Nathan Greenslit Data Analysis Project

Nathan Greenslit

4/17/23

# Edits

1. I would do a lot of visual inspections to see how the different models fit and what they predict. You can try a linear model and a more complex ML one and compare to see if maybe something beyond a linear relationship improves results.
2. Given that you don’t have a lot of data, I’m not sure a PCA makes sense. I think in-depth analysis of the relation between the variables and the outcome and then in addition to the linear models trying a few more complex ones will probably be the better approach.
3. The figures in the manuscript look good, but some are too small to read. Needs breaking up and/or increasing font size

# Summary/Abstract

Each year, plumes of Saharan dust travel across the Atlantic to be deposited in the surface waters of the Caribbean and Gulf of Mexico. Dust aerosols serve as a significant source of nutrients that can elicit a fertilization effect on marine coastal waters, leading to rapid and potentially harmful blooms (e.g. red tides). Previous work has characterized the response of the marine bacteria *Vibrio* to dust input under the oligotrophic settings of the Florida Keys, but less is known regarding settings with higher baseline nutrient levels. This project aims to quantify *Vibrio* population dynamics in response to dust input in coastal sites with higher ambient nutrient levels. Daily sampling took place in Corpus Christi, TX to capture before, during, and after a Saharan dust event and quantitative PCR (qPCR) was used to estimate total counts of *Vibrio.* This study will provide an increased understanding of the conditions that can elicit potentially harmful blooms, highlighting the need for further research to understand the effects of dust deposition in non-oligotrophic waters.

# Introduction

## Saharan Dust

The Sahara Desert is a significant source of atmospheric dust, eliciting an estimated one billion metric tons each year Lenes et al. (2008).

## Microbial Blooms

Previous work conducted in the West Florida Shelf has shown that increased iron availability can stimulate the growth of the nitrogen-fixing cyanobacterium *Trichodesmium*. The high amount of nitrogen produced via fixation can be adequate enough to stimulate toxic red tides, caused by the dinoflagellate *Karenia brevis* (Lenes et al. 2008).

### Vibrio

While it is documented that these dust events can stimulate growth of harmful phytoplankton, emerging evidence shows that heterotrophic bacteria are responding to the episodic influx of limiting resources and substrate, resulting in blooms of bacteria that are associated with disease, which we refer to as harmful bacterial blooms. Among those that respond are bacterium belonging to the genus *Vibrio*, a group of ubiquitous marine opportunistic heterotrophs. While considered conditionally rare (comprising of <1% of the microbial community), under the proper conditions, this genus can rapidly bloom to make up a large percentage of the community over a short period of time (Westrich et al. 2016 ; Westrich et al. 2018). *Vibrio* species are found in brackish to marine waters with salinities ranging from 0-35 and will grow in waters with temperatures ranging from 5°C-40°C (Sampaio et al. 2022). Each year, plumes of Saharan dust travel across the Atlantic via easterly trade winds. These plumes are then deposited in the surface waters of the mid-Atlantic, Caribbean, and Gulf of Mexico. These dust aerosols harbor a wide range of fungi, bacteria, virus-like-particles, minerals, and nutrients (NO3, Fe, PO4). Addition of these nutrients to otherwise oligotrophic settings can result in large and rapid blooms of potentially harmful microbes, presenting a danger to both marine and human health.

*Vibrio* species are excellent opportunists, having multiple advantages that allow them to quickly respond to any newly introduced substrate or nutrients. These advantages include multiple copies of rRNA genes, a rapid doubling time, efficient chemotactic motility, large genomic repertoire (consisting of two circular chromosomes), and multiple Fe-siderophore complexes that allow for rapid uptake of iron into the cell (Jensen, Frost, and Torsvik 2009; Ringgaard et al. 2018 ; Eagon 1962 ; Okada et al. 2005 ; Payne, Mey, and Wyckoff 2016).

Within the *Vibrionaceae* family, there are at least 12 species that are commonly known as human pathogens. Illness can be induced through either a foodborne route via the consumption of raw or uncooked seafood and contaminated drinking water, or through nonfoodborne routes such as wound exposure while swimming. The most common illnesses caused by *Vibrio* include self-limiting diarrhea and cholera (*V. cholerae*), shellfish-induced gastroenteritis (*V. parahaemolyticus, V. vulnificus*), extreme cases of necrotizing fasciitis and septicemia (*V. vulnificus*), and wound and ear infections  (*V. alginolyticus*, *V. vulnificus*) Tran et al. (2013) among others.

