Marine Pollution Bulletin 200 (2024) 116053

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Short-term forecasting of fecal coliforms in shellfish growing waters

Natalie Chazal[a], Megan Carr[a], Andrew K. Leight[b], Sheila M. Saia[a,c], Natalie G. Nelson[a,d,*]

[a] Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC, USA
[b] Cooperative Oxford Laboratory, National Centers for Coastal Ocean Science, National Oceanic and Atmospheric Administration (NOAA), Oxford, MD, USA
[c] State Climate Office of North Carolina, North Carolina State University, Raleigh, NC, USA
[d] Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA

ARTICLE INFO

Keywords:
Water quality
Fecal indicator bacteria
Shellfish sanitation
Random Forest
Forecasting

ABSTRACT

This study sought to develop models for predicting near-term (1–3 day) fecal contamination events in coastal shellfish growing waters. Using Random Forest regression, we (1) developed fecal coliform (FC) concentration models for shellfish growing areas using watershed characteristics and antecedent hydrologic and meteorologic observations as predictors, (2) tested the change in model performance associated when forecasted, as opposed to measured, rainfall variables were used as predictors, and (3) evaluated model predictor importance in relation to shellfish sanitation management criteria. Models were trained to 10 years of coastal FC measurements ($n = 1285$) for 5 major shellfish management areas along the Florida (USA) coast. Model performance varied between the 5 management areas with R^2 ranging from 0.36 to 0.72. Antecedent precipitation variables were among the most important predictors in the day-of forecast models in all management areas. When forecasted rainfall was included in the models, wind components became increasingly important.

1. Introduction

The bivalve shellfish mariculture industry is an increasingly important part of global food systems because of the cost effective and sustainable nature of its farming practices (Theuerkauf et al., 2019; Botta et al., 2020; FAO, 2020; Azra et al., 2021). Bivalves, including oysters and clams, are grown through intertidal, water column, and subtidal culture in uncontrolled estuarine environments. To contain the harvest, farmers use equipment such as floating bags, racks, trays, exclusion nets, or cages to keep their stock contained and keep out predators (Sturmer, 2019). Depending on the source and ploidy, it takes between 12 to 36 months for oysters to reach optimal market size and 12 to 18 months for hard clams to reach optimal market size (University of Florida, 2019; Sturmer, 2019; FDACS, 2022).

The ability to grow bivalves in open estuarine waters without additional inputs (e.g. feed, chemicals) is a key reason why bivalve production is considered sustainable and cost effective. However, farmed bivalve placement in open estuarine waters results in their exposure to changes in ambient estuarine conditions. Bivalve shellfish are filter feeders and are, therefore, sensitive to changes in ambient water quality because they can ingest contaminants from the water column while feeding (Jørgensen, 1990). Once consumed, bivalves concentrate contaminants from the water column in their tissues (Jørgensen, 1990), which necessitates the management of water quality in shellfish growing areas to ensure harvested shellfish are safe for human consumption. Specifically, public health officials are primarily concerned about the presence of fecally-associated pathogens in shellfish tissues, as many such pathogens cause severe human illness if consumed. In the USA, regulatory agencies use fecal indicator bacteria (FIB) concentrations, specifically of fecal coliforms (FC), as proxies for the overall load of fecal pathogens that could threaten human health if present in shellfish. State-level regulatory programs operate under the assumption that shellfish harvested from waters with elevated FC concentrations are contaminated.

FCs are transported to estuarine waters from many sources that are generally associated with human land development and animal excrement. This includes nonpoint source contributions such as the extent of impervious surfaces, livestock production and agricultural operations, domestic pet waste, and large populations of waterfowl and other wildlife within the watershed (Mallin et al., 2000; Mallin et al., 2001; You et al., 2023). Point sources include wastewater treatment plants and industrial discharges (Campos et al., 2013). While point sources often supply FC into waterways through human-controlled discharges, nonpoint source inputs are dependent on transport factors like heavy

* Corresponding author at: Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC, USA.
E-mail address: nnelson4@ncsu.edu (N.G. Nelson).

https://doi.org/10.1016/j.marpolbul.2024.116053
Received 8 September 2023; Received in revised form 5 January 2024; Accepted 13 January 2024
Available online 25 January 2024

rainfall events and higher river stages to move terrestrial sources of FC into waterways. Accordingly, rainfall and stormwater runoff are most commonly considered as the largest drivers of FC loads in coastal waters (Jeng et al., 2005; Coulliette et al., 2009; NSSP, 2019; You et al., 2023). However, the relationship between rainfall and FC loading is not uniform for all waterways along the coast. The sensitivity of an area to nonpoint source FC loading depends on a combination of biogeochemical, hydrologic, and meteorological factors unique to a particular area (Crowther et al., 2001; Chigbu et al., 2005; Campos et al., 2013; You et al., 2023). For example, Coulliette et al. (2009) used a hydrological model incorporating antecedent rainfall levels and distance away from shore and found there were non-linear and linear relationships between different fecal pathogens and rainfall in the Newport River Estuary, NC. They proposed these differential relationships can be explained by tidal influences that change the salinity gradient within the estuary which, in turn, affects the survivability of FIB.

Another major FC transport mechanism within coastal waters is wind-driven erosion (Chigbu et al., 2004; Ufnar et al., 2006; Lewis et al., 2013; NSSP, 2019). Wind causes turbulence in the water column, which resuspends sediments that FCs in the water column have settled into (Kay et al., 2005; Feng et al., 2016). A study characterizing FC sources and transport in the Mississippi Sound found correlations between elevated FC concentrations and significant changes in wind speed and direction (Ufnar et al., 2006). These researchers suggested this relationship is due to sediment disturbance by larger wave action during higher wind activity; especially since FC concentrations were elevated during times of little to no rainfall.

After FCs are transported from sources into waterways, some identified factors that affect the survivability of FCs in estuarine environments include salinity, FC species-specific stressors, and water temperature. FCs are bacteria that live and grow in the gastrointestinal tracts of animals. When they are excreted through feces and enter the environment, the FCs are no longer in the conditions to which they are generally adapted. FCs change their metabolic activity rapidly in response to the stress of changing environments and depending on the species, the introduction into coastal environments can extend or shorten their persistence in a marine environment (Munro et al., 1989; Troussellier et al., 1998). Salinity is one of the major effectors of survivability because of the osmoregulatory shock that the bacteria can tolerate (Munro et al., 1989; Troussellier et al., 1998). Additionally, a negative relationship between FC survival and water temperature has been shown (Faust et al., 1975; Howell et al., 1996; Chigbu et al., 2004; You et al., 2023). This relationship is seen because high water temperatures are generally associated with higher levels of predation and solar radiation, both of which decrease the survival of FC (Sinton, 2005). However, there are often seasonal variations in the importance of water temperature, air temperature, and water column mixing on FC survival and transport that complicate modeling and prediction efforts (Lipp et al., 2001; Chigbu et al., 2005).

Given the complex web of processes contributing to fecal contamination in coastal waters, the management of shellfish sanitation requires consideration of many factors that are formalized in a standardized process overseen by the U.S. Food and Drug Administration's National Shellfish Sanitation Program (NSSP; NSSP, 2019). State shellfish sanitation programs, under the guidance of the NSSP (NSSP, 2019), classify and manage shellfish growing waters as a function of FC concentrations. Shellfish growing areas can be classified as approved, restricted, prohibited, conditionally approved, and conditionally restricted based on "sanitary surveys" of an area, which describe the environmental factors that are current or potential pollution sources, microbiological monitoring data trends, and shellfish stock within the area (NSSP, 2019). Conditionally approved waters are commonly assigned rainfall thresholds, such as 24-h rainfall depths, that are associated with FC concentration exceedances. Rainfall thresholds are determined through observational studies and modified over time as needed. Some states will also use river stage thresholds in tandem with rainfall.

As a protective measure to prevent harvesting contaminated product, temporary closures of conditionally approved shellfish growing areas are issued immediately after these established thresholds are met without collecting observational data to confirm elevated FC concentrations (NSSP, 2019). Because of variation in FC sources and seasonal variation in runoff generation processes, the use of automatic closures following rainfall and/or river stage threshold exceedance could result in shellfish growing areas remaining open for harvest when they should be closed (e.g., smaller rainfall events may result in elevated FC concentrations) and closed when they should remain open. In cases when waters are closed to harvest when they could be safely open, shellfish growers are disenfranchised from the ability to optimally harvest their product (National Sea Grant Law Center, 2019; Evans et al., 2016; Landrum and Ache, 2000). In the case when waters are open when they should be closed, harvested shellfish may be exposed to potentially hazardous water quality conditions, creating public health risks. Issuing closure decisions as a function of direct FC measurement collected immediately following storms as well as throughout the recovery period would avoid potential management inefficiencies, but state shellfish sanitation programs do not have the resources to implement such a system. Instead, predictive FC concentration models may serve as one approach to help managers account for the complex processes contributing to FC concentrations in shellfish waters without the need for intensive water quality sampling resources.

Currently, state shellfish sanitation programs maintain long-term water quality monitoring programs through which they periodically reevaluate the harvest classification of shellfish growing waters. The data generated from shellfish sanitation management presents an opportunity for developing decision-support tools to aid in more precise management of shellfish growing area closures and reopenings (Schmidt et al., 2018; Zimmer-Faust et al., 2018; Wang and Deng, 2019). Previously-developed statistical models predicting short-term FC concentrations have primarily included antecedent rainfall, solar radiation, wind speed/direction, tidal stage, river stage, air and water temperature, salinity, and season as predictors, with performance (R^2) in the range of 0.40–0.74. Though successful at developing models with suitable predictive performance, prior modeling studies were relatively limited in spatial extent, with Schmidt et al. (2018) using two study sites in close proximity in Cornwall, UK, Zimmer-Faust et al. (2018) using 13 sampling stations within a single bay in Oregon, and Wang and Deng (2019) using 3 beaches within 20 km of each other on the coast of Lake Erie. Further work is needed to evaluate whether forecast models developed from shellfish sanitation data scale across distinct management zones and ecoregions. Moreover, prior research has not investigated the use of forecasted, as opposed to observed, rainfall as a FC concentration predictor. Updating models with forecasted rainfall data will potentially allow for more accurate predictions of FC concentrations due to the mechanistic importance of storm water runoff in contributing to elevated FC in coastal systems.

