### **Study Information**

1. **Title (required)** 
   1. *Provide the working title of your study. It may be the same title that you submit for publication of your final manuscript, but it is not a requirement.*

How is accessibility to cities influencing biodiversity change?

1. **Authors (required)**

Daniela Gargya

1. **Description (optional)**
   1. *Please give a brief description of your study, including some background, the purpose of the study, or broad research questions.*

*Background*

Globally, biodiversity is changing in complex ways in a time of accelerating human impact. We know little about how large scale anthropogonic activity, as a prominent contemporary global change driver, is indirectly influencing the current reshuffling of ecological communities. Direct environmental influences, such as forest loss and warming, have been found to catalyse biodiversity changes. Quantifying the indirect impact of global changes of human activity, such as roads and urbanisation, will allow to better unravel different sources of heterogenous biodiversity change. Urbanisation and roads can influence changes in ecological communities by increasing connectivity and favouring urban-adapted species, causing both temporal and spatial turnover. Species’ response to such influences is dependent on ecological aspects such as their mobility. Highly mobile species are found to be affected most by urbanisation. Beyond the effect of urbanisation and roads on biodiversity change, high human population density often coincides with high species richness, illuminating spatial congruence between human development and diverse ecosystems. This can be seen both as a threat and as an opportunity to improved management. Recent global-scale data compilations of biodiversity time-series (BioTIME) and data sets which indirectly capture large-scale anthropogenic activities, will allow me to quantify their influence on ecological communities worldwide. Linking human impact with biodiversity change can provide the needed evidence and predictions for better international policy making in the light of our rapidly changing Anthropocene.

*Purpose of the study*

My aim is to quantify how global change drivers influence ecological communities (and individual taxa) over time and space. Specifically, I will focus on the effects of urbanisation and roads, captured in the metric accessibility to cities, and human population density on temporal and spatial changes on ecological community composition overall and across taxa. These drivers are an indirect representation of large- scale human activity and capture big parts of the alteration of our planet. I will test if these drivers of global change explain heterogenous biodiversity change found across our planet.

*Research questions*

1. Do sites with higher accessibility to cities experience more changes in assemblage composition over time (temporal turnover) than locations with lower accessibility?  
   1. How does the duration of ecological monitoring influence the magnitude of detected temporal turnover trends?
2. How does temporal turnover of ecological communities respond to levels of accessibility across taxa (birds, mammals, terrestrial invertebrates, terrestrial plants)?
3. How is temporal turnover influenced by an interaction between human population density and accessibility?

If time allows look at population level analysis (choosing the 3 most abundant populations from the datasets used above)

1. Do sites with higher accessibility to cities experience more changes in population trends than locations with lower accessibility?  
   1. How does the duration of ecological monitoring influence the magnitude of detected population trends?
2. How do population trends of ecological communities respond to levels of accessibility across taxa (birds, mammals, terrestrial invertebrates, terrestrial plants)?
3. How are population trends influenced by an interaction between human population density and accessibility?
4. **Hypotheses (required)**
   1. *List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.*

Definition of terms used

An ecological assemblage is a collection of species occurring in the same place at the same time, sharing the same phylogeny.

Turnover is the change of species identity of an ecological composition, calculated in the final year of each time series relative to the first.

Population trends are the overall trends in the abundance of an individual species across the duration of each time-series.

For all hypotheses, turnover can be replaced with population trends.

1. Do sites with higher accessibility to cities experience more changes in community composition over time (temporal turnover) than locations with lower accessibility?

*Hypotheses*

* Sites with higher accessibility to cities correspond with greater temporal turnover
* The magnitude of temporal turnover increases for sites which have been monitored over longer durations

I predict greater temporal turnover with greater exposure of accessibility to cities, as an alteration of the natural environment benefits some species, while damaging others, influencing community composition.

I predict greater temporal turnover in sites which have been monitored over longer durations, as the effect of turnover becomes more apparent over time.

1. How does temporal turnover of ecological communities respond to levels of accessibility across taxa (birds, mammals, terrestrial invertebrates, terrestrial plants)?

