Preregistration

Spatio-temporal variation of benthic phototrophic elemental signatures along depth profiles

Kirsten Bevandick[2], Kate Colson[1], Stefano Mezzini[2], William Ou[1], Julee Stewart[3]

[1] University of British Columbia

[2] University of British Columbia (Okanagan campus)

[3] University of Regina

2021-10-06

Study Information

Title	Spatio-temporal variation of benthic phototrophic elemental signatures along depth profiles
Description	Whole-lake ecosystem processes are complex and hard to quantify. Carbon and nitrogen cycling depend on the composition and abundance of a aquatic organisms, which, in turn, can vary spatio-temporally due to many physical and chemical factors. While most studies on whole-lake ecosystem processes focused focused primarily on pelagic production, Gushulak *et al.* (2021) presented the first study on nutrient cycling using depth profiles. Using stable isotope and nutrient concentration profiles of δ^{13}C, δ^{15}N, total C, total N, and the ratio of C to N, they determined

that intermediate lake depths offered the best conditions for maximum phytobenthic production. In this study, our objective is to replicate the study by Gushulak *et al.* (2021) while also accounting for the variation between different locations of the lake.

Hypotheses

Since light availability and disturbance due to currents and mixis affect phytobenthic production, we hypothesize that productivity will be highest at intermediate lake depths, where disturbance is low but the lake is shallow enough to allow large quantities of light to penetrate to the benthic region.

Design Plan

Study type

The proposed study is an observational study. Surface layer soil samples will be collected from various transects within the bed of Gall Lake. Isotopic signatures (δ^{13}C, δ^{15}N) and elemental composition (%C, %N, C:N) will be measured in each sample. For further details see the sampling plan and description of variables below.

Blinding

Personnel who collect samples will be aware of the location, and depth the samples were collected at. Each sample will be labeled with a non-descriptive code to remove all information regarding the samples' locations and depths.

Study design

Superficial soil samples (0-1 cm of depth) will be collected at 1-meter depth intervals (up to 16 m) in Gall Lake, Ontario, Canada, following four orthogonal transects (North, East, South, West). Four independent replicates will be performed along each transect, for a total of $16 \times 4 \times 4 = 256$ soil samples.

Randomization

No randomization will be present in the study. To avoid anthropogenic sources of variation, all samples will be taken by a single individual on the same week. Samples will also be analyzed by a single technician in a single mass spectrometer.

Sampling Plan

Existing data

Registration prior to creation of data. As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized. All data presented here was simulated based on the data from Gushulak *et al.* (2021).

Explanation of existing data

Existing data will not be used in our replication study; all data will be sampled after submitting the preregistration.

Data collection procedures

Data will be collected at the Gall Lake study site (50°11′′ N, 90°42′′ W) located in Ontario, Canada. This undisturbed study site is surrounded by boreal forest with a high abundance of black spruce, jack pine, and poplar and a lower abundance of birch, balsam fir, and larch (Kingsbury, Laird & Cumming, 2012). Data collection will take place in a single week during the summer of 2022.

Surface sediment samples (0-1cm) will be collected using a mini-Glew gravity coring apparatus (Glew, 1991) as was done in Gushulak *et al.* (2021) at sequential depht intervals, determined using depth sounders (Gushulak *et al.*, 2021). The surface sediment sampling will be along three depth transects in each of the basins of Gall Lake, to expand upon the area covered in Gushulak *et al.* (2021) and account for any differences in deposition to the surface sediments due to currents, input, and mixing in Gall Lake. This expansion of coverage will allow for spatial variation among basins, as well as for variation within basins, to be considered in the geochemical analysis of the replication study.

The stable isotope analysis from the surface sediment samples will be conducted using standard methods (Savage, Leavitt & Elmgren, 2004; Bunting *et al.*, 2010) using Thermo Finnigan Delta V isotope ratio mass spectrometer that is equipped with a ConFlow IV dilution inlet system, to ensure consistency with the methods of Gushulak *et al.* (2021).

Sample size

There will be 4 surface sediment samples taken at each of the 16 depth intervals (1-2m, 2-3m, 3-4m, etc.) along the depth transects in each basin of Gall Lake. This

will result in a sample size N=256 for our replication study.

Sample size rationale

We will be working in a Bayesian framework, such that a power analysis is not applicable. Our sample size was therefore determined due to contraints in time and budget.

Stopping rule

We will sample each 1m depth interval (e.g. 1-2m, 2-3m, 3-4m, etc.) until we reach the deepest point of Gall Lake (~17m; (Kingsbury, Laird & Cumming, 2012)).

