**Supplementary Information for “Temperature, resources, and predation interact to shape phytoplankton size-abundance relationships at a continental scale”**

**Appendix S1 Methodological details**

*Justification of modelling choices and exploration of alternatives*

We made choices about the interactions to include in the model, the proxy of the resource availability and other issues about the model, see below:

* This model does not include all possible interactions. We intentionally excluded one possible 2-way interaction (*P* x *R*), two possible 3-way interactions (*P* x *R* x *S, S* x *T* x *R*). We restricted the terms in our model based on the hypotheses described earlier, in part to meaningfully evaluate them with a relatively small dataset of ~1000 points. All results presented are based on this restricted model (Supporting Information Appendix S1, Table S1).
* We used total phosphorous as a proxy for resource availability, this is because phosphorus is more commonly the growth-limiting nutrient in lake ecosystems (Pomati, Shurin, Andersen, Tellenbach, & Barton, 2020).
* There is some uncertainty associated with estimates of the mean phytoplankton volume in each lake. We have ignored this uncertainty in the models presented, but we subsequently explored weighting points by 1/SE of the mean volume estimate. This weighting did not alter any conclusions and so we present the simpler models here.
* Because some phytoplankton species are colonial, our quantification of abundance and volume at the cell level arguably misses important aspects of phytoplankton ecology. We therefore also analyzed our data at the colony level and found similar patterns (Supporting Information Figure S2).

*Data exclusion rules*

We excluded some data to avoid bias in our size-abundance relationships. After these exclusion steps, we were left with data from 1048 lakes. The exclusion rules are reported below:

1. A small number of lakes were visited and measured more than once. To avoid complications associated with non-independence of data points, we included data from only the first visit to each lake.
2. For three lakes, *in situ* temperatures of >60 °C were recorded. We treated these as errors and excluded these temperature values. For reference, the next highest temperature recorded was 35.5 °C.
3. We also excluded zooplankton species that were recorded as being parasitic or carnivorous from our analyses, with the goal of capturing the herbivorous zooplankton. Not all species had an ecological descriptor, so this step only excluded species for which information was available in the dataset.
4. Data from 85 lakes were excluded from the model fitting process because of the absence of measurements of at least one of our study variables.

*Prior Choice*

We set the prior for the main slope of biovolume and density to *Normal*(-0.875, 2), which corresponds to the midpoint between -0.75 and -1, two commonly reported slope values. For the intercept, we set a prior of *Normal*(4, 1), which assumes mean phytoplankton densities ranging from ~100 to 1,000,000 cells per ml (i.e., 10^(4-2) to 10^(4+2)), consistent with previous studies in NLA lakes (Leech, Pollard, Labou, & Hampton, 2018). All other priors were centered at zero with *Normal*(0, 1), except for sigma, which was assigned *Exponential*(1). While these priors were informative based on previous work, they had little effect on the outcomes as confirmed by a prior sensitivity analysis (Figure S1.1).

Table S1. Parameter estimates +/- 95% CrI of the relationship between cell volume (S) and population density (N) of the Bayesian analysis. Fig. 3 is based on this model and does not incorporate the effects of temperature (T), total phosphorous (R), and zooplankton (P). Figure 4, 5, 6 and 7 are based on this model and incorporate the effects of temperature (T), total phosphorous (R), and zooplankton (P). We restricted the model to these terms based on a priori hypotheses and performed no model selection steps thereafter.

|  |  |  |
| --- | --- | --- |
|  | **Total population density** |  |
| *Predictors\** | *Estimates\** | *95% CI\** |
| **(Intercept)\***  **S** | 4.50 (0.02)  -0.97 (0.02) | 4.46 to 4.54  -1.03 to -0.91 |

|  |  |  |
| --- | --- | --- |
|  | **Total population density** |  |
| *Predictors\*\** | *Estimates\*\** | *95% CI\*\** |
| **(Intercept)** | 4.49 (0.02) | 4.46 to 4.53 |
| **S**  **T**  **R**  **P**  **S \* T**  **S \* R**  **S \* P**  **T \* P**  **S \* P \* R**  **S \* T \* P**  **S\* T \* R \* P** | -0.84 (0.03)  0.12 (0.02)  0.7 (0.04)  0.11(0.03)  -0.08 (0.03)  0 (0.04)  -0.05 (0.05)  -0.06 (0.04)  0.01(0.05)  0.01 (0.04)  0.08 (0.04) | -0.90 to -0.79  0.08 to 0.16  0.63 to 0.77  0.05 to 0.18  -0.14 to -0.03  -0.07 to 0.07  -0.14 to 0.04  -0.14 to 0.03  -0.09 to 0.11  -0.07 to 0.08  0 to 0.17 |

\* Model without the effect of T, R and P; \*\*Model that incorporates the effect of T, R and P.

