . .	5
Agenda . . . . .	6
Site Attributes . . . . .	6
Next tasks . . . . .	6
Tasks: . . . . .	7
Site Attributes . . . . .	7
Sample Selection Updates . . . . .	8
Literature Review Updates . . . . .	9
Grant Proposal . . . . .	9
Initial Results by Raahil . . . . .	9
Tasks . . . . .	10
Sample Selection Discussion . . . . .	10
Tasks . . . . .	11
Action Items: . . . . .	11
<b>Division of Tasks:</b>	<b>11</b>
<b>Next scripts</b>	<b>12</b>
Action Items: . . . . .	12
Funding: . . . . .	12
Coding: . . . . .	12
Data: . . . . .	13

Methods: . . . . .	13
Paper/Writing: . . . . .	14

## Project Description

This project aims to produce the first measure of willingness to pay (WTP) for biodiversity. This provides a monetary measure to value biodiversity which can be applied to a wide variety of applications in environmental economics and natural capital accounting. The data relies on E-bird observations of diverse bird species observations. The sample is limited to local, long-time users who are residents (not tourists) in India. The methodology to calculate WTP relies on a revealed preference through the random utility model (RUM). The RUM determines the value of seeking out diverse species of birds by comparing the value of alternative counterfactual bird siting locations and the cost to go to these counterfactuals. Monetary value is determined through the RUM by estimating the travel cost of the reveal preference compared to counterfactual locations.

## Notes

For notes, we present the most recent first so that notes are chronologically most recent.

**i** 6/2/2025

Ryan, Matt, Jovin, Raahil

## TODO:

- **Matt:** Finish species richness site attribute (cluster-time level)
- **Matt:** Finish computing congestion site attribute
- **Matt:** Check why sample size so big. Check number of elements in the choice set for validation.
- **Matt/Ryan:** Compute species richness in radius *around home* (varying at same time unit as other species richness attribute)
- **Jovin:** Fix travel cost. I suggest checking it over with Matt/Ryan
- **Raahil/Ryan:** Work through paper framing

## Agenda

1. Species Richness
  - Maps for imputation can't have the shannon index. Because we don't have abundance counts
  - But we can do an index for the e-bird data hotspots
  - Estimates of bird life and expected species richness

- Monthly values or seasonal values
- 2. Other Site Attributes
  - Have one site data set for temp, rain, and tree cover
  - This is at the hotspot cluster level
- 3. Congestion
  - Ready to go code, just not run yet
  - # of users at hotspot cluster monthly
- 4. Travel Cost
  - Jovin has choice set and code that estimates this for a first attempt
  - Matched to GDP data to impute wages
  - Estimate Euclidean distance between households and hotspot clusters
  - Final data for RUM should be choice set dimensions (for our checks)
- 5. Measure of Experience at the Individual Level
  - Counts of trips (counter)
  - Count for # of different hotspots (counter)
  - So ratio tells us if you have preference for location (“Habit Formation”).
    - Low ratio is little location preference
    - High ratio is high location preference
- 6. Conceptual
  - Birds vs. non-bird still valued as travel cost. So both are still measures of nature recreation.
  - Monetizing value » Add in contribute to GEP or ecosystem service value literature.
  - Novelty: Why is WTP lower in Developing Countries? Welfare for environmental quality in envirodevonomics framework (Env-Dev framing)

**i** 5/16/2025

Ryan, Matt, Jovin, Raahil

## TODO:

- Seperate meetings for Matt/Jovin (coding) and Ryan/Raahil (conception) for our specializations

## Agenda

1. Finalize the site attribute data and create the master dataset
  - Matt and Ryan – complete by next meeting (May 16)
  - We have monthly temperature and rainfall

- We have annual tree cover data
- These are as zonal statistics (average for 10KM radius of hotspot)
  - In E-bird, we have the average distance walked around hotspot
  - Or do we avoid this work because temp, rain, and tree cover does not vary much at such a small spatial scale.
  - Temp, rainfall, and tree are at higher resolutions. So as long as the radius is less than the input data this will not vary.
- We need to think more about mapping of hotspots and hotspot clusters.
- Need to add a fourth site attribute: congestion
  - How frequently other users are using this site?
  - Aggregate number of user trips for each hotspot but will need to do this as a lag (week before)
  - So this is expected congestion.
- We need to transform this into a panel of counterfactual sets appended to the observed trips.
  - Not balanced across counterfactual trips and time
- 2. Calculate travel costs – see notes for guidance; this will be added to the site attribute script
  - For GDP per capita data, use gridded data from here and extract the value closest to each users' home. Then use the formula from Kolstoe/cameron or lloyd-smith to calculate travel costs.
  - Jovin – complete by next meeting (May 16)
  - Collected 2014-2024 GDP per capita data
  - Laptop trouble estimating, working with MPC.
  - Could run the travel cost with a small random sample to make this manageable for Jovin's laptop capacity.
  - Can re-run the pipeline so we are running for only frequent users (1.6M instead of the 3M)
- 3. Species richness
  - Matt and Ryan – bring three ideas for next meeting (May 16)
  - Close to being done (Matt taking the lead)
  - Taking raw e-bird data and getting monthly counts
  - Need to translate from hotspots to hotspot clusters
  - Species ranges (for missing) shapefiles are crashing the code.
  - All of this is to get a species at a hotspot so we can turn this into a richness index
    - Will want this merged as richness for same month/week from previous year
    - In the app, you can see in real-time species count from their favorite places.
    - So likely quite responsive. Last week is best
    - Can also do monthly to replicate results from month last year from Kolsto paper method.

