

Article

The Use of Boosted Regression Trees to Predict the Occurrence and Quantity of *Staphylococcus aureus* in Recreational Marine Waterways

Bridgette F. Froeschke ^{1,*}, Michelle Roux-Osovitz ^{2,†}, Margaret L. Baker ^{2,3} , Ella G. Hampson ^{2,4}, Stella L. Nau ² and Ashley Thomas ²

¹ Department of Environmental Studies, The University of Tampa, 401 W. Kennedy Blvd., Tampa, FL 33606, USA

² Department of Biology, The University of Tampa, 401 W. Kennedy Blvd., Tampa, FL 33606, USA; mosovitz@ut.edu (M.R.-O.); margaret.baker@rice.edu (M.L.B.); ellagh@hawaii.edu (E.G.H.); stella.nau0599@gmail.com (S.L.N.); ashleythomas134@gmail.com (A.T.)

³ Department of Biosciences, Rice University, 6100 Main Street, Houston, TX 77054, USA

⁴ Department of Tropical Plant and Soil Sciences, University of Hawai'i at Mānoa, 2444 Dole Street, Honolulu, HI 96822, USA

* Correspondence: bfroeschke@ut.edu; Tel.: +1-813-257-4004

† These authors contributed equally to this work.

Abstract: Microbial monitoring in marine recreational waterways often overlooks environmental variables associated with pathogen occurrence. This study employs a predictive boosted regression trees (BRT) model to predict *Staphylococcus aureus* abundance in the Tampa Bay estuary and identify related environmental variables associated with the microbial pathogen's occurrence. We provide evidence that the BRT model's adaptability and ability to capture complex interactions among predictors make it invaluable for research on microbial indicator research. Over 18 months, water samples from 7 recreational sites underwent microbial quantitation and *S. aureus* isolation, followed by genetic validation. BRT analysis of *S. aureus* occurrence and environmental variables revealed month, precipitation, salinity, site, temperature, and year as relevant predictors. In addition, the BRT model accurately predicted *S. aureus* occurrence, setting a precedent for pathogen–environment research. The approach described here is novel and informs proactive management strategies and community health initiatives in marine recreational waterways.

Keywords: boosted regression trees (BRT); *S. aureus*; monitoring; aquatic microbes; predictive models



Citation: Froeschke, B.F.; Roux-Osovitz, M.; Baker, M.L.; Hampson, E.G.; Nau, S.L.; Thomas, A. The Use of Boosted Regression Trees to Predict the Occurrence and Quantity of *Staphylococcus aureus* in Recreational Marine Waterways. *Water* **2024**, *16*, 1283. <https://doi.org/10.3390/w16091283>

Academic Editor: Abasiofiok Mark Ibekwe

Received: 9 April 2024

Revised: 18 April 2024

Accepted: 20 April 2024

Published: 30 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The accurate detection and prediction of the occurrence and quantity of pathogenic microbes in marine recreational waterways requires a comprehensive understanding of the interactions among environmental variables, temporal dynamics, and spatial dimensions. Routine monitoring of microbial indicator species abundance (such enteric fecal indicators enterococci and *E. coli* and non-fecal contaminants such as *Staphylococcus aureus*) contributes to effective management strategies, with the goal of protecting both public health and the ecological integrity of vital aquatic ecosystems [1–6]. These indicator species are important as sentinel pathogens, bridging environmental science and human health. *S. aureus* microbial pathogen in particular poses significant risks to both aquatic ecosystems and human health by causing a range of infections from mild to severe [2,7]. *S. aureus* prevalence in rivers, estuaries, and coastal waters underscores the importance of understanding its abundance in these environments, which can serve as an indicator of potential health hazards for both aquatic life and human populations relying on these ecosystems [2,8]. The main sources of *S. aureus* in waterbodies is from human activity [9], stormwater [10–12], coastal streams that drain to the coast [4], and wastewater [2,13]. Changes in the environment due

to natural/seasonal or anthropomorphic activities influence the proliferation of *S. aureus* in water bodies, potentially leading to waterborne disease outbreaks [2,3,8,14].

Current approaches for microbial monitoring involve the identification and quantification of indicator species including *S. aureus* using real-time sampling and end-point detection. This can be labor and time-intensive, expensive, and does not aim to contribute to our understanding of the indicator microbe's ecology [1,15,16]. In addition to monitoring, predictive models are being utilized as proactive water quality management tools for target microbial indicator species. Successful applications of predictive modeling in microbial ecology include the prediction of harmful algal blooms [17] and the estimation of fecal contamination in water bodies [18–20]. The integration of monitoring and predictive modeling approaches can be achieved using statistical models such as the least-angle regression (LARS-lasso) model, boosted regression trees (BRT), and others to increase accuracy and efficacy in advisory efforts [21–25].

Predictive models have also been used to assess the complex interactions between environmental variables and microbial abundance and to facilitate the identification of key factors influencing microbial dynamics [18–20]. The use of predictive models therefore aims at the integration of microbial quantification with user-defined variables that are specific to the habitat under study. Variables include, but are not limited to, spatial, temporal, and environmental variables such as precipitation, temperature, and salinity that can be used to better elucidate the complex interactions within the microbiota of aquatic ecosystems. These variables represent a subset of those that play a crucial role in shaping microbial ecosystems; influencing the survival, proliferation, and transport mechanisms of pathogenic bacterial species such as *S. aureus* [2,19]. Predictive models have been successfully used in both freshwater and marine waterways for predicting the occurrence of fecal microbes [18–21,24,25], suggesting that a modeling approach has the potential to increase our predictive power for other aquatic microbial species as well. For example, using BRT as the predictive model to understand the correlation between environmental variables and *S. aureus* abundance can provide valuable insight into the occurrence and ecosystem dynamics of these bacteria in marine recreational waterways [2,4,18]. Furthermore, these models hold the potential for forecasting disease outbreaks in aquatic environments [26], informing evidence-based decision making for water quality management and public health protection [18].

Of particular interest in this study is the use of the BRT model as it offers a practical methodological approach to gaining valuable insights into the complex interactions between how environmental variables, pollution, and anthropogenic interventions exert discernible impacts on the proliferation and persistence of *S. aureus* in marine recreational waterways [2,4,27]. BRT epitomizes a robust and versatile modeling paradigm that accommodates a heterogeneous array of predictor variables, spanning physical measurements, spatial attributes, and categorical descriptors; this endows the BRT approach with unparalleled versatility in capturing the multifaceted interrelationships between pathogenic bacteria and their environment. Additionally, BRT demonstrates robustness in handling missing data values and outliers, making it suitable for ecological studies where incomplete or anomalous data are common [28–30]. By combining the strengths of regression trees and boosting algorithms, BRT creates an ensemble modeling approach that outperforms traditional single-tree methods in predictive accuracy [28–30]. This ensemble framework, supported by cross-validation techniques, reduces the risk of overfitting, and improves the model's ability to generalize to new datasets [28–30].

This study aimed to utilize the temporal and spatial detection of *S. aureus* coupled with known temporal, spatial, and environmental variables as an example to test how a BRT model can be employed to predict levels of pathogenic bacteria in marine recreational waterways. Firstly, the study sought to establish the viability of isolating and identifying *S. aureus* at seven diverse recreational sites over a two-year period in the Tampa Bay estuary, Florida. Second, the study aimed to predict levels of pathogenic Indicator bacteria, leveraging temporal, spatial, and environmental variables, and *S. aureus* detection

levels as inputs in a BRT model. The aim of the study was to advance our understanding of the relationship between *S. aureus*, environmental parameters, and water quality, ultimately contributing to the development of effective management strategies for protecting human health and ecosystem integrity in marine recreational waterways like the Tampa Bay estuary.