The purpose of this study was to develop predictive FC concentration models that could be used to create regional decision-support tools for shellfish-growers and estuarine resource managers. Our objectives were to (1) develop machine learning-based predictive FC concentration models for shellfish growing areas using watershed characteristics and antecedent hydrologic observations as predictors, (2) test the change in model performance associated with using forecasted, as opposed to observed, rainfall inputs in the models for day-of to two-day-ahead prediction, and (3) identify important variables in the models as well as evaluate whether those variables align with shellfish sanitation management criteria. We also compared forecasted and observed rainfall amounts to assess the uncertainty tradeoffs associated with using predicted rainfall estimates as model inputs. The results of this study can be used to develop short-term (i.e., 1–3 days) forecast tools that can help shellfish growers make informed decisions for their businesses and management agencies make efficient decisions about the timing and spacing of harvest closures.

2. Methods

2.1. Study area: Shellfish waters in Florida, USA

Florida, USA, is home to 8127 ha of shellfish growing areas (SECOORA, 2022) that span diverse ecoregions and climate zones. The Florida Department for Agriculture and Consumer Services (FDACS) implements the NSSP in Florida and manages the state's shellfish sanitation program. The state's conditionally approved shellfish growing waters are managed through temporary closures dictated by threshold exceedances of antecedent rainfall and, in some areas, river stage. While some rainfall thresholds correspond to rain depths within a 24-h period, others span multiple days. In total, there are 1285 spatially fixed sampling water quality stations across five management areas: West Gulf (n = 156), Central Gulf (n = 192), Big Bend (n = 242), South Gulf (n = 249), and Atlantic (n = 442, Fig. 1). FDACS uses management area designations to group sites. Each management area is overseen by a local office (FDACS correspondence). Though defined for administrative purposes, the FDACS management areas also capture differences in physiography, coastal geomorphology, and resource management needs (Hapke et al., 2019). Additionally, the Florida peninsula includes humid subtropical and tropical climate zones and ecosystems (Omernik, 1993; Köppen, 1936); therefore, management areas span distinctive ecoregions with varying rain and temperature patterns.

2.2. Data acquisition and pre-processing

We used FDACS FC monitoring data between 2012 and 2021 in this analysis. Although FDACS's publicly available digitized dataset extends past 2012, we did not consider data collected prior to 2012 because of a shift in FDACS lab procedures resulting in a change in measurement units from most probable numbers (MPN) to colony forming units (CFU). Consequently, the MPN measurements before 2012 cannot be directly compared to CFU measurements taken from 2012 onward (Noble et al., 2003; Gronewold and Wolpert, 2008; Cho et al., 2010).

The NSSP currently allows states to choose from one of two sampling strategies to use in their ambient water quality monitoring programs. The first is systematic random sampling, which aims to capture baseline FC concentration measurements across a range of environmental conditions. The second is adverse pollution condition sampling, which calls for sample collection during periods when point-source pollution events or elevated rainfall and river stage have reached levels that are known to degrade water quality in specific areas. Florida uses the adverse pollution condition sampling strategy, resulting in their monitoring data capturing conditions when FC contamination is expected to be near or at the highest expected concentrations. Per the NSSP, the state is required to collect at least 5 samples annually from each of the spatially fixed water quality sampling stations, amounting to 407,136 total FC observations in the dataset.

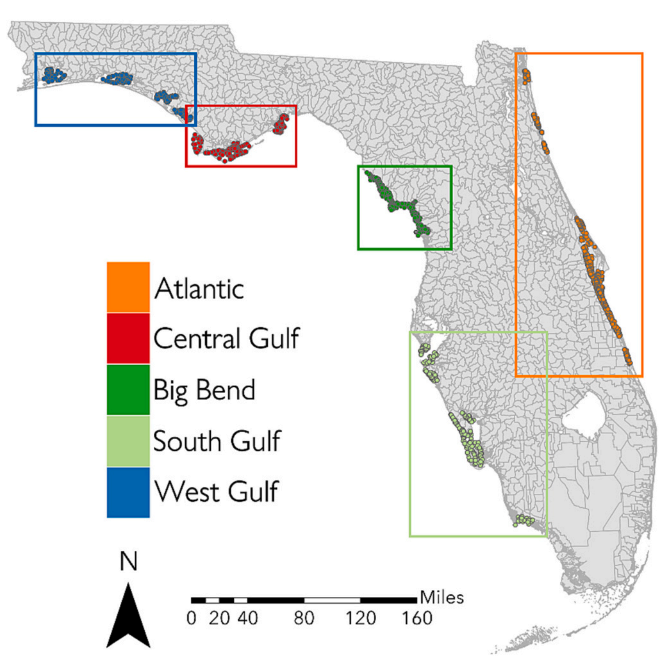

Fig. 1. Study area map of Florida coastline with the FDACS water quality sampling stations colored by the management areas used to split up data for modeling.

2.3. Environmental predictors

To predict daily FC concentrations, we included environmental variables associated with FC dynamics of the non-point sources of FCs, FC transport mechanisms into water bodies, and system mediators contributing to decay times of FC (Table 1). The full suite of predictor variables was included in the initial models and subsequently reduced during the variable selection process (see Section 2.4).

Radar-based, gridded precipitation data (2.5 by 2.5 km resolution pixels) was gathered from NEXRAD (NOAA, 1991), and totaled over U.S. Geological Survey (USGS) 12-digit Hydrological Unit Code (HUC12) subwatershed areas. We determined these totals for 1 day, 3 days, 5 days, and 7 days prior. Rainfall was aggregated over the watershed area to estimate the volume of water expected to runoff into the downstream estuary and transport FC. River stage was measured from USGS gauge sites that had complete and continuous data spanning the 2012 to 2021 range. Additionally, other variables considered in the model that relate to FC transport included watershed channelization characteristics, including the length of natural channelization and the length of artificial channelization within the watershed area as calculated from the USGS National Hydrography Dataset (USGS NHD; USGS, 2017). Increased channelization creates greater opportunity for rapid drainage and transport of stormwater runoff from land to downstream estuarine waters (Falbo et al., 2013); thus, increasing the opportunity for FC transport.

Other meteorological and hydrological parameters including wind speed, wind direction, water temperature, and air temperature were gathered from NOAA CO-OPS stations and were also aggregated on 1-day, 3-day, 5-day, and 7-day totals. Wind speed and direction were aggregated to their latitude (u) components, longitude (v) components, speed, and combined vector components (Grange, 2014).

Watershed characteristics, including total area, were summarized on a HUC12 watershed scale and include as model predictors. Percent land use was obtained from Multi-Resolution Land Characteristics Consortium's National Land Cover Database (MRLC NLCD); classes were aggregated from 20 relatively narrow classes into five broad categories: open water, wetlands, developed, cultivated, and vegetated. The land cover data were aggregated to reduce the total number of variables associated with land cover while still capturing variation in areas that are primarily dominated by natural versus human systems. We associated the FC data to the temporally closest land use dataset which includes 2011, 2013, 2016, and 2019 data (Dewitz and U.S. Geological Survey, 2021). Soil drainage was gathered from the USGS NRC SSURGO dataset and were categorized by USGS groups C, D, B/D, and C/D which correspond to levels of drainage and infiltration that affect runoff potential.

The month time component was included to account for seasonality in rain and wind patterns.

2.4. Forecasted rainfall

To assess short-term prediction capability, we substituted the NEXRAD rainfall observations for forecasted rainfall in the models. The rainfall forecasts were obtained through the U.S. National Weather

Table 1
Environmental variables and their features used to predict FC concentrations. Also indicated is the variables type which represents the mechanism that relates that variable to FC concentrations in coastal waters, the spatial resolution of the dataset used, the features used for each observation, the units of the variable, and the source of the variable's datasets.

Type	Variable	Spatial resolution	Features	Units	Data source
Response	FC Concentration	Discrete Station		CFU	FDACS
Transport	Rainfall	HUC12	1, 3, 5, 7 day totals	kg/m^2	NEXRAD
	River Stage	Discrete Station	1, 3, 5, 7 day totals	ft	USGS
	Wind Speed	Discrete Station	1, 3, 5, 7 day totals	m/s	NOAA CO-OPS
	Wind Direction	Discrete Station	1, 3, 5, 7 day totals	degrees	NOAA CO-OPS
	Channelization (natural)	HUC12		m	USGS NHD
	Channelization (artificial)	HUC12		m	USGS NHD
	Watershed Area	HUC12		acres	FDEP
Source	Land Use and Land Cover	HUC12	2011, 2013, 2016, 2019	% per class	NLCD MRLC
	Soil Drainage	HUC12	C, D, B/D, C/D	% per class	NRCS SSURGO
System Mediators	Air Temperature	Discrete Station	1, 3, 5, 7 day totals	°C	NOAA CO-OPS
	Water Temperature	Discrete Station	1, 3, 5, 7 day totals	°C	NOAA CO-OPS
	Month	NA	NA	NA	NA

Service's National Digital Forecast Database (NDFD) via the Weather Prediction Center (Weather Prediction Center, 2022). The data were gridded quantitative precipitation forecasts (QPF) and provided continuous rainfall predictions over the study area. Forecasts for 1-day, 2-day, and 3-day over USGS HUC12 subwatershed areas were gathered. This was done through determining the total rainfall for the 24-h period for 1-day (i.e., day of), 2-day (next-day), and 3-day (in two days) periods over each of the HUC12 subwatershed areas. Because of the use of different observed and forecasted rainfall variables in the models, we explored correlations between the forecasted and observed rainfall values to characterize the uncertainty introduced to the models with the use of forecasted inputs. Comparisons between forecasted and observed rainfall were made by intersecting the NDFD QPF and NEXRAD datasets and comparing values from overlapping grid pixels, rather than by comparing aggregated watershed totals.

