Salty Seas and Summer Thrills: *Vibrio vulnificus* and the Waters We Love (to Fear)

By: Alicia Acuña, Ankur Bambhrolia, Daniela Ramos, Sheila Resto Rodriguez

Vibrio vulnificus (V. vulnificus) is a gram-negative bacterium commonly found in marine organisms and brackish waters throughout the hydrosphere. Taxonomically, *Vibrio vulnificus* is in the same genus as *V. parahaemolyticus*, and *V. cholerae*, the latter of which is the causative agent of cholera. The coastal conditions *V. vulnificus* flourishes in are the same ones that entice us to enjoy a relaxing day at the beach or indulge in a whimsically wild family vacation. These are often synonymous with swimming and gluttonously indulging in seafood. *V. vulnificus* often enters the body through open wounds or ingestion of contaminated seafood, particularly oysters and other shellfish. Once *V. vulnificus* enters a wound, it can cause necrotizing fasciitis, or flesh-eating disease[11]. Other common symptoms include fever, chills, diarrhea, gastroenteritis and tachycardia[5]. When infected with *V. vulnificus*, immediate antibiotic therapy (doxycycline, ceftazidime, fluoroquinolones) is imperative[6]. In certain cases, surgical debridement or amputation may be necessary[8].

Over the past two decades, the prevalence of *V. vulnificus* has increased and expanded into previously unaffected areas in the northern United States[3], likely due to rising sea surface temperatures brought on by climate change[9]. These rising temperatures, along with changing salinity patterns caused by extreme weather events (drought, hurricanes, etc.) — in addition to chronic environmental stressors such as global warming — have contributed to increased outbreaks of *V. vulnificus*[2, 9]. Consequently, scientists and researchers often use *V. vulnificus* as a "microbial barometer of climate change," indicating that its expanding range correlates with increasing global temperatures[1].

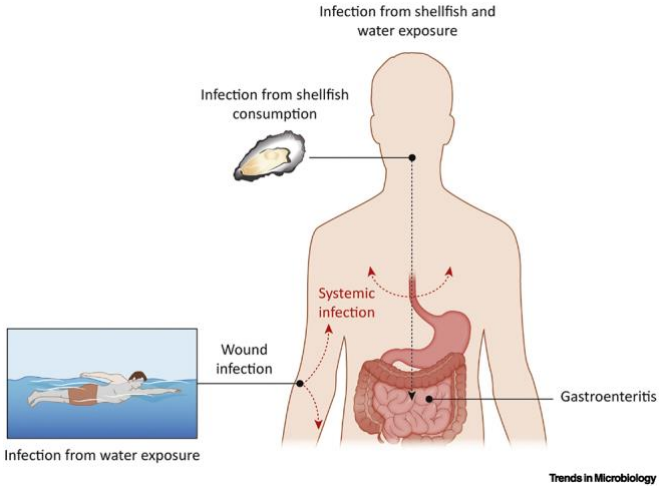


Figure 1: Illustration of routes of *Vibrio vulnificus* infections in humans. The bacterium can infect humans through two different pathways: 1) Wound infection from exposure to contaminated seawater, and 2) Gastrointestinal infections from consumption of contaminated shellfish.

Studies indicate that water salinity and water temperature are the biggest predictors of *V. vulnificus* growth. However, careful analysis of other abiotic and biotic factors including, but not limited to, wind patterns, seagrass bed density, and zooplankton all play a vital role in *Vibrio* distribution. Regarding salinity, *V. vulnificus'* thrives in water with salt concentrations between 5 and 25 parts per thousand. As a result, *V. vulnificus* can also be found in estuaries in addition to sea water. However, there is a limit for temperature and salinity threshold. If the water temperature is too warm or cold, or if the salinity is too high, the pathogen will not thrive. *V. vulnificus* will enter a sessile state when conditions are not favorable or when attached to certain surfaces like oyster shells. Growth is limited in the sessile state, and this metabolic mechanism has given researchers difficulties when culturing *V. vulnificus*, as they often end up with underestimated numbers. Conversely, when conditions are favorable, they enter a mobile, planktonic state in which the bacterium is more active[11].

There are four molecules that contribute to *V. vulnificus's* toxicity: RtxA1, PlpA, VvA, and Vvsp[7]. RtxA1 is a toxin that is secreted by *V. vulnificus* upon apoptosis which contributes to Necro-Fasciitis by killing human cells upon contact. PlpA is a phospholipase enzyme responsible for cell death and blood cell hemolysis. VvA is a toxin that contributes to apoptosis, cell-death, and dysregulation of cell-signaling to its host. Furthermore, the

VvA gene is a key genetic marker for *V. vulnificus*. Lastly, Vvsp is a metalloprotease that is responsible for host tissue damage, and inflammation. Studies indicate that many of the virulent genes have been obtained through Horizontal Gene Transfer (HGT)[8].

