# Manuscript – Draft: Methods

## Data

### Atlas data

- species presence data grouped by time period 1 and 2

- species presence data reduced to those cells that are sampled in both periods

- species presence data reduced to those species that are found in both time periods from the remaining cells   
[these two steps together removed 166640 rows]

### Grid data:

- different scales read in as layers

- homogenized crs between atlases to WGS84

#### Occupancy (AOO)

- AOO was calculated for each species at each scale for each time in each atlas.

- Total number of sampled cells was calculated as the number of cells that are found in both time periods

- Total number of cells per atlas was calculated as the total number of cells in the grids for each atlas

- The following measures were calculated for ((((Species)Scale)Tp)Atlas):

- occupancy area = sum(area per cell)  
 - occupancy ncells = number of unique cell labels  
 - mean area = mean(area per cell)  
 - AOO = occupancy ncells / mean area  
 - relative occupancy area = occupancy area / Total area  
 - relative occupancy ncells = occupancy ncells / Total Ncells sampled  
 - total species richness per atlas = number of distinct species in the atlas

[Data for CZ: 200, EU: 503, JP:208, NY:237 species in each time period]

#### Fractal Dimension (D)

- D was calculated from the Occupancy-Area-Relationship for each species in each time period in each atlas across all scales where the relative occupancy was less than 1 (i.e, excluding saturated species). First, species were completely removed from the data which saturate at the highest resolution. Second, all scales where species relative occupancy ncells was 1 were excluded from the calculation of the fractal dimension D. Then, a linear model with the formula lm(log(AOO)~log(mean\_area)) was fit for each species/time/atlas combination. From this, the slope (b) was extracted and used to compute fractal dimension D as follows: D = -2\*b+2. The transformation is necessary to scale D between 0 and 2 to give a useful indication of its spatial dimension.

#### Telfer Index

- The Telfer index for each species is the standardized residual from a linear regression across all species and is a measure of relative change only as the average real trend across species is obscured (quote from sparta package).

#### Log ratio of temporal change

- Log ratio was calculated as log(AOOtime1 / AOOtime2) at the highest resolution

### Correlation checks

#### Measures calculated from the raw Atlas:

- No correlation higher (lower) than 0.69 (AOO and mean area)

### Data standardization

## External predictors

### Taxonomic homogenization

- To match the atlas data to other predictors gathered from external sources, taxon names had to be homogenized. For this we created a lookup table for several taxonomies (eBird, BirdLife, BirdTree). Nine species were matched by hand because synonyms were not detected from the lookup table.

### Climatic niche breadth

- We calculated climatic niche breadth for each species as the standard deviation along the two major PC axis that include the following CHELSA bio variables averaged across 1981-2010 across the species global range: *mean annual air temperature, mean diurnal air temperature range, isothermality, temperature seasonality, mean daily maximum air temperature of the warmest month, mean daily minimum air temperature of the coldest month, annual range of air temperature, annual precipitation amount, precipitation amount of the wettest month, precipitation amount of the driest month, precipitation seasonality*. These variables are known to characterize the water- and temperature demands of species. Climate raster layers were stacked and aggregated to 10x10km resolution to enhance computation and match roughly the scale of the atlas data.

- First we calculated the global climate space for these variables using PCA (prcomp in R)(for loadings see Table SXX). Climate variables were scaled because of highly different variable dimensions. Second, BirdLife global range maps (subset to: Breeding and resident, native or reintroduce, extant or possibly/probably extant) for each species in the Atlas dataset were projected onto the climate stack and average climate values were extracted for each species at 10x10km square cells across the species global (reduced) distribution. Extracted climate values were used to predict each species global distribution into the climate space of the PCA. Standard deviation of the first and second major axis were computed for each species to characterize the species climatic niche breadth.