2. Materials and Methods

2.1. Detection and Quantitation of *S. aureus* in Marine Waterways

2.1.1. Field Collection and Laboratory Microbial Isolation and Analysis

Water samples for microbial analysis were systematically collected from the following 7 recreational sites in Tampa Bay; Gandy Beach (GB) (−82.59566, 27.87459), Ben T. Davis (BD) (−82.57884, 27.97045), Cypress Pt. Park (CP) (−82.54697, 27.95016), Picnic Island (PI) (−82.55415, 27.85183), Davis Island (DI) (−82.44721, 27.902972), Bahia Beach (BB) (−82.47657, 27.72894), and E. G. Simmons Park Beach (EGS) (−82.47278, 27.74772) (Figure 1). The sampling sites, chosen for their extensive recreational usage, were strategically selected to capture the potential influence of human recreational activities on marine water contamination. Eighteen sampling events were conducted from September 2019 to July 2021. All water samples were collected using sterilized bottles (Fisher Scientific, Atlanta, GA, USA), adhering to EPA standards (<https://www.epa.gov/sites/default/files/2018-10/documents/quality-criteria-water-1986.pdf> accessed on 8 April 2024). including opening and closing the bottles beneath the water surface at knee-deep levels (0.5 m). Notably, all water samples were collected and stored on ice and returned to the lab for processing within three hours of collection. A Hydrolab (Sterling, VA, USA) was used to measure salinity and temperature during each sampling event.

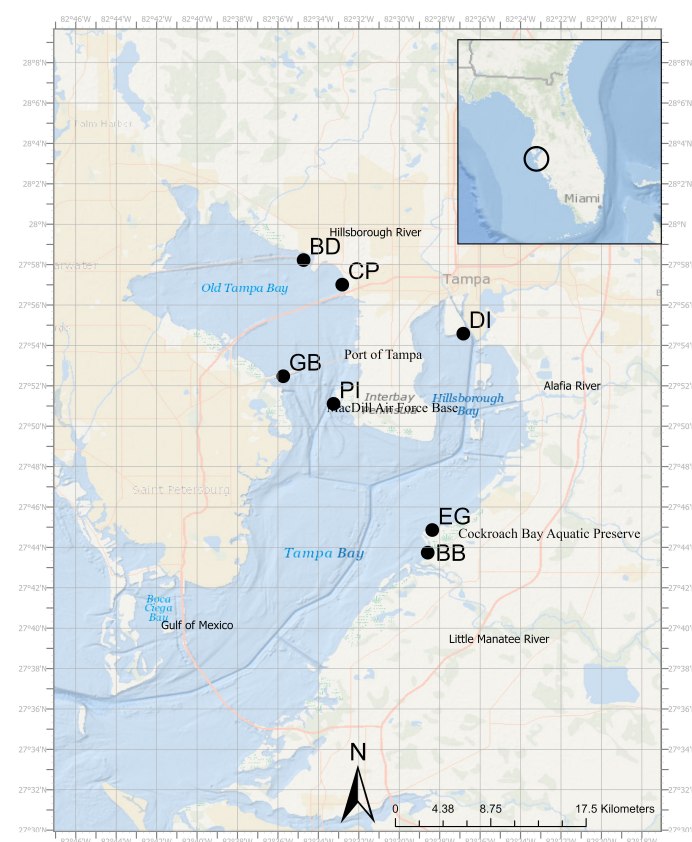


Figure 1. The seven sampling sites in the Tampa Bay estuary, Florida, (GB = Gandy Beach, BD = Ben T. Davis, CP = Cypress Pt. Park, PI = Picnic Island, DI = Davis Island, BB = Bahia Beach, and EG = E.G. Simmons Park Beach). Sampling events were conducted monthly between June 2019 and May 2021, with a total of 18 sampling events.

Water processing methods followed the protocols outlined in the book “Standard Methods for the Examination of Water and Wastewater” [31]. Filtration was executed using a vacuum-operated manifold (Hach Company, Loveland, CO, USA), capturing bacteria on 0.45 µm membrane filters (Hach Company, Loveland, CO, USA). Subsequently, selective, and differential mannitol salt agar (MSA) (Fisher, Waltham, MA, USA) fermentation was used for isolation, followed by an incubation period at 37 °C for 24 h. The resulting yellow colonies were enumerated as *S. aureus*. Overall, 10 random samples per site per sampling event ($n = 70$) were preserved at 4 °C for further biochemical testing. All samples were completed within a 24-h period of refrigeration. For verification purposes, all samples designated for biochemical testing were tested for hemolysis on blood agar (Fisher, Waltham, MA, USA) and validated for the presence of coagulase.

2.1.2. Genetic Validation: DNA Extraction and PCR Amplification of Nuclease Gene

Individual bacterial colonies selected by MSA fermentation and coagulase positive were isolated using a sterile technique and transferred to a 6% InstaGene Matrix (BioRad, Hercules, CA, USA) in nuclease-free water (Fisher, Waltham, MA, USA). Samples were temporarily stored at −20 °C. Thawed samples were vortexed to disperse the bacterial colony and cells were lysed at 98 °C for 8 min. Prior to molecular verification by polymerase chain reaction (PCR), samples were vortexed a second time and centrifuged at 4 °C for 3 min at 12,000 rpm to separate supernatant containing DNA from the InstaGene matrix. *S. aureus* specific thermonuclease gene (Nuc) was amplified using GoTaq Master Mix chemistry (Promega, Madison, WI, USA) with forward primers (5'-GCG ATT GAT GGT GAT ACG GTT-3') and reverse primers (5'-AGC CAA GCC TTG ACG AAC TAA AGC-3') [32]. Specifically, nuclease primers used were designed from a 966 bp sequence of the nuclease A gene from the *S. aureus* Foggi strain to amplify a 279 bp sequence [33,34]. The 279 bp amplicon reports 100% identity to NCBI Accession #CP150135.1. A random sub-sampling of *S. aureus* nuclease PCR amplicons was sequenced to confirm the positive identity of isolates as well as control *S. aureus* N315 strain amplicons (NCBI Accession # NC_002745.2).

Each 20 µL of reaction contained 12.5 µL 2X GoTaq Master Mix, 0.75 µL of forward (10 mM) and 0.75 µL of reverse (10 mM) Nuc primers, 5 µL of template DNA, and 6 µL of nuclease-free water. PCR thermocycling parameters were 94 °C for 10 min followed by 32 cycles of 94 °C for 60 s, 51 °C for 60 s, and 72 °C for 120 s with a final extension at 72 °C for 10 min. PCR reactions were stored at 4 °C until subjected to 1% agarose gel electrophoresis to verify a 279 base pair Nuc gene-specific amplicon. Agarose gels were stained using ethidium bromide (Biorad, Hercules, CA, USA) and imaged using a ChemiDoc (BioRad, Hercules, CA, USA). *S. aureus* N315 strain was cultured, and DNA was extracted as above and used as a positive control in PCR reactions. Samples without successful amplification of the Nuc gene were further subjected to validation using amplification of the 16S gene using forward primers (5'-AGA GTT TGA TCC TGG CTC AG-3') and reverse primers (5'-GGT TAC CTT GTT ACG ACT T-3') [35]. Specifically, the 16S primers used to amplify a region of the 16S sequence were used for phylogenetic analysis of a wide range of bacteria [35], amplifying a 1465 bp sequence. Furthermore, internal 16S sequencing primers are used for microbial identification and current data indicate when that the *Staphylococcus* genus is detected they are not *aureus* species-specific (unpublished data). The GoTaq Master Mix reactions, as detailed above, use the following thermocycling parameters: 94 °C for 5 min followed by 35 cycles of 94 °C for 120 s, 42 °C for 30 s, and 72 °C for 120 s with a final extension at 72 °C for 10 min.