2.5. Modeling

We developed 5 models with different predictor structures (Fig. 2) for each of the 5 management areas along Florida's coast (Fig. 1). In total, we developed 25 different models, which all predicted FC concentrations. All models included the same static variables (e.g., land cover, soil drainage characteristics), but differed in terms of the temporally dynamic variables included (i.e., rainfall, wind, river stage). The first model, referred to as the idealized model (I), used all antecedent rainfall predictors (3-day, 5-day, 7-day) as well as day-of rainfall (referred to as x_0) to predict day-of FC concentrations. We refer to this model as "idealized" as it captures the ideal (yet impossible) predictive

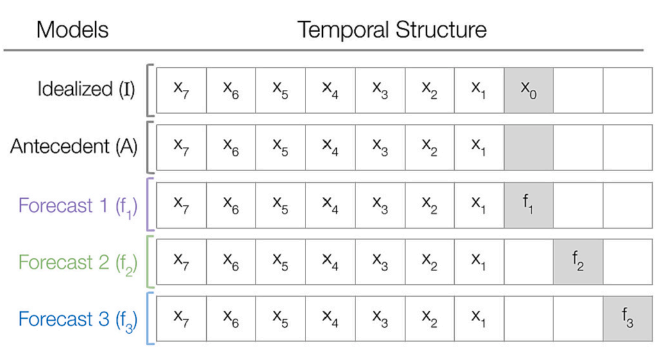

Fig. 2. Dataset alignment schematic showing the antecedent predictors x_1, x_2, … x_7, day-of predictors (x_0), and the forecasted rainfall f_1, f_2, f_3. The gray boxes represent the day of the FC observation (e.g., f_2 was included as a predictor in the model used to estimate the next day's fecal coliform concentrations). Each model was trained separately for the 5 management areas, resulting in 25 models total.

modeling scenario in which the observed day-of rainfall is known. Because the I model includes day-of observed rainfall for predicting day-of FC concentrations, we assumed it included the best possible environmental predictors and would have the lowest error. The second model, referred to as the antecedent model (A), contains all antecedent predictors, and excludes day-of rainfall to predict day-of FC concentrations. This model is representative of the information that a resource manager or grower would have available to them when tasked with making a management decision and corresponds to the data that could be used to create an operational predictive model as part of a decision-support tool. In addition to I and A, three models were created that excluded day-of rainfall but included antecedent and forecasted rainfall predictors. Specifically, day-of rainfall forecasts (f_1) to predict day-of FC concentrations, 2-day (i.e., next-day) rainfall forecasts (f_2) to predict FC concentrations 2 days out, and 3-day (i.e., in two days) rainfall forecasts (f_3) were used to predict FC concentrations 3-days out, respectively. Note that forecasted rainfall variables for the day-of models are referred to as f_1 whereas measured rainfall for the I model is referred to as x_0. Only one forecasted rainfall variable was included in the model at a time and used to predict the FC concentrations corresponding to the day of the forecast. For example, f_3 was used as a predictor in a model that forecasted FC concentrations in two days, but was not used in the models applied to predict day-of or next-day FC concentrations. Building on the A models to forecast next-day and two-day-ahead FC concentrations involved restructuring the dataset so that the antecedent predictors represented the conditions only for the days that had occurred (Fig. 2). For example, if an FC observation occurred on January 5th, the antecedent predictors for the A model would include measurements from January 4th, 3rd, and so on. For the 3-day forecast model (f_3), the antecedent predictor variables would begin 72-h before January 5th, on January 3rd.

Random Forest models were used to produce the predictive FC concentration models. Random Forest is a type of machine learning algorithm used commonly for both regression and classification problems (Breiman, 2001). Broadly speaking, Random Forest models work by creating many decorrelated and individually uninformative decision trees using random subsets of the full dataset to create an ensemble, or "forest", that works together to generate a prediction (Breiman, 2001). The Random Forest algorithm was chosen for its ability to handle the nonlinear relationships between predictors and responses that are commonly seen in estuarine/marine systems (Fan et al., 2015; Zhang et al., 2020), and because of its suitability and robustness with "medium" dataset sizes (Hastie et al., 2009). Models were built in R with the Random Forest method 'rf' within the 'caret' package version 6.0.90 (Kuhn et al., 2022). We used a randomized 80 % training and 20 % testing split with hyperparameter, mtry, tuned automatically using the optimal mtry value from five searches (tuneLength = 5). Mtry dictates the number of variables that are randomly chosen as candidate features used to split data in one of the many trees used to create Random Forests. All model results are reported for the testing data only.

In total, 67 candidate predictor variables were compiled. However, to ensure variables did not include redundant information and reduce multicollinearity between predictor variables, we applied a variable screening procedure from which to reduce the total number of predictors included in the models to the final set of predictor variables used in this study (Table 1). Specifically, we iteratively used variance inflation factors (VIF) to produce a value that corresponds to how collinear a predictor variable is with the rest of the variables. We selected variables within a threshold value of 10 and applied this iterative VIF procedure for each of the 5 management areas. While there are no standard practices for selecting VIF thresholds for multicollinearity, statisticians have suggested the use of 5 or 10 as reasonable maximum values (Montgomery et al., 1992; Zuur et al., 2007). We included all variables, with the exception of forecasted rainfall variables, in this variable screening procedure. To assess each of the models' goodness of fit, we used the coefficient of determination (R^2). We reported root mean squared error (RMSE) and Mean Absolute Error (MAE) to evaluate the error across multiple models. Predictor variable importance was evaluated using incremental node purity scores (Breiman, 2001). Incremental node purity is used to rank features by how well a predictor splits the data in each of the decision trees that are used to train the Random Forest model and is measured by the decrease in magnitude of the Gini Index before and after a feature is used to split the data (Breiman, 2001). To evaluate the error introduced by including forecasted rainfall predictors, RMSE and MAE values were calculated from a pixel-by-pixel comparison of the measured and predicted rainfall for the dates that included FC observations. The resolution of these grids were 2.5 by 2.5 km pixels.

Because the models used unique combinations of watershed, river gauge, and meteorological stations to predict FC concentrations for the associated FDACS water quality sampling stations, we broke the coastline into 116 prediction spaces (Fig. 3). Consequently, these prediction spaces represent the areas that contain unique combinations of predictors, which ultimately result in an integrated model output for any station within that prediction space. In practice, all water quality sampling stations falling within a prediction space will have the same FC concentration prediction from a model for a given day. Additionally, because some prediction spaces did not include water quality sampling stations, no FC concentration predictions were made for those areas. Prediction spaces were determined by creating Voronoi diagrams (Voronoi, 1909) for the USGS Stations and the NOAA CO-OPS Stations and the HUC12 subwatersheds. Voronoi diagrams were chosen for this task because they partition areas based on the proximity to point features. We created the unique prediction spaces by overlapping the Voronoi diagrams.

3. Results

3.1. FDACS monitoring data summary

The FDACS FC concentration observations were summarized with monthly means and total observations per month in each management area (Fig. 4). Overall, the South Gulf management area had the lowest means (3.99 CFU) while the West Gulf had the highest (8.61 CFU). The West Gulf and the South Gulf both had the highest mean FC in 2018, the

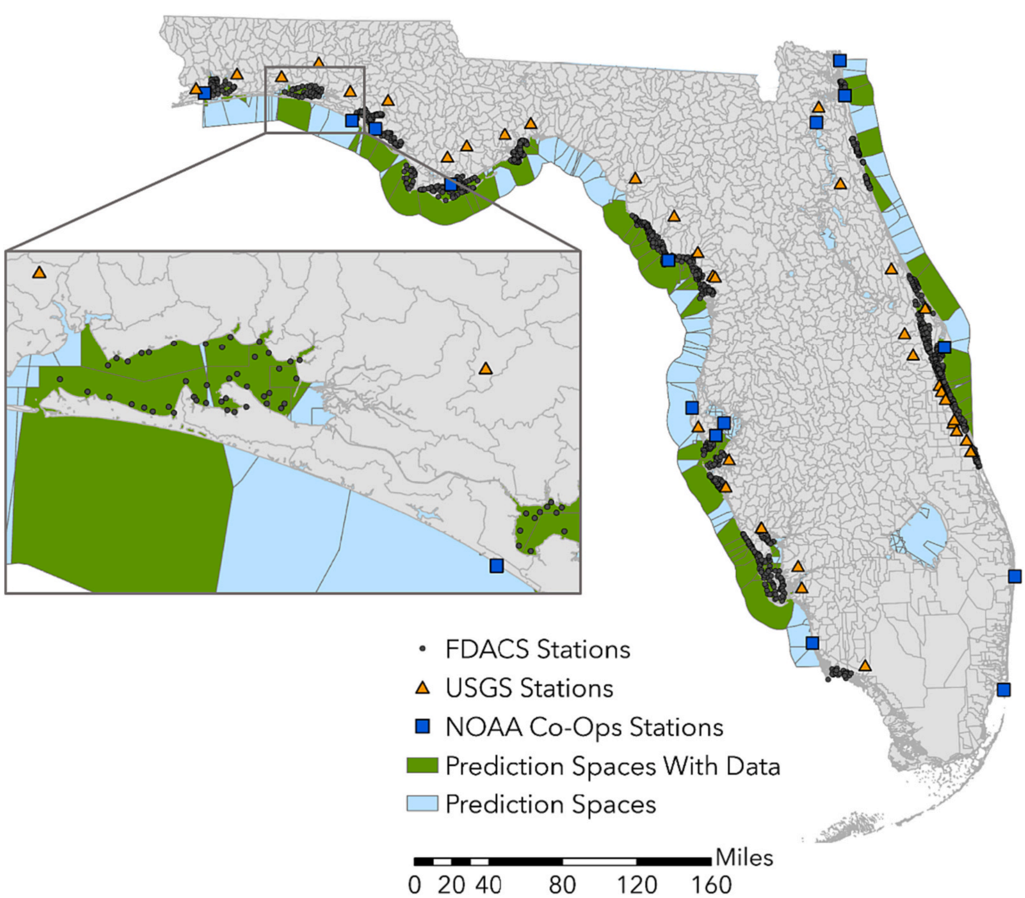

Fig. 3. Map of Florida prediction spaces using the unique combinations of watersheds, USGS Stations, and NOAA CO-OPS stations. Green areas represent areas that contained FDACS water quality sampling stations. Blue areas along the coast represent prediction spaces that did not contain FDACS water quality sampling stations and therefore did not have prediction outputs associated with them. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