### Diversity Metrics

- We calculated gamma species richness (γ) as the total species richness of each atlas in each time period. Species alpha diversity (α) was calculated as the average species richness in all cells in which the species is found at the highest resolution. From these two measures, a species Whittaker’s beta diversity (β) was calculated as:

### Average Species sampling effort

- We calculated the average sampling effort across all cells in which a species occurs to have a measure of over- or under-sampling for each species in each time period at the highest resolution. [Problematic predictor because of NAs for Japan, NY and Europe]

### Spatial Autocorrelation

- We used the ncf R package to calculate Moran’s I from the correlogram of species presence/absences based at the first distance increment, which was based on the original cell size of the specific Atlas. To include adjacent cells in addition to the neighboring cells into the calculation, we set the increment to calculate Moran’s I as: . (See supplementary material for visualization of the estimation of increments)

### Geometries

We aimed to characterize geometric features of our data (country geometries and those of the species range). See table below.

|  |  |  |
| --- | --- | --- |
| Table XX: Geometric indices calculated from the country shapes and species range shapes. Bold indices were chosen for the final model because they captured the most collinearity to the other indices without being collinear to each other. | | |
| **Geometry index** | | **Formula** |
| east-west distance | | Distance from eastern-most to western-most border |
| **north-south distance** | | Distance from northern-most to southern-most border |
| maximum distance | | Maximum distance between any two points on the borders |
| elongation ratio | | E = 1 – S/L (S = short-axis length; L = long-axis length; axis are estimated from the minimum bounding box. It gives a measure of polygon direction in addition to overall elongation |
| circularity | | C = (Perimeter ^2)/ Area, where a circle has C = 1 and everything else C > 1 (🡪 INF) |
| **normalized circularity** | | CNORM = (perimeter^2) / (4 \* pi \* area), where a circle has C = 1 and everything else C < 1 (🡪 0) |
| length of the minimum bounding box | | Length of the smallest rectangle needed to cover the polygon |
| width of the minimum bounding box | | Width of the smallest rectangle needed to cover the polygon |
| elongation of the minimum bounding box | | E = 1 – S/L of minimum bounding box |
| **related circumscribing circle** | | RCC = 1 - A / Ac ; ranges from 0 to 1, where 0 indicates a circular polygon and 1 a highly elongated polygon. It is a measure of polygon elongation |
| linearity index | | Coefficient of determination (r2) calculated from a regression analysis of the x and y coordinates of the exterior convex hull nodes. A measure of how well the polygon can be described by a straight line. It is more efficient than elongation ratio as it does not require finding the minimum bounding box first |
| north bearing of the minimum bounding box | | Angle of the minimum bounding box, measured clockwise from the north angle. |
| bearing along the longest axis | | Angle of the longest axis, measured clockwise from the north angle |
| Center of gravity of the atlas long | | Mean long of all species population centroids in the country |
| Center of gravity of the atlas lat | | Mean lat of all species population centroids in the country |
| Species Centroid lat | | Mean lat of species range inside atlas |
| Species Centroid long | | Mean long of species range inside atlas |
| **Distance centroid-CoG** | | Distance from the species centroid to center of gravity of the atlas |
| **Min dist centroid-border** | | Minimum distance from the species centroid to the border of the atlas |
| **Min dist border species-border atlas** | | Minimum distance from the species range border to the border of the atlas |
| **Max dist centroid-border** | | Maximum distance from the species centroid to the border of the atlas |
| **Max dist border species-border atlas** | | Maximum distance from the species range border to the border of the atlas |
| **Southerness** | | Divided the atlas into half, set values above the centroid to northern and below the centroid to southern and assigned a scaled score where 1 = most northern and 0 = most southern. Extracted the lat and long values for these classes, classified species by the lat and long of the range centroid and transformed the values so that Southerness = 1-Northerness. |
| **Westerness** | | Divided the atlas into half, set values right of the centroid to Eastern, left of the centroid to Western and assigned a scaled score where 1 = most eastern and 0 = most western. Extracted the lat and long values for these classes, classified species by the lat and long of the range centroid and transformed the values so that Westerness = 1-Easterness. |
|  |  | |

### Co-occurrence of species

We used the R package “fossil”(Vavrek, 2011) to calculate co-occurrence metrics from our occupancy data. The package was designed to work with paleontological data, for which abundance is most often difficult to estimate – just like in our atlas data. We computed community presence-absence matrices for each grid cell in each atlas in each time period and estimated a species mean co-occurrence with any other species by estimating pair-wise co-occurrence across all species in the community, and then determining a species average co occurrence to all other species in the community. These values range from -1 to 1, where species with an index of -1 never co-occur with any other species in the matrix, while those with index of 1 tend to always co-occur with all other species. These are two extreme and rather unrealistic examples of the community structure as it is experienced by each species. Intermediate values will give an indication of where on the continuous spectrum of co-occurrence each species is placed.