2.2. BRT Model

This study aimed to explore the relationship between temporal, spatial, and environmental variables, and the quantity of *S. aureus* in the Tampa Bay estuary, Florida. The *S. aureus* quantity was logarithmically transformed using a base of 10 and the resulting transformed data were employed as input for the BRT model. We employed a forward

fit, stage-wise, Gaussian-boosted regression trees model [36]. These parameters included setting the distribution as “Gaussian” to accommodate continuous outcome variables, a tree complexity of 5 to control the complexity of each tree and permit modeling of interactions, a learning rate of 0.005 to determine the weight applied to each tree, and a bag fraction of 0.60 to specify the proportion of the training data used to fit each tree. Unlike traditional regression methods, the BRT model harnesses the strength of both regression trees and boosting algorithms by combining numerous simple tree models. Each of these models comprises a rule-based classifier that groups observations based on predictor variables, allowing for interactions between predictors [27]. To prevent overfitting, a 10-fold cross-validation process was incorporated into the model fitting procedure [28–30]. Additionally, model variability was assessed by refitting the BRT model using 1000 bootstrap iterations with the replacement of the original dataset, ultimately determining mean predicted probabilities and 95% confidence intervals within the study area.

Rstudio software (R version 4.2.3, R Development Core Team) with the ‘gbm’ library supplemented by functions from Elith et al. (2008) [29] was utilized for the analyses. Predictor variables included temporal variables (month and year), site location, and environmental variables as well as precipitation levels in inches within 24 h of sampling, salinity (ppt), and water temperature (°C). Model parameters such as tree complexity, learning rate, and bag fraction were adjusted to optimize model performance while minimizing overfitting. The selection of predictor variables and tuning parameters was based on performance metrics and explained deviance on cross-validated data. BRT model validation involved testing the null hypothesis and utilizing a least-squares linear regression to determine the accuracy of the predicted quantity of *S. aureus* versus the actual calculated quantity.

3. Results and Discussion

3.1. Detection and Quantitation of *S. aureus* in Tampa Bay Estuary

The Tampa Bay estuary, situated in west central Florida, is the largest open water estuary in Florida, encompassing 5698 km² [37]. Understanding the occurrence and ecological role of pathogenic bacteria in the Tampa Bay estuary requires consideration of the estuary’s unique hydrology and recreational use [3,12,13,38]. Four main rivers on the eastern side, the Alafia River, the Hillsborough River, the Little Manatee River, and the Manatee River (Figure 1), contribute freshwater to the Tampa Bay estuary. This freshwater input, particularly notable in northern Tampa Bay, significantly influences water flow patterns [38], (<https://tbep.org/> accessed on 8 April 2024), salinity [39], and potentially the distribution of pathogenic bacteria [3,8].

Fixed sampling sites were strategically selected to represent various regions within the Tampa Bay estuary (Figure 1). These fixed sites served as reliable benchmarks for monitoring changes in water quality over time and were integral to understanding the overall health of the Tampa Bay estuary aquatic environment amidst recreational usage. Bi-monthly water sampling at each site spanned from June 2019 to May 2021 (totaling 18 events). The summary statistics for log₁₀ abundance of *S. aureus* revealed a wide range of values (0.78–5.31), with a positively skewed distribution (Figure 2A).

When examining *S. aureus* levels by site, distinct patterns emerged (Figure 2B). PI exhibited the highest mean log value at 3.11, indicating elevated *S. aureus* levels compared to other sites. PI is positioned between the Port of Tampa and MacDill Airforce Base at the intersection of Hillsborough Bay and Old Tampa Bay (Figure 1). Elevated levels of *S. aureus* detected at PI compared to other sites are likely due to increased public access and designated swimming areas as well as PI’s unique position in the Tampa Bay estuary relative to commercial and recreational use (Figure 1). Conversely, BB had the lowest mean log value at 2.65. BB is in a less densely populated region of Tampa Bay, has restricted access, and is surrounded by Cockroach Bay Aquatic Preserve and areas with lower population density and higher agricultural development (https://tbep.org accessed on 8 April 2024) (Figure 1). BB is north of the Little Manatee River, which contributes to a rich estuary

habitat. The ecological significance of the estuary environment (BB) vs. port environment (PI) could be a contributing factor to the differing microbial ecology at these two sites.

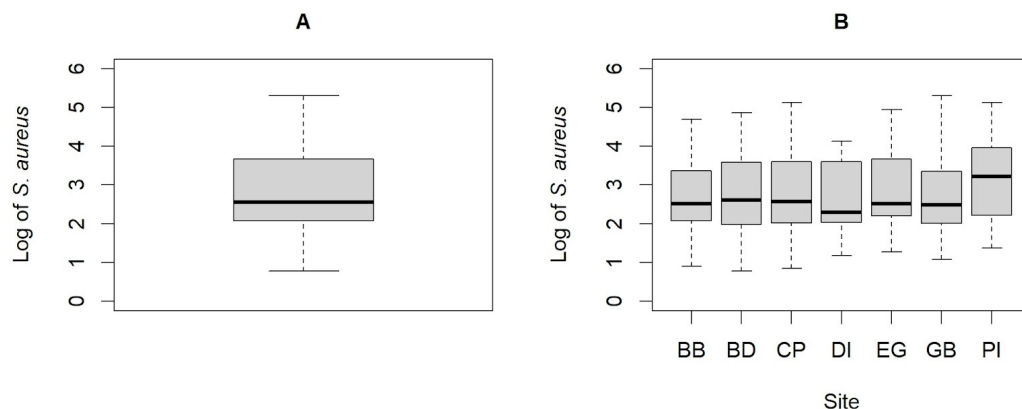


Figure 2. Boxplots illustrating the distribution of log-transformed *Staphylococcus aureus* levels overall (A) and by individual sampling sites (B). The boxplots depict the minimum, first quartile, median, mean, third quartile, and maximum values of log-transformed *S. aureus* levels.

Standard deviations calculated across all seven sites varied, with site EG displaying the lowest variability (0.93) and GB exhibiting the highest variability (1.11) with respect to detected *S. aureus* (Figure 2B). The range of *S. aureus* log values was also site-dependent, with PI having the widest range from 1.38 to 5.12 and DI having the narrowest range from 1.18 to 4.12 (Figure 2B). These site-specific statistics underscored the heterogeneity in *S. aureus* levels between sampling locations, emphasizing the importance of site-specific monitoring and management strategies. Being that each of the seven sites has unique positions in the Tampa Bay estuary and unique seasonal recreational uses, further experiments are needed to elucidate microbial dynamics at each site. Successful detection and quantitation of *S. aureus* across this large marine waterway provides a novel monitoring strategy both spatially and temporarily.

3.2. Testing of Trained BRT Model to Predict Occurrence of *S. aureus* in the Tampa Bay Estuary

The BRT model employed in this study to predict the occurrence of *S. aureus* was configured with specific parameters to optimize its predictive performance. The resulting regression model provided good predictive accuracy (R-squared value of 0.67; Figure 3). This level of explanatory power provides substantial predictive modeling of ecological data. Additionally, the linear equation derived from the regression model ($y = 1.77 + 0.37 \times x$) provided a concise representation of the relationship between the predictor and response variables, offering valuable insights into the underlying dynamics of *S. aureus* bacterial abundance (Figure 3). Overall, the BRT model with an R-squared value of 0.67 represents a highly effective tool for predicting *S. aureus* bacteria levels in aquatic environments, facilitating informed decision-making and management strategies for water quality assessment and public health protection.