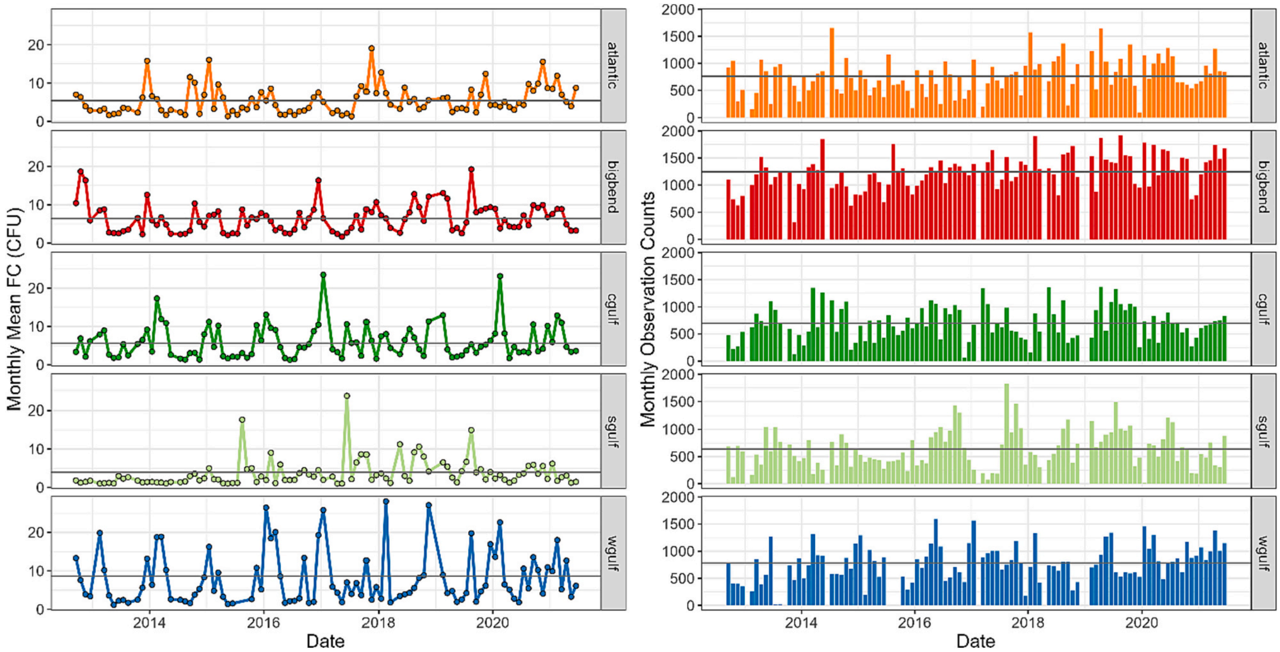

Fig. 4. Summarized monthly mean fecal coliform (FC) concentration sampling results between 2012 and 2021. Summaries include mean monthly FC concentrations (CFU) represented for each management area (left) and the total number of FC samples taken within the management area for each month within the study period (right). Gray lines represent the averages for each plot. Abbreviations: Central Gulf (cgulf), South Gulf (sgulf), West Gulf (wgulf).

Central Gulf in 2017, the Atlantic in 2021, and the Big Bend in 2012 (Fig. 4). Sampling in each of the management areas was generally consistent across the years with Big Bend having the highest sampling effort overall and South Gulf having the lowest (Fig. 4). Most management areas showed peak FC concentrations during the winter months (November – February) with the exception of the South Gulf management area, which showed peak mean FC concentrations between August and October (Fig. 5).

3.2. Model performance

The performance of the models varied by structure and management area (Table 2). The Central Gulf, Big Bend, South Gulf, and Atlantic management areas performed similarly with R^2 values ranging from 0.36 to 0.46. The best overall performance was seen in the A model in the West Gulf region ($R^2 = 0.72$). The overall worst performing model was the f_2 model in the South Gulf region ($R^2 = 0.36$). There was negligible change in RMSE (differences range from 0.31 to 1.56 in.) across models within the same management areas.

To explore the model error further, we used the regulatory threshold of FC concentrations associated with shellfish harvest closures, 14 CFU (NSSP, 2019). We categorized the measured FC concentration from the models into exceedance (>14 CFU) and safe levels (≤14 CFU). The RMSEs were recalculated for each of these categories within each of the management areas (Table 3). The exceedances consistently had a much higher RMSE than the safe levels. This indicates that the models were able to predict FC concentrations within the safe ≤14 CFU limit better than in exceedance conditions.

3.3. Variable importance scores

We used standardized, incremental node purity scores to summarize the feature importance for each of the models in each of the management areas (Fig. 6). Generally, antecedent precipitation was the most important predictor of FC concentrations in the I and A models for all regions except the Big Bend. In the Big Bend management area, day-of river stage (River Stage 0 Day) was the most important predictor in

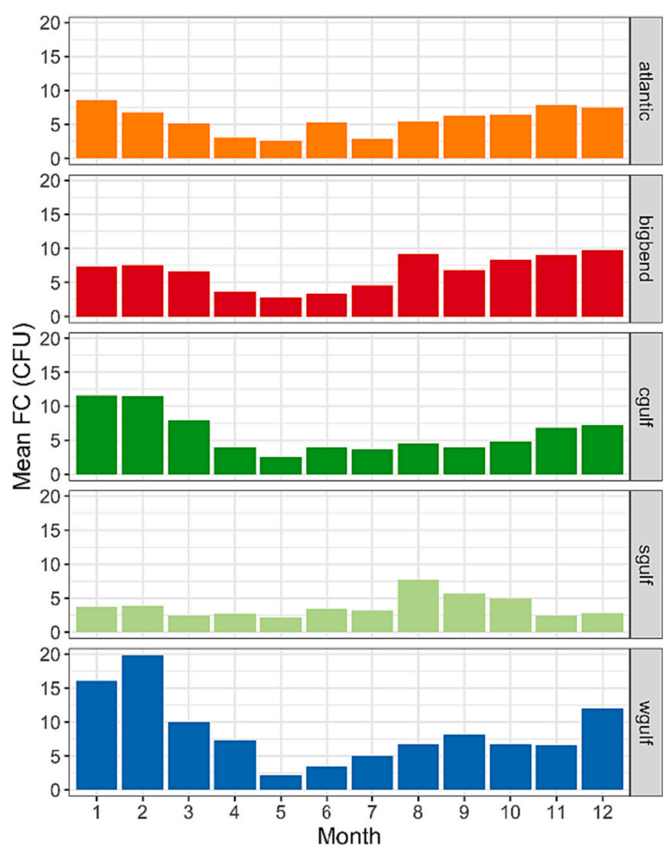

Fig. 5. Summarized fecal coliform (FC) concentration sampling results (CFU) between 2012 and 2021 averaged to a monthly scale for each management area. Abbreviations: Central Gulf (cgulf), South Gulf (sgulf), West Gulf (wgulf).

Table 2
R^2, RMSE, and number of testing observations (n) for each of the models built in this study for all the study areas.

	Idealized (I)			Antecedent (A)			Forecast 1 (f_1)			Forecast 2 (f_2)			Forecast 3 (f_3)		
	R^2	RMSE	n	R^2	RMSE	n	R^2	RMSE	n	R^2	RMSE	n	R^2	RMSE	n
West Gulf	0.70	9.48	7106	0.72	9.29	7124	0.68	9.52	6354	0.68	9.95	6313	0.69	9.69	7098
Central Gulf	0.41	8.48	8606	0.44	8.61	8614	0.43	9.09	7946	0.42	8.55	7939	0.44	8.01	8549
Big Bend	0.40	10.4	10,798	0.43	10.4	10,852	0.38	11.5	9992	0.46	10.1	9779	0.44	9.94	10,633
South Gulf	0.39	6.61	5356	0.40	6.53	5379	0.43	6.36	4884	0.36	6.22	4763	0.38	6.16	5259
Atlantic	0.42	9.14	8029	0.40	9.17	8067	0.38	9.05	7335	0.40	9.36	7189	0.42	9.06	7933

Table 3
RMSE and number of testing observations (n) for each of the models built in this study for all the study areas, categorized by the data either exceeding (E) or being within safe (S) regulatory threshold limits (14 CFU).

