Mapping Optimal Locations for Oyster Aquaculture: I. Remote Sensing and Field Observations

Final Report

March 2024

GLO Contract No.: 22-045-006-D103

Prepared by:
Mohamed Ahmed, Principal Investigator
Meghan Bygate, Graduate Research Assistant

Texas A&M University-Corpus Christi
6300 Ocean Dr., Unit 5850
Corpus Christi, Texas 78412
Phone: 361-825-3278
Email: Mohamed.ahmed@tamucc.edu

Submitted to:
Texas General Land Office
1700 Congress Ave.
Austin, TX 78701-1495

This report was funded in part by a Texas Coastal Management Program grant approved by the Texas Land Commissioner, providing financial assistance under the Coastal Zone Management Act of 1972, as amended, awarded by the National Oceanic and Atmospheric Administration (NOAA), Office for Coastal Management, pursuant to NOAA Award No. NA21NOS4190136. The views expressed herein are those of the author(s) and do not necessarily reflect the views of NOAA, the U.S. Department of Commerce, or any of their subagencies.

Executive Summary

Due to the growing demand for high-quality oysters, it is becoming increasingly important to identify new, optimal sites with the best biophysical conditions for oyster reefs in Matagorda Bay. Oyster Habitat Suitability Index (OHSI) models evaluate habitat quality and may greatly aid restoration projects. Environmental and water quality indicators (WQIs) are used to generate these OHSIs. In this study, we extracted WQIs from remote-sensing data during the period from 2014 to 2023 (Section I), then used these data to produce five OHSI models for Matagorda Bay (Section II).

Remote-sensing datasets offer a unique opportunity to observe spatial and temporal trends in WQIs—such as chlorophyll-a, salinity, and turbidity—across various aquatic ecosystems. In this study, we used available in-situ WQI measurements (chlorophyll-a: 17, salinity: 478, and turbidity: 173) and Landsat-8 surface-reflectance data to examine the capability of empirical and machine-learning (ML) models to retrieve these indicators over Matagorda Bay, Texas, between 2014 and 2023. We employed 36 empirical models to retrieve chlorophyll-a (12 models), salinity (2 models), and turbidity (22 models) and four ML families—deep neural network, distributed random forest, gradient boosting machine, and generalized linear model—to retrieve salinity and turbidity. We used the Nash-Sutcliffe efficiency coefficient, correlation coefficient (r), and normalized root mean square error to assess the performance of empirical and ML models. Results indicate that (1) the empirical models were minimally effective when applied over Matagorda Bay without calibration; (2) once calibrated over Matagorda Bay, the performance of the empirical models improved significantly; (3) ML models outperformed calibrated empirical models when used to retrieve turbidity and salinity; and (4) the deep neural network family outperformed all other ML families when used to retrieve salinity.

We then generated five OHSI models over Matagorda Bay during the period from 2018 to 2023 on both monthly and annual scales. Each model used five physical parameters (model inputs): salinity, turbidity, water temperature, depth, and water velocity. OHSI1 is a classification-based overlay model, classifying habitat quality in each model input based on specific thresholds that define unsuitable, moderate, and optimal habitat conditions. OHSI2 is a regression-based nonlinear model that regressed these inputs against total live oyster count to depict habitat quality. Both OHSI1 and OHSI2 were implemented through unweighted and weighted versions. OHSI3 is a ML-based model that uses total live oyster count as a target. OHSI scores from all models demonstrated spatial and temporal variability in habitat quality. Across all models, optimal habitat was generally found in shallower regions, near freshwater sources, and in areas with high water flow. Unsuitable habitat was typically identified in the primary bay, characterized by greater water depths, higher salinity, and slower water flow. OHSI1 indicated that spring and fall were optimal for oysters, while winter was unsuitable. OHSI2 identified winter months as optimal and considered summer unsuitable. Conversely, OHSI3 found the most suitable conditions in summer and determined that spring and late fall were unsuitable.

The developed approach provides a reference context, a structured framework, and valuable insights for using empirical and ML models and Landsat-8 data to retrieve WQIs over aquatic ecosystems. The modeled WQI data could be instrumental in expanding the footprint of in-situ observations and improving current efforts to conserve, enhance, and restore important habitats in aquatic ecosystems. OHSI models generated from this study offer valuable insights into long-term habitat restoration activities in Matagorda Bay.

Table of Contents

List of Figures

List of Tables

Section (I)*

Monitoring Water Quality Indicators over Matagorda Bay, Texas, Using Landsat-8

(*) based on article that is currently under review (Bygate, M., Ahmed, M., 2024, Monitoring Water Quality Indicators over Matagorda Bay, Texas Using Landsat-8. Remote Sensing).

1.1. Abstract

Remote sensing datasets offer a unique opportunity to observe spatial and temporal trends in water quality indicators (WQIs), such as chlorophyll-a, salinity, and turbidity, across various aquatic ecosystems. In this study, we used available in-situ WQI measurements (chlorophyll-a: 17, salinity: 478, and turbidity: 173) along with Landsat-8 surface reflectance data to examine the capability of empirical and machine-learning (ML) models in retrieving these indicators over Matagorda Bay, Texas, between 2014 and 2023. We employed 36 empirical models to retrieve chlorophyll-a (12 models), salinity (2 models), and turbidity (22 models) and four ML families—deep neural network (DNN), distributed random forest, gradient boosting machine, and generalized linear model—to retrieve salinity and turbidity. We used the Nash-Sutcliffe efficiency coefficient (NSE), correlation coefficient (*r*), and normalized root mean square error (NRMSE) to assess the performance of empirical and ML models. Results indicate that (1) the empirical models displayed minimal effectiveness when applied over Matagorda Bay without calibration, (2) once calibrated over Matagorda Bay, the performance of the empirical models experienced significant improvements (chlorophyll-a: NRMSE: 0.91 ± 0.03, *r*: 0.94 ± 0.04, NSE: 0.89 ± 0.06; salinity: NRMSE: 0.24 ± 0, *r*: 0.24 ± 0, NSE: 0.06 ± 0; turbidity: NRMSE: 0.15 ± 0.10, *r*: 0.13 ± 0.09, NSE: 0.03 ± 0.03), (3) ML models outperformed calibrated empirical models when used to retrieve turbidity and salinity, and (4) the DNN family outperformed all other ML families when used to retrieve salinity (NRMSE: 0.87 ± 0.09, *r*: 0.49 ± 0.09, NSE: 0.23 ± 0.12) and turbidity (NRMSE: 0.63± 0.11, *r*: 0.79 ± 0.11, NSE: 0.60 ± 0.20). The developed approach provides a reference context, a structured framework, and valuable insights for using empirical and ML models and Landsat-8 data to retrieve WQIs over aquatic ecosystems. The modeled WQI data could be used to expand the footprint of in-situ observations and improve current efforts to conserve, enhance, and restore important habitats in aquatic ecosystems.