### Previous Work

Previous work has characterized *Vibrio* blooms in response to Saharan dust input in the oligotrophic setting of the Caribbean and subtropical Atlantic (Barbados and Florida Keys, respectively), with surface water concentrations of total *Vibrio* increasing by five to thirty times that found during non-dust conditions and returning to baseline levels within 24-48 hours. *Vibrio* composition within the larger microbial community also shifted following dust deposition, with initial levels of <1.4% to a peak of 19.8% of the bacterial community (Westrich et al. 2016). A similar phenomenon was also observed in the surface waters of the tropical and open ocean mid-Atlantic with *Vibrio* populations increasing 1.5-fold in the mid-Atlantic following deposition (Westrich et al. 2018) In a follow up study in the Florida Keys, episodic dust events during a daily time series promoted a succession of bacterial responses, with declines in *Prochlorococcus* coinciding with initial increases in bacteria belonging to the order *Vibrionales*, followed by subsequent shifts in response of different bacterial groups (Borchardt et al. 2020). To date, our understanding of *Vibrio* and other microbial response to dust input has focused on oligotrophic settings, where dust input is considered to be a critical source of limiting nutrients. However, the composition of desert dust can be complex and may have the potential to deliver critical resources that could be exploited by opportunistic microbes (like *Vibrio*) and elicit a growth response despite higher baseline nutrient levels. Given that human exposure is likely in these coastal waters, it is important to elucidate the relationship between dust input and microbial blooms in these higher nutrient environments.

## Present Research

This study aims to (1) quantify *Vibrio* population dynamics and composition, and (2) characterize potential microbial blooms and community shifts in response to dust input in coastal sites with higher ambient nutrient levels. The findings from this study will provide an increased understanding of the conditions that can elicit potentially harmful blooms (i.e., *Vibrio*), highlighting the need for further research to understand the effects of dust deposition in non-oligotrophic waters. We predict that (1) *Vibrio* will be among the first to respond to the influx of dust constituents and (2) that the fertilization effects of dust deposition (and possible successional changes) will be dampened at sites with higher baseline nutrient levels, as dust derived nutrients become less critical for growth.

With a heightened risk of *Vibrio* exposure in coastal waters, it is vital to understand microbial dynamics in nearshore environments. If dust is proven to elicit a microbial response, these events can be used to serve as an “early warning” of increased exposure risk to better protect public health.

## Questions/Hypotheses to be addressed

How do Saharan dust events influence Vibrio populations in high nutrient coastal waters?

*What I am looking for* I expect to see a more dramatic growth response in the Gulf (low nutrient), whereas the higher nutrient sites may exhibit a dampened growth response since they already have high baseline nutrient levels. If we still see a growth response despite already having supportive background nutrient levels at these sites, this may suggest that there are other constituents in the dust that can elicit a growth response. Nutrients, dust input, salinity, and temperature will be the primary factors of interest.

*How I will analyze it* - Look at Vibrio growth over time series (estimated from qPCR) - Examine relationships between dust input and Vibrio growth - Examine influence of site-specific environmental parameters (Temperature, Salinity, Nutrients) on growth response - Run linear models on factors like: dust x growth and nutrients x growth - NMDS to see which parameters have the strongest influence on growth

# Methods

## Description of Study Area

Sampling took place at three locations in Corpus Christi, TX: Blind Oso Bay, a residential canal system on Padre Island, and the Gulf. Sites were chosen based on differing baseline nutrient levels. Blind Oso Bay is a shallow tributary that is popular amongst wadefisherman and kayakers. The Bay is often impacted by fluctuating salinities and high levels of nutrients and chlorophyl due to its proximity to a wastewater treatment plant (**wetz?**) Additionally, the Bay has persistent issues with high levels of fecal indicator bacteria, placing it on the U.S. impaired waters list ([Nicolau & Hill, 2014](https://www.tceq.texas.gov/downloads/water-quality/tmdl/oso-bay-oyster-harvesting-assessment-103/103-oso-bay-monitoring-report-2013.pdf), [Texas Commission on Environmental Quality, 2022](https://www.tceq.texas.gov/downloads/water-quality/assessment/integrated-report-2022/2022-303d.pdf)). The residential canal system is also impacted by high nutrients and chlorophyl from storm water runoff, and salinity is primarily driven by precipitation. The Gulf site is characterized by constant salinities (~36) and low nutrient levels.

## Sample Collection and Processing

#### Collection

A high resolution (daily) time series was conducted starting on July 7th and ending on July 19th, 2022 capturing before, during, and after a Saharan dust event. Samples were collected in 1L autoclaved Polypropylene Bottles and immediately placed in a cooler on ice. Following transport to the laboratory, the samples were concentrated onto 0.2µm pore size, hydrophobic polycarbonate membranes (25mm in diameter) and stored at -80°C until DNA Extraction using a ZymoBIOMICS DNA Miniprep Kit (Cat#:4300).

