

Causal Inference Framework for Ocean Microbial Community Responses to Warmer Temperature

(Minh) Viet Tran ^{1,2,3}, Christian L. Müller ^{1,2,3,4}

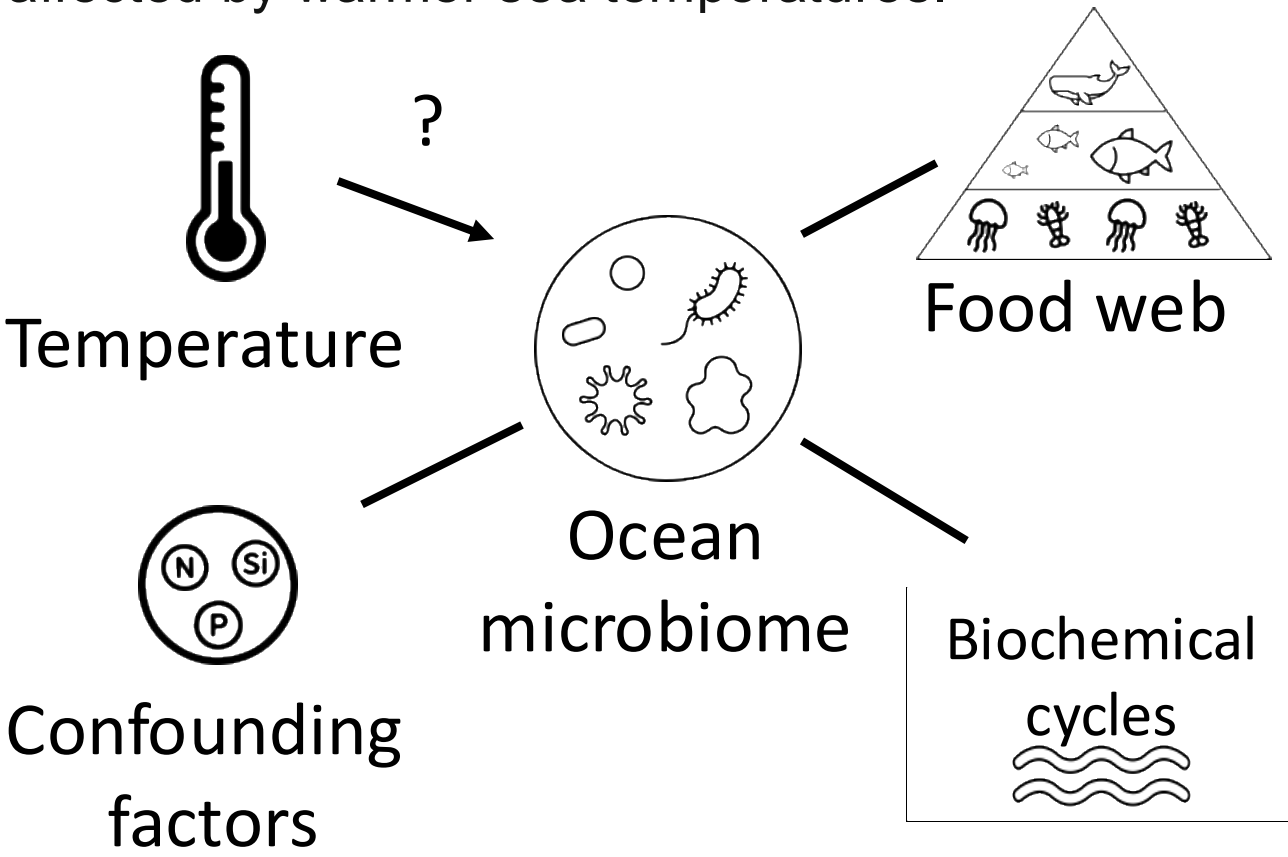
¹ Helmholtz Munich, Germany
² Ludwig-Maximilians-Universität München, Germany
³ Munich Center for Machine Learning (MCML), Germany
⁴ Flatiron Institute, USA

Abstract

Understanding how ocean temperature influences microbial communities is critical for forecasting ecological responses to climate change. This project outlines a framework for applying the Rubin Causal Model (RCM) to observational microbiome sequencing data using matching methods. The aim is to estimate the causal effect of temperature on microbial taxa, while controlling for confounding environmental variables.