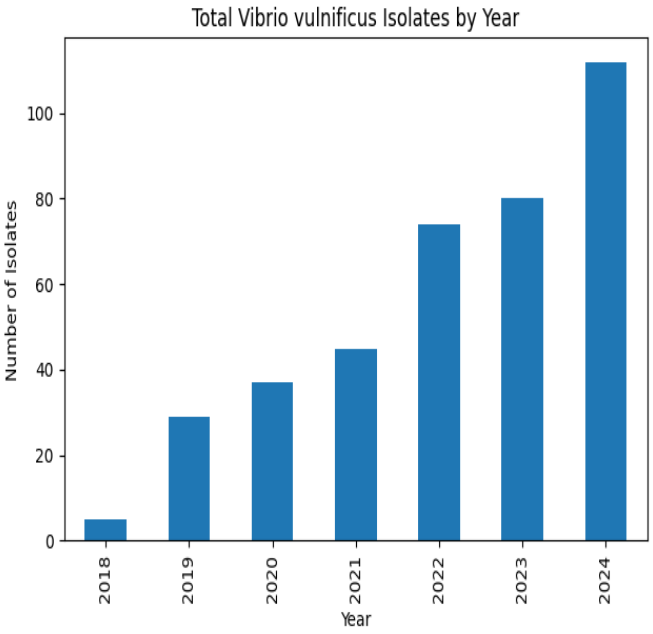
Per the Centers for Disease Control and Prevention (CDC) there are between 150-200 cases of *V. vulnificus* reported annually[7]. In the United States, *V. vulnificus* accounts for 9.8% of all *Vibrio* infections and has the highest fatality rate, approximately 50% for septicemia[4]. Among all foodborne diseases, *V. vulnificus* has the highest fatality rate[7]. Individuals with compromised immune systems, such as those with cancer, liver disease, or diabetes, are at greater risk for severe complications[10]. In addition to health risks, *Vibrio* infections are responsible for approximately $320 million in health-related costs annually in the United States, with *V. vulnificus* accounting for $28 million[1]. Archer et al. gathered *V. vulnificus* cases (1988-2018) from the Cholera and Other Vibrio Illness Surveillance (COVIS) system and developed a model to predict future *V. vulnificus* infections with 7 Global Climate Models (GCM). According to their projections, by 2041-2060, the distribution of *V. vulnificus* infection is expected to spread to populous areas around New York[1]. These predictive studies highlight the importance of understanding the transmission dynamics and environmental factors influencing *Vibrio* species*(spp.)* today. The increasing prevalence of *V. vulnificus* and other pathogenic *Vibrio spp.*, driven by climate change, demonstrates a growing public health concern domestically and globally. *Vibrio spp.* is crucial to the nitrogen and carbon cycles, making them ecologically essential in aquatic environments[3]. Therefore, research should focus on understanding the biology of *Vibrio spp*. to mitigate the risk of potential outbreaks. For our study, we aim to examine how environmental factors, such as water temperature and salinity affect the prevalence of *V. vulnificus* in the eastern coastal waters of the United States.

V. vulnificus remains prevalent in the marine environment and affects those who interact with it. Susceptible populations include chefs, swimmers, and seafood consumers. As aforementioned, *V. vulnificus* possesses the highest fatality rate of all foodborne diseases[8]. By understanding how *V. vulnificus* spreads and how the infection and mortality rates change, we aim to provide critical insight for medical professionals and policy makers to prevent further outbreak of this disease.

Methods

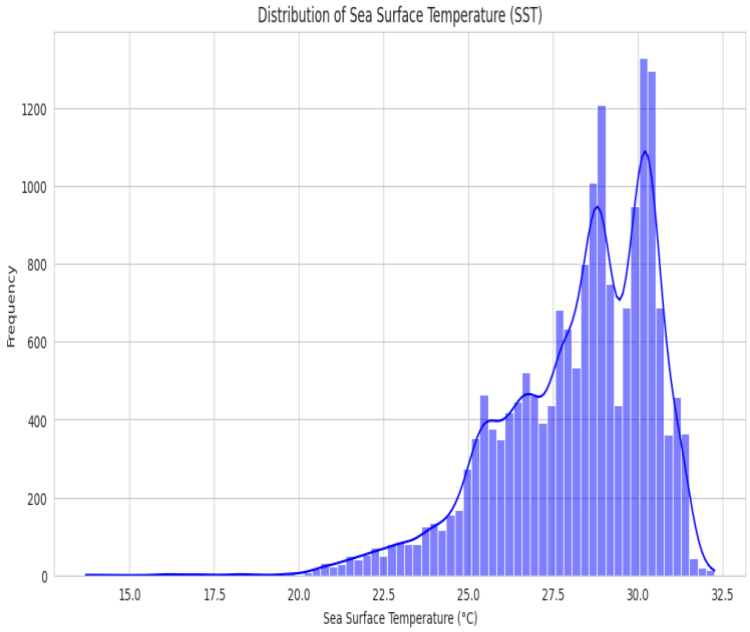
Data Collection:

Data was collected from the Centers for Disease Control and Prevention (CDC), the National Oceanic and Atmospheric Administration (NOAA), and the National Aeronautics and Space Administration (NASA). From the CDC, we gathered data from the BEAM (Bacteria, Enterics, Ameba, and Mycotics) Dashboard, which tracks illnesses caused by a range of pathogens (e.g., bacteria, parasites, and viruses). For our research purposes, we exported the complete dataset and proceeded to filter by pathogen and serotype/species (Vibrio and vulnificus/vulnificus Ctx–, respectively). Once this was completed, we retained the following