Vertical profiles of salinity, temperature, pH, and dissolved oxygen were obtained using a YSI ProPlus sonde. Samples were collected for chlorophyl *a* analysis, inorganic nutrients, dissolved and particulate organic matter, and microbial analysis.

#### Vibrio Enumeration

Total *Vibrio* concentrations were quantified using a SYBR Green quantitative PCR (qPCR) method. Estimates were made using genus-specific quantitative polymerase chain reaction (qPCR) PCR product using Vibrio group specific primers targeting a variable region of the 16S rRNA gene, 567F, 5′GGCGTAAAGCGCATGCAGGT3′ and 680R, 5′-GAAATTCTACCCCCCTCTACAG-3. Master mix and PCR conditions were derived from (Westrich et al. 2016). Briefly, 5µL of 2X SYBR PowerUp Master Mix (Applied Biosystems, Foster City, CA) were added to primers with a final concentration of 0.16µM and PCR water for a final volume of 10µL. Reactions were run in triplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA) with the following cycling conditions: 2 minutes at 50◦C for UDG activation and 95◦C for 2 min to activate AmpliTaq polymerase and UP, followed by 40 cycles of 95◦C for 3 seconds for denaturation and 60◦C for 30 seconds for annealing and extension. Each run was followed by a dissociation step (60C to 95C by 0.5C increments) to determine a melt curve for analysis of specificity. Each test was run with positive and negative controls in triplicate. Cycle threshold values for each qPCR test was compared to a standard curve representing 10^1 to 10^6 gene copies per reaction volume.

## Data import and cleaning

This project works with three primary data sets. (1) Enumeration of Vibrio bacteria as copies/mL from quantitative PCR (copies\_master.csv), (2) Dust concentrations as aerosol optical thickness (AOT) derived from the Naval Research Lab (nrl\_conc.csv), and (3) Temperature and Salinity across the daily time series (ysi\_2.csv). Below is a brief summary of how each data set was cleaned. More details can be found in the supplementary files, containing the code and comments describing what each line does. Please refer to the project\_README.md file located in the NathanGreenslit-MADA-Project folder for details on the repository contents and instructions on reproducibility

#### Cleaning of Total Vibrio Enumeration data set

Quantitative PCR provides enumeration results as “cycle threshold or Cq” values. This depicts the cycle number in which enough of the DNA target was present to be amplified and thus detected. Lower Cq values correspond to higher target concentrations (as it takes less cycles to amplify), and vice verse for higher Cq values. The cleaning of this data set consisted of converting these Cq values into something that can be used to quantify Vibrio (copies of target per mL of seawater). To do so, I needed to take into account the amount of water sample that was concentrated on a filter, the total amount of DNA that was eluted following a DNA extraction, and the amount of DNA template added to each mix for qPCR (to name a few). Taking these into account, we are able to calculate our way from Cq values to copies of Vibrio per mL of seawater. qPCR was conducted in triplicate, so the last step was to take the average of the three replicates to have a final value per sample. A date column was also added for each sample.

#### Cleaning of Dust Concentration (AOT) Data set

Dust concentration (AOT) was collected at time points 0 hour, 6 hour, 12 hour, and 18 hour. In this script, I made different data sets based on time points as well as a data set containing the summed dust concentration (of all time points) per day. For downstream statistical analysis, it will be helpful to have these different data sets to compare to time of sample collection.

#### Cleaning of Environment data set

This data set did not require any cleaning.

# Results

## Exploratory/Descriptive analysis

#### Dust AOD and Vibrio Enumeration

|  |
| --- |
| Figure 1: Total Daily Vibrio at All Sites |

The top portion of the figure depicts total enumerated Vibrio as copies per mL, with color by site. There is a noticeable shift in copies in the Gulf site, a less noticeable but still present shift in Blind Oso Bay, and practically no changes in the Canal site. The bottom portion of the figure depicts the dust concentration (AOD) across the daily time series. This figure depicts the summed dust concentration 24hr prior to sample collection (time points 13hr–> 7hr the next day) across the daily time series. During the time series, two small periods of higher dust AOD occurred on the 9th and 13th respectively. A much larger spike in AOD was observed on the 16th. Initial shift in copies per mL at the Gulf and Blind Oso occur 24-48hr following initial dust introduction on the 13th, and experience another shift around 24-48hr of higher dust input on the 16th.