	Idealized (I)			Antecedent (A)			Forecast 1 (f_1)			Forecast 2 (f_2)			Forecast 3 (f_3)		
	Type	RMSE	n	Type	RMSE	n	Type	RMSE	n	Type	RMSE	n	Type	RMSE	n
West Gulf	E	21.7	965	E	21.5	949	E	20.9	885	E	22.6	906	E	21.7	984
	S	5.48	6141	S	5.33	6175	S	5.86	5469	S	5.49	5407	S	5.78	6114
Central Gulf	E	24.9	714	E	25.7	700	E	27.3	665	E	25.4	635	E	23.6	678
	S	4.72	7892	S	4.74	7914	S	4.70	7281	S	4.81	7304	S	4.67	7871
Big Bend	E	25.1	1265	E	24.9	1293	E	28.6	1221	E	24.3	1189	E	24.4	1200
	S	6.22	9533	S	6.17	9559	S	6.05	8771	S	5.94	8590	S	5.93	9433
South Gulf	E	29.8	195	E	28.5	210	E	30.4	165	E	29.6	146	E	28.9	173
	S	3.45	5161	S	3.36	5169	S	3.09	4719	S	3.48	4617	S	3.31	5086
Atlantic	E	27.0	694	E	27.6	683	E	27.8	602	E	28.4	611	E	26.9	696
	S	4.75	7335	S	4.62	7384	S	4.49	6733	S	4.54	6578	S	4.52	7237

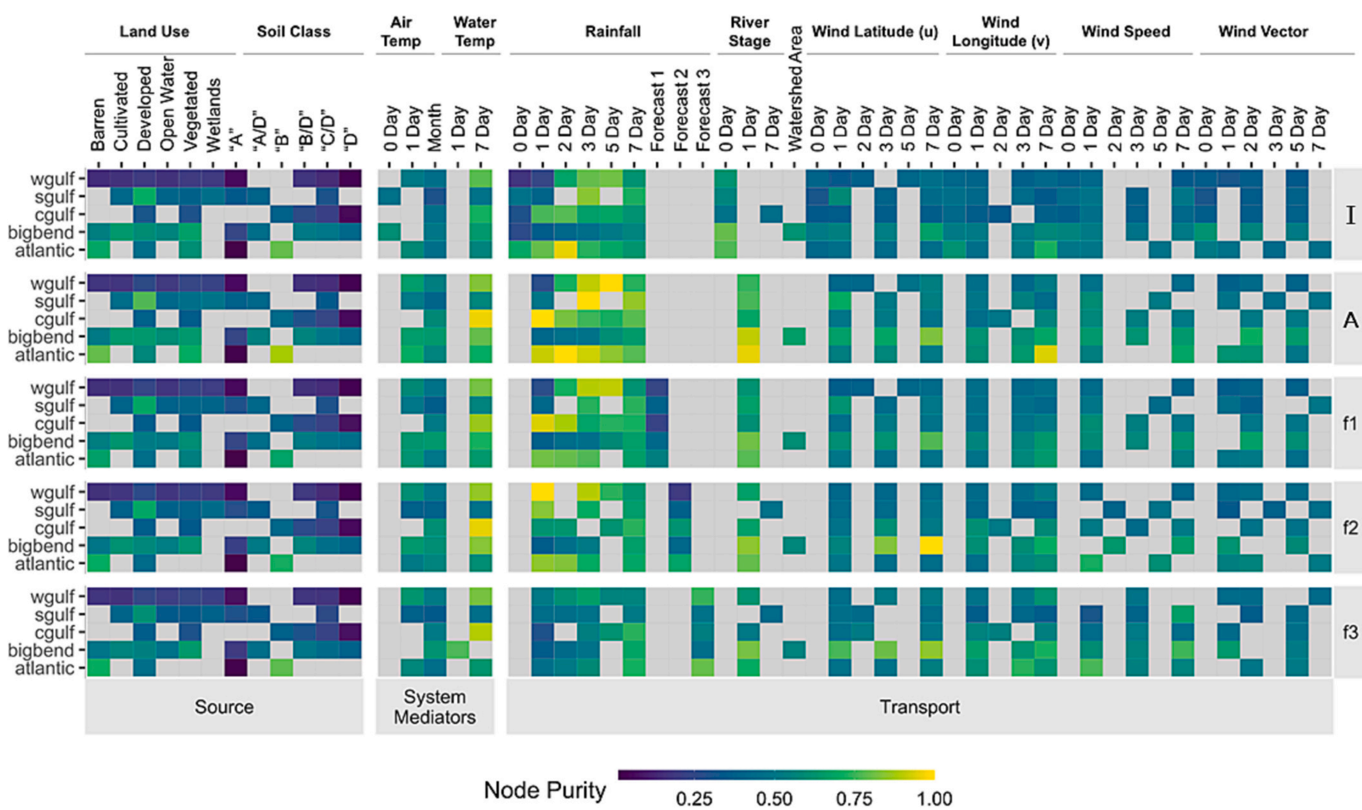

Fig. 6. Heatmap of the variable importance scores in all models. Management areas are listed on the left and model structure type is listed on the right. Columns correspond to candidate predictor variables considered in the analysis. Variables are organized by their mechanistic categories (Source, System Mediators, Transport). The gray cells are variables that were removed by pre-processing VIF procedures or were not considered for a particular model structure (e.g., the forecasted variables were not considered in the I and A models, see Section 3.4). Abbreviations: Central Gulf (cgulf), South Gulf (sgulf), West Gulf (wgulf).

the I model, while the day before river stage (River Stage 1 Day) was the most important predictor in the A model. In both the Atlantic and the Big Bend management areas, there was an increase in importance in the 7-days-before wind component variables (Wind Lat. 7 Day, Wind Long. 7 Day). In the f_1 model, which used day-of rainfall forecasts, we saw a slight change in the overall variables that were important in predicting FC concentrations. However, there were more differences seen in the 2- and 3-day forecast models, most notably the increased importance in water temperature and wind components in the 3-day forecast models (f_3). We saw that 3-day forecasted rainfall was highly important in the Atlantic management area compared to the other management areas. In the Atlantic and Big Bend management areas, there was increased importance distributed among the antecedent wind component predictors in the f_1, f_2, and f_3 models. The increase in variable importance aside from wind was also seen in the remaining predictors in the forecasted models relative to the I and A models, but to a lesser extent.

3.4. Forecasted rainfall performance

Because the main differences within the model structures corresponded to the inclusion or exclusion of rainfall forecasts as predictors, we evaluated the error that forecasted rainfall introduced into the models by comparing the day-of (x_0) observed rainfall to the forecasted rainfall (f_1, f_2, f_3; Fig. 7).

These metrics revealed the relationship between observed and forecasted rainfall values for all data from f_1 models were relatively poor with an increase in error for the 2-day and 3-day predictions. The pattern of increasing error as we predicted rainfall farther out held true for all the management areas.

4. Discussion

The models created in this study performed sufficiently well to be integrated into decision-support systems, with R^2 values in the range of 0.36 to 0.72 and RMSE values ranging from 6.22 to 11.5. Other studies that utilize machine and statistical learning methods to predict FC concentrations specifically in shellfish growing waters perform similarly or slightly better than our top performing model, despite the differences in modeling techniques and data structures. Zimmer-Faust et al. (2018) compared the prediction performances of five different models, with the best performer being Classification Trees (R^2 ranging from 0.5 to 0.67 for each management area). Wang and Deng (2019) used an Artificial Neural Network to predict $\log_{10}$ transformed FCs along the Louisiana Gulf Coast with an R^2 of 0.40.

Evaluating the performance of the 5 management area models developed in this study relative to other models previously created to support shellfish sanitation management is challenging because prior studies predicted $\log_{10}$ transformed, as opposed to non-transformed, FC concentrations as the model output. Our models performed worse than studies that utilize $\log_{10}$ transformations of FC concentrations in regression models within coastal waters (Lin et al., 2008; Wang et al., 2022). However, the use of the $\log_{10}$ transformation in regression models, similar to the use of geometric means, will generally bias the model output towards the prediction of lower concentrations (Haas, 1996). Meaning that, although overall performance may improve when using $\log_{10}$ transformed FC concentrations, the predictive performance in exceedance categories will be lower once the model output is back-transformed to the original units (Brooks et al., 2016). When we compare our results to the few studies that utilize Random Forest regression models to predict non-log-transformed FC concentrations in coastal waters more broadly (Parkhurst et al., 2005), our models demonstrated better performance. Parkhurst et al. (2005) used Random Forest regression models to predict FC concentrations using two months of daily water quality samples from five coastal beaches used for recreational swimming. The reported R^2 values were 0.30, 0.73, 0.10, 0.60, and 0.23 for models that predicted raw FC concentrations (Parkhurst et al., 2005). The differences in Parkhurst et al. (2005) and the current study is our inclusion of step variable selection step before model training, our data's regulatory source, and our data's larger temporal and spatial extents.

While the overall model performance was similar to previously reported studies, variations in performance were observed among the 5 management areas. Overall, the models for the West Gulf performed better, with consistently higher R^2 values [0.68, 0.72], than the Central Gulf, Big Bend, South Gulf, and Atlantic, which all had comparable R^2 values in the range of [0.36, 0.72] (Table 2). The West Gulf contained the highest percentage of exceedance data (Table 3) and had the highest mean FC concentrations (Fig. 4) and the dataset was more evenly