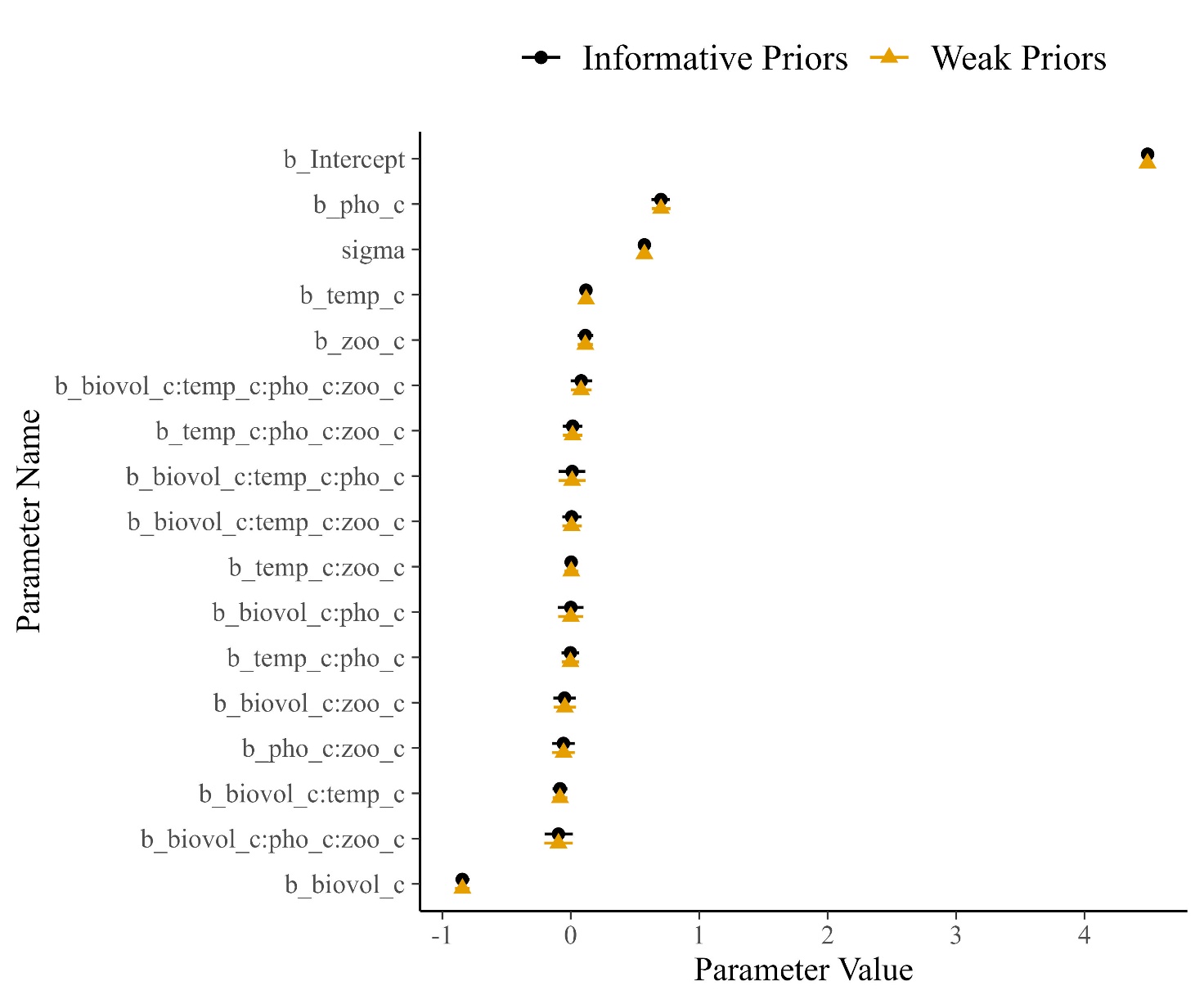


Fig. S1.1. Prior sensitivity check showing model parameters estimated with informative priors (as described in the text) versus parameters estimated with weak priors (by doubling the standard deviations in the prior). Similarity of the points indicates that the inference comes almost entirely from the data, not the priors.