#### Environmental Parameters

|  |
| --- |
| Figure 2: Environmental parameters across daily time series |

Temperature and salinity at the Canal site is relatively high and consistent as compared to the other sites. Blind Oso Bay experiences some shifts in temperatures and drastic fluctuations in salinity. The Gulf site had a drop in temperature (possibly due to upwelling) and relatively consistent ocean salinities.

\*need to add details on other environmental variables

## Statistical analysis

### Check Distributions:

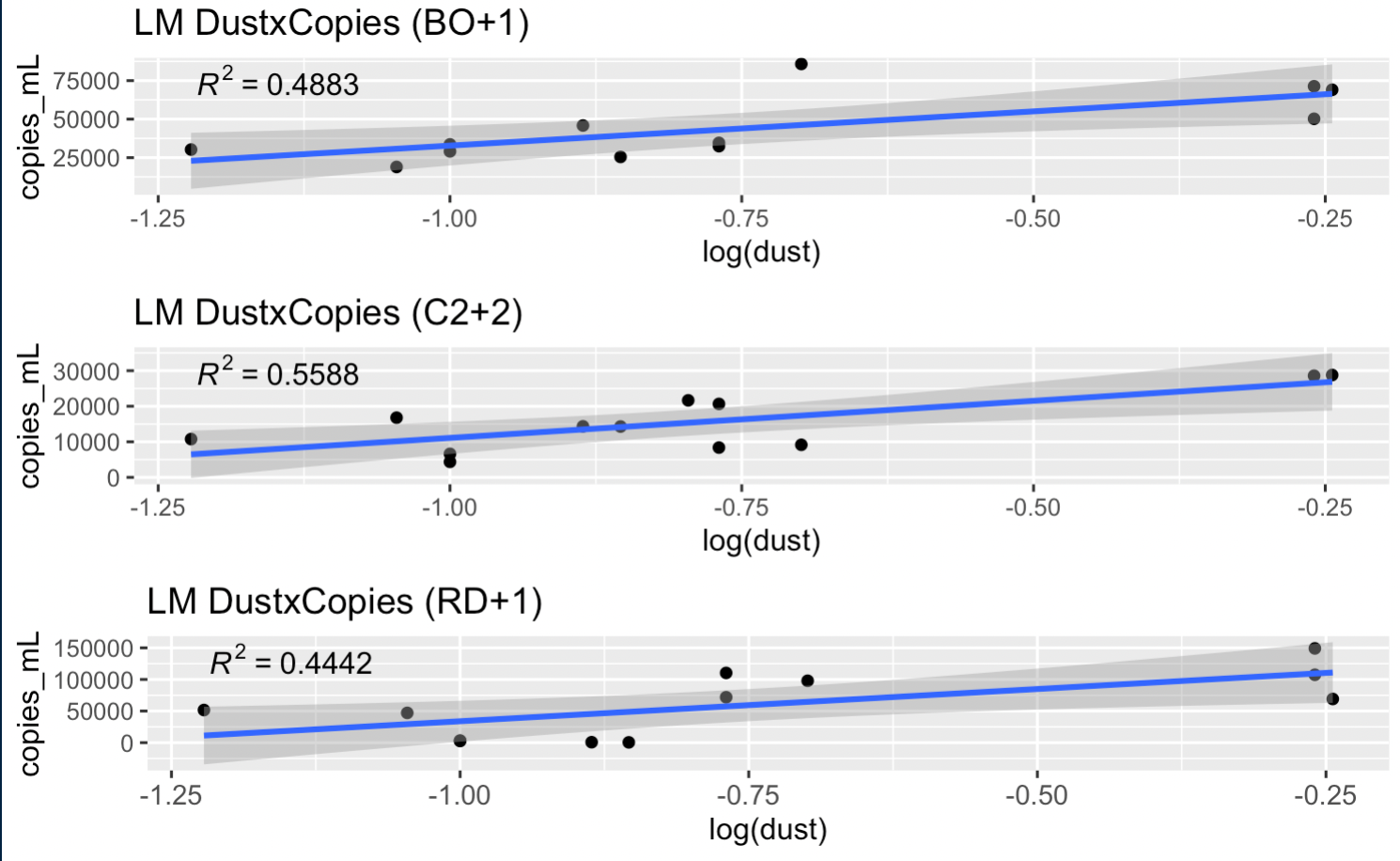
Shapiro-Wilk’s tests were conducted to assess distribution of the data. Any data that did not have a normal distribution was log-transformed and the test was run again. Variables that continue to fail normality will have non-parametric tests run on them.