1.2. Introduction

Collecting and analyzing water quality (WQ) data is crucial for assessing spatiotemporal trends in the health of aquatic ecosystem. WQ has a substantial impact on the wellbeing of the human population and the species that rely on these ecosystems (Bugica et al., 2020; Silva et al., 2022; Strobl & Robillard, 2008; Wilber & Bass, 1998). Water quality indicators (WQIs) are specific parameters used to assess the WQ in various ecosystems (Kumar & Dua, 2009; Misaghi et al., 2017). These indicators include parameters such

as temperature (Kannel et al., 2007; Mishra et al., 2021), dissolved oxygen (DO) (Kannel et al., 2007; Mishra et al., 2021), pH (hydrogen ion concentration) (Kannel et al., 2007; Mishra et al., 2021), salinity (Mishra et al., 2021), turbidity (Hossain et al., 2021), suspended solids (Kannel et al., 2007; Lim & Choi, 2015; Mishra et al., 2021), chlorophyll-a (Lim & Choi, 2015), nitrogen (Lim & Choi, 2015; Mishra et al., 2021), phosphorus (Lim & Choi, 2015; Mishra et al., 2021), and heavy metals (Poshtegal & Mirbagheri, 2023). The ultimate purpose of monitoring WQIs is to identify changes in aquatic conditions and to provide data to decision makers that they can use to inform their environmental management and protection practices (AL-Fahdawi et al., 2015; Behmel et al., 2016; Ighalo & Adeniyi, 2020).

In-situ observations involve collecting measurements of WQIs directly at a location of interest within an aquatic ecosystem. These observations provide a common approach to monitor WQIs across ecosystems (González et al., 2018; Lim & Choi, 2015). However, these observations are relatively expensive and time consuming, and are spatially and temporally limited (AL-Fahdawi et al., 2015; Peterson et al., 2020; Silva et al., 2022). Some areas are difficult to access for sampling purposes and/or have very limited resources. In addition, periodic WQ sampling stations might also miss the short-lived magnitudes, trends, and gradients in WQIs. Moreover, permanent field WQ stations are vulnerable to damage from extreme climate events.

Remote-sensing datasets (e.g., visible, thermal, radar) in the public domain provide a unique and cost-effective opportunity to monitor WQIs, especially in large and inaccessible aquatic ecosystems (Goddijn-Murphy et al., 2009; Lim & Choi, 2015; Pahlevan et al., 2022). These space-based observations offer a wealth of data with excellent spectral, spatial, and temporal resolutions that facilitate the calculation of WQIs. Landsat images, for example, have been extensively used to estimate optical and non-optical WQIs such as turbidity (Carpenter & Carpenter, 1983; González-Márquez et al., 2018; Peterson et al., 2020; Sharaf El Din, 2020), total nitrogen (TN) and total phosphorus (TP) (Lim & Choi, 2015; Peterson et al., 2020; Sharaf El Din, 2020; Vakili & Amanollahi, 2020; Wei et al., 2022), chlorophyll-a (Markogianni et al., 2017; Peterson et al., 2020; Sudheer et al., 2006; Zhang et al., 2022), DO (Peterson et al., 2020), total suspended solids (TSS) (Sharaf El Din, 2020), salinity (Ansari & Akhoondzadeh, 2020; Bayati & Danesh-Yazdi, 2021), and ammonium (Markogianni et al., 2018). Moderate Resolution Imaging Spectroradiometer images have been used to retrieve salinity (Hu et al., 2004; Wong et al., 2007), chlorophyll-a (Hu et al., 2004; Huang et al., 2011; Kim et al., 2017; Liu et al., 2021), colored dissolved organic material (Hu et al., 2004; Schaeffer et al., 2015; Yu et al., 2016), TSS (Kim et al., 2017; Mathew et al., 2017), TN (Arıman, 2021; Mathew et al., 2017; Singh et al., 2013; Yu et al., 2016), and TP (Arıman, 2021; Mathew et al., 2017; Singh et al., 2013). Sentinel-2 images have been used to retrieve salinity (Bayati & Danesh-Yazdi, 2021; Hossen et al., 2022), chlorophyll-a (Hossen et al., 2022; Sent et al., 2021; Toming et al., 2016; Virdis et al., 2022), TN (Guo et al., 2021; Hossen et al., 2022), TP (Guo et al., 2021; Hossen et al., 2022), colored

dissolved organic material (Sent et al., 2021; Toming et al., 2016; Virdis et al., 2022), turbidity (Sent et al., 2021; Torres-Bejarano et al., 2021; Virdis et al., 2022), and electrical conductivity (EC) (Torres-Bejarano et al., 2021). Thermal infrared data has been used to retrieve water temperature (Ding & Elmore, 2015; Gorokhovich et al., 2022; Shi & Hu, 2021). Radar data were used to retrieve EC (Shareef, 2015; Shareef et al., 2016), salinity (Shareef, 2015; Shareef et al., 2016), and total dissolved salts (Shareef, 2015; Shareef et al., 2016). Previous efforts that used remote-sensing products to monitor WQIs over multiple ecosystems commonly used empirical models (Ansari & Akhoondzadeh, 2020; Carpenter & Carpenter, 1983; He et al., 2021; Lim & Choi, 2015; Markogianni et al., 2018; Sharaf El Din, 2020). However, determining the appropriate model for extracting WQI in a specific area of interest, along with the factors to be considered while selecting such a model, presents a significant challenge.

With the rise of artificial intelligence techniques, modeling capabilities have made great advances (Trinh et al., 2017; Wei et al., 2020). Machine learning (ML) algorithms were used to analyze complex linear and nonlinear patterns between remote-sensing datasets and WQI data (Lim & Choi, 2015; Peterson et al., 2020; Sudheer et al., 2006; Zhang et al., 2022). Several studies used convoluting neural networks, artificial neural networks, and deep neural networks (DNNs) to retrieve salinity (Ansari & Akhoondzadeh, 2020; Bayati & Danesh-Yazdi, 2021), TN (Jakovljević et al., 2018; Pu et al., 2019; Zhang et al., 2022), TP (Jakovljević et al., 2018; Pu et al., 2019; Zhang et al., 2022), chlorophyll-a (Peterson et al., 2020; Sudheer et al., 2006), DO (Jakovljević et al., 2018; Peterson et al., 2020), and turbidity (Bormudoi et al., 2022; Jakovljević et al., 2018). Other ML families, such as Extreme Gradient Boosting, Gradient Boosting Machine (GBM), Regression Tree, and Multi-Layer Perceptron have been implemented to retrieve EC (Krishnaraj & Honnasiddaiah, 2022), DO (Krishnaraj & Honnasiddaiah, 2022; Wagle et al., 2020), pH (Krishnaraj & Honnasiddaiah, 2022; Wagle et al., 2020), TN (Li et al., 2022), TP (Li et al., 2022; Wagle et al., 2020), chlorophyll-a (Wagle et al., 2020), and turbidity (Wagle et al., 2020). However, determining the most suitable ML model family for a specific area of interest remains challenging.