Graph 1. Total count of Vibrio cases by year, from 2018 to 2024

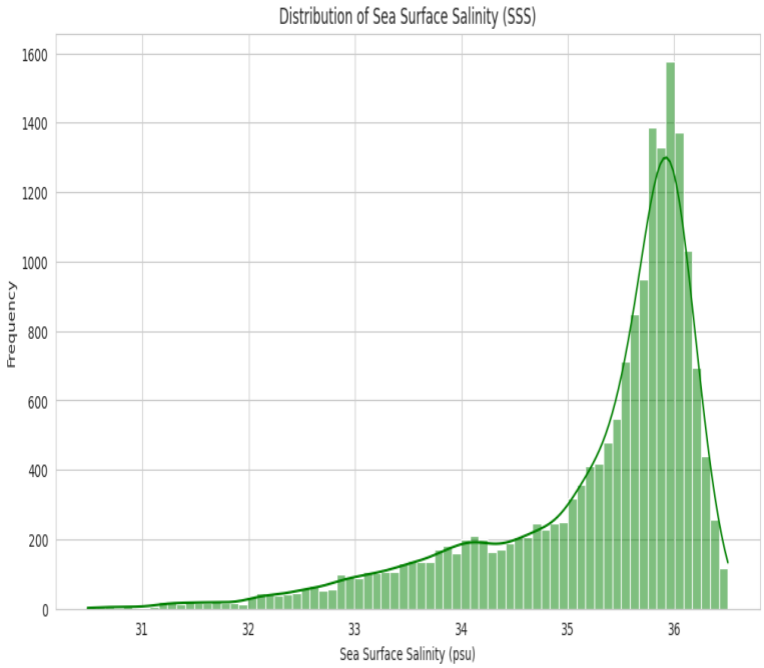
columns: Year, Month, State, Source, Pathogen, Serotype/Species, and Number of Isolates.



Graph 2. Distribution of Sea Surface Temperature, or SST, in the dataset. The histogram displays a concentration of SST values mainly between 25°C and 30°C, with a significant peak around 28°C.

After finalizing the preprocessing steps for the *V. vulnificus* data, we obtained sea surface temperature (SST) data from the NOAA OI SST V2 High Resolution Dataset. We limited the data to a date range of January 1990 to December 2024 and focused on the Eastern U.S. region. The dataset included monthly mean SST, longitude, latitude, and time columns. We converted longitude coordinates to the standard format and split the time column into separate month and year columns. We also removed any rows with missing SST values. Additionally, we used a GitHub CSV file containing state boundary boxes to assign a state to each row based on latitude and longitude ranges. Lastly, we used the state column to create a geographic region column.

To obtain salinity data, we created a NASA Earth Data account and downloaded 150 files from the OISSS_L4_multimission_monthly_v2_2.0 dataset, dates ranging from 2011 to 2024. These files were converted from netCDF format to CSV, then concatenated into a single file. We removed rows with missing salinity (SSS) values and filtered the dataset to only include data from the Eastern U.S. Furthermore, time values were split into month and year, and state and region columns were added using the same boundary box CSV used for the SST data.



Graph 3. Distribution of Sea surface salinity (SSS)

All processed datasets were uploaded to MySQL Workbench and joined for analysis. First, the SST and SSS data were joined using the columns: month, year, latitude, and longitude. Then the combined SST/SSS dataset was joined with the *V. vulnificus* dataset on month, year, and state. Since multiple geographic coordinates (latitude/longitude) shared the same month, year, and state values, the joined table resulted in many-to-one relationships. To address this issue, we aggregated the data by grouping by state, month, year, source, and serotype/species, and averaged the SST, SSS, and number of isolate values within each group. This approach allowed us to retain important categorical differences while standardizing the dataset for more accurate analysis.

Software Tools:

Data cleansing and analysis were completed using the following software tools: Python (Pandas, NumPy, SciPy, Scikit-learn), MySQL Workbench, and Google Colab. For visualizations, we used seaborn. To complete our time series analysis, we utilized three different models such as SARIMA, XGBoost, Random Forest, etc.

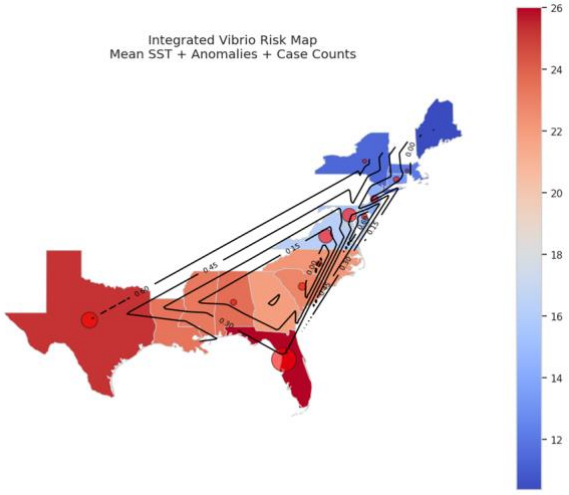


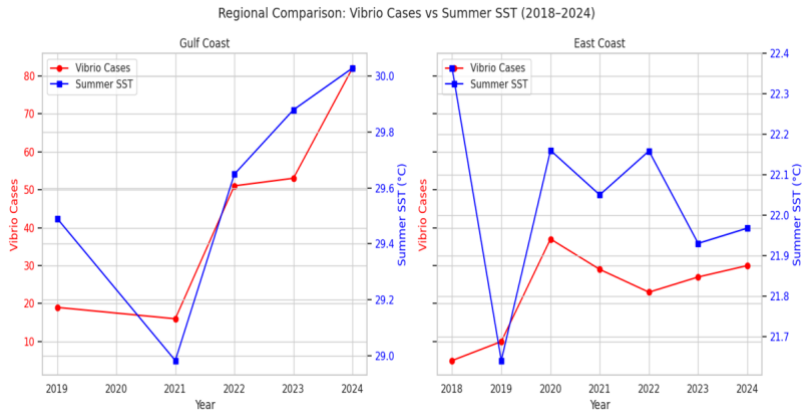
Figure 2. Shows the correlation between SST (color gradient), SST anomalies (contour lines), and case counts across U.S. coastal regions (bubble sizes).