### Cross-correlation Analysis:

Previous literature has shown that Vibrio respond 12-24 hours following dust deposition events. Based on this, we need to look at lags between the two times series for dust data and growth data. To do this, we can set up a cross-correlation analysis, which allows us to examine the lag/lead relationship between the two variables.

CCF plots showed that BO has a significant lag relationship a value of -1. In other words, dust occurs, and one day later, we see a response in copies\_mL. C2 has a lag at -2, and RD at -1.

### Linear Regression Models



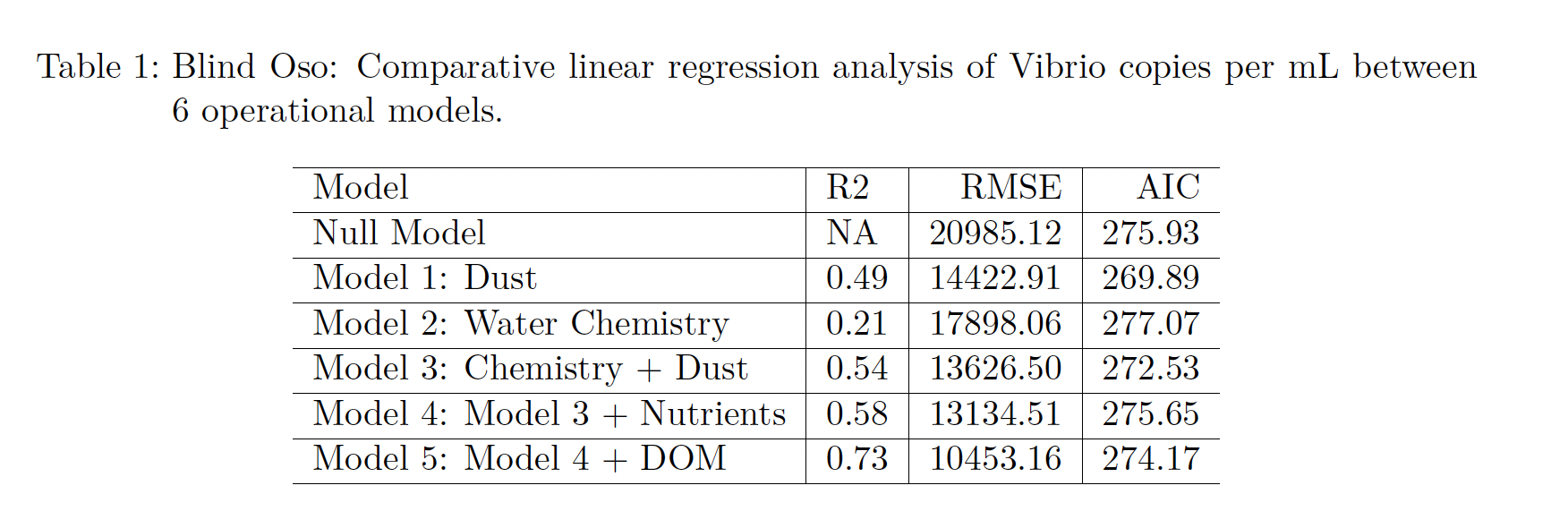
Above are linear regressions that examine the relationship between dust concentration and copies\_mL. As mentioned above, the lags identified may not correspond to the most significant relationship between the two variables.