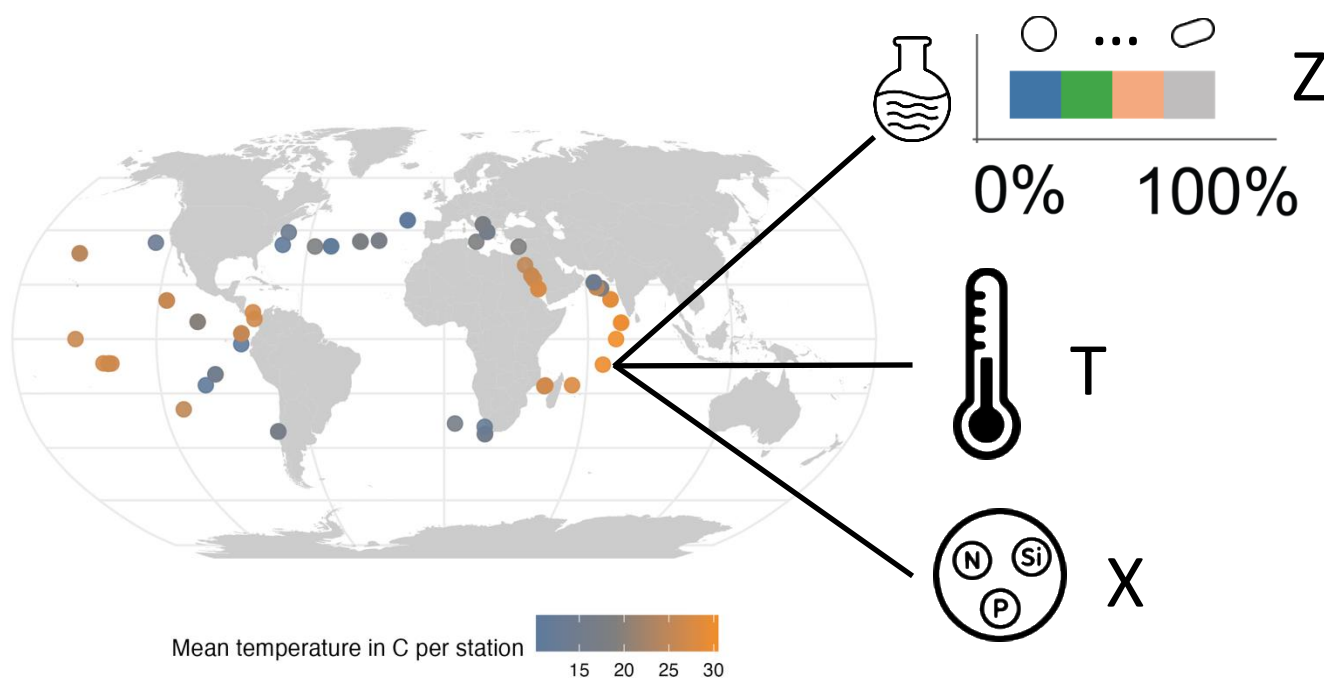
Motivation

The ocean microbiome is critical for the marine food web, biochemical cycles, and climate regulation. This study investigates how it is affected by warmer sea temperatures.



Dataset

Data: Tara Oceans sequencing data [1]
Treatment (T): High (>25°C) vs low (<19°C) temperature
Covariates (X): Salinity, nutrient concentration, water column depth



Outcomes of interest

Let Z be the compositional data matrix with n rows and p taxa. We are interest in two types of outcomes:

- Community summary statistics:
 - Observed richness:
$$\sum_{j=1}^p \mathbb{I}(Z_{ij} > 0)$$
 - Shannon diversity:
$$-\sum_{j=1}^p Z_{ij} \ln(Z_{ij})$$
- Individual taxa: For each taxa Z_{ij} we applied LinDA [2], a robust regression model to test for a significant association with the treatment variable while controlling for multiple testing.