Results

Data Analysis:

We used correlational analyses to better understand the relationship between environmental factors (SST and SSS) and *V. vulnificus* cases. These variables were examined under different conditions including by year, season, and geographic region (Gulf states vs Eastern states). Additionally, we investigated the environmental factors together and independently; for example, SST & SSS vs. *V. vulnificus,* SST vs. *V. vulnificus*, and SSS vs. *V. vulnificus*. This allowed us to quantify the degree to which these variables co-vary. The regional analysis was crucial for us to identify spatial variability in the environmental drivers of *V. vulnificus*. For example, examining the summer SST by region demonstrated a more accurate portrayal of mean temperature ranges. The graph below shows that, in Gulf states, the temperature starts just below 29°C, while in the East Coast region, it begins around 21.6 °C. Without this separation by region, the average temperature would not be an accurate representation of either region.

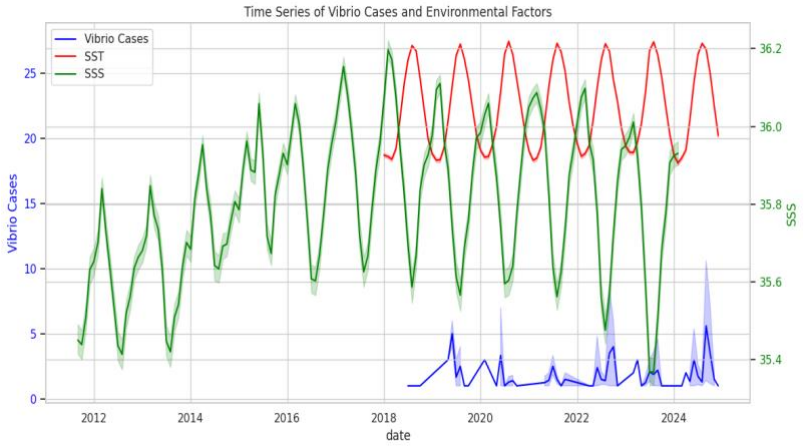
This heatmap demonstrates the relationship between SST and SSS anomalies and *V. vulnificus* counts. From this visualization, we observed a pattern suggesting that SST anomalies (higher than the expected temperature) can contribute to an increase in *V. vulnificus* cases. Given the temporal nature of our data, we also employed a time series analysis. We generated time series plots to visualize the trends in *V. vulnificus* incidence, SST, and SSS over time. This allowed us to identify patterns such as seasonality and long-term changes, and to visually assess how environmental factors relate to changes in *V. vulnificus* cases.



Graph 4: Vibrio cases vs. Summer SST by region

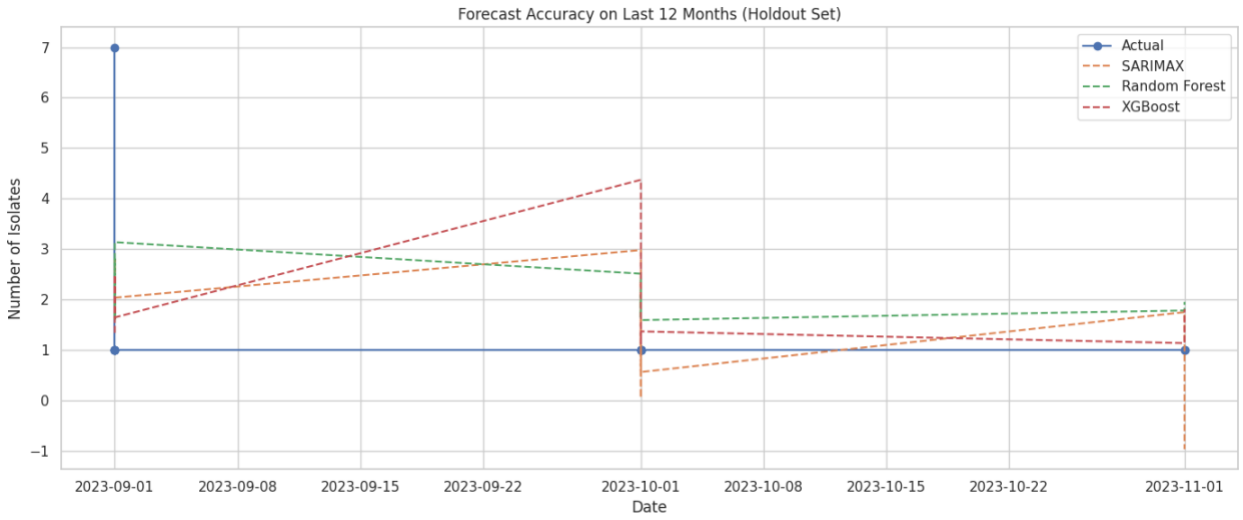
Lastly, we utilized machine learning models to project 60 months into the future, based on the collected historical data. We employed three models (SARIMAX, XGBOOST, Random Forest Regression, etc.) to observe the best fit for our needs, given that the data has seasonal trends and is nonlinear. All three models can be used for forecasting time series data. XGBOOST captures nonlinearity in data; however, this can break the temporal structure of the data. Random Forest regression creates lag features (i.e. shifts the data back by a set number of months, years, days etc.) and creates trees to determine an output forecasted value. The lag can potentially capture seasonality trends without overfitting the data. Finally, SARIMAX— (Seasonal Autoregressive Integrated Moving Average with eXogenous regressors)— takes seasonality and external factors

(e.g. SSS and SST) when forecasting values. This model potentially fits well with the vibrio data since we want to see how SSS and SST influences vibrio cases and it takes seasonality into account.