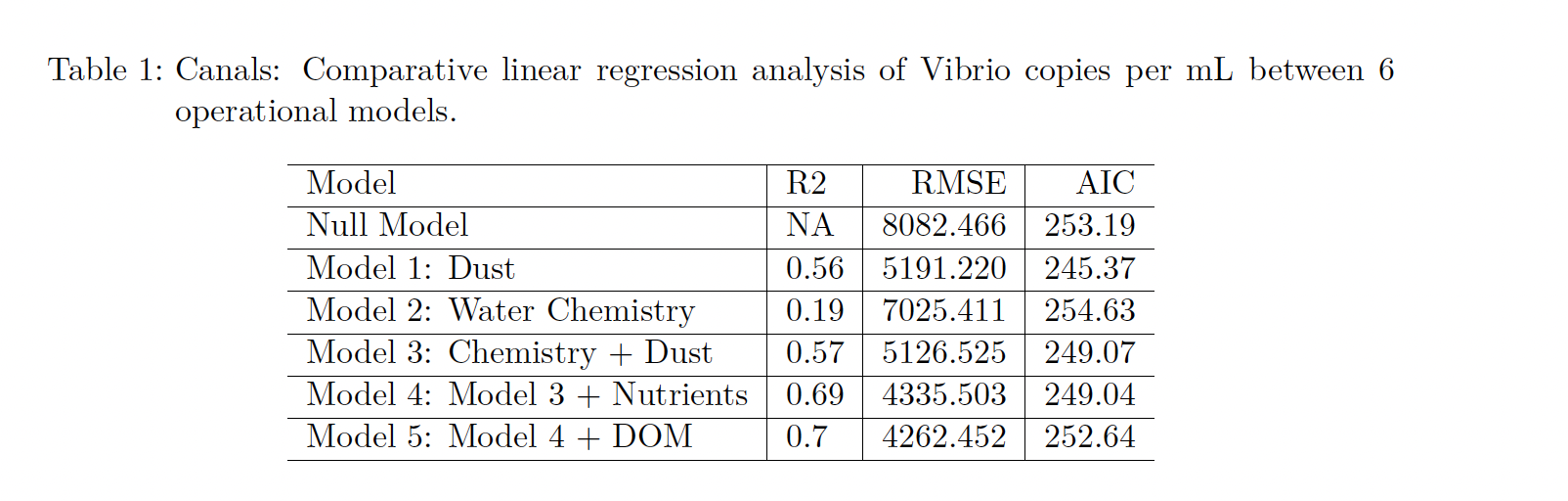
| Site | Model Type | RMSE | AIC |
| --- | --- | --- | --- |
| BO | Null | 20472.82 | 275.9331 |
| BO | Actual | 14422.91 | 269.8923 |
| C2 | Null | 8472.898 | 253.1876 |
| C2 | Actual | 5191.22 | 245.3679 |
| RD | Null | 50386.56 | 296.9821 |
| RD | Actual | 36132.11 | 291.933 |

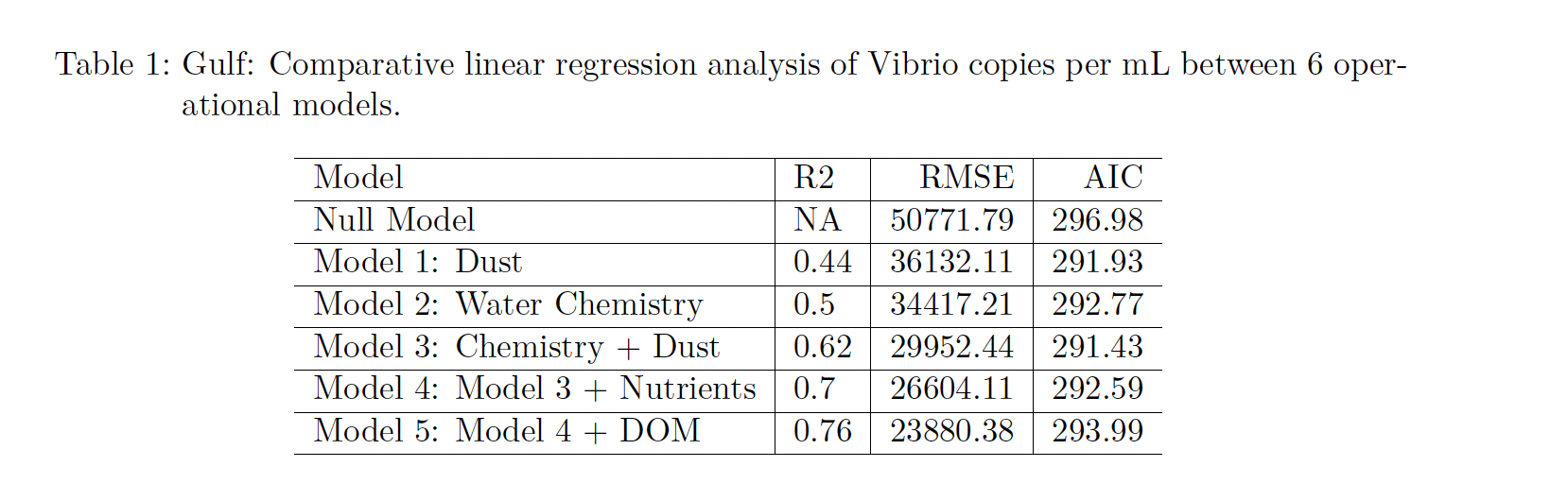
A few things to note:

* My stats code contains some chunks that create predicted values based off of the current data set. Since my data set is small, I am not able to portion of ~25% to use to test my model as it could impact the strength of my model. Therefore, plotting the predicted vs. actual will give an R2 of 1.00. I am using AIC as a metric of model performance and comparing data produced models to null models.
  + Will need to interpret findings as exploratory and hypothesis generating, and will need to be careful about trying to draw generalizable conclusions.