Rubin causal model (RCM)

RCM [3] frames causal inference as a "missing data" problem: We can only observe one of the potential outcomes.

Unit i	$Y_i(1)$	$Y_i(0)$	Individual causal effect τ
1	100	?	?
2	?	100	?
3	1000	?	?
4	?	300	?

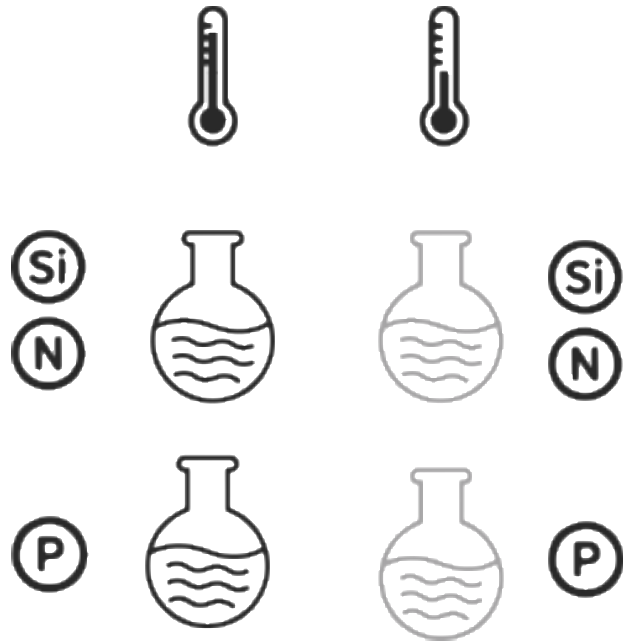
Instead of estimating the individual causal effect we rely on estimating the average treatment effect on the treated (ATT):