*Hypothesis*

* Ecological communities of taxa respond differently to high and low levels of accessibility.

I predict both positive and negative trends to be present.

1. How is temporal turnover influenced by an interaction between human population density and accessibility?

*Hypothesis*

* The relationship between accessibility to cities and temporal turnover is steeper, when human population density is higher

I predict the relationship between accessibility and temporal turnover to be steeper, when human population density is higher, as high HPD increases the pressures on the natural systems, leading to higher turnover.

### **Design Plan**

*In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.*

1. **Study type (required)**

Other

1. **Blinding (required)**
   1. *Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.*

No blinding is involved in this study.

1. **Is there any additional blinding in this study?**

N/A

1. **Study design (required)**
   1. *Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.*

In my analysis I will test the relationships between accessibility to cities and 1) temporal turnover, 2) spatial turnover, 3) taxa’s individual response 4) the interaction of human population density on temporal turnover, and 5) and the interaction of human population density on spatial turnover. My analysis will include observational studies of assemblage compositions over time from the terrestrial realms only, with no manipulative experiments included. I will include data from 173 studies, from 9427 unique location, totaling 2773780 observations. The time-series used represent repeated studies of species abundance and identity of all species found within an ecological community. The data collection of BioTIME was consistent within studies but not between studies.

The BioTIME database is limited in its even representation of different taxa and latitudes. It underrepresents reptiles and amphibians and the tropics and polar regions (see Figure 1).

I will extract the mean accessibility score over ~100km² grid cells around each location of available biodiversity time-series.

1. **Randomization (optional)**
   1. *If you are doing a randomized study, how will you randomize, and at what level?*

N/A

### **Sampling Plan**

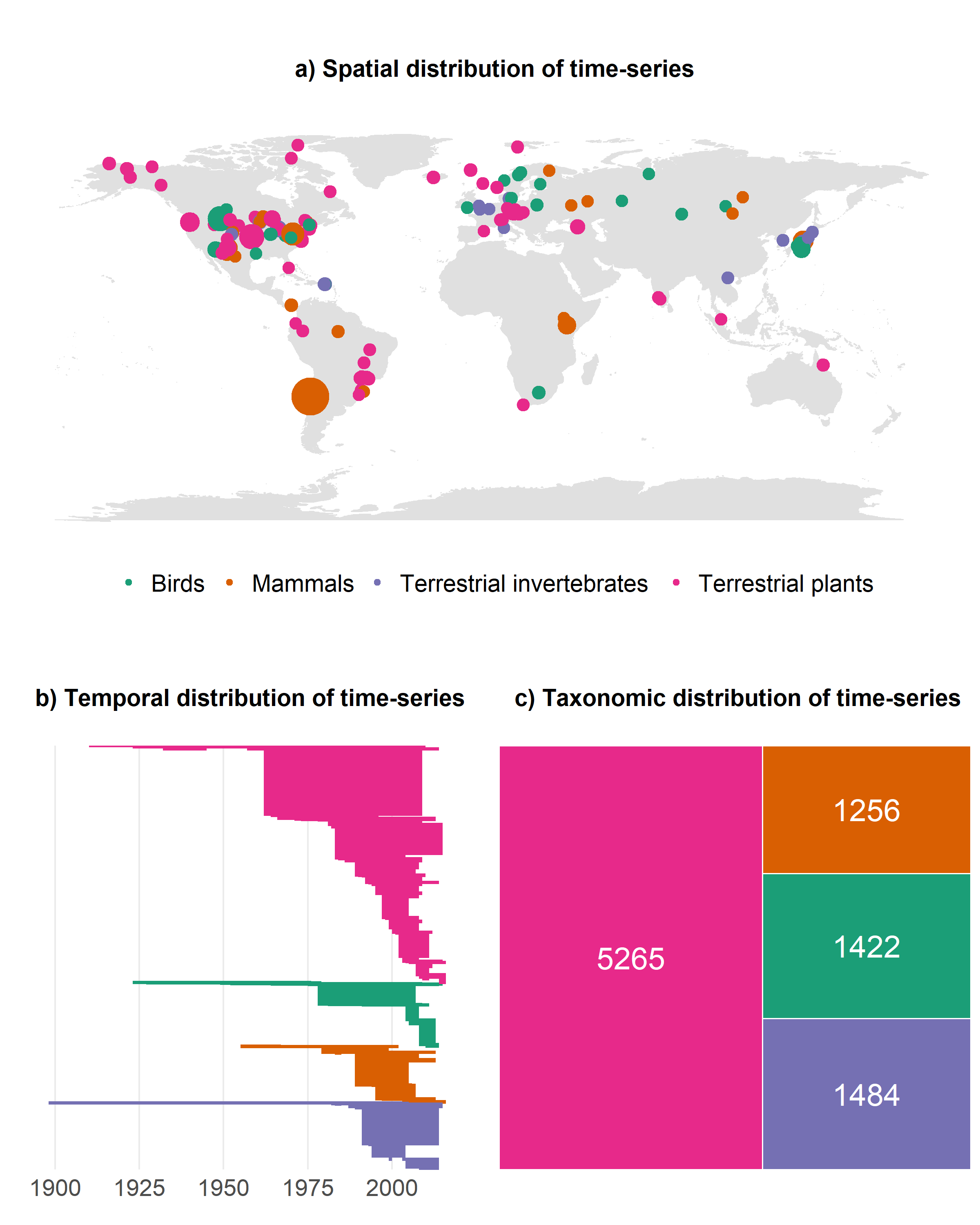
*In this section we’ll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.*