Our study demonstrated a robust predictive capability in estimating *S. aureus* bacterial levels in a marine recreational waterway. These findings underscore the effectiveness of the BRT model in capturing the complex interplay between environmental variables and *S. aureus* bacterial abundance. Furthermore, our study contributes to the growing body of literature on predictive modeling in microbial ecology [18], emphasizing the importance of integrating advanced statistical techniques with ecological principles to enhance our understanding of microbial dynamics in aquatic systems [18,27]. By leveraging predictive modeling, our research offered valuable insights into the factors driving bacterial abundance in marine waters (discussed below), facilitating evidence-based decision making for water quality management and public health protection.

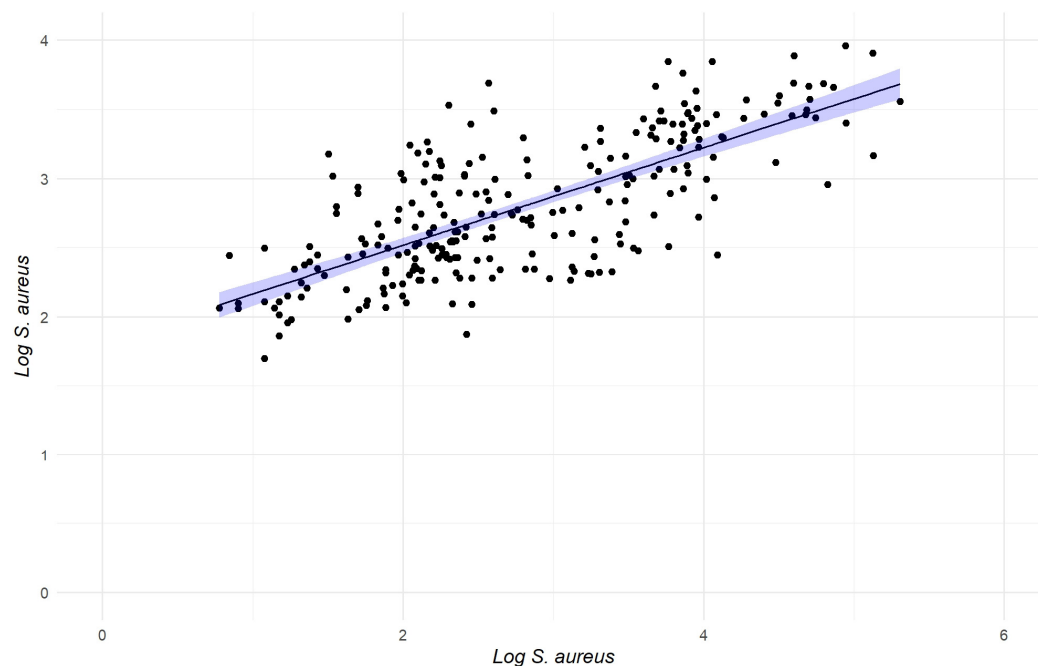


Figure 3. Correlation between predicted log-transformed *S. aureus* concentrations, derived from boosted regression trees (BRT) model, and the observed log concentrations. The line model was $y = 1.77 + 0.37 \times x$, with an R^2 value of 0.67.

3.3. The BRT Model Predicts the Influence of Temporal Variables on the Occurrence of *S. aureus* in the Tampa Bay Estuary

Month, as a temporal variable, emerged with the highest relative importance (36.59%, Figure 4), indicating a discernible seasonal pattern in the abundance of *S. aureus* in marine waters (Figure 5A). Throughout the year, predicted levels of *S. aureus* remained consistently high (Figure 5A), exhibiting a gradual increase from January, followed by declines in May, July, and September (Figure 5A). Notably, September marked a significant decline, while October and November witnessed a sharp increase in *S. aureus* abundance (Figure 5A). The temporal variable, year of collection, had a lower relative importance of 6.97% (Figure 5B) in predicting the occurrence of *S. aureus* in the Tampa Bay estuary (Figure 5B). The model predicted an exponential increase in the occurrence of *S. aureus* in 2020 and an exponential decrease in 2021 (Figure 5B).

Taken together, there are temporal patterns in occurrence and abundance of *S. aureus* over short-term (month) and long-term (year) scales (Figure 5). The distinct short-term seasonal fluctuations in *S. aureus* abundance, typically peaking during warmer months (June and August), are characterized by elevated water temperatures, and increased recreational activities. These trends coincide with periods of heightened human activity and environmental disturbances, such as storm runoff and nutrient inputs, which can foster bacterial growth and persistence in marine ecosystems [2,40]. Conversely, colder months often see reduced *S. aureus* levels (January–May) due to decreased microbial activity and limited opportunities for bacterial dissemination. Understanding these seasonal patterns is vital for implementing targeted monitoring and management strategies to mitigate potential health risks associated with *S. aureus* contamination in marine recreational waterways.

It is possible that the long-term temporal patterns seen are due to the increase in recreational uses during the COVID-19 pandemic. Specifically, a rapid increase in the occurrence of *S. aureus* in February 2020 (Figure 5, month 2) was followed by sustained high levels of *S. aureus* in March and April 2020 (Figure 5, months three and four). Additionally, a decrease in *S. aureus* follows in May 2020, coinciding with the reopening of Florida (Figure 5, month five). To our knowledge, the data presented here are the first reporting monitoring *S. aureus* in the Tampa Bay estuary during the pandemic. Further studies to

assess the model's accuracy over long temporal scales and over a wider range of recreational sites would lend additional support to the strength of the BRT model approach.

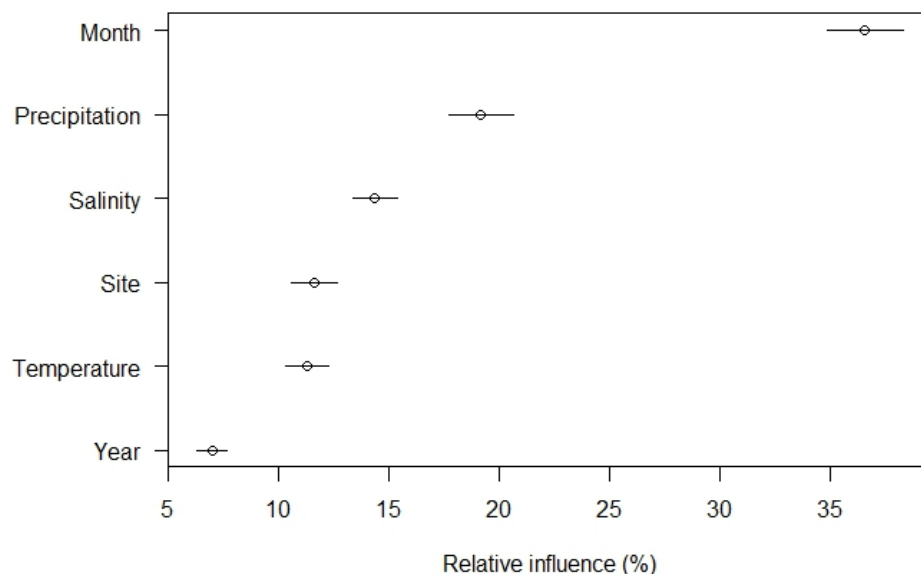


Figure 4. Relative influence, along with ± 1 standard deviation (SD), for pivotal predictive parameters, ranked in descending order of magnitude for the BRT model.