Matagorda Bay, also called the Lavaca-Colorado estuary, is one of seven major estuarine systems located on the Texas coast (Figure 1; inset a). As the second largest estuary in Texas, Matagorda Bay is known for its vital role in supporting diverse ecosystems, serving as a critical habitat for marine life, and providing a foundation for commercial and recreational activities (Aguilar, 2017; Brody et al., 2004; Wilber & Bass, 1998).

In this study, turbidity, chlorophyll-a, and salinity WQIs were retrieved from Landsat-8 images over Matagorda Bay during the period from 2014 to 2023. Because they exhibit complex spatiotemporal patterns and trends, these three WQIs have been established in several studies as significant markers of health of Matagorda Bay (Onabule et al., 2020; Ward & Armstrong, 1980). Along with water depth, current conditions, and temperature, several studies have found these WQIs to be the most significant in controlling

the growth of oysters in Matagorda Bay (Aguilar, 2017; Kinsey & Montagna, 2005; Marshall et al., 2019; Wilber & Bass, 1998). Both empirical and ML models were used. Specific questions to be addressed in this study include: (1) Which empirical model(s) could perform well in extracting chlorophyll-a, salinity, and turbidity over Matagorda Bay? (2) Which model(s) witness a significant performance enhancement once recalibrated over Matagorda Bay? (3) Do ML models usually outperform empirical models in extracting WQIs? (4) Which ML model family performs well in extracting salinity and turbidity over Matagorda Bay?

A total of 36 empirical models, documented in previous research, were applied to retrieve turbidity, chlorophyll-a, and salinity over Matagorda Bay, Texas (Figure 1). Specifically, 12 models were employed for chlorophyll-a, 2 models for salinity, and 22 models for turbidity estimation. In addition, four distinct ML algorithm families—DNN, GBM, Distributed Random Forest (DRF), and Generalized Linear Model (GLM)—were employed to retrieve salinity and turbidity over Matagorda Bay.

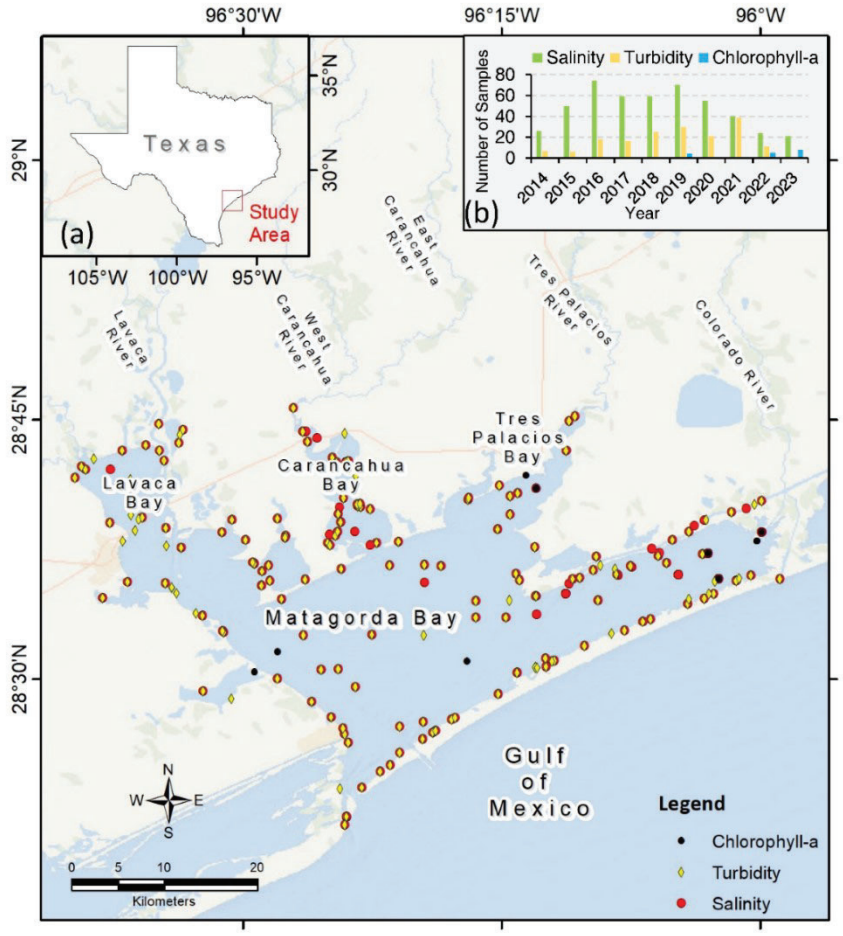

Figure 1: Spatial distribution of Matagorda Bay along the Texas Gulf Coast, and locations of the in-situ WQIs measurements used in this study. (a) Location of the study area along the Texas coast. (b) Yearly distribution of in-situ WQIs measurements during the study period (2014–2023).

1.3. Data

1.3.1. Study Area

Matagorda Bay (area: 1,070 km^2) is comprised of a primary bay (Matagorda Bay) and several subsystems, including Lavaca Bay, Tres Palacios Bay, and East Matagorda Bay (Figure 1) (Brody et al., 2004; McBride, 2022; Ward & Armstrong, 1980). Matagorda Bay holds diverse ecological significance as an estuarine ecosystem, nurturing a wide range of flora and fauna and functioning as a habitat and breeding ground for various fish, shellfish, and aquatic organisms that enhance local biodiversity and support fisheries (Caillier, 2023; Olsen, 2019; Renaud & Williams, 2023; Wetzel & Armstrong, 1987). This ecological richness translates into substantial economic value, driven by thriving fisheries, recreational pursuits, and the potential for shipping and transportation. Both commercial and recreational fishing bolster the local economy, and the bay attracts tourists for activities like boating, birdwatching, and outdoor exploration (Brody et al., 2004; Haby, 2012; Ropicki et al., 2016).

Oysters are commercially harvested in Matagorda Bay (Culbertson, 2008). They also play a crucial role in maintaining WQ, supporting local fisheries, and enhancing the overall ecological resilience of Matagorda Bay (Kim & Montagna, 2009; Kinsey & Montagna, 2005). Studies have highlighted the oyster reefs in Matagorda Bay as important nurseries for various commercially valuable species, such as spotted seatrout and red drum (Grabowski et al., 2012). The threats these oyster populations face include habitat degradation, reduced water flow, and disease, which are directly related to the WQ in Matagorda Bay (Aguilar, 2017; Kinsey & Montagna, 2005). Monitoring WQIs is significant for developing effective conservation measures that are crucial to preserving the ecological and economic benefits that oysters in Matagorda Bay provide.