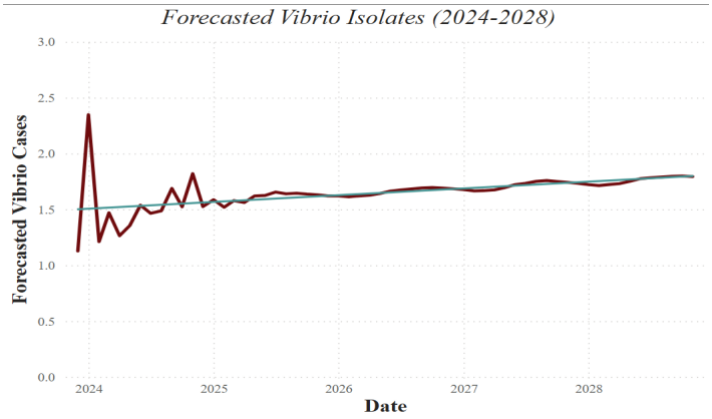


Graph 5. Time series plot visualizes the dynamics of Vibrio vulnificus cases (in blue) alongside the key environmental factors, sea surface temperature (SST) (in red) and sea surface salinity (SSS) (in green), over the period from 2010 to 2024.

While comparing the three models, we observed the Mean Absolute Error (MAE), and the Root Mean Squared Error (RMSE). MAE provides a sense of how far the model predictions are off on average compared to the actual data (i.e., +/- 2.5 Celsius). RMSE is more sensitive to outliers, which is useful if the large errors are costly (i.e., public health risks due to misjudging temperature). If RMSE >> MAE, then there are probably a few forecasts with large errors.



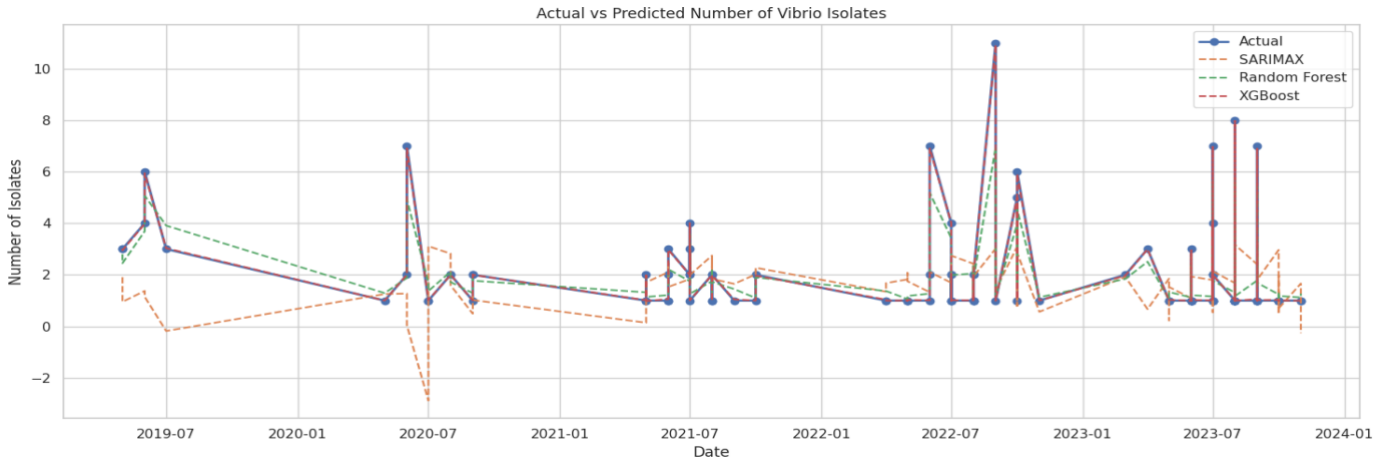
Graph 6. When the models are fitted into the holdout set (data not used to train the model) the XGBOOST model overfits the data. The closest model that fits the actual data is the SARIMAX model. This verification is further supported by the MAE and RMSE



Graph 7. This graph zooms in on the forecasted vibrio values. The Forecasted Vibrio cases are seasonal and are trending upwards as indicated by the blue trend line. However, the trend is not as drastic as the initial forecasted values.

To forecast future *V. vulnificus* outbreaks, we tested 4 different time series models on SSS and SST: ARIMA, SARIMA, Prophet and Random Forest Regression (RGR). Prophet is a time series forecasting model that uses forecast seasonal data. Furthermore, ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) are statistical models used in time series forecasting to predict future values by modeling trends, seasonality (in SARIMA), and patterns based on past observations and errors. When adding xogenous features (e.g. SSS and SST) to forecast a particular value, the SARIMA model becomes the SARIMAX model. To compare the accuracy of these four models, we compared the MAE and the RMSE values for each feature. For SSS, the Random Forest model yielded the lowest MAE and RMSE values, so it was used to generate forecasted SSS values. Similarly, the