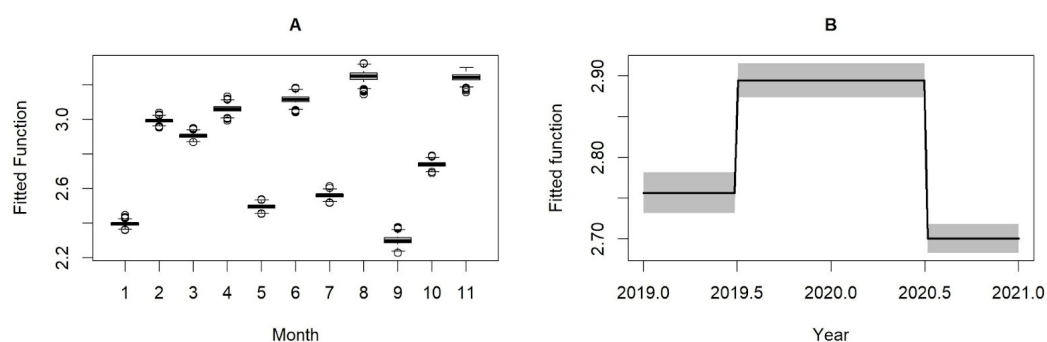


Figure 5. Temporal patterns and trends in *S. aureus* levels predicted by the BRT model, with bootstrapped confidence intervals enhancing prediction reliability. The y-axis represents the fitted function values from the model. (A) Predicted values (2.2–3.1) of *S. aureus* by month. (B) Predicted values (2.7–2.9) of *S. aureus* by year.

Furthermore, we hypothesize that the rapid decline in *S. aureus* occurrence in 2021 (Figure 5B) as well as the monthly overall decline in July and September (Figure 5A) were potentially due to an increase in the presence of *Karenia brevis* (the dinoflagellates responsible for red tide) in the Tampa Bay estuary. According to Florida Fish and Wildlife Harmful Algal Bloom (FWC-HAB) monitoring in 2019 and 2020, *K. brevis* was not present or present at very low levels; however, from 2021 June through to October FWC-HAB, medium to high levels of *K. brevis* in the Tampa Bay estuary and surrounding marine waterways were reported (<https://myfwc.com/research/redtide/statewide/> accessed on 8 April 2024). Although little is known from the field regarding interactions between *S. aureus* and *K. brevis*, the research in bacterioplankton ecology provides evidence that microbial communities are significantly impacted by plankton bloom dynamics [41–45]. Recent studies off the west coast of Florida focused on the influence of *K. brevis* abundance and chemical metabolites on microbial communities [46], providing evidence for the importance of further studying dynamic interactions within the bacterioplankton community, especially known pathogens such as *S. aureus*.

3.4. The BRT Model Predicts the Influence of Environmental Variables on the Occurrence of *S. aureus* in the Tampa Bay Estuary

The environmental variable, precipitation, had the second-highest relative importance (19.18%; Figure 4) in predicting *S. aureus* levels in the Tampa Bay estuary. The BRT model indicated that the highest occurrence of *S. aureus* was predicted when precipitation exceeded 1 inch within a 24-h period of sampling in Tampa Bay (Figure 6A). This finding aligns with previous research suggesting that stormwater could serve as a source of human pathogenic bacteria in recreational water bodies [2,8,14]. Notably, results from Froeschke et al. (2019) [8], focusing on the Hillsborough River that discharges into the Tampa Bay estuary, reported higher levels of *S. aureus* during wet weather events, indicating potential contamination correlating with high precipitation levels. Studies conducted at selected beaches in California also reported a higher detection rate of *S. aureus* in water discharge samples during wet weather events, suggesting increased contamination [2]. Overall, these findings underscore the potential impact of stormwater on bacterial contamination levels in marine waters and highlight the need for effective stormwater management strategies to mitigate public health risks associated with waterborne pathogens.

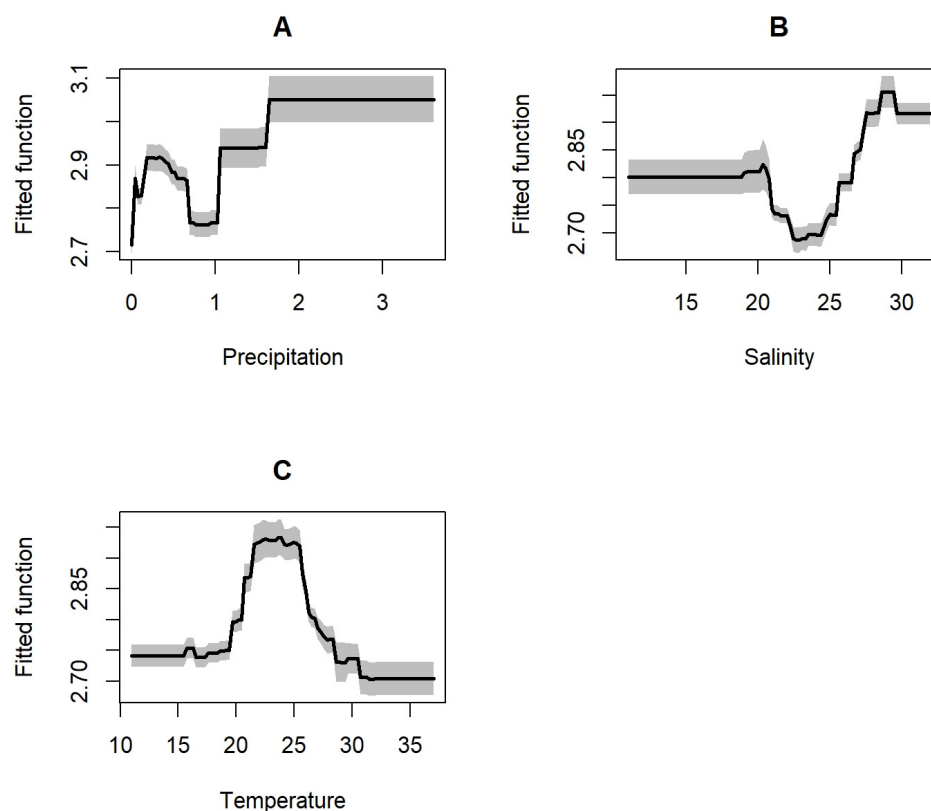


Figure 6. Environmental patterns and trends in *S. aureus* levels predicted by the BRT model, with bootstrapped confidence intervals enhancing prediction reliability. The y-axis represents the fitted function values ranging from 2.7 to 3.1. (A) Predicted values of *S. aureus* by precipitation levels. (B) Predicted values of *S. aureus* by salinity. (C) Predicted values of *S. aureus* by temperature.

Salinity was the environmental variable that had the third highest relative influence on the occurrence of *S. aureus* in the Tampa Bay estuary (14.35%; Figure 4). Within the salinity range of 20 to 25 ppt, the BRT model revealed the lowest occurrence of *S. aureus*, followed by an exponential increase at 25 ppt and stabilization at 29 ppt (Figure 6B). These findings underscore the significant impact of salinity on the abundance and distribution of this pathogen in aquatic environments. In estuarine bays, freshwater inflow acts as a critical regulator of salinity levels, profoundly influencing the equilibrium of marine

ecosystems [47]. Adequate freshwater input is essential for maintaining optimal salinity levels, which are vital for sustaining diverse habitats and aquatic life.

