# Metadata Code – Static Predictors

## 0\_1\_Atlas\_prep.qmd

Commented Quarto file which handles the atlas data. The script is used to calculate occupancy, occupancy-area-relationship, telfer index of change and log ratio of AOO change (AOO = area of occupancy).

The data is subset to remove biases: cells that were sampled twice, species that are found in both time periods (thus, have change data). NAs in the data are also homogenized to the same NA-string. Switched spherical geometry off and transformed atlas Coordinate-Reference-System (CRS) to WGS84.

Saturated scales (relative occupancy N cells = 1) are excluded from calculation of OAR. Species with less than 2 scales without saturation are excluded completely (this was the case for 21 species from sampling period 1 in CZ and 24 species from sampling period 2 in CZ).

Negative fractal dimension for species with only 1 cell (max = 3) of occurrence; slope == 1 or slightly above 1 (e.g.,1.013)

R-Version:

* R version 4.3.2 (2023-10-31 ucrt) -- "Eye Holes"

### Required packages in R:

* dplyr, rstatix, plyr
* sf
* sparta package (GitHub) – telfer
* tidyr (pivot\_wider)

### Required data:

* Bird atlas data for CZ, NY, JP, EU

### Produces the following files:

#### 1) presence\_data\_raw.rds

= a merged data set of the first and second sampling period of all atlases (full data).

* I added a column where the sampling period is bifactorial: either 1 or 2. (it’s not a spatial object)

#### 2) presence\_data\_reduced.rds

= Reduced version of 1), subset to cells that were sampled twice and species that are found in both time periods.

#### 3) presence\_data\_final.rds

= Modified version of 2) that includes columns for (at all scales and time periods separately calculated for each species in each atlas):

* Total N cells sampled
* Total N cells Atlas
* Sampled and unsampled cells together (thus NAs in species columns if cell was not sampled)
* Fixed Nas (in some atlases, Nas were “NA” in others “<NA>”).
* Added grain column (10 for CZ, 5 for NY, 20 for JP, 50 for EU). It’s the approximate side length of the cell

#### 4) Occupancy\_table.csv

Reduced version of 3) with measures of occupancy:

* Occupancy area = sum(area) occupied by each species
* mean area = mean occupied area of a cell
* Occupancy N cells = Number of distinct cells occupied by the species
* AOO = Number of occupied cells \* mean area of the cells
* relative occupancy area = Occupancy area / Total area
* relative occupancy N cells = occupancy N cells / Total N cells sampled
* total species richness atlas (gamma diversity) = Number of distinct species in the atlas
* *Scale* column where the scale at which the whole atlas is covered by only 1 grid cell is scale = 1, and all other scales are proportions of this.

#### 5) Big\_table\_CZ\_JP\_NY\_EU.csv

Big table with all measures calculated for all species in all sampling periods in all atlases (i.e., occupancy, fractal dimensions, telfer, log-ratio)

#### 6) Change\_Data.csv

Subset of 5) only with Species, Dataset, Sampling Period and Log Ratio.

## 0\_2\_AtlasPredictors.qmd:

Script to create predictors that can be easily calculated from the atlases. Note that the climatic niche of species is calculated from the global range and has thus its own designated script (0\_3\_Climate\_Niche\_prep.qmd).

Since the data needs a slightly different format for these calculations than they did in the previous script, we have to read the raw atlas data in, remove cells and species that were only sampled once and bind them into a list (instead of a data frame).

For this script, the scales were reduced to the one with the highest resolution. These are different between atlases.

### Required R-Packages:

* dplyr, rstatix,
* sf (atlas grids)
* ape (ladderize and read phylogenetic tree)
* phyloregion (Fair Proportion Evol. Distinctness)
* fossil (co-occurrence)
* geosphere, geodata, terra, tidyverse, tidyterra (Geometries)

### Required Data:

* Bird atlas data (CZ, JP, NY, EU)
* BirdTree (single tree from Weeks et al 2022– is subsample of BirdTree)
* AVONET\_final.csv (Tax. Homogenized version of AVONET to the species in our data)

From 0\_1\_Atlas\_prep.qmd script:

* presence\_data\_final.rds
* presence\_sf\_list.rds
* Big\_table\_CZ\_JP\_NY\_EU.csv

From 0\_3\_Climate\_Niche\_prep.qmd script:

* Niches\_df.rds
* RangeSizeBOTW\_df.rds

From 0\_2b\_Create\_TaxLookupTable.R

* Tax\_lookup\_BirdLife\_eBird\_BirdTree\_Atlas.csv OR
* Tax\_lookup\_BirdLife\_eBird\_BirdTree\_AtlasBOTW.csv (by hand merged with Birds of the World taxonomy)

|  |  |  |
| --- | --- | --- |
| Table XX: All predictor variables calculated | | |
|  | Atlas predictors | Species predictors |
|  | Dataset |  |
|  | Total area |  |
|  | Time period |  |
|  | Grain |  |
|  |  |  |
|  |  |  |
|  |  |  |

### Produces the following files:

* Presence\_sf\_list.rds (list of sf objects for each atlas and sampling period)
* Diversity\_AvgEffort.rds (Dataframe with alpha, beta, gamma richness and mean sampling effort per species)
* SAC\_df.rds (Spatial autocorrelation)
* Geometries.rds (Geometric features of atlas and species ranges)
* Coocc\_df\_final.rds (Co-occurrence indices per species)
* All\_predictors.rds (final dataframe with all predictors ready for predictor selection)