- We aim to use the simpson and shanon index from **vegan** package
- 4. Individual Attributes
  - With individual fixed effects, we partial out any time-invariant
  - From the e-bird data:
    - How frequently you went in the past year
    - Average distance traveled in the past year
    - Is there any measure of learning over time? Value of recreation is not static
      - \* Experience index
      - \* Cumulative experience (simple counter of trips up until that point)
      - \* Could weight this by how much you walk, and taxonomy of bird level data (raw text file of trip-species level)
      - \* Could collapse species count to trip level.

## Project Management

- Meetings need to be in person.
- Jovin update meeting calendar
- Start assigning tasks instead of asking for volunteers
- Need to come up with a specialization of tasks.
- Coding and Data: Matt (lead), Jovin, Ryan (assist)
- Conceptualize: Ryan (lead), and Raahil
  - How our model departs from workhouse.
  - Literature review for our novelty
- Some tasks:
  - Google maps driving time
  - inaturalist species richness mapping

**i** 5/15/2025

Ryan, Matt

## Coding Notes

- Ryan is taking lead on estimating travel cost.
- We need a separate hotspots shapefile for running zonal stats
- We need unique hotspot site IDs and hotspot cluster IDs
- Explore alternatives to making hotspot clusters
- Name for the master data
- Need to add attributes to hotspot clusters (script 4)
- Mapping of hotspot cluster ID to hotspot IDs

Raahil, Matt

## Agenda

1. Site attributes data progress
  - In progress
2. Individual attributes data progress
3. Coding sample selection progress
  - Done, flexible to time period and years (e.g. quarterly for at least 8 years)

## Site Attributes

- Still gathering geospatial data
  - [MODIS VCF forest cover](#)
  - [Copernicus ERA5-Land](#)
    - \* [ecmwfr](#), R API interface
- Run zonal statistics on buffer around hotspots
- If data becomes too large to easily share, we can start using Globus and MSI's tier 2 data storage

## Next tasks

1. Merge site and individual attributes to make master dataset
2. Calculate travel cost
  - Start with linear distance for now
  - Calculate cost
    - Refer to Lloyd-Smith and Cameron/Kolstoe
    - Travel cost = 2 (round trip) x distance x 1/3 x wage (from GDP)
3. Species Richness
  - Decide on method:
    - Weitzman index
    - iNaturalist
    - Simpson index
    - Shannon index
4. Looking ahead
  - Look at code for estimating mixed logit with big data
    - [mixl](#), C++ based R package
    - [xlogit](#), GPU-Accelerated Python Package
  - How to properly weight the WTP for aggregation
  - Research how to include fixed effects in utility functions (alternative specific constant)

- User fixed effects
- hotspot fixed effects

**i** 4/25/2025

Matt, Ryan

### Tasks:

- Matt: Download data and share
- Ryan: Start writing pseudocode in the new scripts.
- Both: Play with ClimateR

### Site Attributes

- Limitation:
  - It takes time to build GIS data. So a lot of the attributes are not updated for 2024/25
  - ClimateR
- Division of the attributes to collect
  - Protected area (binary or distance) – fuzzy borders leads to spillovers
    - \* Matt
    - \* WDPA (<https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA>)
  - Tree cover – higher willingness to pay for forest
    - \* Matt
    - \* <https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/>
    - \* <https://data.globalforestwatch.org/>
    - \* <https://glad.umd.edu/dataset/global-2010-tree-cover-30-m>
    - \* Worse comes to worse we use the ESA data
    - \* Or we could use NDVI data as a proxy.
  - Rainfall (monthly) - seasonality
    - \* Ryan
    - \* [https://ftp.cpc.ncep.noaa.gov/GIS/GRADS\\_GIS/GeoTIFF/GLB\\_DLY\\_PREC/](https://ftp.cpc.ncep.noaa.gov/GIS/GRADS_GIS/GeoTIFF/GLB_DLY_PREC/)
    - \* CHRIPS SPI data
    - \* <https://mikejohnson51.github.io/climateR/index.html>
  - Temperature (monthly) - seasonality
    - \* Ryan
    - \* GISS Surface Temperature Analysis (GISTEMP v3)
    - \* Surface air temperature (no ocean data), 250km smoothing (9 MB)
    - \* Might need the land mask