Freshwater inputs create fluctuations in salinities within estuary environments that affect microbial communities, especially at the marine-freshwater boundaries and in large coastal bays [48–50]. According to the Tampa Bay Estuary program (<https://tbep.org/> accessed on 8 April 2024), the salinity between the lower and northern regions of the Tampa Bay estuary ranges from 25–38 ppt and 18–32 ppt, respectively, with the lowest salinities seen in Hillsborough Bay (15–30 ppt) (see Figure 1). The salinity gradients throughout the Tampa Bay estuary are regularly monitored by The Tampa Bay Estuary Program (<https://tbep.org/> accessed on 8 April 2024) and the salt balance in the bay can be modeled both horizontally and vertically [39]. The wide breadth of data available regarding salinity as an environmental variable could be harnessed in future studies, using BRT to accurately predict microbial community occurrence related to tidal flux, freshwater input, and storm runoff and provide valuable information for public health and microbial ecology.

Temperature also influenced the occurrence of *S. aureus* in the Tampa Bay estuary (11.29%, Figure 4). According to the BRT model, there was an exponential increase in *S. aureus* occurrence within temperature ranges of 21 °C to 25 °C, followed by a substantial decrease at 26 °C (Figure 6C). This underscores the critical role of temperature in shaping the dynamics of *S. aureus* populations in marine waterways. Temperature affects the growth, survival, and activity of bacteria, including *S. aureus*, with optimal ranges promoting proliferation and potentially higher contamination levels. Understanding these temperature preferences is crucial for assessing the risk of *S. aureus* contamination in marine environments and implementing targeted mitigation strategies. Additionally, temperature fluctuations could influence microbial community composition and ecosystem dynamics [51], further highlighting the importance of considering temperature dynamics in monitoring and managing water quality in coastal regions.

3.5. The BRT Model Predicts the Influence of Spatial Variables on the Occurrence of *S. aureus* in the Tampa Bay Estuary

The spatial variable, site, had a relative importance value of 11.62% (Figure 4). The BRT model successfully predicted the occurrence of *S. aureus* (Figure 2) and reflected the patterns observed in the data (Figure 7). Specifically, the BRT model indicated that the highest levels of *S. aureus* were present at PI, EGS, and GB (Figure 7), while BD, CP, DI, and BB clustered together with the lowest levels of *S. aureus* (Figure 7). The BRT model indicates variation in *S. aureus* across different sites, underscoring the unique characteristics of grouped locations (Figure 7). For example, high levels of *S. aureus* occur at sites where there is increased recreational use by ease of water access and designated swim areas with increased sandy beach vs. rocky shoreline (PI, EGS, and GB) whereas lower levels of *S. aureus* occur at sites that have restricted or private access and rocky shoreline vs. sandy beach (BD, CP, DI, and BB). The areas of lower *S. aureus* are likely frequented by fishermen and recreational boaters and the areas of higher *S. aureus* are likely to have more foot traffic, swimmers, and beach use. It is worth noting that spatial variability detected by the BRT model in *S. aureus* abundance might also be attributed to differences in environmental factors such as nutrient availability and water circulation patterns in addition to human activities, which can vary significantly between sites and influence bacterial growth and distribution. Understanding these site-specific patterns is crucial for targeted monitoring and management efforts to mitigate the risk of *S. aureus* contamination in marine recreational waterways.

The results for each variable aligned with the biology of *S. aureus*, reflecting known ecological patterns of this bacterial species. The success in predicting log abundance, as evidenced by high prediction accuracy and comparisons with linear regression, represents a notable advancement in environmental monitoring and public health considerations. These findings, anchored in the biological significance of each variable, provide valuable insights for future research, and contribute to a deeper understanding of microbial dynamics in complex ecosystems. Moreover, elucidating the factors driving *S. aureus* abundance can

inform targeted interventions and management strategies to mitigate potential health risks associated with this pathogen, underscoring the importance of integrating ecological knowledge into public health initiatives.

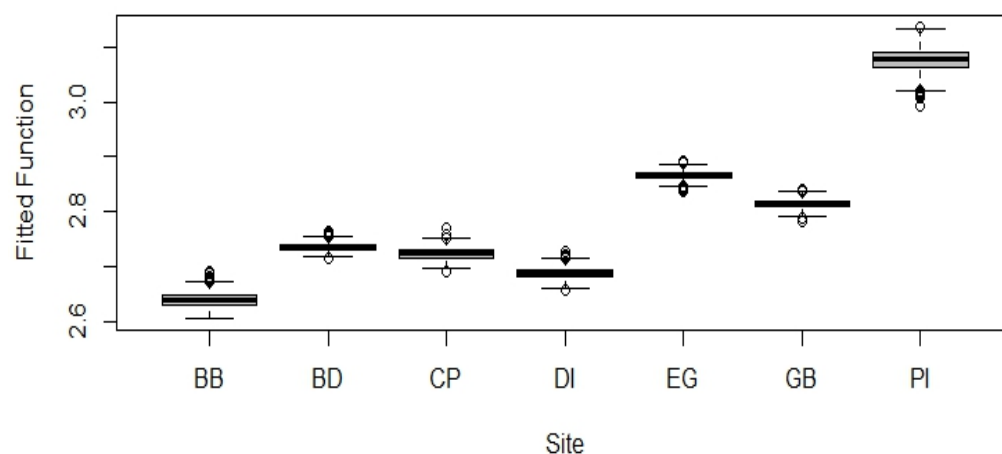


Figure 7. Spatial patterns and trends in *S. aureus* levels predicted by the BRT model, with bootstrapped confidence intervals enhancing prediction reliability. The *y*-axis represents the fitted function values ranging from 2.6 to 3.1.

4. Conclusions

Our study developed a BRT model of predictive levels of *S. aureus* in the Tampa Bay estuary, Florida. The model demonstrated good predictive performance and provided insight into the environmental and human-use factors influencing the observed patterns. Specifically, the BRT model identified the relative contributions of temporal, environmental, and spatial variables on the abundance of *S. aureus*. Month, precipitation, and salinity emerged as primary contributors to *S. aureus* occurrence, with distinct seasonal patterns observed throughout the year. Stormwater from precipitation events was identified as a potential source of bacterial contamination, highlighting the importance of effective stormwater management strategies. Additionally, salinity levels played a crucial role in shaping *S. aureus* abundance, with optimal ranges influencing bacterial growth and distribution. Furthermore, spatial variability in *S. aureus* abundance underscored the unique characteristics of different sites within the Tampa Bay estuary, emphasizing the need for site-specific monitoring and management efforts. By providing a robust framework for understanding how environmental factors influence pathogen abundance, this methodology contributes to the advancement in our knowledge of pathogen occurrence and abundance and provides insights into understanding marine microbial ecology and epidemiology.

Author Contributions: The work presented in this manuscript is novel research conducted in collaboration with all co-authors. B.F.F. and M.R.-O. contributed equally to this work. Conceptualization, methodology, validation, and supervision were carried out in collaboration with B.F.F. and M.R.-O. Resource allocation and funding acquisition by B.F.F. and M.R.-O. Experimental investigation and data curation were conducted by B.F.F., M.R.-O., M.L.B., E.G.H., S.L.N. and A.T. R software and formal analysis was conducted by B.F.F. B.F.F. and M.R.-O. contributed equally to writing—original draft preparation and visualization. All authors contributed to writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Tampa Bay Estuary Research Foundation, grant number GR2996. The APC was funded by The University of Tampa College of Natural and Health Sciences.