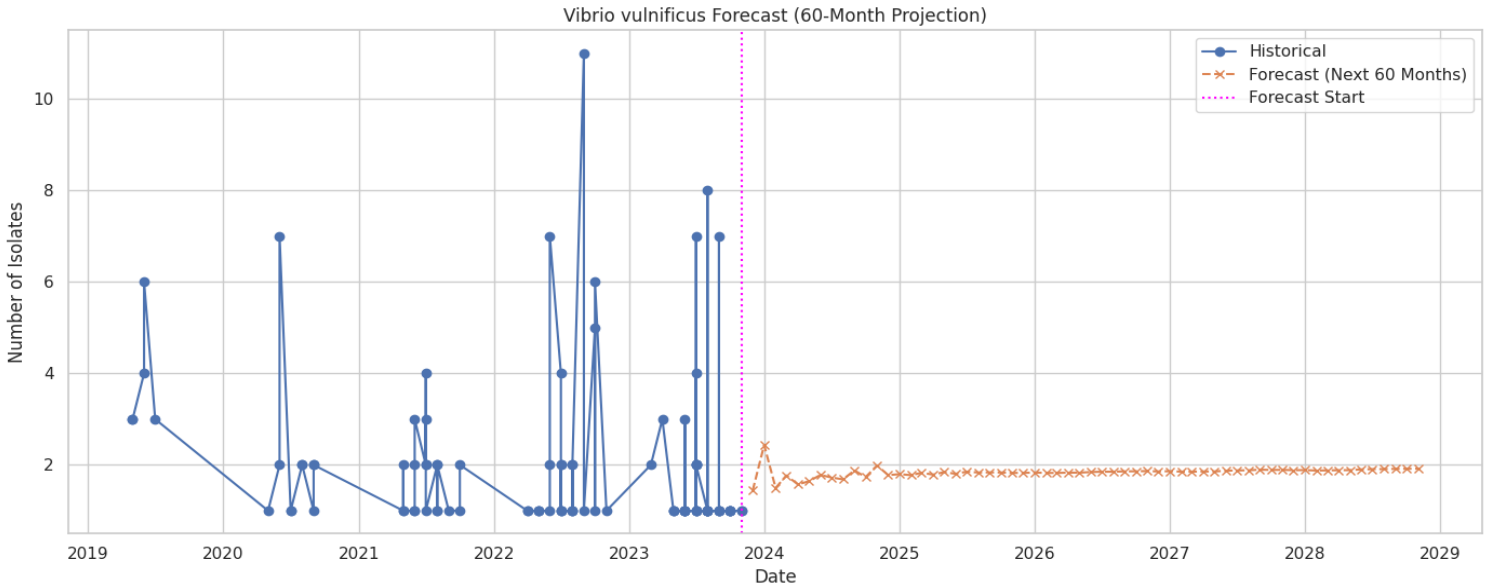
Prophet model produced the lowest MAE and RMSE values for SST and was therefore used to forecast SST values. The difference in model fitting between these two features likely stems from the wavelength and frequency of the seasonality trends (Graph 5). SST has a stable cyclic trend, meaning that the wavelength and frequency are the same. This indicates that the SST data is *univariate*. Prophet captures univariate data very well. Conversely,



Graph 8. Test of different Forecasting Models

SSS values trend up and down through time, making the data multivariate. As a result, Random Forest fit the SSS data better than Prophet.

After forecasting future values, we then used historical data and forecasted SST and SSS to project vibrio cases for the next 60 months. Since external variables influence the forecasted *V. vulnificus* cases, we used SARIMAX, RFR, and XGBoost models and created validation scores. Of the three models, SARIMAX had the lowest MAE and RMSE values, indicating it was the best model to forecast *V. vulnificus* cases for the next 60 months. The resulting graph indicates that *V. vulnificus* cases will fluctuate between 1 to 2 confirmed cases per month. This seasonal pattern mirrors the SSS and SST values, indicating a noticeably clear relationship between *V. vulnificus*, SSS and SST.



Graph 9. The line plot above forecasts the next 60 months of Vibrio outbreaks. On average, per the SARIMAX model, we can expect about 2 Vibrio cases per year.

As aforementioned, SARIMAX is an extension of the ARIMA and SARIMA models—the main difference being that SARIMAX accounts for exogenous variables that may influence the desired forecasted variable in addition to the forecasted variables' own historical data.

SARIMAX is broken down into two main components: regular (non-seasonal component) and a seasonal component—both of which model how the data changes over time. In our model, we used an order component of (p =1, d =0, q =1). This means that we added a one-month lag feature to predict the next Vibrio count (p = 1), assumed data is stationary (d = 0) and 1 past error (p =1), which is the difference between actual and forecasted value. Collectively, this part of the model uses the previous month, both directly (Vibrio Counts) and indirectly (through past errors). Furthermore, the seasonal component uses an order of (P = 1, D = 0, Q = 1, s = 12). P represents the seasonal auto regression meaning that model uses data value from exactly 1 year ago. D = 0 assumes that there is no strong *change* in seasonal trend. Q = 1 tells the model to use the prediction error from one month ago. Setting P and Q to one align the predicted value from its counterpart from a year ago. Finally, the s value represents how long the data needs to be repeated. Since we have a yearly cycle, the value is set to 12.

In summation, the Autoregressive terms (p = 1, P = 1) tells the model what happened last year, and what happened recently. The Moving Average terms (q =1, Q =1) helps the model adapt to the data on a seasonal and short-term basis. Setting the differencing (d = 0, D = 0) indicates that the data does not have strong trends, rather it has seasonal spikes. Manipulating these orders—also known as hyperparameters— can affect the model fit.

Three different SARIMAX models were tested. One where only a one-month lag was incorporated, one where only a 12-month rolling mean was calculated, and a final model where both a 12-month rolling window and a one-month lag was incorporated. The SARIMAX models' performance was assed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). AIC and BIC use statical measures to compare the goodness of fit and complexity of multiple models. AIC balances how well the model fits the data given the complexity of the model (i.e. number of parameters). BIC penalizes complexity more severely compared to AIC based on the number of observations and the number of parameters. This is meant to give an indication of overfitting. The parameters that resulted in the lowest AIC (319.086) and BIC (340.063) scores was the model that incorporated the rolling mean and lag1. This indicates that Vibrio is influenced by both short-term and long-term environmental stressors.

Discussion:

Our primary hypothesis proposed a link between increased *Vibrio vulnificus* (*V. vulnificus*) prevalence and warmer water temperatures, along with low-to-moderate salinity. Our findings, particularly from the time series analysis and partial regression plots, largely support the role of Sea Surface Temperature (SST) in driving *V. vulnificus* cases. The partial effect plots clearly illustrate a positive correlation: as SST rises, so do case numbers. This aligns with existing research indicating that elevated ocean temperatures create more favorable conditions for *V. vulnificus* proliferation. Notably, our SST vs. Vibrio cases plot visually reinforces this, demonstrating increased case numbers in warmer regions.