- \* <https://data.giss.nasa.gov/gistemp/>
- Distance to the coast – species gradient to the coast
  - \* Ryan
  - \* Dissolve the districts data.
  - \* Make a polyline and measure distance to the nearest line
- Species richness from the previous week (expectation) – and any other measures of biodiversity
  - \* Both Matt and Ryan
  - \* Weitzman index using ebird data to generate diversity score
  - \* [https://scholar.harvard.edu/files/weitzman/files/on\\_diversity.pdf](https://scholar.harvard.edu/files/weitzman/files/on_diversity.pdf)
  - \* Validate with inaturalist data on birds for india
  - \* Could test a couple of different indices
- How we are sharing the site attribute data
  - Matt will have the GDrive link.
- Updating the config files
  - Once we have data download
- Start script to merge site attributes to sites
  - What to do about 4.demographics. Seems like that should be site attributes.
  - Notably, some of the data will need to become panel.
    - \* Rainfall
    - \* Temperature
    - \* Protected Areas
    - \* Tree cover
    - \* species richness (we need to dig some that has a time dimension)
  - Static:
    - \* Distance to coast

**i** 4/18/2025

Raahil, Matt, Jovin, Ryan

## Sample Selection Updates

- different biodiversity levels by incorporating interaction terms in locals and tourists. WTP for nature tourism vs WTP for nature recreation or biodiversity
  - If we estimate both, then we can compare both measures. We can also place it within the larger literature of many WTP for both subsets of ecosystem services and the population using those services.
  - Broadly compare conservation, nature recreation, tourism, or biodiversity



- We need to frame this as a development paper as the lambda in the Envirodevonomics paper
- Frequency of observations are consistent over time and potentially visits near protected areas.
  - Need to capture seasonanlity (time fixed effects)
  - Need to consider some aspect of congestion (can calculate that as a site attribute)
  - If you go two months a year you are less likely to be a tourist

## Literature Review Updates

- We can uses species richness as an explanatory variable in the model. But could add other biodiversity measures beyond bird diversity.
- Need to think about how to frame this in the developing context instead of the environment context
- Very few WTP for environmental quality in development economics space
- Would be great if our WTP measure could be placed within a welfare model so we can make counterfactual statements
  - Need to have homogenous agents to aggregate for welfare gains » e-bird users are not representative of population
  - Could overcome this by population demographic weights to make e-bird users representative
  - But welfare statements are a big improvement over the other e-bird papers
- Could explore the WTP measure of space and time. So we go beyond the cross-sectional estimates in the literature
  - If we have 1st and 2nd moment, then we can assign individuals an individual WTP.

## Grant Proposal

- Jovin will provide an update.

## Initial Results by Raahil

- We are getting coverage of the country
- Some overlap of population density and users » likely representative
- Most hotspots and users are in the south (consistent with e-birder users)
- Hotspots data could be weighted by visits
- Should plot protected areas overlapping hotspots
- Demographic t-test shows e-bird users are richer on average. But we can use this to back out our inverse weight for the welfare calculations

- E-bird users are selecting into this where there are richer places (so can afford this hobby)

## Tasks

- How do we distinguish ourselves from the Pacific Northwest paper Koleset 2017
- First pass of the sample: users who use this at least 2 months out of the year
- Make two new scripts:
  - Merging in site attributes (Can merge in satellite data – radius around the centroid and take the mean – and district data about the site attributes – merge on district ID)
    - \* Protected area (binary or distance) – fuzzy borders leads to spillovers
    - \* Tree cover – higher willingness to pay for forest
    - \* Rainfall (monthly) - seasonality
    - \* Temperature (monthly) - seasonality
    - \* Distance to the coast – species gradient to the coast
    - \* Species richness from the previous week (expectation) – and any other measures of biodiversity
  - Estimating the RUM via the mixed multinomial logit
- Email tag for grant proposals
- Make a script into a single PDF

**i** 3/31/2025

Raahil, Matt, Jovin

## Sample Selection Discussion

- We want to ensure that the sample of ebird observers is representative
  - Remove tourists
  - Remove infrequent users
- Potential strategies
  - Limit to users that log at least a certain number of times per month
  - Exclude users that post outside of protected areas
  - Exclude users that include tourism tags in text variable(e.g. “ecolodge”)