### Multi-variate Linear Regressions







## Full analysis

# Discussion

## Summary and Interpretation

## Strengths and Limitations

**One important thing to note**: Our data set contains dust data for all of 2022. Our copies\_mL data only contains values from 7/7/22 –> 7/19/22. Therefore, these CCA were made with only this time series. This limits our analysis, as we have dust data prior that can explain copies\_mL on the earlier dates 7/7 or 7/8. But we cannot incorporate this into the CCA as we have NA for copies\_mL for those early days. Below we will run linear models that DO contain these prior dust days. And since we have identified a lag, we can shift the dust data to line up with the copies\_mL that it corresponds with directly - and we no longer have NAs. But this ultimately means that the lag identified above. may not be the best fit for our model below, so some additional work is needed that tests other lags in the code.

## Conclusions

# References

Baker-Austin, Craig, James D. Oliver, Munirul Alam, Afsar Ali, Matthew K. Waldor, Firdausi Qadri, and Jaime Martinez-Urtaza. 2018. “Vibrio Spp. Infections.” *Nature Reviews Disease Primers* 4 (1): 1–19. <https://doi.org/10.1038/s41572-018-0005-8>.

Ben-Haim, Y., F. L. Thompson, C. C. Thompson, M. C. Cnockaert, B. Hoste, J. Swings, and E. Rosenberg. 2003. “Vibrio Coralliilyticus Sp. Nov., A Temperature-Dependent Pathogen of the Coral Pocillopora Damicornis.” *International Journal of Systematic and Evolutionary Microbiology* 53 (1): 309–15. <https://doi.org/10.1099/ijs.0.02402-0>.

Borchardt, Trace, Kelsey V. Fisher, Alina M. Ebling, Jason R. Westrich, Peng Xian, Christopher D. Holmes, William M. Landing, Erin K. Lipp, Michael S. Wetz, and Elizabeth A. Ottesen. 2020. “Saharan Dust Deposition Initiates Successional Patterns Among Marine Microbes in the Western Atlantic.” *Limnology and Oceanography* 65 (1): 191–203. <https://doi.org/10.1002/lno.11291>.

d’Almeida, Guillaume A. 1986. “A Model for Saharan Dust Transport.” *Journal of Climate and Applied Meteorology* 25 (7): 903–16. <https://doi.org/10.1175/1520-0450(1986)025<0903:AMFSDT>2.0.CO;2>.

Eagon, R. G. 1962. “*PSEUDOMONAS NATRIEGENS* , A MARINE BACTERIUM WITH A GENERATION TIME OF LESS THAN 10 MINUTES.” *Journal of Bacteriology* 83 (4): 736–37. <https://doi.org/10.1128/jb.83.4.736-737.1962>.

Griffin, Dale W, Virginia H Garrison, Jay R Herman, and Eugene A Shinn. n.d. “African Desert Dust in the Caribbean Atmosphere: Microbiology and Public Health.”

Haldar, S., A. Maharajan, S. Chatterjee, S. A. Hunter, N. Chowdhury, A. Hinenoya, M. Asakura, and S. Yamasaki. 2010. “Identification of Vibrio Harveyi as a Causative Bacterium for a Tail Rot Disease of Sea Bream Sparus Aurata from Research Hatchery in Malta.” *Microbiological Research* 165 (8): 639–48. <https://doi.org/10.1016/j.micres.2009.12.001>.

Jensen, Sigmund, Petter Frost, and Vigdis L. Torsvik. 2009. “The Nonrandom Microheterogeneity of 16S rRNA Genes in *Vibrio* *Splendidus* May Reflect Adaptation to Versatile Lifestyles.” *FEMS Microbiology Letters* 294 (2): 207–15. <https://doi.org/10.1111/j.1574-6968.2009.01567.x>.

Kellogg, Christina A., Dale W. Griffin, Virginia H. Garrison, K. Kealy Peak, Nelson Royall, Raymond R. Smith, and Eugene A. Shinn. 2004. “Characterization of Aerosolized Bacteria and Fungi From Desert Dust Events in Mali, West Africa.” *Aerobiologia* 20 (2): 99–110. <https://doi.org/10.1023/B:AERO.0000032947.88335.bb>.

Lenes, J. M., B. A. Darrow, J. J. Walsh, J. M. Prospero, R. He, R. H. Weisberg, G. A. Vargo, and C. A. Heil. 2008. “Saharan Dust and Phosphatic Fidelity: A Three-Dimensional Biogeochemical Model of Trichodesmium as a Nutrient Source for Red Tides on the West Florida Shelf.” *Continental Shelf Research* 28 (9): 1091–1115. <https://doi.org/10.1016/j.csr.2008.02.009>.