Data Availability Statement: The original data presented in the study are openly available in Tampa Bay Water Atlas at <https://tampabay.wateratlas.usf.edu/> accessed on 8 April 2024.

Acknowledgments: We extend our appreciation to the following dedicated students who contributed to the field collections and laboratory analysis: Tala Bleau, Zoe Golden, Hayley Kudzmas, Shannon Twyman, Brittany Small, Brendon Smith, Christian Gibbs, Mathew Olszewski, Michelle Satine, and Katie Sherman. Their tireless efforts and commitment were invaluable to the success of this study, and we are deeply thankful for their contributions. We thank John Froeschke for reviewing the document.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Lugo, J.L.; Lugo, E.R.; de la Puente, M. A systematic review of microorganisms as indicators of recreational water quality in natural and drinking water systems. *J. Water Health* **2020**, *19*, 20–28. [\[CrossRef\]](#)
2. Goodwin, K.D.; McNay, M.; Cao, Y.; Ebentier, D.; Madison, M.; Griffith, J.F. A multi-beach study of *Staphylococcus aureus*, MRSA, and enterococci in seawater and beach sand. *Water Res.* **2012**, *46*, 4195–4207. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Hassard, F.; Gwyther, C.L.; Farkas, K.; Andrews, A.; Jones, V.; Cox, B.; Brett, H.; Jones, D.L.; McDonald, J.E.; Malham, S.K. Abundance and Distribution of Enteric Bacteria and Viruses in Coastal and Estuarine Sediments—A Review. *Front. Microbiol.* **2016**, *7*, 1692. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Viau, E.J.; Goodwin, K.D.; Yamahara, K.M.; Layton, B.A.; Sassoubre, L.M.; Burns, S.L.; Tong, H.-I.; Wong, S.H.C.; Lu, Y.; Boehm, A.B. Bacterial pathogens in Hawaiian coastal streams—Associations with fecal indicators, land cover, and water quality. *Water Res.* **2011**, *45*, 3279–3290. [\[CrossRef\]](#)
5. Curiel-Ayala, F.; Quiñones-Ramírez, E.I.; Pless, R.C.; González-Jasso, E. Comparative studies on *Enterococcus*, *Clostridium perfringens* and *Staphylococcus aureus* as quality indicators in tropical seawater at a Pacific Mexican beach resort. *Mar. Pollut. Bull.* **2012**, *64*, 2193–2198. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Valeriani, F.; Giampaoli, S.; Buggiotti, L.; Gianfranceschi, G.; Romano Spica, V. Molecular enrichment for detection of *S. aureus* in recreational waters. *Water Sci. Technol.* **2012**, *66*, 2305–2310. [\[CrossRef\]](#)
7. Kuehnert, M.J.; Kruszon-Moran, D.; Hill, H.A.; McQuillan, G.; McAllister, S.K.; Fosheim, G.; McDougal, L.K.; Chaitram, J.; Jensen, B.; Fridkin, S.K.; et al. Prevalence of *Staphylococcus aureus* nasal colonization in the United States, 2001–2002. *J. Infect. Dis.* **2006**, *193*, 172–179. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Froeschke, B.F.; Williams, A.; Seferian, R. Spatial distribution and antibiotic susceptibility of *Staphylococcus aureus* from discharge into the Hillsborough River, Tampa, FL. *Fla. Sci.* **2019**, *82*, 1–8.
9. Topić, N.; Cenov, A.; Jozić, S.; Glad, M.; Mance, D.; Lušić, D.; Kapetanović, D.; Mance, D.; Vukić Lušić, D. *Staphylococcus aureus*—An Additional Parameter of Bathing Water Quality for Crowded Urban Beaches. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5234. [\[CrossRef\]](#)
10. Noble, R.T.; Weisberg, S.B.; Leecaster, M.K.; McGee, C.D.; Dorsey, J.H.; Vainik, P.; Orozco-Borbón, V. Storm effects on regional beach water quality along the southern California shoreline. *J. Water Health* **2003**, *1*, 23–31. [\[CrossRef\]](#)
11. Selvakumar, A.; Borst, M. Variation of microorganism concentrations in urban stormwater runoff with land use and seasons. *J. Water Health* **2006**, *4*, 109–124. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Selvakumar, A.; O'Connor, T.P. Seasonal variation in indicator organisms infiltrating from permeable pavement parking lots at the Edison Environmental Center, New Jersey. *Water Environ. Res.* **2022**, *94*, e10791. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Rosenberg Goldstein, R.E.; Micallef, S.A.; Gibbs, S.G.; Davis, J.A.; He, X.; George, A.; Kleinfelter, L.M.; Schreiber, N.A.; Mukherjee, S.; Sapkota, A.; et al. Methicillin-resistant *Staphylococcus aureus* (MRSA) detected at four U.S. wastewater treatment plants. *Environ. Health Perspect.* **2012**, *120*, 1551–1558. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Tosić, M.; Restrepo, J.D.; Izquierdo, A.; Lonin, S.; Martins, F.; Escobar, R. An integrated approach for the assessment of land-based pollution loads in the coastal zone. *Estuar. Coast. Shelf Sci.* **2018**, *211*, 217–226. [\[CrossRef\]](#)
15. Zhang, S.; Li, X.; Wu, J.; Coin, L.; O'Brien, J.; Hai, F.; Jiang, G. Molecular Methods for Pathogenic Bacteria Detection and Recent Advances in Wastewater Analysis. *Water* **2021**, *13*, 3551. [\[CrossRef\]](#)
16. Wade, T.J.; Sams, E.; Brenner, K.P.; Haugland, R.; Chern, E.; Beach, M.; Wymer, L.; Rankin, C.C.; Love, D.; Li, Q.; et al. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: A prospective cohort study. *Environ. Health* **2010**, *9*, 66. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Anderson, D.M.; Cembella, A.D.; Hallegraeff, G.M. Progress in understanding harmful algal blooms: Paradigm shifts and new technologies for research, monitoring, and management. *Ann. Rev. Mar. Sci.* **2012**, *4*, 143–176. [\[CrossRef\]](#) [\[PubMed\]](#)
18. de Brauwere, A.; Gourgue, O.; de Brye, B.; Servais, P.; Ouattara, N.K.; Deleersnijder, E. Integrated modelling of faecal contamination in a densely populated river-sea continuum (Scheldt River and Estuary). *Sci. Total Environ.* **2014**, *468–469*, 31–45. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Heasley, C.; Sanchez, J.J.; Tustin, J.; Young, I. Systematic review of predictive models of microbial water quality at freshwater recreational beaches. *PLoS ONE* **2021**, *16*, e0256785. [\[CrossRef\]](#)
20. Wang, J.; Deng, Z. Modeling and predicting fecal coliform bacteria levels in oyster harvest waters along Louisiana Gulf coast. *Ecol. Indic.* **2019**, *101*, 212–220. [\[CrossRef\]](#)