We observed a weak negative correlation between sea surface salinity (SSS) and the number of *V. vulnificus* isolates, hinting that lower salinity levels might be associated with a higher prevalence of the bacteria. This finding is particularly interesting given that *V. vulnificus* is known to prefer brackish waters with a salinity range of 5 to 25 ppt. However, the predominantly higher salinity range (35 to 38 ppt) in our limited East Coast

and Gulf Coast data makes it difficult to fully assess the true impact of salinity on case occurrences in these regions. The generalizability of our salinity findings is further limited by the geographic focus of our data and our reliance on aggregated information from the CDC, NOAA, and NASA, which could introduce biases due to inconsistencies in reporting and spatial resolution. While our results strongly suggest that higher sea surface temperature (SST) is a key driver of Vibrio cases in these areas, the complex relationship with salinity requires further investigation across different coastal regions and *Vibrio* species.

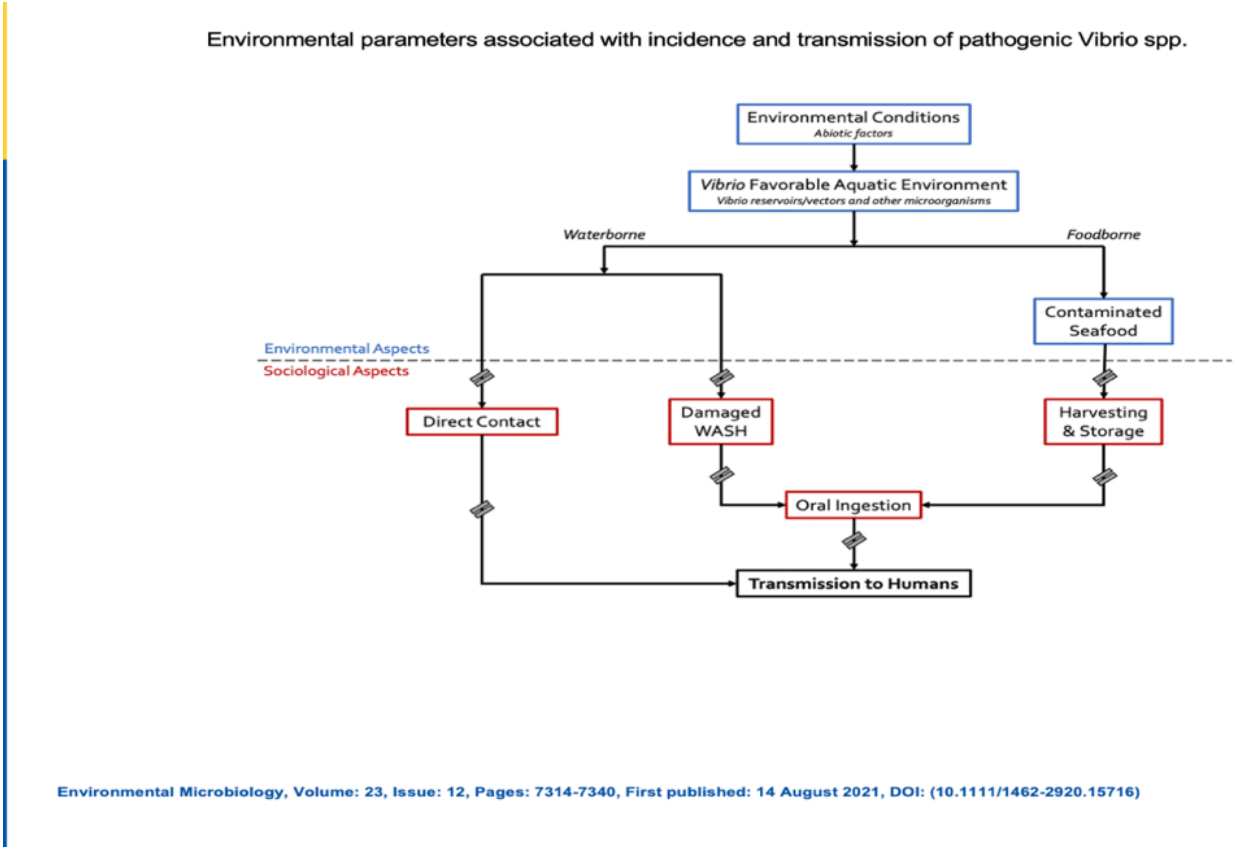
Conclusion and Limitations

To address these limitations and enhance future research, a more integrated approach is needed, incorporating a wider range of environmental factors and higher-resolution spatial data. This could involve the use of high-resolution satellite imagery, expanded microbial monitoring programs for more localized *V. vulnificus* measurements, and detailed meteorological data to account for the influence of weather patterns. Additionally, to improve the accuracy of our models, techniques such as ridge regression or principal component analysis could be used to mitigate the effects of multicollinearity identified in our regression outputs. The limitations in our time series analysis could be addressed by fine-tuning model hyperparameters or exploring more advanced time series models to better account for seasonality, given the limited number of data points. Furthermore, a Generative Pre-Trained Transformer (GPT) may yield more fruitful results than statical modeling. GPTs run through data multiple times (epoch) at a set learning rate. By setting the rate to a small value, in conjunction with a high epoch rate, the GPT may output accurate forecast compared to statical modeling. This approach would require a decoder only GPT as all input and output values are numerical. This can be further expanded to incorporate multiple input variables that were not included in the SARIMAX model (Region, State, pathogen source etc.) to produce a more nuanced forecast of SSS, SST and Vibrio cases.