$$ATT = E(Y_i(1) - Y_i(0) | T_i = 1).$$

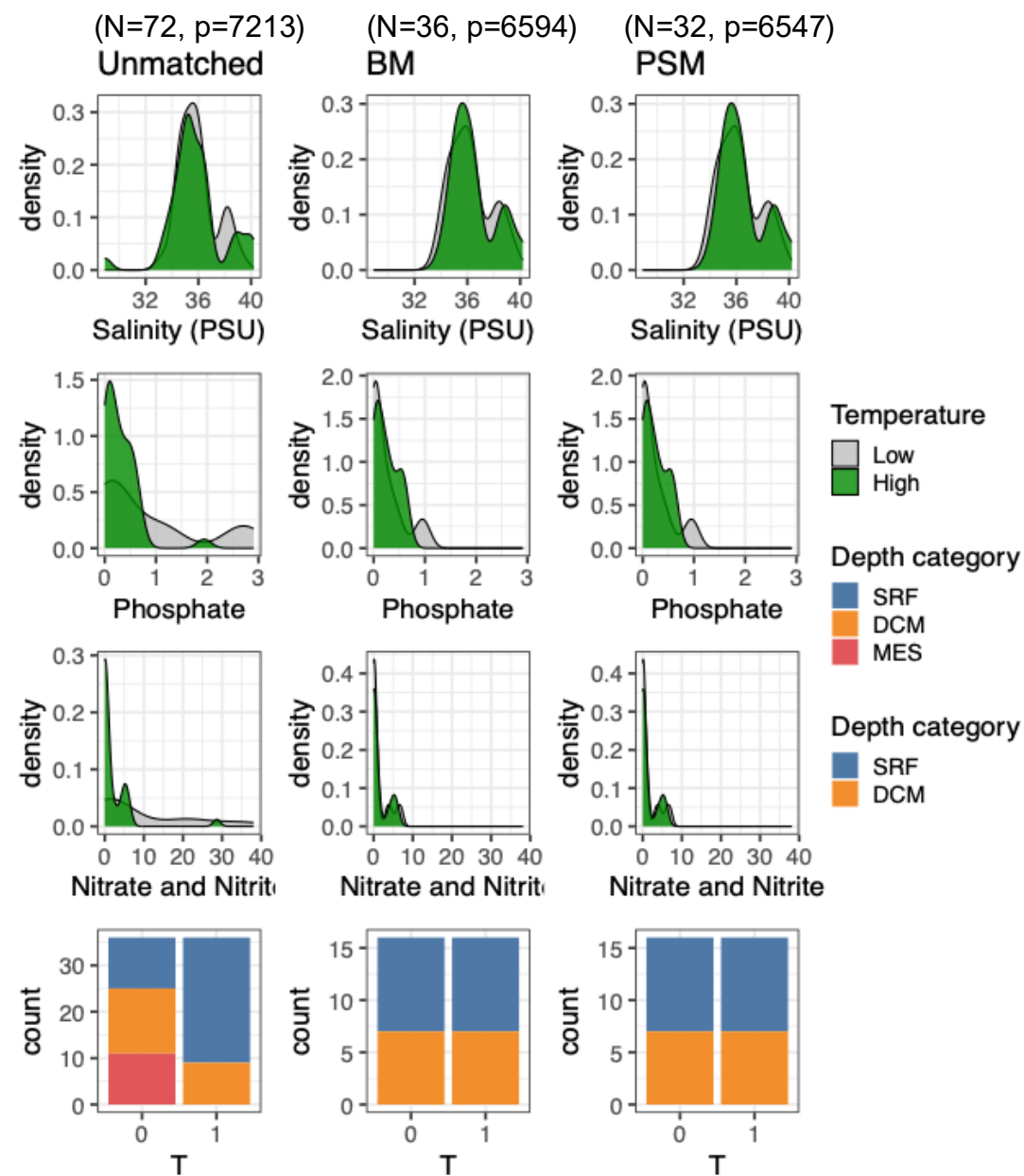
Matching strategies

We use two different matching strategies:

- Propensity score matching (PSM) [3]
Perform 1:1 nearest neighbor matching on PS
- Bipartite matching (BM) [5]:
Finding the maximum # of matched pairs in a bipartite graph



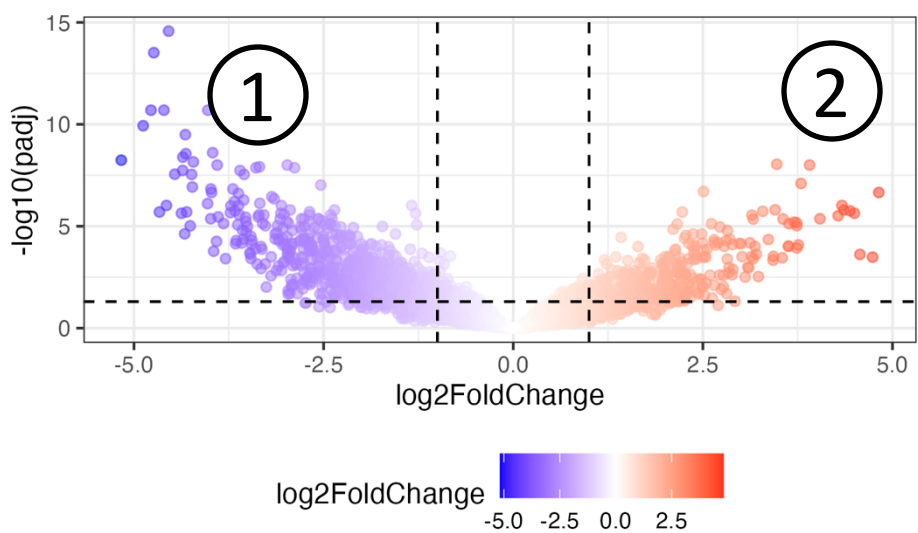
Matching results in better covariate balance