Matagorda Bay's ecosystem is significantly shaped by its connectivity to neighboring water bodies, primarily the Gulf of Mexico for the main bay and river connections for the secondary bays (Kim & Montagna, 2009). The estuary is fed by four major rivers: the Colorado River, the Lavaca River, the Carancahua River, and the Tres Palacios River (Aguilar, 2017; Palmer et al., 2011). The Colorado River is the primary source of freshwater, supplying an estimated 34% of total freshwater inflow for bay system (Kim & Montagna, 2009; Kucera et al., 2002; McBride, 2022; Montagna, 1994; Wetzel & Armstrong, 1987). Freshwater input plays a pivotal role in determining the overall health of Matagorda Bay (Aguilar, 2017; Kim & Montagna, 2009; Palmer et al., 2011; Wetzel & Armstrong, 1987; Wilber & Bass, 1998); it supplies nutrients to the bay (McBride, 2022; Wilber & Bass, 1998), regulates salinity (Aguilar, 2017; Kim & Montagna, 2009; McBride, 2022; Montagna, 1994; Wilber & Bass, 1998), and is responsible for the transportation of sediment (Kim & Montagna, 2009; Wilber & Bass, 1998).

Table 1: Statistics for in-situ WQIs (salinity, turbidity, and chlorophyll-a) collected over Matagorda Bay, at locations shown in Figure 1, from 2014 to 2023. Statistical measurements include maximum (Max.), minimum (Min.), and average and standard deviation (Avg. ± Std. Dev.).

WQI	Units	No. of Samples	Max.	Min.	Avg. ± Std. Dev.
Salinity	psu	478	35.91	0.10	17.37 ± 8.22
Turbidity	NTU	173	91.00	2.00	25.18 ± 18.60
Chlorophyll-a	µg/L	17	25.50	0.04	5.11 ± 7.55

1.3.2. In-Situ Data

We gathered in-situ measurements over Matagorda Bay for the period from 2014 to 2023 (Inset b; Figure 1). This data was collected from the public databases of the Texas Commission on Environmental Quality (TCEQ) (TCEQ, 2023), Lower Colorado River Association (LCRA) (LCRA, 2023), and Texas Wildlife Parks Department (TWPD). TCEQ sampling procedures have been adopted by both LCRA and TWPD (LCRA, 2022, 2023). For sampling within estuaries, water samples are obtained at a depth of 0.30 m. Salinity was measured using a multiprobe instrument and recorded to the nearest 0.10 psu (practical salinity unit) (TCEQ, 2012). Turbidity is measured using the benchtop turbidity meter instrument and reported to the nearest 0.02 NTU (nephelometric turbidity unit) (TCEQ, 2012; Texas Secretary of State, 2023). Chlorophyll-a concentrations were measured on collected water samples (TCEQ, 2012) using a spectrophotometer and recorded in µg/L (Dunne, 1999).

The locations of in-situ measurements are shown in Figure 1. A total of 17 measurements were collected for chlorophyll-a, 478 for salinity, and 173 for turbidity. These are the only measurements that align in time with the dates of Landsat-8 acquisition during the investigated period (2014–2023). Table 1 provides statistics for these measurements. Table 1 indicates that over Matagorda Bay, chlorophyll-a concentrations ranged from 0.04 µg/L in August 2023 to 25.50 µg/L in September 2019. Salinity ranged from 0.10 psu in May 2021 to 35.91 psu in October 2022. Turbidity ranged from 2.0 NTU in November 2017 to 91.0 NTU in July 2020. The average (± standard deviation) is estimated at 17.37 ± 8.22 psu, 25.18 ± 18.60 NTU, and 5.11 ± 7.55 µg/L, for salinity, turbidity, and chlorophyll-a, respectively.

1.3.3. Remote-Sensing Data

In this study, we used Landsat-8 (Collection 2, Level 2, Tier 1) images from 2014 to 2023 over Matagorda Bay. These images cover path 26 and row 40 of the Landsat worldwide reference system. Landsat-8 launched on February 11, 2013, as a joint mission between U.S. Geological Survey and the National Aeronautics and Space Administration (Danbara, 2014; Lim & Choi, 2015). Landsat-8 carries two instruments, the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIR). These instruments

record data over 11 spectral bands. OLI records apparent radiance in 9 optical bands in wavelengths ranging from optical to shortwave infrared at a spatial resolution of 30 m. The TIR instrument records in thermal infrared wavelengths at a 120-m spatial resolution. Landsat-8 has 16-day temporal resolution and a 185-km swath width. A total of five Landsat-8 images were used to retrieve chlorophyll-a. Salinity used 121 images, while turbidity used 43 images.

1.4. Methods

1.4.1. Surface Reflectance

Surface reflectance data were extracted from Landsat-8 images in a Google Earth Engine (GEE) environment. In the GEE editor, the scale factor was applied to Landsat-8's band 1 through band 7 (B1–B7) to generate surface reflectance data. These surface reflectance data were corrected for the temporally, spatially, and spectrally varying scattering and absorbing effects of atmospheric gases, aerosols, and water vapor (Danbara, 2014; Lim & Choi, 2015). For modelling purposes, surface reflectance data were extracted over locations of in-situ observations (Figure 1) and times where in-situ observations align with the dates of Landsat-8 acquisition during the investigated period (2014–2023). Due to the inherently low surface reflectance of water, some pixels showed negative surface reflectance values (Kuhn et al., 2019; Mondejar & Tongco, 2019). These pixels were deleted and not considered for further analyses. Surface reflectance values for pixels in cloud-contaminated areas were also removed. Figure 2 displays scatterplots depicting the relationship between surface reflectance data across each band (B1–B7) and the WQIs. The correlation coefficient between chlorophyll-a and surface reflectance, from all bands, was estimated to be 0.89 ± 0.07 (average ± standard deviation). B6 and B7 exhibit the highest positive correlation (0.96 and 0.98, respectively), with chlorophyll-a while B1 and B2 demonstrate the lowest positive correlations (0.80 and 0.81, respectively). The correlation coefficient between salinity and surface reflectance is weak. It estimated at -0.11 ± 0.04. B4 and B5 show the maximum negative correlation of -0.16, whereas B7 shows the minimum negative correlation (-0.05). The correlation coefficient between turbidity and surface reflectance is estimated to be 0.05 ± 0.04. B4 exhibits the highest correlation of 0.12, whereas B7 demonstrates the lowest correlation of -0.02.

Table 2 presents statistics for surface reflectance data generated from B1 through B7 during the periods when in-situ observations correspond with Landsat-8 acquisition dates. The surface reflectance data used for salinity retrieval displayed mean values ranging from 0.04 to 0.11, with corresponding standard deviations ranging from 0.03 to 0.07. For turbidity, the surface reflectance data revealed average values spanning from 0.06 to 0.12, accompanied by standard deviations ranging from 0.04 to 0.09. The surface reflectance data used for chlorophyll-a retrieval demonstrated average and standard deviation values ranging from 0.07 to 0.15 and 0.09 to 0.16, respectively.