Registration prior to analysis of the data

1. **Explanation of existing data (optional)**
   1. *If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.*

All my data exist already. I will be the first to integrate the different databases together in this way. The BioTIME database is openly accessible on their website (<http://biotime.st-andrews.ac.uk/>). The accessibility to cities is available through the malariaatlas research website (<https://malariaatlas.org/research-project/accessibility_to_cities/>).

I have looked at meta-data only to figure out sample sizes, but I have not conducted any of the analysis. I visualized the meta-data of the BioTIME time-series in Figure 1.

  
*Figure 1: Visualization of meta-data of BioTIME time-series across space, time and taxa.*

1. **Data collection procedures (required)**
   1. *Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that don’t include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.*

All my data is readily available online and accessible to anyone, since all databases are open source. I will not conduct any first-hand data collection myself. I will use the Google Earth Engine (GEE), which is an open source online global spatial analysis platform to extract accessibility scores from the accessibility to cities database. I will use the biodiversity time series data from the BioTIME database. All further analysis will be conducted in R. I will include all available data that meet my criteria.

1. **Sample size (required)**
   1. *Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?*

I will combine records from 173 studies, from 9427 unique location, totaling 2773780 observations of community composition as well as the accessibility to cities score and human population density at each site. The sample sizes for different taxa can be found in Table 1. The spatial and temporal scales for sample sizes can be found in Table 2. As the BioTIME database is continuously updated, more data might become available. I will use all the data available as of the 03/03/2020.

Table 1: Sample sizes on different levels of observation for taxa and total.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taxa** | **Studies** | **Plots** | **Observations** |
| Birds | 35 | 1422 | 1299941 |
| Mammals | 22 | 1256 | 385663 |
| Terrestrial invertebrates | 21 | 1484 | 192008 |
| Terrestrial plants | 95 | 5265 | 896168 |
| **Total** | **173** | **9427** | **2773780** |

Table 2: Mean spatial and temporal scales of sample sizes.

|  |  |  |
| --- | --- | --- |
| **Spatial scale** | Mean plots per study | 54 |
|  | Mean observation per plot | 15240 |
| **Temporal scale** | Mean years per study | 17.5 |
|  | Mean data points per study | 11.1 |

1. **Sample size rationale (optional)**
   1. *This could include a power analysis or an arbitrary constraint such as time, money, or personnel.*

N/A – I will use all data available which fit my criteria: part of the terrestrial realm, at least 2 survey points in time, minimum time-series duration of 5 years and having at least 20 studies per taxa.