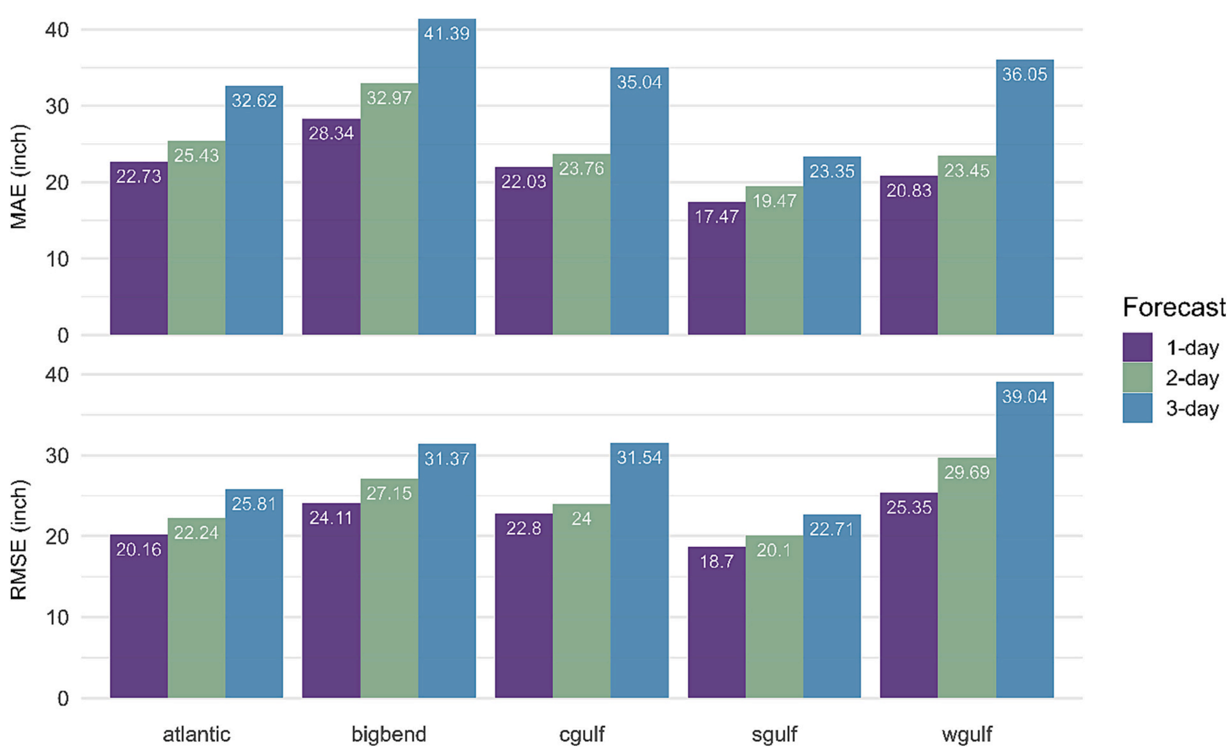

Fig. 7. Bar plot illustrating the RMSE and MAE values calculated from the observed vs. forecasted rainfall (inches) for all three prediction periods (1-day, 2-day, 3-day) within each management area.

distributed between exceedances and safe FC concentrations. Therefore, West Gulf model FC concentration predictions were not overly biased towards the safe, lower FC concentrations levels. Also, the variations in the performance between the 5 management areas could be due to the differences in the mechanisms driving local FC concentration dynamics and whether those dynamics were accurately captured within the model. For example, compared to the West Gulf's humid, subtropical climate, the South Gulf falls under a tropical climate classification (PRISM Climate Group, 2014; Köppen, 1936), which means that the quantity and timing of rainfall over the course of a year differ in the South Gulf relative to the other study regions. This variation in rainfall amount and timing in the South Gulf could impact how antecedent rainfall and river stage forecasts can be used to predict current and future FC conditions. The annual total rainfall remained similar between management areas (PRISM Climate Group, 2014) with the exception of the West Gulf region, which had elevated rainfall totals in 2018 to 2020 compared to the other management areas. However, further work is needed to explore the explanatory value of climatic characteristics relative to the varying FC dynamics of each region.

Among the most important predictor variables for most models were antecedent rainfall and river stage. The importance of rainfall and river stage transport mechanisms for predicting FC concentrations in coastal waters has been corroborated by other studies (Crowther et al., 2001; Chigbu et al., 2005; Jeng et al., 2005; Coulliette et al., 2009; Campos et al., 2013; NSSP, 2019; You et al., 2023). FDACS sets rainfall and, in some cases, river stage thresholds for temporary closures of conditionally managed shellfish growing areas in Florida. The alignment of the variable importance results with current management criteria suggests that the predictors and models are capturing real, as opposed to spurious, relationships between hydrologic predictors and FC concentrations. Furthermore, the land use and soil class predictors were relatively unimportant overall, indicating that variables capturing FC transport drivers, as opposed to sources, were the most predictive of FC concentrations. Land use and soil classifications are important sources to consider for understanding FC dynamics in coastal systems. Because we use land use and soil classifications summarized on a HUC12 scale, it is possible that the relationships between more fine-scale, localized land use and soil characteristics were not captured because of the summarization on the larger scale.

Across models, the variable importance of predictors also highlighted the role of antecedent rainfall in predicting FC concentrations, even in the models that include forecasted rainfall. The I model was the only model to include day-of observations (Fig. 2), with day-of rainfall expected to be highly predictive of day-of FC concentrations given the importance of “first flush” effects (Krometis et al., 2007; Parker et al., 2010). However, in comparing the I model to the other models, we found that the day-of rainfall information was not as important as antecedent 1- to 7-day rainfall measurements. The importance of lagged rainfall may reflect the time needed to generate stormwater runoff following a rain event and transport FC to estuarine waters. While many of Florida's conditionally approved shellfish growing areas have multi-day rainfall thresholds, other states often include 24-h rainfall totals as triggers for temporary harvest closures of conditionally approved shellfish growing areas (NSSP, 2019). The importance of longer lag times as FC concentration predictors in our models indicates that windows of time spanning >24 h should be further explored when structuring conditional management protocols.

A notable exception to the overarching importance of antecedent rainfall predictors was exhibited in the Big Bend models, for which antecedent rain predictors were relatively unimportant and wind variables (u and v components as well as 1- and 7-day wind speeds) were among the more important predictors (Fig. 6). This could be due to the known relationship between wind and elevated FC concentrations in the Gulf Coast area (Chigbu et al., 2004; Ufnar et al., 2006; Lewis et al., 2013; NSSP, 2019). This relationship involves wind both increasing the distance FCs can travel laterally in the body of water and increase the wave action which re-suspends FCs in the sediment. Additionally, the Big Bend is relatively undeveloped, particularly compared to most coastal areas, and there are extensive seagrass beds and marshes due to the region having minimal wave energy (Mattson et al., 2006). As a result, there is less potential for stormwater runoff generation from impervious surfaces or through artificial drainage systems, which could partly explain the lack of importance of antecedent rainfall in the Big Bend models.

Surprisingly, the forecasted rainfall variables were not important predictors in the f_1, f_2, and f_3 models (Fig. 6). Similar to the I and A models, antecedent rainfall and river stage were the most important variables in the f_1 and f_2 models, with wind components being secondary in importance in f_1 and f_2 and most important in f_3. Error in forecasted rainfall is suspected to be a key reason why these predictors were not important in the f_1, f_2, and f_3 models. We compared observed and forecasted rainfall for 1-day (day-of), 2-day (next-day), and 3-day (in two days) and found that error was high overall (RMSE = [18.70, 39.04] inches and MAE = [17.47,41.39] inches), and highest for 3-day forecasts (Fig. 7). The NDFD QPF product is produced from deterministic meteorological models that are used to predict large-scale weather patterns and are not optimized for individual locales (Weather Prediction Center, 2022). While the Weather Prediction Center has been reducing the biases in these models for decades (Weather Prediction Center, 2022), inaccurate forecasts associated with imperfect parameterization and uncertainty in the measurement of current conditions can still occur (Xu et al., 2022). To make the forecasts useful for water quality forecasting, additional and localized bias correction (Ruth et al., 2009) may be needed, particularly since we noted differences in rainfall forecast error across the different management areas. For example, rainfall forecast error was lowest in the South Gulf relative to other regions. Additionally, across all forecast period in the West Gulf region, there were higher RMSE values (RMSE = 25.35, 29.69, 39.04 in. for 1-, 2-, and 3-day periods, respectively) than MAE values (20.83, 23.45, 36.05 in.). Because larger errors have greater influence on RMSE than MAE, the disparity between RMSE and MAE tells us that within the West Gulf region there were relatively higher errors between the forecasted and observed rainfall. Conversely, the Atlantic management area showed the opposite relationship of higher MAE values than RMSE values, indicating the greater presence of relatively small errors in this region.

Considering both model performance and variable importance, we ultimately identified the A model as being the best for day-of FC concentration forecasting relative to the other model structures we evaluated. Because performance remained similar among the sets of models, variable importance trends were considered more discriminatory for determining the optimal model framework. The I, A, and f_1 models predict day-of FC concentrations and are thus directly comparable, while the f_2, and f_3 models predict next-day and two-day-ahead FC concentrations. Though the I and A models, which did not use forecasted rainfall as a predictor, performed similarly to the f_1 model, variable importance trends from the I and A models aligned more closely with our mechanistic understanding of primary FC concentration drivers. The I model represents an idealized scenario in which day-of rainfall is already known when forecasting FC concentrations, which is impossible in practice. However, developing the I model proved to be valuable, as we determined that the A model performed comparably to the I model ([0.40, 0.72] and [0.39, 0.70], respectively), demonstrating that day-of rainfall is not necessarily required to make useful FC concentration predictions. Additionally, given the uncertainty introduced to the FC concentration predictions via the use of forecasted rainfall as a predictor, we argue that the A model is more reliable than the f_1 model. Importantly, because we did not evaluate whether the A model effectively predicted next-day or two-day-ahead FC concentrations, we cannot conclude whether the use of forecasted rainfall as predictors is justified for the f_2 and f_3 models, but the error documented in the 2- and 3-day rainfall forecasts (Fig. 7) suggests that their use in FC concentration forecasting introduces too much uncertainty (Seo, 1998).

5. Conclusions

This study utilized shellfish sanitation data to develop Random Forest regression models to predict FC concentrations for sites spanning 5 management areas along the Florida coast. Predictors included a suite of different meteorological, hydrological, and geological variables to represent the mechanisms behind the sources of FC, transport of FC from the sources to the waterways, and the factors that affect survivability of FC in marine systems. We then updated these models with 1-, 2-, and 3-day rainfall forecast predictors to determine the performance as a short-term forecast tool. While the performance of the models varied by management area and model structure, we determined that the performance of the predictive models was adequate to inform decision support tools. Furthermore, the most important predictors in the models were the same as the variables currently used by shellfish sanitation managers to make closure decisions, indicating the models generated predictions in accordance with real mechanisms driving FC fate and transport. Specifically, the A model was recommended to be the best performing model for decision support. Further analysis of localized bias correction (Ruth et al., 2009) is likely needed if forecasted rainfall is going to be used in predictive coastal FC concentration models.