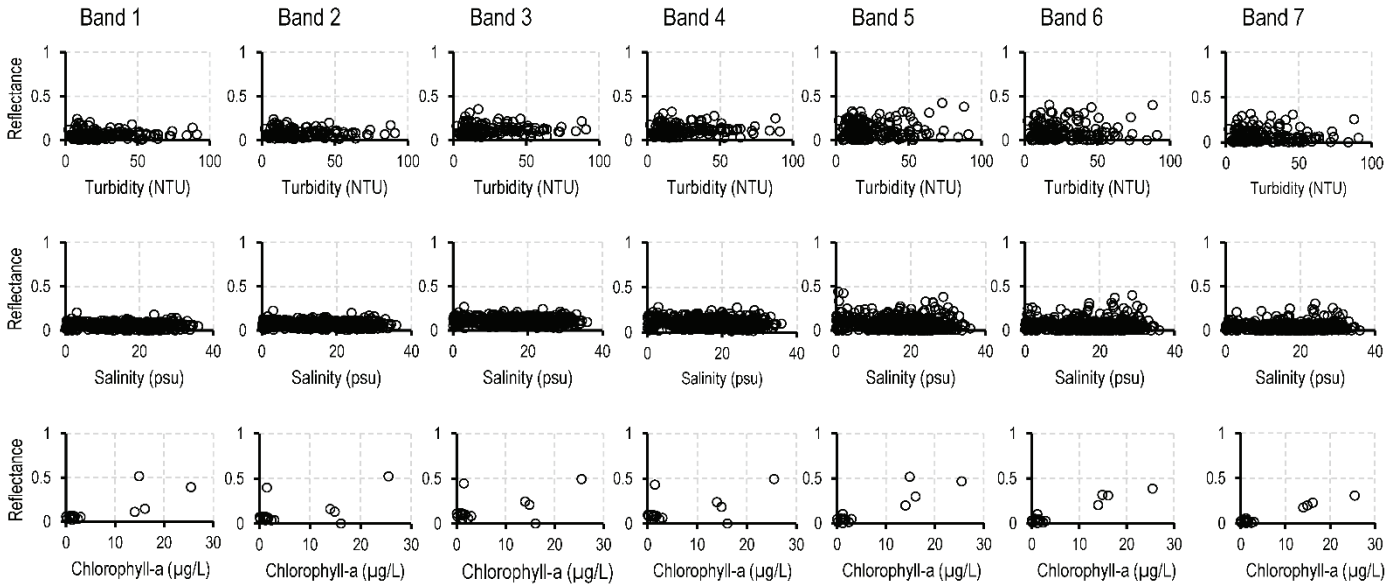

Figure 2: Scatterplots of turbidity (top panel), salinity (middle panel), and chlorophyll-a (bottom panel) and Landsat-8 derived surface reflectance data at each band (B1 – B7).

1.4.2. Empirical Models

A total of 36 empirical models, documented in previous studies, were applied to Landsat-8 surface reflectance data to extract WQIs data for Matagorda Bay. We reported all empirical models that were published and used to retrieve WQIs from Landsat-8 data. These models along with their geographic origins and sources are presented in Table 3. Salinity retrieval used 2 models (Sn1 and Sn2), turbidity retrieval involved 22 models (T1–T22), and chlorophyll-a estimation used 12 models (C1–C12) (Table 3). The empirical models for chlorophyll-a used B1, B2, B3, B4, B5, B6, and B7. Salinity made use of B1, B2, B3, and B4, while turbidity models employed B1, B2, B3, B4, and B5. These models were initially used with their original weights (Table 3), i.e., uncalibrated models. Additionally, we performed recalibration specifically tailored to Matagorda Bay, resulting in the generation of new weights for each of these models (i.e., calibrated models). Empirical models over Matagorda Bay were calibrated using the ordinary least squares package in a geographic information system environment (Cabral et al., 2011; Peprah & Mensah, 2017) .

Table 2: Statistics for surface reflectance data extracted at in-situ locations. Statistical measurements include minimum (Min.), maximum (Max.), and average and standard deviation (Avg. ± Std. Dev).

WQI		Salinity (478)	Turbidity (173)	Chlorophyll-a (17)
B1	Min.	0.00	0.00	0.02
	Max.	0.20	0.24	0.52
	Avg. ± Std. Dev	0.05 ± 0.03	0.06 ± 0.04	0.10 ± 0.14
B2	Min.	0.01	0.01	0.04
	Max.	0.23	0.24	0.52
	Avg. ± Std. Dev	0.07 ± 0.03	0.08 ± 0.04	0.12 ± 0.13
B3	Min.	0.02	0.04	0.06
	Max.	0.27	0.35	0.49
	Mean	0.11 ± 0.04	0.12 ± 0.05	0.15 ± 0.13
B4	Min.	0.01	0.03	0.04
	Max.	0.28	0.32	0.49
	Avg. ± Std. Dev	0.11 ± 0.05	0.11 ± 0.06	0.14 ± 0.13
B5	Min.	0.00	0.00	0.00
	Max.	0.44	0.42	0.52
	Avg. ± Std. Dev	0.08	0.12	0.12
	St. Dev.	0.07	0.09	0.16
B6	Min.	0.00	0.00	0.00
	Max.	0.40	0.40	0.39
	Avg. ± Std. Dev	0.06 ± 0.06	0.10 ± 0.09	0.10 ± 0.12
	St. Dev.	0.06	0.09	0.12
B7	Min.	0.00	0.00	0.00
	Max.	0.31	0.32	0.31
	Avg. ± Std. Dev	0.04 ± 0.04	0.07 ± 0.07	0.07 ± 0.09

Table 3: Empirical models coefficients published in previous studies along with the locations they were applied to and their respective sources. Models over Matagorda Bay were calibrated.

WQI	Model ID	Equation	Location	Source
Chlorophyll-a	C1	$-0.025 + 1.029(B3) + 0.643(B5)$	Krishnagiri Reservoir, India	(Elangovan & Murali, 2020)
		$1.03 - 20.11(B3) + 59.62(B5)$	Matagorda Bay	—
	C2	$0.2 - 1.504(B3) - 1.321(B1) - 2.567(B7) + 0.06(B6) + 4.668(B5)$	Krishnagiri Reservoir, India	(Elangovan & Murali, 2020)
		$0.70 - 18.98(B1) - 16.66(B3) + 84.92(B5) - 173.05(B6) + 215.55(B7)$	Matagorda Bay	—
	C3	$0.011 + 1.091(B6) + 0.133(B7)$	Krishnagiri Reservoir, India	(Elangovan & Murali, 2020)
		$-0.57 - 26.82(B6) + 114.46(B7)$	Matagorda Bay	—
	C4	$0.010 - 0.468(B1) + 1.525(B5)$	Krishnagiri Reservoir, India	(Elangovan & Murali, 2020)
		$0.14 - 45.55(B1) + 81.13(B5)$	Matagorda Bay	—
	C5	$-0.004 + 0.977(B3) - 0.02(B1)$	Krishnagiri Reservoir, India	(Elangovan & Murali, 2020)
		$-8.88 - 106.68(B1) + 163.85(B3)$	Matagorda Bay	—
	C6	$14027.14e^{-3.36[(B1+B4)/B3])}$	Barataria Basin, Mississippi	(Vargas-Lopez et al., 2021)
		$-7.84 - 117.05(B1) + 76.32(B3) - 96.43(B4)$	Matagorda Bay	—
	C7	$-38.621 + 92.050[B2/(B1 + B2 + B3)]$	Trichonis Lake, Greece	(Markogianni et al., 2018)

