Preregistration

Understanding environmental disequilibrium in freshwater bioindicator communities across the Northern Hemisphere

Cindy Gao[1], Lars Iversen[1]

[1] McGill University

22. September 2025

Study Information

Understanding environmental disequilibrium in freshwater bioindicator communities across the Northern Hemisphere	Understanding environmental disequilibrium in freshwater bioindicator communities across the Northern Hemisphere
Description	Species' responses to environmental change are frequently lagged, especially under accelerating global change. Such lags have not been investigated in bioindicator species, which are species that are used to infer the state of an ecosystem. If bioindicators are exhibiting such lags, then they may not respond as expected. My project aims to investigate if mayfly, caddisfly, and stonefly (i.e. EPT) species, commonly used freshwater bioindicators, are lagging behind water chemistry and climate change using a community response diagram framework.

Hypotheses If EPT species are lagging behind their enivronment, then the environment inferred from the species present at a given site will be significantly different than the observed environment relative to null communities sampled from the regional pool.

Design Plan

We will create a dataset of EPT species abundances, climate, and water quality measurements for stream and river sites across Europe, Canada, and the United States. For every site and year, we will infer the enivornmental variable measurements based off the species composition and compare this inferred environment against the observed environment relative to null communities sampled from the regional pool.

Study type **Other**. Data synthesis and analysis project. Existing data is collated and used in a large-scale statistical analyses.

Blinding No blinding is involved in this study.

Study design

Randomization None.

Sampling Plan

Existing data **Registration prior to analysis of the data**. As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

Explanation of existing data	We will be using existing long-term water quality monitoring data collected by national and provincial programs, open-access repositories, and published studies. Access to the data has been limited by the fact that this is the first time I will have ever worked on a project of this nature.
Data collection procedures	Data will be synthesized from existing long-term water quality monitoring programs at the federal and provincial level, open-access repositories, and published studies.
Sample size	Our data contains 24 major watersheds across Canada at the moment. We will eventually have approximately 1250 watersheds across the Northern Hemisphere.
Sample size rationale	Our goal is to conduct a comprehensive assessment of EPT species for lagged responses. We therefore would like to have as many samples as possible to detect spatial and temporal patterns, which is why we are choosing to amalgamate a large dataset.
Stopping rule	We will terminate data collection in October 2026. This is because our analysis pipelines have already been written, and we would like to amass as much data as possible. This stopping date is arbitrary.

Variables

Variables include EPT species abundance, dissolved oxygen, total phosphorus, conductivity, turbidity, total dissolved solids, alkalinity, mean annual temperature, and precipitation. Measured variables will be 'degree of mismatch'.

Manipulated variables	None.
Measured variables	The measured variable is degree of mismatch, or environmental disequilibrium, at each site. This is taken to be the numerical difference between observed and inferred environments.

Indices None.

Analysis Plan

We aim to use the comclim R package to estimate degree of mismatch at each site in our dataset.

Statistical models We will use community climate diagrams to estimate site-specific enivronmental disequilibrium. The measurement of interest is numerical environmental disequilibrium values taken to be the absolute deviation between the inferred and observed environment. ## Transformations

Data will be standardized, but otherwise will not be transformed.

Inference criteria We will use the standard $p<0.05$ for determining if the deviation between the observed and inferred environments are significantly different relative to null communities sampled from the regional pool of species. ## Data exclusion

Outliers in water chemistry as determined by experts at Environment and Climate Change Canada will be removed from the data.

Missing data Missing data will be imputed using MICE.

Exploratory analyses (optional) None.

References

Li, L., Zheng, B., & Liu, L. (2010). Biomonitoring and Bioindicators Used for River Ecosystems: Definitions, Approaches and Trends. Procedia Environmental Sciences, 2, 1510–1524.

Blonder, B., Moulton, D. E., Blois, J., Enquist, B. J., Graae, B. J., Macias-Fauria, M., McGill, B., Nogué, S., Ordonez, A., Sandel, B., & Svenning, J.-C. (2017). Predictability in community dynamics. Ecology Letters, 20(3), 293–306.