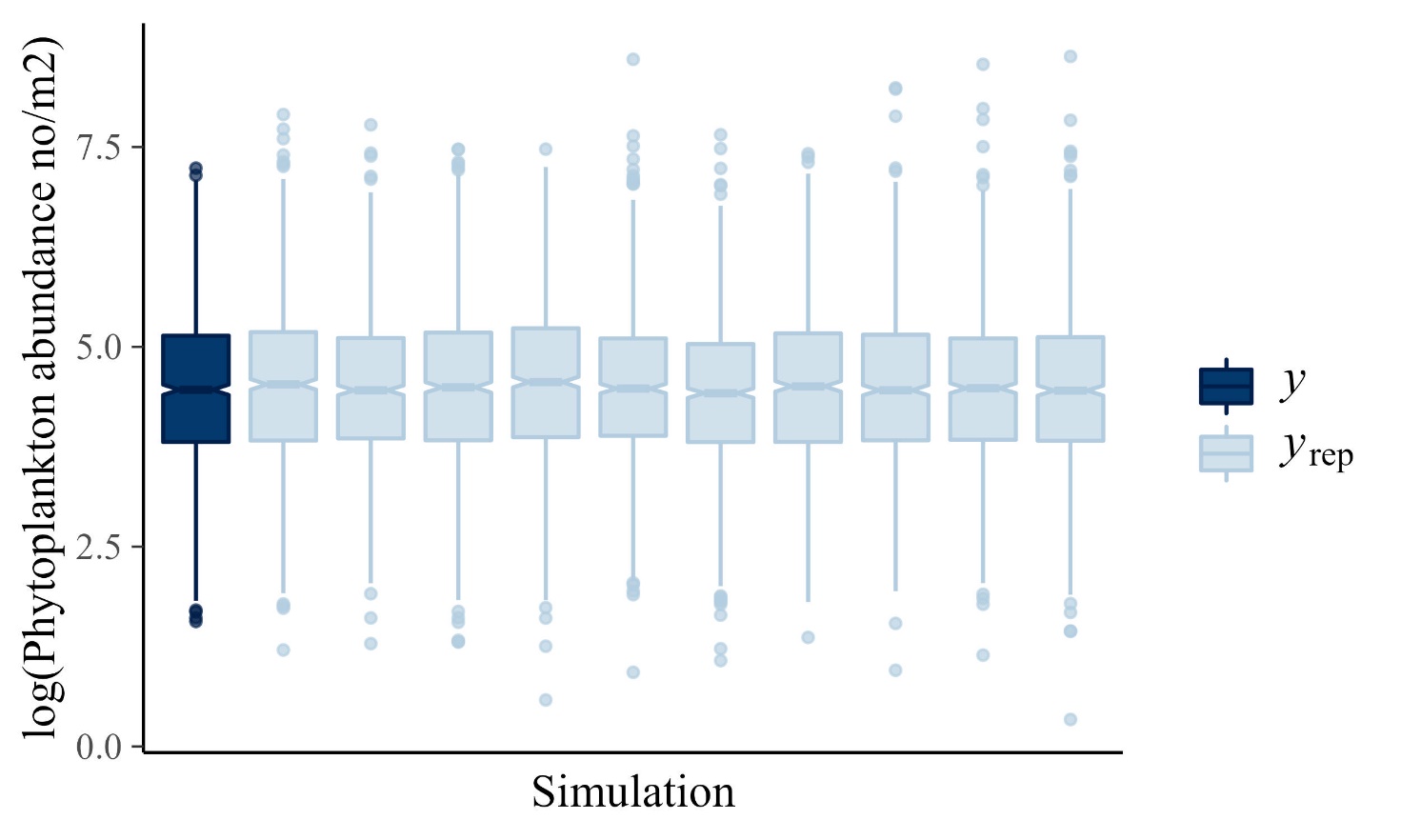


Fig. S1.2. Posterior predictive check comparing the raw data (*y*) to 10 simulated datasets from the posterior distributions (*y*rep). Similarity in the boxplots indicates that the model generates data that are similar to the original data set, which indicates good model fit (Gabry et al. 2017).