		$+2239.647[(B1+B2)/2]$		
		$-6.27 + 243.63(B1) - 395.91(B2) + 209.61(B3)$	Matagorda Bay	—
	C8	$2.41\left(B4/B3\right) + 0.187$	Java Sea, Cirebon	(Buditama et al., 2017)
		$-5.30 + 196.76(B3) - 141.61(B4)$	Matagorda Bay	—
	C9	$65.7 - 4932.7(B4)$	Lake Chivero, Zimbabwe	(Masocha, Dube, et al., 2018)
		$-2.02 + 50.86(B4)$	Matagorda Bay	—
	C10	$-26.50 + 67.4\left(B5/B4\right)$	Lake Chivero, Zimbabwe	(Masocha, Dube, et al., 2018)
		$1.78 - 45.48(B4) + 80.70(B5)$	Matagorda Bay	—
	C11	$7.7555\log\left(B3/B1\right) + 1.1738$	Jordan Lake, North Carolina	(Yang & Anderson, 2016)
		$-8.88 - 106.68(B1) + 163.85(B3)$	Matagorda Bay	—
	C12	$0.7354log\left(B5/B3\right) + 1.5972$	Jordan Lake, North Carolina	(Yang & Anderson, 2016)
		$1.03 - 20.11(B3) + 59.62(B5)$	Matagorda Bay	—
Salinity	Sn1	$-71820(B4)^2 + 1334.6(B4) + 4564.5(B3)^2 - 235.67(B3) - 24340(B2)^2 + 1187.3(B2) + 32232(B1)^2 - 1333.5(B1) + 39.97$	Arabian Gulf	(Zhao & Temimi, 2016)
		$16.85 + 73.85(B1) - 110.32(B2) + 164.33(B3) - 127.86(B4)$	Matagorda Bay	—
	Sn2	$39.664 - 1233.1(B1) + 1067.5(B2) - 189.58(B3) + 1640.8(B4) + 23823(B1)^2 - 17844(B2)^2 + 1944.7(B3)^2 - 94613(B4)^2$	Arabian Gulf	(Zhao et al., 2017)
		$16.85 + 73.85(B1) - 110.32(B2) + 164.33(B3) - 127.86(B4)$	Matagorda Bay	—
Turbidity	T1	$289.1\frac{B4\pi}{1-((B4\pi)/16.86)}$	Damariscotta River and Harpswell Sound Bay, Maine	(Snyder et al., 2017)
		$20.73 + 39.71(B4)$	Matagorda Bay	
	T2	$380.32(B4) - 1.7826$	Cam Ranh Bay (CRB) and Thuy Trieu Lagoon (TTL), Vietnam	(Quang et al., 2017)
		$20.73 + 39.71(B4)$	Matagorda Bay	—
	T3	$-8.1043\left(B5/B4\right) + 7.2697$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$20.77 + 45.11(B4) - 5.52(B5)$	Matagorda Bay	—
	T4	$297.86(B3) - 2.8208$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$22.41 + 23.87(B3)$	Matagorda Bay	—
	T5	$604.54(B5) - 2.2241$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$23.80 + 11.63(B5)$	Matagorda Bay	—
	T6	$424.54(B2) - 4.4504$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$23.58 + 20.23(B2)$	Matagorda Bay	—
	T7	$504.31(B1) - 7.1769$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$24.56 + 9.53(B1)$	Matagorda Bay	—
	T8	$12.895\left(B5/B1\right) - 3.158$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$24.04 - 7.40(B1) + 13.64(B5)$	Matagorda Bay	—
	T9	$6.3388\left(B4/B3\right) - 2.4028$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$25.99 - 342.64(B3) + 347.67(B4)$	Matagorda Bay	—
	T10	$5.5354\left(B5/B2\right) - 1.0947$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$23.37 - 9.22(B2) + 9.12(B5)$	Matagorda Bay	—
	T11	$3.5623\left(B5/B3\right) + 2.8059$	CRB and TTL, Vietnam	(Quang et al., 2017)
		$22.38 - 18.28(B3) + 5.70(B5)$	Matagorda Bay	—
	T12	4.21 - 74.26(B2) – 14.84(B3) + 267.24(B4) – 126.89(B5)	Tsegn-Wen and Nan-Haw Reservoir, Taiwan	(Liu & Wang, 2019)
		$24.16 - 210.41(B2) - 250.12(B3) + 430.58(B4) - 13.66(B5)$	Matagorda Bay	—
	T13	$-590 + 1445.9\left(B4/B2\right)$	Lake Chivero, Zimbabwe	(Masocha, Dube, et al., 2018)
		$19.95 - 311.32(B2) + 265.60(B4)$	Matagorda Bay	
	T14	$-51.1 + 137.5\left(B4/B2\right)$	Lake Chivero, Zimbabwe	(Masocha, Dube, et al., 2018)
		$19.95 - 311.32(B2) + 265.60(B4)$	Matagorda Bay	—
	T15	$-1.1+5.8\left(B2/B4\right)$	Ramganga River, India	(Allam et al., 2020)
		$19.95 - 311.32(B2) + 265.60(B4)$	Matagorda Bay	

	T16	$3.896 - 4.186(^{B2}/_{B3})$	Ramganga River, India	(Allam et al., 2020)
		$21.20 - 92.09(B2) + 96.75(B3)$	Matagorda Bay	—
	T17	$-138.2 - 1718(^{B4}/_{B3}) + 695.1e^{(^{B4}/_{B3})}$	Mississippi River, Mississippi	(Pereira et al., 2017)
		$25.99 - 342.64(B3) + 347.67(B4)$	Matagorda Bay	—
	T18	$20.981(^{B3}/_{B2}) - 8.901$	Tseng-Wen reservoir, Taiwan	(Liu & Wang, 2019)
		$21.20 - 92.09(B2) + 96.75(B3)$	Matagorda Bay	—
	T19	$102.56(B3 + B4) - 5.5003$	Tseng-Wen reservoir, Taiwan	(Liu & Wang, 2019)
		$25.99 - 342.64(B3) + 347.67(B4)$	Matagorda Bay	—
	T20	$90.319(B2 + B3 + B4) - 10.775$	Tseng-Wen reservoir, Taiwan	(Liu & Wang, 2019)
		$23.79 - 214.67(B2) - 234.23(B3) + 405.99(B4)$	Matagorda Bay	—
	T21	$20.254ln(B2 + B3) + 46.009$	Tseng-Wen reservoir, Taiwan	(Liu & Wang, 2019)
		$21.20 - 92.09(B2) + 96.75(B3)$	Matagorda Bay	—
	T22	$14.735\ln(B2 + B3 + B4) + 30.802$	Tseng-Wen reservoir, Taiwan	(Liu & Wang, 2019)
		$23.79 - 214.67(B2) - 234.23(B3) + 405.99(B4)$	Matagorda Bay	—

1.5. Machine Learning (ML) Models

1.5.1. ML Model Families

In this study, the open-source H2O-AML (automated ML) platform was employed (accessible at: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html). H2O-AML provides user-friendly, fully automated supervised learning algorithms, catering to both those well-versed in the field and those without expertise (Truong et al., 2019). Within the realm of H2O-AML, various ML families are used in this study, including DNN, DRF, GBM, and GLM.