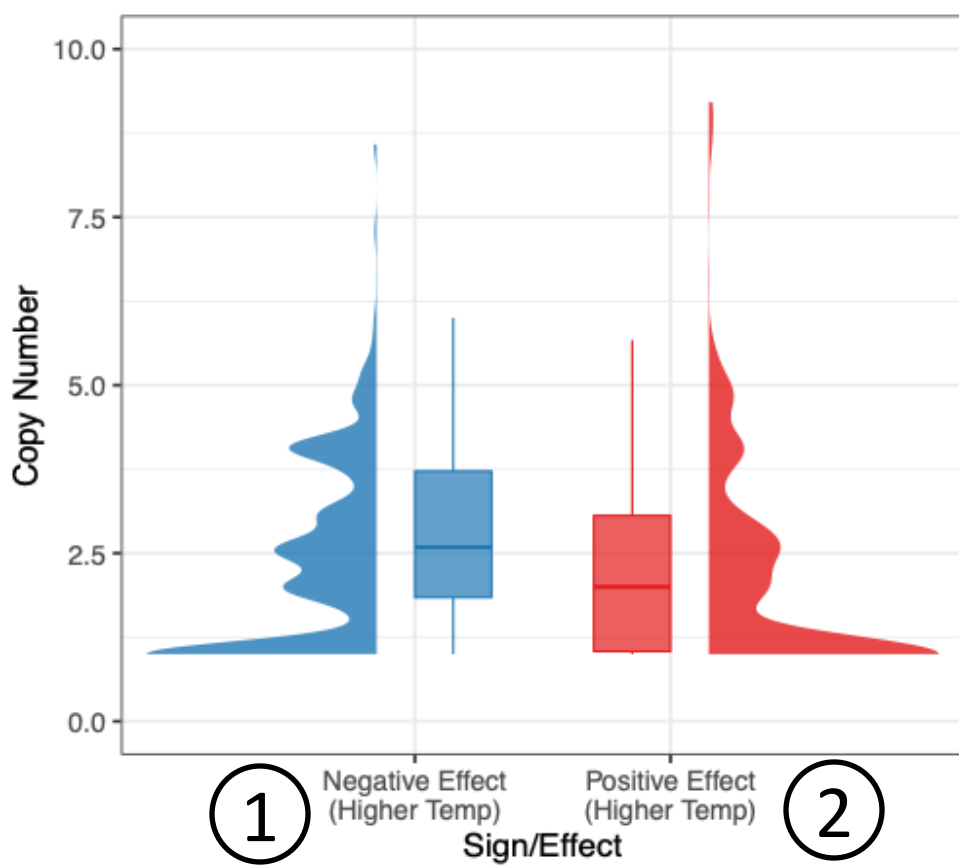
Effect on individual taxa

We observe a higher proportion of negatively affected taxa after matching.

Method	Negative	Positive
Baseline	1718	1546
PSM	① 986	432 ②
BM	613	239



Distribution shift indicates that slow-growing bacteria (low copy number) are favored, confirming [6].



Negative effect on diversity

We observe a consistent decrease of the diversity in terms of observed richness and in the Shannon diversity.

Method	Richness	Shannon
Baseline	-801.9	-0.23
PSM	-426.5	-0.23
BM	-514.3	-0.15

Discussion / outlook

- Extending the idea to continuous treatment variables and a larger dataset
- Conducting a more formal sensitivity analysis

References

[1] Shinichi Sunagawa, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, Georg Zeller, Daniel R Mende, Adriana Alberti, et al. Structure and function of the global ocean microbiome. Science, 348(6237):1261359, 2015.
[2] Huijuan Zhou, Kejun He, Jun Chen, and Xianyang Zhang. Linda: linear models for differential abundance analysis of microbiome compositional data. Genome biology, 23(1):95, 2022.
[3] Guido W Imbens and Donald B Rubin. Causal inference in statistics, social, and biomedical sciences. Cambridge university press, 2015.
[4] Paul R Rosenbaum and Donald B Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39(1):33–38, 1985.
[5] Alice J Sommer, Annette Peters, Martina Rommel, Josef Cyrys, Harald Grallert, Dirk Haller, Christian L Müller, and Marie-Abèle C Bind. A randomization-based causal inference framework for uncovering environmental exposure effects on human gut microbiota. PLoS computational biology, 18(5):e1010044, 2022.
[6] Clare I Abreu, Martina Dal Bello, Carina Bunse, Jarone Pinhassi, and Jeff Gore. Warmer temperatures favor slower-growing bacteria in natural marine communities. Science Advances, 9(19):eade8352, 2023.