**Suppelementary References**

Arel-Bundock, V. (2022). modelsummary: Data and model summaries in R. *Journal of Statistical Software,* 103(1), 1–23.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2017). Visualization in Bayesian workflow. *arXiv preprint arXiv:1709.01449*.

Jeffrey B. A. (2021). Ggthemes: Extra Themes, Scales and Geoms for ‘ggplot2’. R package version 4.2.4. <https://CRAN.R-project.org/package=ggthemes>

Leech, D. M., A. I. Pollard, S. G. Labou, and S. E. Hampton. 2018. Fewer blue lakes and more murky lakes across the continental US: Implications for planktonic food webs. *Limnology and Oceanography*, 63(6), 2661-2680.

Long, J. A (2022). Interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.6, <https://cran.r-project.org/package=interactions>.

Lüdecke D (2022). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.11, [https://CRAN.R-project.org/package=sjPlot](https://cran.r-project.org/package=sjPlot).

Lüdecke, D. (2018a). sjmisc: Data and variable transformation functions. *Journal of Open Source Software*, 3, 754.

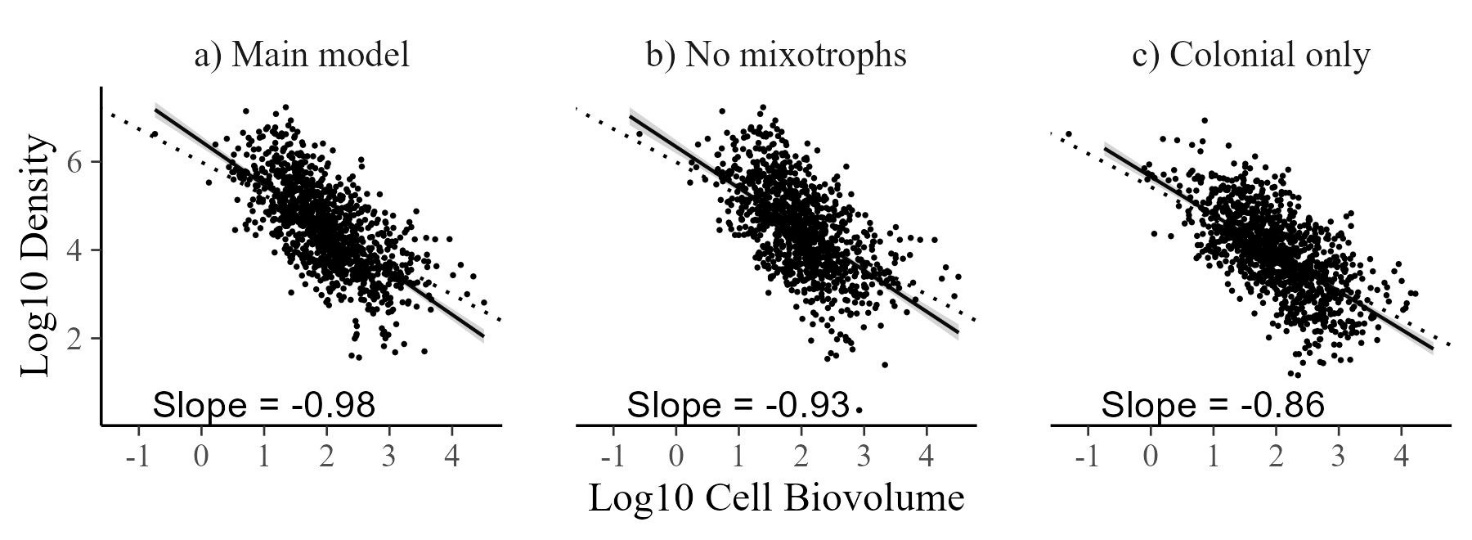
Lüdecke, D. (2018b). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3, 772.