DNN is a feedforward network that uses multiple hidden layers composed of neurons to analyze complex relationships between inputs and target features (Kamilaris & Prenafeta-Boldú, 2018; Mathew et al., 2021; Oyebisi & Alomayri, 2023; Tang et al., 2020). DRF combines multiple weak decision trees to produce a strong ensemble forest (Asgari et al., 2022; Shrivastav & Kumar, 2022). The GBM family generates an ensemble model using parallel regression trees (Natekin & Knoll, 2013; Shrivastav & Kumar, 2022). The GLM algorithm produces regression models using exponential distributions (Osawa et al., 2011; Pekár & Brabec, 2018). Comprehensive descriptions regarding structures and hyperparameters for H2O-AML families can be found on the H2O-AML website (https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html).

1.5.2. ML Model Setup and Structure

For both turbidity and salinity, the input data for the ML models were comprised of seven Landsat-8 bands (Figure 3). The organization of the input data aimed to facilitate a more straightforward comparison between ML and empirical models. Due to limited data availability, ML models were not created for chlorophyll-a.

To guarantee equal consideration of all input variables, the following equation was employed to normalize model inputs within the range of 0 to 1 (Figure 3):

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

where x_{norm} represents the normalized value for a specific input x, while x_{max} and x_{min} represent the maximum and minimum recorded values of x, respectively.

For each model family (DNN, DRF, GBM, GLM), the input and target data were divided into training (64%), validation (16%), and testing (20%) sets (Figure 3). The process of dividing the input/target data followed a random format. This method of generating random data ensures unique input values for each model run.

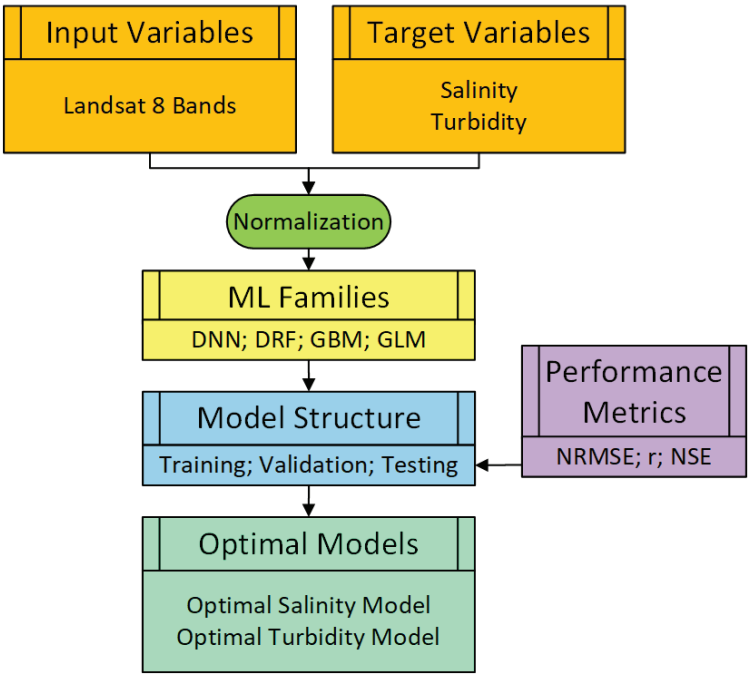

Figure 3: Input and target variables and structures of the ML families used to retrieve salinity and turbidity over Matagorda Bay.

To prevent overfitting in each model, early stopping criteria were enforced, employing the mean-squared error as the stopping metric. This involved setting a stopping round value of 5 and a stopping tolerance of 0.0001. This option specifies the tolerance value by which a ML model must improve before training cease. In this case, the moving average for last 6 simulation rounds is calculated, where the first

moving average is reference value for other 5 moving averages to compare. The model will stop training if the ratio between the best moving average and reference moving average is more or equal 0.0001. These stopping options are used to increase model performance by restricting the number of models that get built. To improve the simulation accuracy, we ran each simulation for 50 runs. The model with the highest testing performance, based on the Nash-Sutcliffe Efficiency (NSE) coefficient, was selected as an optimal model. An optimal model was selected for salinity and for turbidity from each ML model family (e.g., DNN, DRF, GBM, GLM) (Figure 3).

1.5.3. Model Performance Measures

Performance for both empirical and ML models was assessed using the NSE coefficient, correlation coefficient (*r*), and normalized root mean square error (NRMSE). These measures were used to evaluate the performance of similar models worldwide (Elangovan & Murali, 2020; González-Márquez et al., 2018; Liu & Wang, 2019; Masocha, Mungenge, et al., 2018). The standard variation in the 50 ML model simulations was used to quantify errors in each performance metric.

The NSE measures the relative magnitude of residual variance to the variance of observed data. NSE values range from -∞ to 1, with optimal performance at 1:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X^{pred.} - X^{obs.})^2}{(X^{obs.} - \widehat{X^{Obs.}})^2}. \quad (2)$$

The *r* value measures the strength of linear relationship between predicted and observed data with value ranges between -1 and 1. An *r* value of 0 means there is no correlation, and positive (negative) values mean positive (negative) correlation, with 1 (-1) indicating perfect positive (negative) correlations between predicted and observed values:

$$r = \frac{\sum_{i=1}^{n}\left(X^{pred.} - \widehat{X^{pred.}}\right)(X^{obs.} - \widehat{X^{obs.}})}{\sqrt{\sum_{i=1}^{n}\left(X^{pred.} - \widehat{X^{pred.}}\right)^2 \Sigma\left(X^{obs.} - \widehat{X^{obs.}}\right)^2}}. \quad (3)$$

The NRMSE is the RMSE normalized by standard deviation of the observed data with value ranges from 0 to ∞, given by this equation:

$$NRMSE = \frac{1}{\sigma}\sqrt{\frac{\sum_{i=1}^{n}(X^{pred.} - X^{obs.})^2}{n}}, \quad (4)$$

where $X^{pred.}$ and $X^{obs.}$ represent predicted value and observed (e.g., in-situ) value, respectively, and $\widehat{X^{pred.}}$ and $\widehat{X^{obs.}}$ are mean values; *n* represents the input data size; and σ is the standard deviation of observed data.