In addition to local bias correction of rainfall forecasts, future work could explore the addition of forecasted river stage, such as from the National Water Model (Blodgett, 2022), in FC concentration prediction, given the understood importance of elevated river stage as a primary driver of elevated FC concentrations in addition to stormwater runoff from rainfall (Faust, 1976; Ahn et al., 2005; Chigbu et al., 2005; Wilkinson et al., 2006; Vidon et al., 2008; Kay et al., 2010). However, any future work that incorporates additional forecasted predictor variables should consider more rigorous uncertainty quantification to not just provide a forecasted FC concentration, but perhaps a risk probability.

While the models we developed could be useful as decision-support tools, the NSSP does not currently allow for FC forecasts to be used in the temporary closure decision making process. However, the results of this study indicate that models can be used to integrate the complex network of factors influencing increased water quality risks beyond simply using rainfall and river stage thresholds. Moreover, because we modeled FC concentrations without regard for shellfish sanitation specific thresholds for closure, these models can be used for applications outside of shellfish harvesting. Projects that aim to monitor coastal water quality for activities like beach recreation can leverage the large spatial and temporal extents of data produced from shellfish sanitation programs as they have been established for much longer than many programs monitoring similar parameters and span a large portion of state coastlines.

CRediT authorship contribution statement

Natalie Chazal: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Megan Carr:** Methodology, Software, Writing – review & editing. **Andrew K. Leight:** Methodology, Validation, Writing – review & editing. **Sheila M. Saia:** Validation, Writing – review & editing, Methodology. **Natalie G. Nelson:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data analyzed in this study are publicly available through government agencies, with data sources referenced in text.

Acknowledgements

We thank Jill Fleiger and Michelle Smith of the Florida Department of Agriculture and Consumer Services for sharing data and providing insights on Florida's shellfish sanitation program. We would also like to thank Aaron Mayhew, John Schmidt, and Mark Klein at the National Weather Service's Southeast River Forecast Center for their invaluable help with attaining and utilizing the rainfall forecast data. We declare no conflicts of interest. This work was supported by USDA NIFA project 2019-67021-29936.

References

Ahn, J.H., Grant, S.B., Surbeck, C.Q., DiGiacomo, P.M., Nezlin, N.P., Jiang, S., 2005. Coastal water quality impact of stormwater runoff from an urban watershed in Southern California. Environmental Science & Technology 39, 5940–5953. https://doi.org/10.1021/es0501464 (American Chemical Society).

Azra, M.N., Okomoda, V.T., Tabatabaei, M., Hassan, M., Ikhwanuddin, M., 2021. The contributions of shellfish aquaculture to global food security: assessing its characteristics from a future food perspective. Front. Mar. Sci. 8, 365. https://doi.org/10.3389/fmars.2021.654897.

Blodgett, D., 2022. National Water Model V2.1 Retrospective for Selected NWIS Gage Locations, (1979-2020). U.S. Geological Survey Data Release. https://doi.org/10.5066/P9K5BEJG.

Botta, R., Asche, F., Borsum, J.S., Camp, E.V., 2020. A review of global oyster aquaculture production and consumption. Mar. Policy 117, 103952. https://doi.org/10.1016/j.marpol.2020.103952.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brooks, W., Corsi, S., Fienen, M., Carvin, R., 2016. Predicting recreational water quality advisories: a comparison of statistical methods. Environ. Model. Software 76, 81–94. https://doi.org/10.1016/j.envsoft.2015.10.012.

Campos, C.J.A., Kershaw, S.R., Lee, R.J., 2013. Environmental influences on faecal indicator organisms in coastal waters and their accumulation in bivalve shellfish. Estuaries and Coasts 36, 834–853. https://doi.org/10.1007/s12237-013-9599-y.

Chigbu, P., Gordon, S., Strange, T., 2004. Influence of inter-annual variations in climatic factors on fecal coliform levels in Mississippi Sound. Water Res. 38, 4341–4352. https://doi.org/10.1016/j.watres.2004.08.019.

Chigbu, Paulinus, Gordon, S., Tchounwou, P.B., 2005. The seasonality of fecal coliform bacteria pollution and its influence on closures of shellfish harvesting areas in Mississippi sound. Int. J. Environ. Res. Public Health 2, 362–373. https://doi.org/10.3390/ijerph2005020023 (Molecular Diversity Preservation International).

Cho, K.H., Han, D., Park, Y., Lee, S.W., Cha, S.M., Kang, J.-H., Kim, J.H., 2010. Evaluation of the relationship between two different methods for enumeration fecal indicator bacteria: colony-forming unit and most probable number. J. Environ. Sci. 22, 846–850. https://doi.org/10.1016/S1001-0742(09)60187-X.

Coulliette, A.D., Money, E.S., Serre, M.L., Noble, R.T., 2009. Space/time analysis of fecal pollution and rainfall in an eastern North Carolina estuary. Environmental Science & Technology 43, 3728–3735. https://doi.org/10.1021/es803183f (American Chemical Society).

Crowther, J., Kay, D., Wyer, M.D., 2001. Relationships between microbial water quality and environmental conditions in coastal recreational waters: the fylde coast, UK. Water Res. 35, 4029–4038. https://doi.org/10.1016/S0043-1354(01)00123-3.

Dewitz, J., U.S. Geological Survey, 2021. National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021). U.S. Geological Survey Data Release. https://doi.org/10.5066/P9KZCM54.

Evans, K.S., Athearn, K., Chen, X., Bell, K.P., Johnson, T., 2016. Measuring the impact of pollution closures on commercial shellfish harvest: the case of soft-shell clams in Machias Bay, Maine. Ocean & Coastal Management 130, 196–204. https://doi.org/10.1016/j.ocecoaman.2016.06.005.

Falbo, K., Schneider, R.L., Buckley, D.H., Walter, M.T., Bergholz, P.W., Buchanan, B.P., 2013. Roadside ditches as conduits of fecal indicator organisms and sediment: implications for water quality management. J. Environ. Manage. 128, 1050–1059. https://doi.org/10.1016/j.jenvman.2013.05.021.

Fan, J., Ming, H., Li, L., Su, J., 2015. Evaluating spatial-temporal variations and correlation between fecal indicator bacteria (FIB) in marine bathing beaches. J. Water Health 13, 1029–1038. https://doi.org/10.2166/wh.2015.031.

Faust, Maria A., 1976. Coliform bacteria from diffuse sources as a factor in estuarine pollution. Water Res. 10, 619–627. https://doi.org/10.1016/0043-1354(76)90143-3.

Faust, M.A., Aotaky, A.E., Hargadon, M.T., 1975. Effect of physical parameters on the in situ survival of Escherichia coli MC-6 in an estuarine environment. Appl. Microbiol. 30, 800–806. https://doi.org/10.1128/am.30.5.800-806.1975.

FDACS, 2022. Aquaculture Best Management Practices Manual. Incorporated Into Rule 5L-3.004, F.A.C.

Feng, Z., Reniers, A., Haus, B.K., Solo-Gabriele, H.M., Kelly, E.A., 2016. Wave energy level and geographic setting correlate with Florida beach water quality. Mar. Pollut. Bull. 104, 54–60. https://doi.org/10.1016/j.marpolbul.2016.02.011.

Food and Agriculture Organization (FAO), 2020. The state of world fisheries and aquaculture 2020. Sustainability in action. https://doi.org/10.4060/ca9229en.