Pomati, F., Shurin, J. B., Andersen, K. H., Tellenbach, C., & Barton, A. D. (2020). Interacting temperature, nutrients and zooplankton grazing control phytoplankton size-abundance relationships in eight Swiss lakes. *Frontiers in Microbiology,* 10, 3155.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Wickham, H., François, R., Henry, L., & Müller, K. (2022). dplyr: A Grammar of Data Manipulation. <https://dplyr.tidyverse.org>.

**Appendix S2 Comparing models without mixotrophs and with only colonial organisms**

****

Figures S2.1. Model outcomes are not dependent on the presence of mixotrophs or colonial organisms. a) Result from the main text, b) same model but with mixotrophic taxa removed, and c) same model but with only colonial taxa. In each case, the slope does not include the predicted value of -0.75 (see Table S2 for full parameter estimates and credible intervals).

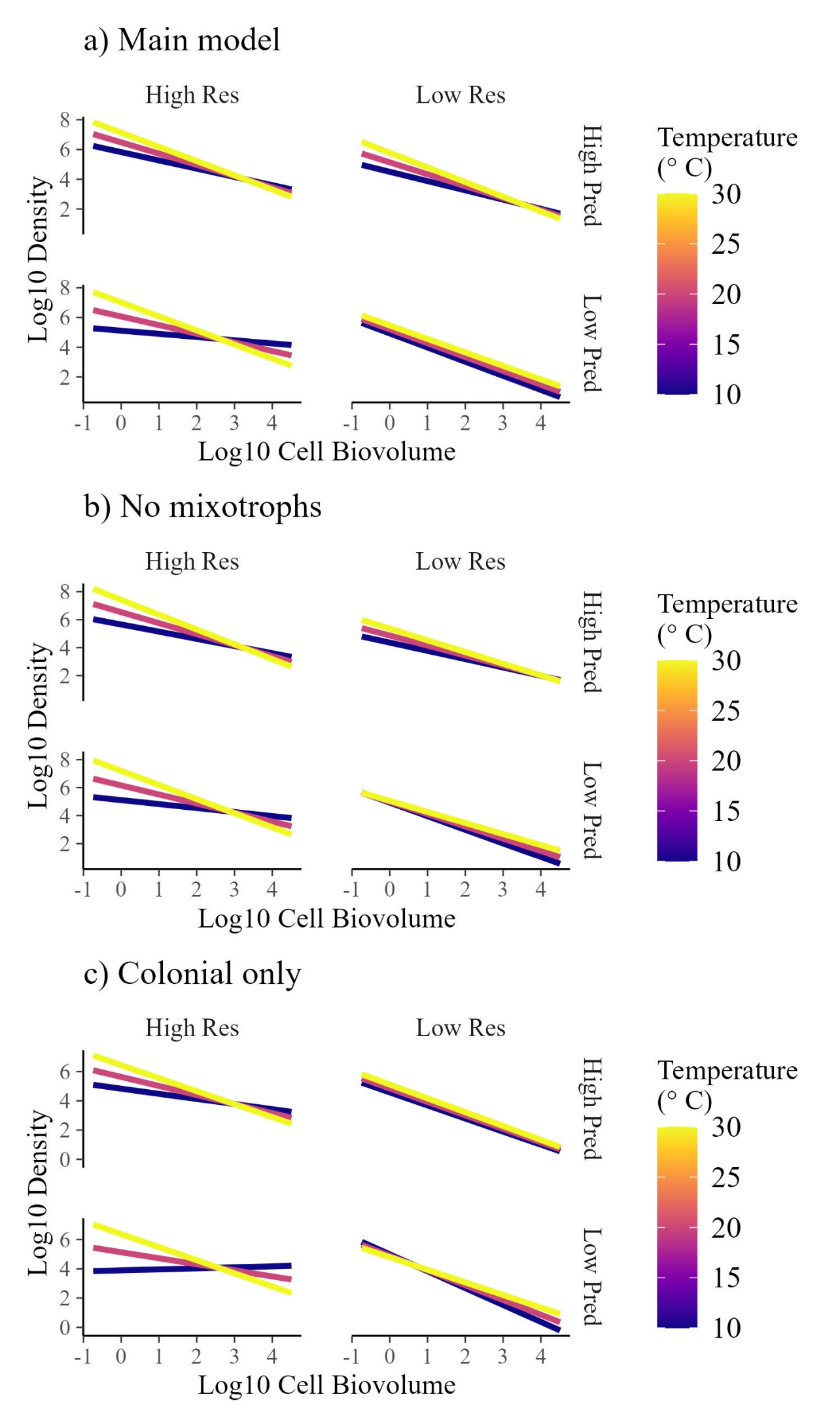
****

Fig. S2. Temperature (T), total phosphorous (P), and zooplankton predation (Z) interact to alter size-abundance relationships across 1048 lake phytoplankton communities. The fitted model is based on a) the data in the main text, b) data without mixotrophs, or c) colony level size and density, rather than cell size and density. These ares multiple regression model predicting total log10 population density (cells per mL) with mean log10 colony volume (μm3), temperature (°C) and log10 nitrate concentration (mg per L), as well as interactions between those predicted. Although all predictors are treated as continuous in the model, we ‘slice’ the regression surface at 3 zooplankton biomass levels (50 μg L-1, 500 μg L-1 and 5000 μg L-1), 2 total phosphorous concentrations (3 mg L-1, and 300 mg L-1), 2 temperatures (10°, and 30°C) to visualize changes in slope driven by these predictors. See Table S2 for model parameter values.