In summary, our study reinforces the significant role of warmer sea temperatures in driving *V. vulnificus* cases. However, our findings also highlight the intricate nature of marine disease dynamics and the need for ongoing research. Future efforts should prioritize improving model accuracy, incorporating a broader spectrum of climatic and biological factors, and addressing the inherent limitations in data availability and quality. A more comprehensive understanding of these factors is crucial for developing effective strategies to mitigate the public health risks posed by *V. vulnificus* in the context of a changing climate. For example, the European Centre for Disease Prevention and Control (ECDC) maintains a daily *Vibrio* risk index that monitors favorable environmental conditions for *Vibrio* growth. They assess variables such as sea surface temperature and salinity using satellite remote sensing in the Baltic Sea and surrounding areas [4]. When the risk is categorized as medium or higher, notifications are then reported on the Communicable Disease Threats Report (CDTR). This typically occurs during the summer months, when the sea surface temperatures rise and *Vibrio* risk increases. The ECDC *Vibrio* map is one of few surveillance models that utilize real-time environmental variables; however, geographically it is limited.

In the United States, the National Center for Coastal Ocean Science (NCCOS) has several predictive models for *Vibrio* species, primarily focusing on V. *parahaemolyticus* and *V. vulnificus.* For *V. parahaemolyticus,* several coastal regions (e.g. Alaska, Tampa Bay, Pacific Northwest) are closely monitored to assess the risk for potential shellfish contamination and support regulatory measures to mitigate transmission from raw or undercooked seafood. In contrast, *V. vulnificus* is monitored in Chesapeake Bay to provide water safety guidance through a forecasted model.

These models offer opportunities for improving public health interventions and raising awareness about the dangers of *Vibrio spp.*, specifically *V. vulnificus*, which is the most lethal. For example, during warmer months, advisory notices and public outreach can inform individuals about increased risk; this awareness may encourage them to take precautions to avoid exposure such as limiting raw seafood consumption or refraining from recreational swimming. Additionally, during high-risk periods, seafood industries could limit the transportation of seafood to prevent dormant *Vibrio* from becoming active, or ensure shellfish are thoroughly cooked at 63 °C to kill the bacteria. Therefore, enhancing these predictive models will not only help researchers identify emerging hotspots, but also support sociological efforts to reduce transmission.



Graph 10. Explains the environmental and sociological factors that influence pathogenic Vibrio spp. The double slashes are areas of opportunity for public health intervention. (Brumfield et al.,2021)

References

[1] Archer, E. J., Baker-Austin, C., Osborn, T. J., & others. (2023). Climate warming and increasing *Vibrio vulnificus* infections in North America. *Scientific Reports, 13*(3893). https://doi.org/10.1038/s41598-023-28247-2

[2] Baker-Austin, C., Trinanes, J., Taylor, N. G. H., Hartnell, R., Siitonen, A., & Martinez-Urtaza, J. (2020). Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Trends in Microbiology, 28*(1), 81–82. https://doi.org/10.1016/j.tim.2019.08.006

[3] Brumfield, K. D., Chen, A. J., Gangwar, M., Usmani, M., Hasan, N. A., Jutla, A. S., Huq, A., & Colwell, R. R. (2023). Environmental factors influencing occurrence of *Vibrio parahaemolyticus* and *Vibrio vulnificus*. *Applied and Environmental Microbiology, 89*(6), e0030723. https://doi.org/10.1128/aem.00307-23

[4] Brumfield, K. D., Usmani, M., Chen, K. M., Gangwar, M., Jutla, A. S., Huq, A., & Colwell, R. R. (2021). Environmental parameters associated with incidence and transmission of pathogenic *Vibrio* spp. *Environmental Microbiology, 23*(12), 7314–7340.

[5] Cleveland Clinic. (2025). *Vibrio vulnificus*. https://my.clevelandclinic.org/health/diseases/24884-vibrio-vulnificus

[6] Haftel, A., & Sharman, T. (2025). *Vibrio vulnificus* infection. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK554404/

[7] Hughes, M. J., Flaherty, E., Lee, N., Robbins, A., & Weller, D. L. (2024). Notes from the field: Severe *Vibrio vulnificus* infections during heat waves — Three eastern U.S. states, July–August 2023. *MMWR Morbidity and Mortality Weekly Report, 73*(4), 84–85. https://doi.org/10.15585/mmwr.mm7304a3

[8] Ma, L.-C., Li, M., Chen, Y.-M., Chen, W.-Y., Chen, Y.-W., Cheng, Z.-L., Zhu, Y.-Z., Zhang, Y., Guo, X.-K., & Liu, C. (2024). Genomic insight into zoonotic and environmental *Vibrio vulnificus*: Strains with T3SS2 as a novel threat to public health. *Microorganisms, 12*(11), 2375. https://doi.org/10.3390/microorganisms12112375

[9] Oliver, J. D. (2012). *Vibrio vulnificus*: Death on the half shell. A personal journey with the pathogen and its ecology. *Microbial Ecology, 65*(4), 793–799. https://doi.org/10.1007/s00248-012-0140-9

[10] Sheahan, M., Gould, C. A., Neumann, J. E., Kinney, P. L., Hoffmann, S., Fant, C., Wang, X., & Kolian, M. (2022). Examining the relationship between climate change and vibriosis in the United States: Projected health and economic impacts for the 21st century.

[11] Waidner, L. A., & Potdukhe, T. V. (2023). Tools to enumerate and predict distribution patterns of environmental *Vibrio vulnificus* and *Vibrio parahaemolyticus*. *Microorganisms, 11*(10), 2502. https://doi.org/10.3390/microorganisms11102502