21. Cyterski, M.; Shanks, O.C.; Wanjugi, P.; McMinn, B.; Korajkic, A.; Oshima, K.; Haugland, R. Bacterial and viral fecal indicator predictive modeling at three Great Lakes recreational beach sites. *Water Res.* **2022**, *223*, 118970. [CrossRef] [PubMed]
22. Christensen, V.G.; Stelzer, E.A.; Eikenberry, B.C.; Olds, H.T.; LeDuc, J.F.; Maki, R.P.; Saley, A.M.; Norland, J.; Khan, E. Cyanotoxin mixture models: Relating environmental variables and toxin co-occurrence to human exposure risk. *J. Hazard Mater.* **2021**, *415*, 125560. [CrossRef] [PubMed]
23. Francy, D.S.; Brady, A.M.G.; Stelzer, E.A.; Cicale, J.R.; Hackney, C.; Dalby, H.D.; Struffolino, P.; Dwyer, D.F. Predicting microcystin concentration action-level exceedances resulting from cyanobacterial blooms in selected lake sites in Ohio. *Environ. Monit. Assess.* **2020**, *192*, 513. [CrossRef] [PubMed]
24. Jones, R.M.; Liu, L.; Dorevitch, S. Hydrometeorological variables predict fecal indicator bacteria densities in freshwater: Data-driven methods for variable selection. *Environ. Monit. Assess.* **2013**, *185*, 2355–2366. [CrossRef]
25. Whitman, R.L.; Nevers, M.B. Summer *E. coli* patterns and responses along 23 Chicago beaches. *Environ. Sci. Technol.* **2008**, *42*, 9217–9224. [CrossRef] [PubMed]
26. Richiardi, L.; Pignata, C.; Fea, E.; Bonetta, S.; Carraro, E. Are Indicator Microorganisms Predictive of Pathogens in Water? *Water* **2023**, *15*, 2964. [CrossRef]
27. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
28. Elith, J.; Graham, C.H.; Anderson, R.P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **2006**, *29*, 129–151. [CrossRef]
29. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef]
30. Elith, J.; Leathwick, J. Boosted Regression Trees for Ecological Modeling. 2011. Available online: <https://www.semanticscholar.org/paper/Boosted-Regression-Trees-for-ecological-modeling-Elith-Leathwick/ac744d455dff6c7036e4ae5c6dfbf4b04d7f7e97> (accessed on 4 December 2023).
31. Lipps, W.C.; Braun-Howland, E.B.; Baxter, T.E. *Standard Methods for the Examination of Water and Wastewater*, 24th ed.; APHA Press: Washington, DC, USA, 2023; ISBN 0-87553-299-3.
32. Fang, H.; Hedin, G. Rapid screening and identification of methicillin-resistant *Staphylococcus aureus* from clinical samples by selective-broth and real-time PCR assay. *J. Clin. Microbiol.* **2003**, *41*, 2894–2899. [CrossRef]
33. Shortle, D. A genetic system for analysis of staphylococcal nuclease. *Gene* **1983**, *22*, 181–189. [CrossRef] [PubMed]
34. Brakstad, O.G.; Aasbakk, K.; Maeland, J.A. Detection of *Staphylococcus aureus* by polymerase chain reaction amplification of the nuc gene. *J. Clin. Microbiol.* **1992**, *30*, 1654–1660. [CrossRef] [PubMed]
35. Weisburg, W.G.; Barns, S.M.; Pelletier, D.A.; Lane, D.J. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **1991**, *173*, 697–703. [CrossRef] [PubMed]
36. De'ath, G.; Lough, J.M.; Fabricius, K.E. Declining Coral Calcification on the Great Barrier Reef. *Science* **2009**, *323*, 116–119. [CrossRef] [PubMed]
37. Yates, K.K.; Greening, H.; Morrison, G. *Integrating Science and Resource Management in Tampa Bay, Florida*; USGS: St. Petersburg, FL, USA; p. 280, Federal 1348. Available online: <https://pubs.usgs.gov/circ/1348/> (accessed on 8 April 2024).
38. Geurink, J.; Water, T.; Basso, R.; Tara, P.; Trout, K.; Ross, M. Improvements to Integrated Hydrologic Modeling in the Tampa Bay, Florida Region: Hydrologic Similarity and Calibration Metrics. In Proceedings of the Joint Federal Interagency Conference 2006, Reno, NV, USA, 2–6 April 2006.
39. Zhu, J.; Weisberg, R.H.; Zheng, L.; Qi, H. On the salt balance of Tampa Bay. *Cont. Shelf Res.* **2015**, *107*, 115–131. [CrossRef]
40. Whitman, R.L.; Harwood, V.J.; Edge, T.A.; Nevers, M.B.; Byappanahalli, M.; Vijayavel, K.; Brandão, J.; Sadowsky, M.J.; Alm, E.W.; Crowe, A.; et al. Microbes in beach sands: Integrating environment, ecology and public health. *Rev. Environ. Sci. Bio/Technol.* **2014**, *13*, 329–368. [CrossRef] [PubMed]
41. Arandia-Gorostidi, N.; Krabberød, A.K.; Logares, R.; Deutschmann, I.M.; Scharek, R.; Morán, X.A.G.; González, F.; Alonso-Sáez, L. Novel Interactions between Phytoplankton and Bacteria Shape Microbial Seasonal Dynamics in Coastal Ocean Waters. *Front. Mar. Sci.* **2022**, *9*, 901201. [CrossRef]
42. Seymour, J.R.; Amin, S.A.; Raina, J.-B.; Stocker, R. Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. *Nat. Microbiol.* **2017**, *2*, 17065. [CrossRef] [PubMed]
43. Sipler, R.E.; McGuinness, L.R.; Kirkpatrick, G.J.; Kerkhof, L.J.; Schofield, O.M. Bacteriocidal effects of brevetoxin on natural microbial communities. *Harmful Algae* **2014**, *38*, 101–109. [CrossRef]
44. Wang, Y.; Lin, H.; Huang, R.; Zhai, W. Exploring the plankton bacteria diversity and distribution patterns in the surface water of northwest pacific ocean by metagenomic methods. *Front. Mar. Sci.* **2023**, *10*, 1177401. [CrossRef]
45. Jones, K.L.; Mikulski, C.M.; Barnhorst, A.; Doucette, G.J. Comparative analysis of bacterioplankton assemblages from *Karenia brevis* bloom and nonbloom water on the west Florida shelf (Gulf of Mexico, USA) using 16S rRNA gene clone libraries. *FEMS Microbiol. Ecol.* **2010**, *73*, 468–485. [CrossRef] [PubMed]
46. Patin, N.V.; Brown, E.; Chebli, G.; Garfield, C.; Kubanek, J.; Stewart, F.J. Microbial and chemical dynamics of a toxic dinoflagellate bloom. *PeerJ* **2020**, *8*, e9493. [CrossRef] [PubMed]
47. Possamai, B.; Hoeinghaus, D.J.; Odebrecht, C.; Abreu, P.C.; Moraes, L.E.; Santos, A.C.A.; Garcia, A.M. Freshwater Inflow Variability Affects the Relative Importance of Allochthonous Sources for Estuarine Fishes. *Estuaries Coasts* **2020**, *43*, 880–893. [CrossRef]

-
48. Izabel-Shen, D. Understanding response of microbial communities to saltwater intrusion through microcosms. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 929–933. [[CrossRef](#)] [[PubMed](#)]
 49. Crump, B.C.; Bowen, J.L. The Microbial Ecology of Estuarine Ecosystems. *Annu. Rev. Mar. Sci.* **2024**, *16*, 335–360. [[CrossRef](#)] [[PubMed](#)]
 50. Wang, H.; Chen, F.; Zhang, C.; Wang, M.; Kan, J. Estuarine gradients dictate spatiotemporal variations of microbiome networks in the Chesapeake Bay. *Environ. Microbiome* **2021**, *16*, 22. [[CrossRef](#)]
 51. Philippot, L.; Griffiths, B.S.; Langenheder, S. Microbial Community Resilience across Ecosystems and Multiple Disturbances. *Microbiol. Mol. Biol. Rev.* **2021**, *85*, e00026-20. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.