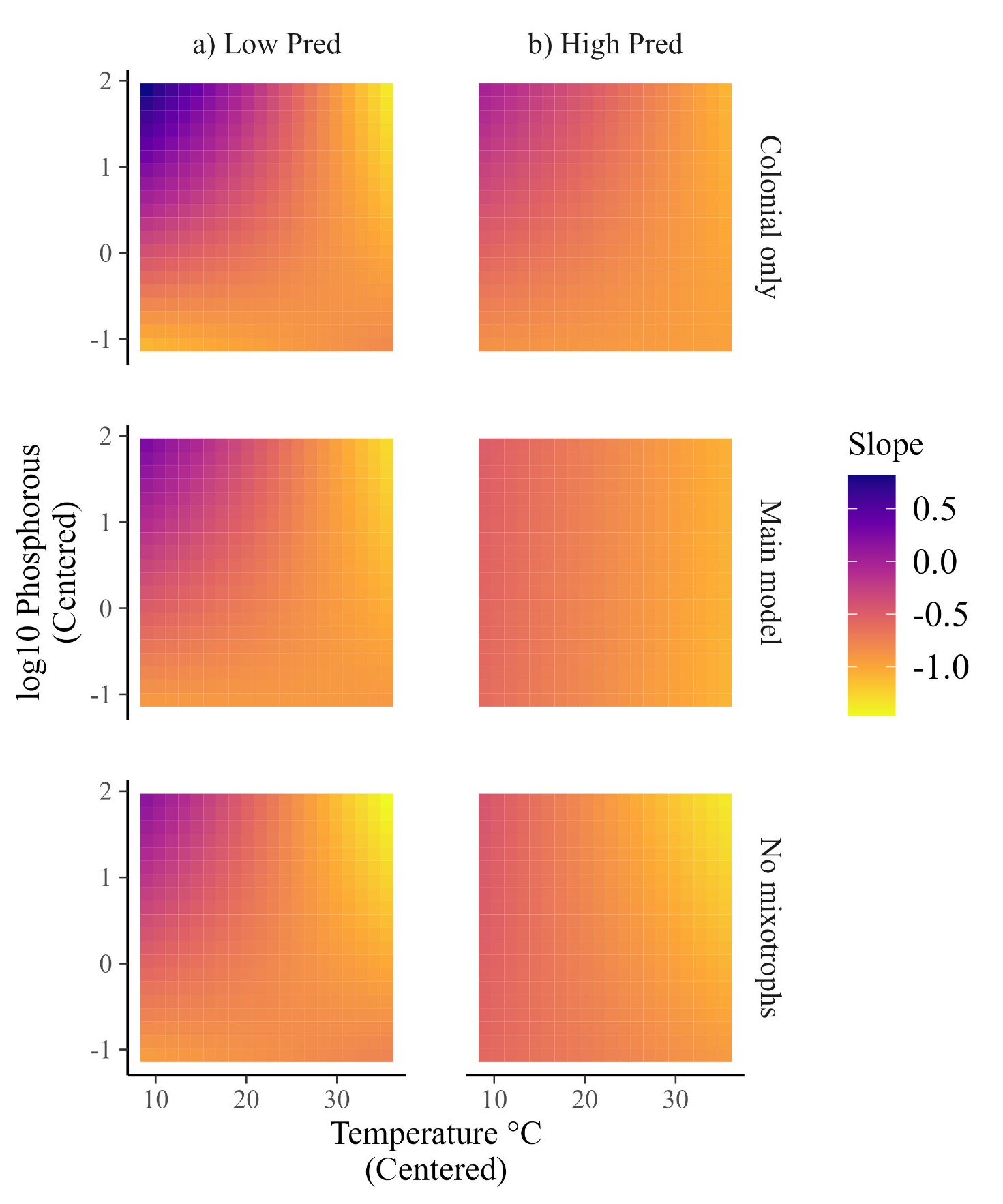


Fig. S3. Interaction plot showing the predicted slope values given variable predicator values using three different subsets of data. “Colonial only” are data with only colonial taxa, “Main model” are the same results as in the main text, and “No mixotrophs” are data without mixotrophic taxa. Slope values are the posterior means.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table S2. Parameter estimates +/- 95% CrI of the relationship between cell volume (S) and population density given different data groupings. "Main model" is the same model in the main text. "Colonial taxa only" includes only colonial phytoplankton, and "No mixotrophs" includes only non-mixotrophic taxa. The similarity in the estimates indicates minimal differences between models. | | | | |
|  |  | Main model | Colonial taxa only | No mixotrophs |
| Univariate | **(Intercept)** | 4.5 (4.5 to 4.5) | 3.9 (3.9 to 4) | 4.5 (4.4 to 4.5) |
|  | **S** | -1 (-1 to -0.9) | -0.9 (-0.9 to -0.8) | -0.9 (-1 to -0.9) |
|  |  |  |  |  |
| Multivariate | **(Intercept)** | 4.5 (4.5 to 4.5) | 3.9 (3.9 to 4) | 4.5 (4.4 to 4.5) |