Variables

Manipulated variables

Depth

For phytobenthos analyses, we will manipulate the depth that surface sediment samples are collected. Four sediment samples will be collected for every 1-m depth interval of Gall Lake. Samples for this categorical variable will range from a 1-m to 16-m depth and will be determined with the use of a depth sounder (Gushulak *et al.*, 2021). In R scripts, this manipulated variable will be named `depth_m`.

Measured variables

Stable Isotopic Analysis

The single outcome variables for the stable isotopic ratio of nitrogen and carbon will be measured from freeze-dried sediment subsamples of Gall Lake. Samples will be placed in a Thermo Finnigan Delta V isotope ratio mass spectrometer that has a ConFlow IV dilution inlet system as described by Gushulak *et al.* (2021) Following the calibration procedure for laboratory standards explained in Bunting *et al.* (2010) and Savage, Leavitt & Elmgren (2004) , the isotope values will be analyzed with the use of atmospheric gas. The standard notation for the stable isotopic ratio of nitrogen and carbon are δ^{15}N and δ^{13}C, respectively. In R scripts,

these single outcome variables will be named `d15n` and `d13c` for nitrogen and carbon, respectively.

Indices

Nitrogen and Carbon Content

Using generalized additive models, we will be determining the percent content of nitrogen and carbon in the depth intervals of Gall lake by manipulating the stable isotope content of these elements. Specifically, gamma distributions as described by Mushet *et al.* (2020) will be used to determine the percent content. In R scripts, the manipulated variables will be named `perc_n` and `perc_c` for nitrogen and carbon, respectively.

equation?

Analysis Plan

Statistical models

Similarly to Gushulak *et al.* (2021), the isotope data was analyzed using Generalized Additive Models (GAMs) via the `mgcv` package (Wood, 2011, 2017). The amount of ^{15}N and ^{1}3C (`d15n` and `d13c` in the `R` scripts, respectively) were modeled using a Gaussian conditional distribution with an *identity* link function since both parameters take any real value (i.e. they can be both positive and negative). The percentages of nitrogen and carbon in the samples (`perc_n` and `perc_c`, respectively) were converted to proportions (`frac_n` and `frac_c`) so they coould be modeled with a beta distribution with a *logit* link function. (There is no distribution in the `mgcv` package for numbers between 0 and 100). Finally, the proportion of carbon to nitrogen (C:N, `c_n_ratio` in the scripts) were modeled using a GAM with a gamma conditional distribution and a *log* link function. A gamma distribution was most appropriate since the ratio of two positive (non-zero) numbers is strictly greater than zero. The mean effects and their Bayesian credible intervals (CIs) were estimated on the link scale using normal approximation ($\pm$1.96 standard deviations) and back-transformed to response values using the inverse-link function. For example, the mean C:N ratios and their CIs was estimated on the *log* scale and

then back-transformed to ratios by exponentiating the estimate. (See the following section for more information on GAMs and transformations.)

All five models accounted for a shared global trend between replicates and trends within-replicate. For simplicity, the smoothness parameter was assumed to be the same between replicates (see model "GS" in Pedersen *et al.*, 2019).

```
# stable nitrogen isotope
m_d15n <- gam(d15n ~
                s(depth_m, k = 15) + # average effect of depth
                s(depth_m, replicate, k = 10, bs = 'fs'), # effect of replicate
              family = gaussian(link = 'identity'), # I(k) = k
              data = isotopes,
              method = 'REML') # optimiziation method for the smoothness parameter

# stable carbon isotope
m_d13c <- gam(d13c ~
                s(depth_m, k = 15) +
                s(depth_m, replicate, k = 10, bs = 'fs'),
              family = gaussian(link = 'identity'),
              data = isotopes,
              method = 'REML')

# fraction of carbon in the samples
m_frac_c <- gam(frac_c ~
                  s(depth_m, k = 15) +
                  s(depth_m, replicate, k = 10, bs = 'fs'),
                family = betar(link = 'logit'), # logit(0) = -Inf, logit(1) = Inf
                data = isotopes,
                method = 'REML')

# fraction of nitrogen in the samples
m_frac_n <- gam(frac_n ~
                  s(depth_m, k = 15) +
                  s(depth_m, replicate, k = 10, bs = 'fs'),
```

```
                family = betar(link = 'logit'),
                data = isotopes,
                method = 'REML')

#C:N ratio in the samples
m_c_n <- gam(c_n_ratio ~
               s(depth_m, k = 15) +
               s(depth_m, replicate, k = 10, bs = 'fs'),
             family = Gamma(link = 'log'), # log(0) = -Inf, log(Inf) = Inf
             data = isotopes,
             method = 'REML')
```

