

`combine_databases.R` put together tetradensity with body mass databases for vertebrates.

`make_density_model.R` creates one model per taxonomic class of density. Models are selected, i.e. AIC was used to select the model with the most parsimonious set of predictors.

Materials and methods

First draft of the part of the methods I was involved in.

To calculate energy fluxes, species biomass densities and the predator-prey trophic interactions are required. Our workflow can be divided into four steps: 1) we retrieved body mass information for amphibians (EuropAmphib), birds (Elton), mammals (Elton), and reptiles (Slavenko). 2) We obtained biomass densities of species by fitting a statistical model similar to Santini et al. (2018a) to the TetraDENSITY dataset (Santini et al., 2018b). 3) We predicted biomass densities for vertebrate species in Europe and combined this information with their predator-prey interactions, obtained from TETRA-EU 1.0 (Maiorano et al., 2020). 4) for each $10 \times 10 km^2$ cell in Europe, we then calculated the energy fluxes for the local food web using species local biomass densities and local temperature to estimate metabolic rates.

Taxonomic harmonization (`harmonize-taxonomy.R`)

As original sources for taxonomic names were different across datasets, our first step was to harmonize the species names against a common taxonomic backbone. As datasets comprised multiple taxonomic groups and had regional to global scope, we chose GBIF, a multi-taxa, global backbone, to harmonize taxonomic names (Grenie et al., 2021). GBIF was accessed through `rgbif` (3.6.0) on December 2021. We first appended all taxonomic names from all datasets into one list of 25,688 unique species names, that was then queried in GBIF to obtain the accepted taxonomic name. When a first iteration on GBIF (using `name_backbone()`) did not returned an accepted name, we ran a second step (using `name_usage()`) where we used the GBIF key for the taxonomic name to query the database. A total of 2,243 species names were re-assigned, including changes in taxonomic families (Supplementary material *backbone.csv*).

Body mass inference (`combine_databases.R`)

As some species lacked body mass information, necessary to calculate both species densities and energy fluxes, we performed a multiple imputation by chain equations using the package *mice* (ref). Multiple imputation had the advantage of obtaining unbiased uncertainty estimates compared to single imputation techniques (Nakagawa, 2015). Imputations were performed using a Bayesian linear regression (`mice(..., method = "norm")`) using taxonomic family as

predictor. All chains had similar mean and standard deviation, the influence of missing data on estimate uncertainty was low (0.077), and the average relative efficiency was high (0.993), all indicating robust imputations.

Biomass density predictions (`make_density_models.R`)

To predict species biomass densities, we replicated the analysis performed in Santini et al. (2018b). In particular, we subsetted TetraDENSITY to retain only locations within Europe, as this was the target area of our predictions. We obtain the climatic predictors NPP, etc (*to add and explain what they are*). We then built for each taxonomic class separately the linear model:

$$\log_{10}(\text{density}) = \log_{10}(\text{mass}) + \log_{10}(\text{mass})^2 + \log_{10}(\text{NPP}) + \log_{10}(\text{NPP})^2 + \text{PCV} + \text{PCV}^2$$

Quadratic terms were included as they were shown to play an important role in determining species densities (Santini et al., 2018b). We then used a multi-model averaging approach using package MuMIn (`dredge()` and `model.avg()`). In particular, we averaged coefficient estimates across all model that had $\Delta\text{AIC} < 2$ from the best model; we used the full average, i.e. including a coefficient as zero when was not present in a model, as conditional average can lead to overestimates of model parameters. We then used the climatic, body mass, and species distribution data to predict, using the fitted linear models, species individual densities for each $10 \times 10 \text{ km}^2$ cell. As some taxonomic orders were present in TETRA-EU, but not in TetraDENSITY (e.g. Chiroptera), our models lacked coefficient estimates for such orders; we obtained these by averaging the coefficient estimated of the other orders present in the taxonomic class. After species densities of individuals were obtain, we calculated species biomass densities by multiplying the individual densities by the species body mass.

Changes to Santini et al. (2018b) model (for supplement, if needed)

The statistical model of Santini et al. (2018b) contains a random effect structure to *factor out* the effect of taxonomic units and site replicates from vertebrate density estimates. This was achieved by including as random intercept the nested effect of *order/family/species*, which explained a large proportion of the total variance in the data. Our models included only fixed effects (i.e. they were not mixed-models) and we included only taxonomic order as predictor of species densities. This was because we were not interested, contrary to Santini et al. (2018b), in an overall response of vertebrate densities to climatic factors, but our aim was to predict species densities accurately for each site, without making statistical inference on the model results. As such, including taxonomic information as fixed effect allowed us to obtain estimates of between-family differences that gave more accurate predictions compared to including taxonomic information as a random intercept.

Fluxweb (`flux-per-pixel.R`)

Using species biomass densities, we calculated the energy losses due to metabolic rates of species as:

$$e^{0.71 \cdot \log_{10}(mass) + A - \frac{0.69}{boltz \cdot (273.15 + temp)}}$$

where $A = 19.50$ for endotherms and $A = 18.05$ for ectotherms (refs). The efficiency of energy extraction from a resource was set equal to 0.906 for all species. Using the TETRA-EU metaweb and the distribution ranges of species, we built a local food web subsetting the European metaweb retaining only the species that were considered present in the focus cell. When food webs consisted of only two species, flux calculations were skipped. As some cycles were present in the European metaweb (e.g. *Vulpes vulpes* and *Bubo bubo* predating on each other), we removed one of such link randomly and independently for each cell in order to avoid computational issues when calculating fluxes. We then calculated fluxes using package fluxweb (Gauzens et al., 2018; `fluxing()`).

References

- Maiorano, L., Montemaggiore, A., Ficetola, G. F., O’connor, L., & Thuiller, W. (2020). TETRA-EU 1.0: a species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, 29(9), 1452-1457.
- Santini, L., Isaac, N. J., Maiorano, L., Ficetola, G. F., Huijbregts, M. A., Carbone, C., & Thuiller, W. (2018). Global drivers of population density in terrestrial vertebrates. *Global Ecology and Biogeography*, 27(8), 968-979.
- Santini, Luca, Nick JB Isaac, and Gentile Francesco Ficetola. “TetraDENSITY: A database of population density estimates in terrestrial vertebrates.” *Global Ecology and Biogeography* 27.7 (2018): 787-791.
- Grenié, M., Berti, E., Carvajal-Quintero, J., Dädlow, G. M. L., Sagouis, A., & Winter, M. (2021). Harmonizing taxon names in biodiversity data: a review of tools, databases, and best practices. *Methods in Ecology and Evolution*.
- Nakagawa, S. (2015). Missing data: mechanisms, methods and messages. *Ecological statistics: Contemporary theory and application*, 81-105.
- Gauzens, B., Barnes, A., Gilling, D. P., Hines, J., Jochum, M., Lefcheck, J. S., ... & Brose, U. (2019). fluxweb: An R package to easily estimate energy fluxes in food webs. *Methods in Ecology and Evolution*, 10(2), 270-279.