|  | **S** | -0.8 (-0.9 to -0.8) | -0.8 (-0.9 to -0.8) | -0.8 (-0.9 to -0.8) |
|  | **T** | 0.1 (0.1 to 0.2) | 0.1 (0.1 to 0.1) | 0.1 (0.1 to 0.2) |
|  | **R** | 0.7 (0.6 to 0.8) | 0.7 (0.6 to 0.8) | 0.8 (0.7 to 0.8) |
|  | **P** | 0.1 (0.1 to 0.2) | 0.1 (0 to 0.2) | 0.1 (0.1 to 0.2) |
|  | **S\*T** | -0.1 (-0.1 to 0) | -0.1 (-0.1 to 0) | -0.1 (-0.2 to 0) |
|  | **S\*R** | 0 (-0.1 to 0.1) | 0.1 (0 to 0.2) | -0.1 (-0.2 to 0) |
|  | **T\*R** | 0 (-0.1 to 0.1) | 0 (-0.1 to 0.1) | 0 (-0.1 to 0.1) |
|  | **S\*P** | 0 (-0.1 to 0) | -0.1 (-0.2 to 0) | 0 (-0.1 to 0.1) |
|  | **T\*P** | 0 (0 to 0.1) | 0 (0 to 0.1) | 0 (-0.1 to 0.1) |
|  | **R\*P** | -0.1 (-0.1 to 0) | 0 (-0.1 to 0.1) | -0.1 (-0.1 to 0) |
|  | **S\*T\*P** | 0 (-0.1 to 0.1) | 0 (0 to 0.1) | 0 (-0.1 to 0.1) |
|  | **S\*T\*P** | 0 (-0.1 to 0.1) | 0 (-0.1 to 0.1) | 0 (-0.1 to 0.1) |
|  | **S\*R\*P** | -0.1 (-0.2 to 0) | -0.1 (-0.2 to 0) | -0.1 (-0.2 to 0) |
|  | **T\*R\*P** | 0 (-0.1 to 0.1) | 0 (-0.1 to 0.1) | 0 (-0.1 to 0.1) |
|  | **S\*T\*R\*P** | 0.1 (0 to 0.2) | 0.1 (0 to 0.2) | 0.1 (0 to 0.2) |