Thus, each model had two predictors: a predictor which accounted for the mean effect of lake depth (`s(depth_m, k = 15)`) and one which used factor smooths (`bs = 'fs'`) for each replicate. Both of the predictors used thin plate regression splines (the default for the `s()` function in `mgcv`). The number of knots (`k`) for each smooth was allowed to be reasonably high, since the models were fit via penalized maximum likelihood and thus over-fitting the data was unlikely (Wood, 2011; Simpson, 2018). The smoothness parameters were optimized using Restricted Marginal Likelihood (`method = 'REML'`), rather than the default Generalized Cross Validation (GCV), since REML does not over-fit as often as GCV (Reiss & Ogden, 2009; Wood, 2011).

Although credible intervals were estimated, this study was performed with a purely Bayesian approach, so there was no interest in statistical significance and Frequentist null-hypothesis testing. Rather, the models were used to compare similarities between our estimates and the estimates by Gushulak *et al.* (2021).

Transformations

The only data transformation that was performed was the conversion of percent carbon and nitrogen (between 0% and 100%) to proportions (between 0 and 1). Note that since the transformation $a = \frac{b}{100}$ is a linear transformation, Jensen's inequality does not apply here (Jensen, 1906).

Modeling the data with GAMs removes the need for transforming data which violates the normality assumptions which linear models depend on: GAMs estimate transformed mean responses ($g[\mathbb{E}(Y)]$, where $g(\cdot)$ is the link function) rather than

the mean transformed response ($\mathbb{E}[g(Y)]$), so Jensen's inequality (Jensen, 1906) does not apply here. ## Inference criteria

Data exclusion	Enter your response here.
Missing data	Enter your response here.
Exploratory analyses (optional)	N/A

Other

Other (Optional)	Enter your response here.

References

Bunting L., Leavitt P.R., Weidman R.P. & Vinebrookeb R.D. (2010). Regulation of the nitrogen biogeochemistry of mountain lakes by subsidies of terrestrial dissolved organic matter and the implications for climate studies. *Limnology and Oceanography* **55**, 333–345. https://doi.org/10.4319/lo.2010.55.1.0333

Glew JohnR. (1991). Miniature gravity corer for recovering short sediment cores. *Journal of Paleolimnology* **5**. https://doi.org/10.1007/BF00200351

Gushulak C.A.C., Haig H.A., Kingsbury M.V., Wissel B., Cumming B.F. & Leavitt P.R. (2021). Effects of spatial variation in benthic phototrophs along a depth gradient on assessments of whole-lake processes. *Freshwater Biology*, fwb.13820. https://doi.org/10.1111/fwb.13820

Jensen J.L.W.V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* **30**, 175–193. https://doi.org/10.1007/BF02418571

Kingsbury M.V., Laird K.R. & Cumming B.F. (2012). Consistent patterns in diatom assemblages and diversity measures across water-depth gradients from eight Boreal lakes from north-western Ontario (Canada): Diatom assemblages and diversity in relation to lake depth in boreal lakes. *Freshwater Biology* **57**, 1151–1165. https://doi.org/10.1111/j.1365-2427.2012.02781.x

Mushet G.R., Laird K.R., Leavitt P.R., Maricle S., Klassen A. & Cumming B.F. (2020). Bottom-Up Forces Drive Increases in the Abundance of Large Daphnids in Four Small Lakes Stocked with Rainbow Trout (Oncorhynchus mykiss), Interior British Columbia, Canada. *Ecosystems* **23**, 873–890. https://doi.org/10.1007/s10021-019-00443-0

Pedersen E.J., Miller D.L., Simpson G.L. & Ross N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ* **7**, e6876. https://doi.org/10.7717/peerj.6876

Reiss P.T. & Ogden T.R. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 505–523. https://doi.org/10.1111/j.1467-9868.2008.00695.x

Savage C., Leavitt P.R. & Elmgren R. (2004). Distribution and retention of effluent nitrogen in surface sediments of a coastal bay. *Limnology and Oceanography* **49**, 1503–1511. https://doi.org/10.4319/lo.2004.49.5.1503

Simpson G.L. (2018). Modelling Palaeoecological Time Series Using Generalised Additive Models. *Frontiers in Ecology and Evolution* **6**, 149. https://doi.org/10.3389/fevo.2018.00149

Wood S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* **73**, 3–36

Wood S.N. (2017). *Generalized Additive Models: An Introduction with R*, 2nd edn. Chapman; Hall/CRC.