## Tasks

- Matt/Jovin: descriptive analysis on data to determine distributions of frequent users
  - Complete by 4/18

**i** 3/21/2025

Raahil, Matt, Ryan, Jovin

### Action Items:

- Everyone confirms that we can run the first 3 basic scripts
- Starting the annotated bibliography with 10 most important papers:
  - WTP for environmental quality
  - WTP in developing countries using travel cost
  - Review of mixed logit methodology
- Next week we need to talk about sample selection in depth next week

### Division of Tasks:

- Review mixed logit method
- (Matt) Review of WTP in developing countries using travel cost methods
- (Ryan) Review of WTP methods for environmental quality
- (Ryan) Method to estimate welfare effects (some exogenous shock)
- (Raahil + Jovin + Matt) Determining the sample population criteria
  - important early step
- (Raahil) Gathering data for population demographics (SHRUG, DHS, etc.)
- Ways to annotate bibliography:
  - paper title + journal, research question, research design (e.g. stated preference), data/sample size, result (i.e. WTP estimate)
- We need to think about how we contribute.
- We need to think about sample selection.
  - Criteria based on usage, overlap with protected areas, etc.
  - How do we remove tourists?
  - Who is the average person? Estimates will be sensitive
- Need to gridded map of ebird observations

## Next scripts

4. Merge attributes of the counterfactual sites. Need to collect site level data and/or individual level data
  - Site attributes: biodiversity, accessibility, ruggedness, congestion, temperature, rainfall, regional development
  - For first run, just do biodiversity.
  - Biodiversity comes from either diversity of birds or other attributes of ecosystem diversity
  - Expected biodiversity is why you go. So we can use historical measures of biodiversity
5. Run the mixed logit to estimate the WTP values
6. Welfare analysis of WTP values

**i** 02/28/25

Raahil, Matt, Ryan

### Action Items:

- Ryan: Get the correct e-bird file from Matt (DONE)
- Ryan: Merge code files into scripts folder. (DONE)
- Ryan: Initiate a document for literature review with template example (DONE)
- We need to specialize and define tasks. Volunteer over email. Tasks: Lit review, sample selection, characterize the sample, theory person (DONE)
- And set up a recurring meeting time. Need a poll. Try for next two weeks from now. (DONE)
- Jovin: Provide your github username so Ryan can make you a collaborator on the github repo. (DONE)

### Funding:

- No longer pursue IonE funding

### Coding:

- Matt was able to replicate everything
  - Created a YAML to process file names
  - No longer needs new file paths
- Ryan stuck on loading data
- Need to all use the same exact file download for e-bird

- script 1: processes e-bird data, determines user homes, and calculates home for those who don't report it.
- script 2: Creates the hotspot observations. Measures distance from home to hotspot.
- script 3: Most important. 12,000 hotspots so too many counterfactuals. This script reduces the counterfactual set by clustering hotspots nearby (likely parks) and only consider hotspots within a distance of home.
- The final data set is a panel of person (user) ID by trip ID appended to all counterfactual trips for this user.
- When we actually estimate this, we will need to plan for a no-trip counterfactual.
- One of the main things we need to do is make the computation easier for us. Otherwise it becomes too computationally intensive.
- Keep all the scripts in scripts folder.

### Data:

- We will need to bring in site attribute data
- So data on landscape measurements.
- Matt says that not all intermediate files are being save
- Specifically, the master file is not saving
- Need to identify our sample population
  - Need a careful (defensible) data-driven criteria for sample
  - Needs to be regular, local users (not temporary or tourists)
  - Determine this based on usage frequency and mobility patterns
  - Consider overlaying protected areas polygons on top of the observations/hotspots
  - Buffers may create a censoring problem. So we would just test this against larger and larger buffers
- Need to find ways to characterize the users (demographics) without having any information about them.
  - Important for welfare calculations.
  - Try to match with DHS data
  - Or overlap home location with SHRUG data set
  - Econometrically, take attributes about where they live against the average for the country, how do high and low WTP birders fall within the distribution
  - Potentially use this as weights for welfare calculations.

### Methods:

- If we get a random shock to bird population, then we can think about the causal effect of increasing biodiversity

### **Paper/Writing:**

- For the ML cluster, will want to read up later to find out if there is a better way or more common practice in economics.
- Create a lit review. Three groups:
  1. estimates of WTP in developing countries generally using the discrete choice travel cost method (Ted Miguel water quality paper).
  2. WTP for environmental quality papers and what design they are using to estimate (e.g., contingent valuation, stated preferences – something from ecology)
  3. Review of any updated versions of the mixed logit model – new adaptations of the model from class
- For this: question, data used, design, and finding (dollar estimates would be great).
- Create a document in the docs folder