Grange, S., 2014. Technical Note: Averaging Wind Speeds and Directions. https://doi.org/10.13140/RG.2.1.3349.2006.
Gronewold, A.D., Wolpert, R.L., 2008. Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. Water Res. 42, 3327–3334. https://doi.org/10.1016/j.watres.2008.04.011.
Haas, C.N., 1996. How to average microbial densities to characterize risk. Water Res. 30, 1036–1038. https://doi.org/10.1016/0043-1354(95)00228-6.
Hapke, C.J., Kramer, P.A., Fetherston-Resch, E.H., Baumstark, R.D., Druyor, R., Fredericks, X., Fitos, E., 2019. Florida coastal mapping program—overview and 2018 workshop report. In: USGS Numbered Series 2019–1017. Florida Coastal Mapping Program—Overview and 2018 Workshop Report. Vol. 2019–1017. Open-File Report. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/ofr20191017.
Hastie, T., Tibshirani, R., Friedman, J., 2009. Random Forests. In: Hastie, T., Tibshirani, R., Friedman, J. (Eds.), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics. Springer, New York, NY, pp. 587–604. https://doi.org/10.1007/978-0-387-84858-7_15.
Howell, J.M., Coyne, M.S., Cornelius, P.L., 1996. Effect of sediment particle size and temperature on fecal Bacteria mortality rates and the fecal coliform/fecal streptococci ratio. J. Environ. Qual. 25, 1216–1220. https://doi.org/10.2134/jeq1996.00472425002500060007x.
Jeng, H.A.C., Englande, A.J., Bakeer, R.M., Bradford, H.B., 2005. Impact of urban stormwater runoff on estuarine environmental quality. Estuar. Coast. Shelf Sci. 63, 513–526. https://doi.org/10.1016/j.ecss.2004.11.024.
Jørgensen, C.B., 1990. Bivalve Filter Feeding: Hydrodynamics, Bioenergetics, Physiology and Ecology. Olsen & Olsen.
Kay, D., Stapleton, C.M., Wyer, M.D., McDonald, A.T., Crowther, J., Paul, N., Jones, K., et al., 2005. Decay of intestinal enterococci concentrations in high-energy estuarine and coastal waters: towards real-time T90 values for modelling faecal indicators in recreational waters. Water Res. 39, 655–667. https://doi.org/10.1016/j.watres.2004.11.014.
Kay, D., Anthony, S., Crowther, J., Chambers, B.J., Nicholson, F.A., Chadwick, D., Stapleton, C.M., Wyer, M.D., 2010. Microbial water pollution: a screening tool for initial catchment-scale assessment and source apportionment. Science of The Total Environment 408, 5649–5656. https://doi.org/10.1016/j.scitotenv.2009.07.033 (Special Section: Integrating Water and Agricultural Management Under Climate Change).
Köppen, W., 1936. Handbuch der Klimatologie: in fünf Banden. Das geographische System der Klimate. In: II. Teil. Das geographische System der Klimate. Verlag von Gebrüder Borntraeger, pp. 98–167. https://doi.org/10.1515/9783111491530-003.
Krometis, L.-A.H., Characklis, G.W., Simmons, O.D., Dilts, M.J., Likirdopulos, C.A., Sobsey, M.D., 2007. Intra-storm variability in microbial partitioning and microbial loading rates. Water Res. 41, 506–516. https://doi.org/10.1016/j.watres.2006.09.029.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., et al., 2022. caret: Classification and Regression Training (version 6.0-90).
Landrum, K.E., Ache, B., 2000. The Economic Impact of Oyster Closures and Their Implications for the Shellfish Challenge in the Baratariaterrebonne National Estuary, Louisiana. Water Environment Federation, pp. 796–807.
Lewis, D.J., Atwill, E.R., M. das G. C. Pereira, and R. Bond., 2013. Spatial and temporal dynamics of fecal coliform and associated with suspended solids and water within five northern California estuaries. J. Environ. Qual. 42, 229–238. https://doi.org/10.2134/jeq2011.0479.
Lin, B., Syed, M., Falconer, R.A., 2008. Predicting faecal indicator levels in estuarine receiving waters – an integrated hydrodynamic and ANN modelling approach. Environ. Model. Software 23, 729–740. https://doi.org/10.1016/j.envsoft.2007.09.009.
Lipp, E.K., Kurz, R., Vincent, R., Rodriguez-Palacios, C., Farrah, S.R., Rose, J.B., 2001. The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. Estuaries 24, 266–276. https://doi.org/10.2307/1352950.
Mallin, M.A., Williams, K.E., Esham, E.C., Lowe, R.P., 2000. Effect of human development on bacteriological water quality in coastal watersheds. Ecol. Appl. 10, 1047–1056. https://doi.org/10.1890/1051-0761(2000)010[1047:EOHDOB]2.0.CO;2.
Mallin, M., Ensign, S., McIver, M., Shank, G., Fowler, P., 2001. Demographic, landscape, and meteorological factors controlling the microbial pollution of coastal waters. Hydrobiologia 460, 185–193. https://doi.org/10.1023/A:1013169401211.
Mattson, R., Frazer, T., Hale, J., Blitch, S., Ahijevych, L., 2006. Florida Big Bend. USGS.
Montgomery, D.C., Peck, E.A., Vining, G.G., 1992. Introduction to Linear Regression Analysis. John Wiley & Sons.
Munro, P.M., Gauthier, M.J., Breittmayer, V.A., Bongiovanni, J., 1989. Influence of osmoregulation processes on starvation survival of Escherichia coli in seawater. Applied and Environmental Microbiology 55, 2017–2024. https://doi.org/10.1128/aem.55.8.2017-2024.1989 (American Society for Microbiology).
National Sea Grant Law Center, 2019. Overcoming Impediments to Shellfish Aquaculture through Legal Research and Outreach: Case Studies 118.
NOAA National Weather Service (NWS) Radar Operations Center, 1991. NOAA next generation radar (NEXRAD) level 2 base data. NOAA National Centers for Environmental Information. 2022 https://doi.org/10.7289/V5W9574V. Accessed November 22.
Noble, R.T., Weisberg, S.B., Leecaster, M.K., McGee, C.D., Ritter, K., Walker, K.O., Vainik, P.M., 2003. Comparison of beach bacterial water quality indicator measurement methods. Environ. Monit. Assess. 81, 301–312. https://doi.org/10.1023/A:1021397529041.
NSSP, 2019. National Shellfish Sanitation Program (NSSP) Guide for the Control of Molluscan Shellfish: 2019 Revision: 502.
Omernik, J.A., 1993. Ecoregions: A Spatial Framework for Environmental Management. Book Chapter. PB-94-209764/XAB; EPA-600/A-94/170. Environmental Protection Agency, Corvallis, OR (United States). Environmental Research Lab.
Parker, J.K., McIntyre, D., Noble, R.T., 2010. Characterizing fecal contamination in stormwater runoff in coastal North Carolina, USA. Water Res. 44, 4186–4194. https://doi.org/10.1016/j.watres.2010.05.018.
Parkhurst, D.F., Brenner, K.P., Dufour, A.P., Wymer, L.J., 2005. Indicator bacteria at five swimming beaches—analysis using random forests. Water Res. 39, 1354–1360. https://doi.org/10.1016/j.watres.2005.01.001.
PRISM Climate Group, 2014. Oregon State University. https://prism.oregonstate.edu (data created 4 Feb, accessed 16 Jan 2023).
Ruth, D.P., Glahn, B., Dagostaro, V., Gilbert, K., 2009. The performance of MOS in the digital age. Weather and forecasting 24. Am. Meteorol. Soc. 504–519. https://doi.org/10.1175/2008WAF2222158.1.
Schmidt, W., Evers-King, H.L., Campos, C.J.A., Jones, D.B., Miller, P.I., Davidson, K., Shutler, J.D., 2018. A generic approach for the development of short-term predictions of Escherichia coli and biotoxins in shellfish. Aquac. Environ. Interact. 10, 173–185. https://doi.org/10.3354/aei00265.
SECOORA, 2022. SECOORA data portal. https://portal.secoora.org/. Accessed March 1 2022.
Seo, D.-J., 1998. Real-time estimation of rainfall fields using radar rainfall and rain gage data. J. Hydrol. 208, 37–52. https://doi.org/10.1016/S0022-1694(98)00141-3.
Sinton, L.W., 2005. Biotic and abiotic effects. In: Belkin, S., Colwell, R.R. (Eds.), Oceans and Health: Pathogens in the Marine Environment. Springer US, Boston, MA, pp. 69–92. https://doi.org/10.1007/0-387-23709-7_3.
Sturmer, L., 2019. Getting Started. Florida Shellfish Aquaculture Online Resource Guide.
Theuerkauf, S.J., Jr, J.A.M., Waters, T.J., Wickliffe, L.C., Alleway, H.K., Jones, R.C., 2019. A global spatial analysis reveals where marine aquaculture can benefit nature and people. PLOS ONE 14 (Public Library of Science: e0222282). https://doi.org/10.1371/journal.pone.0222282.
Troussellier, M., Bonnefont, J.-L., Courties, C., Derrien, A., Dupray, E., Gauthier, M., Gourmelon, M., et al., 1998. Responses of enteric bacteria to environmental stresses in seawater. Oceanol. Acta 21, 965–981. https://doi.org/10.1016/S0399-1784(99)80019-X.
Ufnar, D., Ufnar, J.A., Ellender, R.D., Rebarchik, D., Stone, G., 2006. Influence of coastal processes on high fecal coliform counts in the Mississippi sound. J. Coast. Res. 22, 1515–1526. https://doi.org/10.2112/04-0434.1.
University of Florida, 2019. About the Industry. Florida Shellfish Aquaculture Online Resource Guide.
USGS, 2017. National Hydrography Dataset. U.S. Geological Survey data release.
Vidon, P., Tedesco, L.P., Wilson, J., Campbell, M.A., Casey, L.R., Gray, M., 2008. Direct and indirect hydrological controls on E. coli concentration and loading in Midwestern streams. J. Environ. Qual. 37, 1761–1768. https://doi.org/10.2134/jeq2007.0311.
Voronoi, G., 1909. Nouvelles applications des paramètres continus à théorie des formes quadratiques. Deuxième Mémoire. Recherches sur les parallélloèdres primitifs. Journal für die reine und angewandte Mathematik 1909, 67–182. https://doi.org/10.1515/crll.1909.136.67 (Berlin, New York: Walter de Gruyter, Berlin/New York).
Wang, J., Deng, Z., 2019. Modeling and predicting fecal coliform bacteria levels in oyster harvest waters along Louisiana Gulf coast. Ecol. Indic. 101, 212–220. https://doi.org/10.1016/j.ecolind.2019.01.013.
Wang, S., Peng, H., Liang, S., 2022. Prediction of estuarine water quality using interpretable machine learning approach. J. Hydrol. 605, 127320 https://doi.org/10.1016/j.jhydrol.2021.127320.
Weather Prediction Center, 2022. Product Information: Quantitative Precipitation Forecasts (QPFs). National Weather Service.
Wilkinson, J., Kay, D., Wyer, M., Jenkins, A., 2006. Processes driving the episodic flux of faecal indicator organisms in streams impacting on recreational and shellfish harvesting waters. Water Res. 40, 153–161. https://doi.org/10.1016/j.watres.2005.11.001.
Xu, L., Chen, N., Yang, C., Yu, H., Chen, Z., 2022. Quantifying the uncertainty of precipitation forecasting using probabilistic deep learning. Hydrology and Earth System Sciences 26, 2923–2938. https://doi.org/10.5194/hess-26-2923-2022 (Copernicus GmbH).
You, S., Huang, X., Xing, L., Lesperance, M., LeBlanc, C., Moccia, L.P., Mercier, V., Shao, X., Pan, Y., Zhang., X., 2023. Dynamics of fecal coliform bacteria along Canada's coast. Mar. Pollut. Bull. 189, 114712 https://doi.org/10.1016/j.marpolbul.2023.114712.
Zhang, X., Zhi, X., Chen, L., Shen, Z., 2020. Spatiotemporal variability and key influencing factors of river fecal coliform within a typical complex watershed. Water Res. 178, 115835 https://doi.org/10.1016/j.watres.2020.115835.
Zimmer-Faust, A.G., Brown, C.A., Manderson, A., 2018. Statistical models of fecal coliform levels in Pacific Northwest estuaries for improved shellfish harvest area closure decision making. Mar. Pollut. Bull. 137, 360–369. https://doi.org/10.1016/j.marpolbul.2018.09.028.
Zuur, A., Ieno, E.N., Smith, G.M., 2007. Analyzing Ecological Data, 2007th edition. Springer, New York; London.