# Supplementary material

## Data Processing

### PCA Loadings:

|  |  |  |  |
| --- | --- | --- | --- |
| Table SXX: PCA Loadings for calculating the global climate space in which species occurrences are projected to estimate climatic niches. | | | |
| Bio\_Name | CHELSA\_1981-2010\_Bio\_Var\_name | PC1 | PC2 |
| bio1 | mean annual air temperature | 0.36 | 0.26 |
| bio2 | mean diurnal air temperature range | -0.15 | 0.48 |
| bio3 | isothermality | 0.28 | 0.18 |
| bio4 | temperature seasonality | -0.36 | 0.14 |
| bio5 | mean daily maximum air temperature of the warmest month | 0.28 | 0.41 |
| bio6 | mean daily minimum air temperature of the coldest month | 0.39 | 0.14 |
| bio7 | annual range of air temperature | -0.34 | 0.25 |
| bio12 | annual precipitation amount | 0.35 | -0.13 |
| bio13 | precipitation amount of the wettest month | 0.33 | 0.03 |
| bio14 | precipitation amount of the driest month | 0.27 | -0.30 |
| bio15 | precipitation seasonality | 0.01 | 0.55 |
| Proportion of variance | | 0.51 | 0.22 |
| Cumulative proportion | | 0.51 | 0.73 |

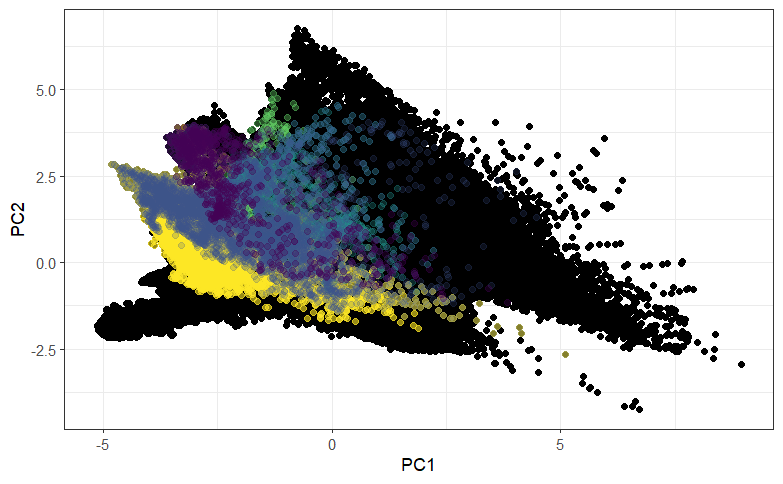


Figure S1: Example of climate space for some species (colored points) against the global climate space (black points)

## Spatial Autocorrelation

A grid with a green square and red arrow

Description automatically generated

c

a

Figure SXX: Estimation of increment for calculation of Moran’s I on the gridded atlas data.

## Collinearity in the predictors

We calculated Pearson correlation coefficients for a closer selection of predictor variables that we chose on the basis of our hypotheses for further analysis. Variables that exceeded a threshold of Pearson > 0.8 were removed from the final set of predictors that went into the models. These variables are listed in Table XX.

|  |  |
| --- | --- |
| Table SXX: Highly correlated predictors that were removed from the analysis | |
| **Atlas predictors** | **Species predictors** |
| Related circumscribing circle | Related circumscribing circle |
| Longitude center of gravity | Longitude centroid |
| Latitude center of gravity |  |
| East-West distance | East-West distance |
| North-South distance |  |
| Maximum distance | Maximum distance |
|  | Max distance from the border of the range to the border of the atlas |
|  | Max distance from the centroid to the border of the atlas |
| Linearity index |  |
| Total area |  |
| Gamma species richness |  |
| grain |  |
| Elongation ratio |  |
|  | Relative occupancy # cells |
|  | Mean probability to co-occurr |

A close-up of a computer screen

Description automatically generated

Figure SXX: Correlation matrix of predictor variables to predict temporal change. Those labels marked in red are highly correlated and some of them are finally removed (see Table SXX).