Mills, Matthew M., Celine Ridame, Margaret Davey, Julie La Roche, and Richard J. Geider. 2004. “Iron and Phosphorus Co-Limit Nitrogen Fixation in the Eastern Tropical North Atlantic.” *Nature* 429 (6989): 292–94. <https://doi.org/10.1038/nature02550>.

Okada, Kazuhisa, Tetsuya Iida, Kumiko Kita-Tsukamoto, and Takeshi Honda. 2005. “Vibrios Commonly Possess Two Chromosomes.” *Journal of Bacteriology* 187 (2): 752–57. <https://doi.org/10.1128/JB.187.2.752-757.2005>.

Payne, Shelley M., Alexandra R. Mey, and Elizabeth E. Wyckoff. 2016. “Vibrio Iron Transport: Evolutionary Adaptation to Life in Multiple Environments.” *Microbiology and Molecular Biology Reviews* 80 (1): 69–90. <https://doi.org/10.1128/MMBR.00046-15>.

Ramírez-Camejo, Luis A., Anabella Zuluaga-Montero, Vernon Morris, José A. Rodríguez, María T. Lázaro-Escudero, and Paul Bayman. 2022. “Fungal Diversity in Sahara Dust: Aspergillus Sydowii and Other Opportunistic Pathogens.” *Aerobiologia* 38 (3): 367–78. <https://doi.org/10.1007/s10453-022-09752-9>.

Richards, Gary P., Michael A. Watson, David S. Needleman, Karlee M. Church, and Claudia C. Häse. 2015. “Mortalities of Eastern and Pacific Oyster Larvae Caused by the Pathogens Vibrio Coralliilyticus and Vibrio Tubiashii.” Edited by M. W. Griffiths. *Applied and Environmental Microbiology* 81 (1): 292–97. <https://doi.org/10.1128/AEM.02930-14>.

Ringgaard, Simon, Wen Yang, Alejandra Alvarado, Kathrin Schirner, and Ariane Briegel. 2018. “Chemotaxis Arrays in Vibrio Species and Their Intracellular Positioning by the ParC/ParP System.” Edited by Victor J. DiRita. *Journal of Bacteriology* 200 (15). <https://doi.org/10.1128/JB.00793-17>.

Rosenberg, Eugene, and Leah Falkovitz. 2004. “The *Vibrio* *Shiloi* / *Oculina* *Patagonica* Model System of Coral Bleaching.” *Annual Review of Microbiology* 58 (1): 143–59. <https://doi.org/10.1146/annurev.micro.58.030603.123610>.

Sampaio, Ana, Vanessa Silva, Patrícia Poeta, and Florin Aonofriesei. 2022. “Vibrio Spp.: Life Strategies, Ecology, and Risks in a Changing Environment.” *Diversity* 14 (2): 97. <https://doi.org/10.3390/d14020097>.

Savoie, Dennis L., Joseph M. Prospero, and Eric S. Saltzman. 1989. “Non-Sea-Salt Sulfate and Nitrate in Trade Wind Aerosols at Barbados: Evidence for Long-Range Transport.” *Journal of Geophysical Research* 94 (D4): 5069. <https://doi.org/10.1029/JD094iD04p05069>.

Tran, L, L Nunan, Rm Redman, Ll Mohney, Cr Pantoja, K Fitzsimmons, and Dv Lightner. 2013. “Determination of the Infectious Nature of the Agent of Acute Hepatopancreatic Necrosis Syndrome Affecting Penaeid Shrimp.” *Diseases of Aquatic Organisms* 105 (1): 45–55. <https://doi.org/10.3354/dao02621>.

Westrich, Jason R., Alina M. Ebling, William M. Landing, Jessica L. Joyner, Keri M. Kemp, Dale W. Griffin, and Erin K. Lipp. 2016. “Saharan Dust Nutrients Promote *Vibrio* Bloom Formation in Marine Surface Waters.” *Proceedings of the National Academy of Sciences* 113 (21): 5964–69. <https://doi.org/10.1073/pnas.1518080113>.

Westrich, Jason R., Dale W. Griffin, Douglas L. Westphal, and Erin K. Lipp. 2018. “Vibrio Population Dynamics in Mid-Atlantic Surface Waters During Saharan Dust Events.” *Frontiers in Marine Science* 5 (February): 12. <https://doi.org/10.3389/fmars.2018.00012>.