

Metadata from “Data and scripts: Catchment scale deforestation increases the uniqueness of subtropical stream communities”

Authors: Fabiana Schneck, Luis M. Bini, Adriano S. Melo, Danielle K. Petsch, Victor S. Saito, Simone Wengrat, Tadeu Siqueira

1 Brief description of biological and environmental variables

Sampling region: Sampling was carried out in southeast Brazil between latitudes 23°49'S and 24°20'S. We sampled 100 perennial second- and third-order streams within 20 catchments (five streams per catchment). Streams were isolated from each other. Samplings occurred between September and November 2015. Streams were within catchments in pristine forests to agricultural ones.

Instream environmental variables: At each stream, we selected one riffle for sampling. We measured current velocity (m s^{-1}) and depth (cm) at nine random locations, whereas stream width (m) and particle size classes were estimated at three locations. We visually estimated particle size classes and classified them in: % of mud (<0.25 mm), sand (0.25–2 mm), gravel (>2 –16 mm), pebble (>16 –64 mm), cobble (>64 –256 mm), and boulder (>256 –1024 mm). We also estimated the percentage of the canopy cover of riparian vegetation at three random points. We measured electrical conductivity (mS cm^{-1}), pH, and dissolved oxygen (mg L^{-1}) at each site using a Horiba U-50 multiparameter probe. We obtained water samples to determine total nitrogen (mg L^{-1}) and total phosphorus ($\mu\text{g L}^{-1}$) in laboratory.

Land-use variables: We analyzed land use by visual interpretation of multispectral RapidEye imagery, at 5m spatial resolution, obtained from the Brazilian Ministry of Environment. We measured the proportion of land cover classes (native forest,

secondary forest, planted forests, pasture, agriculture, urban areas, mining, bare soil, water bodies and mixed agricultural-natural land uses) within a 400-m-radius buffer around each sampling site. Detailed information on how land-use variables were obtained is available at <https://doi.org/10.5281/zenodo.5650136>

Diatom's sampling and identification: At each stream, we sampled 10 stones of similar size and removed 25 cm² of periphytic material from each stone. The material from the 10 stones was pooled and preserved in 4% formalin. We cleaned the samples using acid and then mounted permanent slides using Naphrax®. We counted approximately 500 valves from each site using a light microscope with 1000× magnification. We identified diatoms mostly as species.

Insect's sampling and identification: We sampled insects using a kick-net with a mesh size of 0.5 mm. Sampling took 2 minutes per stream, divided in four 30-seconds sampling units obtained in the main microhabitats of the riffle (i.e., areas with different current velocities, depths, and substrate particle sizes). The samples were preserved in ethanol. We counted and identified organisms from the orders Ephemeroptera, Odonata, Plecoptera, Megaloptera, Trichoptera, and Coleoptera to the genus level.

2. Data files accompanying this metadata

env20_5.txt: a matrix with the values of each instream variable (columns) estimated in each stream (rows). The first and second columns represent catchment and stream identity, respectively. It also contains latitude and longitude of sampled streams in decimal degrees.

streams_buffer_400m.csv: a matrix with the values of each land-use variable (columns) estimated around each stream (rows). The three first columns represent and

ID number, catchment and stream identity, respectively. Note that variables not related to land-use classes are also in this dataset.

diatoms20_5.txt: a matrix with the abundance of each species (columns) sampled in each stream (rows). The first and second columns represent catchment and stream identity, respectively.

diatoms_traits20_5.txt: a dataframe with five traits (columns) for each diatom species (rows). Traits and respective categories according to columns order:

- (i) length/width ratio (measurement of at least 20 cells of each species whenever possible);
- (ii) size class: from 1 to 5, being 1 the smallest and 5 the greatest (based on biovolume);
- (iii) functional guild: low profile (low); high profile (high); motile; planktonic;
- (iv) life form: colonial (col); noncolonial (non_col);
- (v) mode of attachment: unattached; prostrate; tube-forming; vertical.

diatoms_phylo20_5.txt: a dataframe with diatom species (rows) classified into genus, family, order, subclass, class, and subdivision levels (columns).

inverts20_5.txt: a matrix with the abundance of each genus (columns) sampled in each stream (rows). The first and second columns represent catchment and stream identity, respectively.

inverts_traits20_5.txt: a dataframe with six traits (columns) for each insect genus (rows). Traits and respective categories according to columns order:

- (i) refuge building: no refuge (no_ref); fixed nets and retreats (fix_net); portable shelter of sand, debris and/or wood (por_shel_wood); portable shelter of leaf parts (por_shel_leaf);
- (ii) body shape: hydrodynamic (hydrod); not hydrodynamic (not_hydrod);
- (iii) locomotion: burrowers; climbers/crawlers; sprawlers; clingers; swimmers; skaters;
- (iv) functional feeding guild: collector-gatherers (coll_gat); collector-filterers (coll_filt); herbivores; predators; shredders;
- (v) respiration: tegumental (tegum); gill; air;
- (vi) body size: small: <9 mm (1); medium: 9–16 mm (2); large: >16 mm (3).

inverts_phylo20_5.txt: a dataframe with insect genera (rows) classified into subfamily, family, superfamily, suborder, and order levels (columns).