1. **Stopping rule (optional)**

N/A

### **Variables**

*In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.*

1. **Manipulated variables (optional)**
   1. *Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.*

N/A

1. **Measured variables (required)**
   1. *Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.*

Response variable:

* *Temporal turnover*: changes in species composition due to replacement over time, calculated as the comparison in the final year of each time-series relative to the first.
* *Population trends: overall trend in population abundance over time.*

Explanatory variables:

* The mean accessibility to cities scores for the 100km² grid cells around the location of the biodiversity record
* The interaction term between accessibility and population density
* Taxa
* Duration

Random variables:

* Broad grid cell
* Study ID/plot

Examples of all variables can be found in the sample data csv file in the appendix.

1. **Indices (optional)**
   1. *If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.*

The accessibility to cities score is derived from multiple layers of input data. For more details on the method behind the database, see Weiss et al., 2018.

### **Analysis Plan**

*You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.*

*A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.*

1. **Statistical models (required)**
   1. *What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions, subgroup analyses, pairwise or complex contrasts, or follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Remember that any test not included here must be noted as an exploratory test in your final article.*

I will use hierarchical models in a Bayesian framework in R. First, I will calculate temporal turnover trends. The outcome of these calculations will be used subsequently and tested against accessibility scores. The specific structure of the statistical model is outlined below. The model includes weakly informative priors.

temporal turnover ~ accessibility + duration + accessibility:human population density, (accessibility|taxa); random = grid cell, study ID/plot

The models will be based on a zero one inflated beta distribution.

I will run the model first with both random intercepts and slopes but if no model convergence can be achieved, I will run them with only random intercept.

I will initially run models with one chain of 100 000 iterations with a warmup of 10 000 iterations and will increase these metrics if convergence is not achieved. I will assess convergence visually by examining the trace plots.

1. Do sites with higher accessibility to cities experience more changes in community composition over time (temporal turnover) than locations with lower accessibility?

where temporal turnover is calculated as the change in community composition due to species replacement, based on the final year of each biodiversity time-series relative to the first. Turnover is based on Jaccard’s dissimilarity metric, where zero means no change in community composition and one means complete change of species.

1. How does temporal turnover of ecological communities respond to levels of accessibility across taxa (birds, mammals, terrestrial invertebrates, terrestrial plants)?

For details on model structure, distribution and priors, see beginning of section.

1. How is temporal turnover influenced by an interaction between human population density and accessibility?

For details on model structure, distribution and priors, see beginning of section.

Additional analysis

* Accessibilities sensitivity to cell size
* Temporal mismatch of data: analysis with less data, e.g. including only data from 1970 – 2010

1. **Transformations (optional)**
   1. *If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.*

Scaling accessibility and dissimilarity between 0 and 1.

HPD between 0 and 1.

(Coding scheme for categorical value of human population density)

Grid cell random effect

Partitioning of studies?

1. **Inference criteria (optional)**
   1. *What criteria will you use to make inferences? Please describe the information you will use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?*

As I will be using a Bayesian framework, my inferences will be based on the posterior distribution of each fixed effect. They will be considered significant if the lower and upper 95% credible intervals don’t overlap zero. I will conclude all results, regardless of the direction or magnitude of the effect size.

1. **Data exclusion (optional)**
   1. *How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled? Will you use any awareness check?*

I will include all data that fit my criteria.

1. **Missing data (optional)**
   1. **How will you deal with incomplete or missing data?**

There is no missing data per se as the accessibility dataset is global. There are temporal mismatches though between the biodiversity time-series and the one-point nominal measuring point of accessibility in 2015. The durations of the time-series vary and go back to early 20th century. The accessibility score can be considered a cumulative variable. In order to analyze the influence of the temporal mismatch of the BioTIME time series with the accessibility score, I will conduct a sensitivity analysis which will only include data from the BioTIME data set from 1970 - 2010.

1. **Exploratory analysis (optional)**
   1. *If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.*

N/A

### **Other**

1. **Other (Optional)**
   1. I*f there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.*

N/A

**Appendix**

Appendix 1: Sample data csv file