1.6. Results

1.6.1. Empirical Models

Figure 4 and Table 4 present performance metrics of the 36 uncalibrated and calibrated empirical models implemented in this study. Performance varied in the 12 uncalibrated empirical chlorophyll-a models (Figure 4a; Table 4), with NRMSE ranging from 0.32 to 1475.10, *r* from -0.88 to 0.96, and NSE from -23527276.57 to -0.12. Among these uncalibrated models, model C8 had the highest performance, based on NSE values, (NRMSE: 0.32, *r*: 0.65, NSE: -0.12); model C7 displayed the lowest performance (NRMSE: 1475.10, *r*: 0.80, NSE: -23527276.57) (Figure 4a; Table 4). Statistical assessments of the uncalibrated chlorophyll-a models indicate poor performance (NRMSE: 126.91 ± 424.70, *r*: 0.42 ± 0.67, NSE: -1961890 ± 6791336). After calibration over Matagorda Bay, the chlorophyll-a models demonstrated enhanced performance (NRMSE: 0.91 ± 0.03, *r*: 0.94 ± 0.04, NSE: 0.89 ± 0.06). These calibrated models also demonstrated comparable performance when compared to one another (Figure 4a; Table 4). The NRMSE for the calibrated models ranged from 0.85 to 0.96, with corresponding *r* values from 0.88 to 0.99, and NSE values from 0.88 to 0.98. Notably, calibrated model C2 exhibited the highest performance (optimal model) with NRMSE, *r*, and NSE values of 0.96, 0.99, and 0.98, respectively. The scatterplot of the calibrated model C2 demonstrates an overall acceptable performance (Figure 5a). Calibrated model C9 displayed the lowest performance, with NRMSE, *r*, and NSE values of 0.85, 0.88, and 0.77, respectively.

The performance of the uncalibrated empirical salinity models varied across the matrices, with NRMSE ranging from 180.47 to 180.48, *r* from -0.16 to 0.16, and NSE from -32638.62 to -32635.55 (Figure 4b; Table 4). Notably, neither of the uncalibrated empirical salinity models produced statistically significant performance (NRMSE: 180.47 ± 0.01, *r*: 0.16 ± 0.22, NSE: -32635 ± 2.18). The calibrated empirical salinity model exhibited a slight improvement with NRMSE, *r*, and NSE values of 0.24, 0.24, and 0.06, respectively (Figure 4b; Table 4). However, the scatterplot for the calibrated salinity model indicates poor performance (Figure 5b). It is important to note that since the same bands (B1–B4) were used to generate the uncalibrated salinity models, only one calibrated model was generated for salinity.

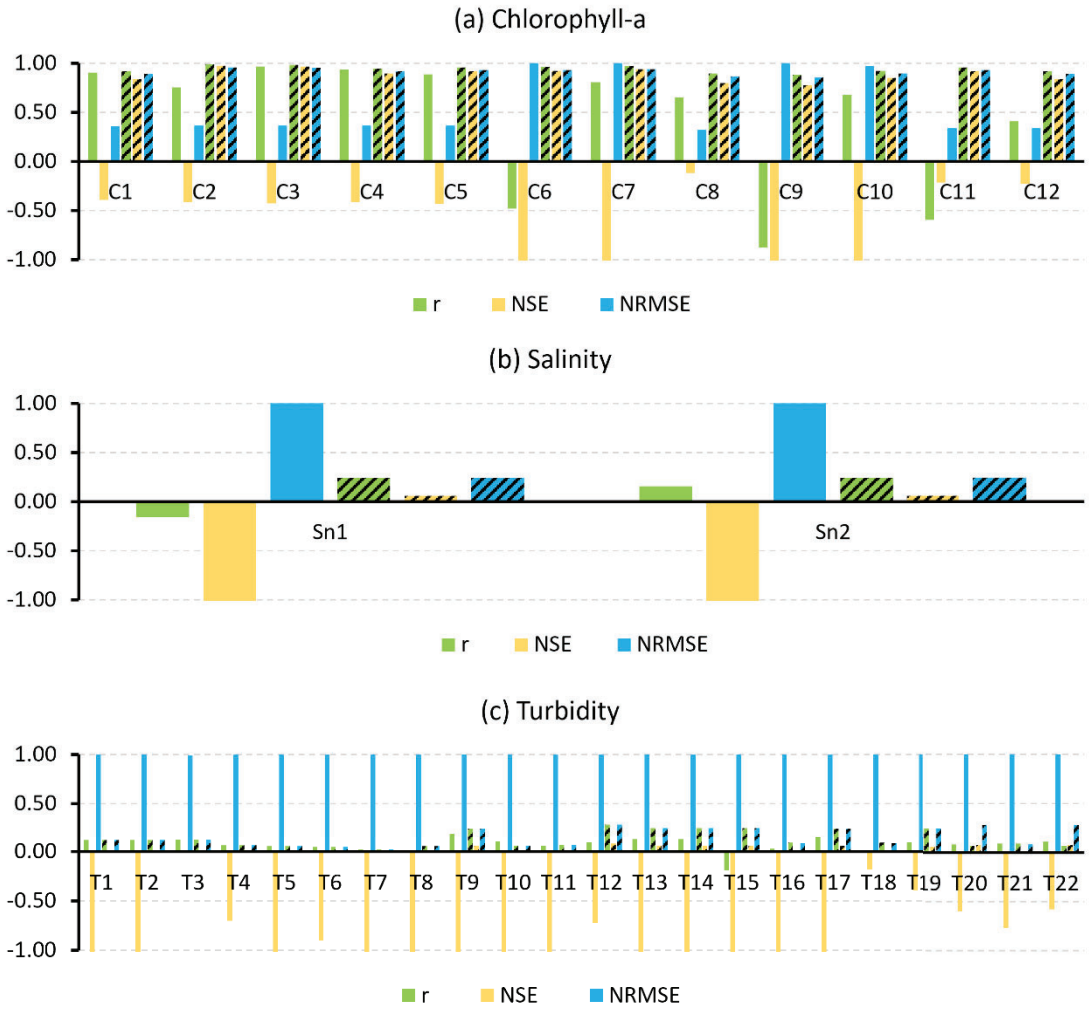

Figure 4: Performance statistics for (a) chlorophyll-a, (b) salinity, and (c) turbidity uncalibrated (solid colors) and calibrated (dashed lines) empirical models. Note the y-axis is fitted to ±1 for display purposes; actual higher and lower values are listed in Table 4.

Table 4: Performance metrics for 36 uncalibrated and calibrated empirical models (equations showed in Table 3) applied over Matagorda Bay.

WQI	Model ID	Uncalibrated Models			Calibrated Models		
		NRMSE	*r*	NSE	NRMSE	*r*	NSE
Chlorophyll-a	C1	0.36	0.90	-0.39	0.89	0.92	0.84
	C2	0.36	0.75	-0.41	0.96	0.99	0.98
	C3	0.36	0.96	-0.43	0.95	0.98	0.96
	C4	0.36	0.93	-0.42	0.92	0.94	0.89
	C5	0.36	0.89	-0.43	0.93	0.96	0.92

	C6	7.00	-0.48	-528.50	0.93	0.96	0.92
	C7	1475.10	0.80	-23527276.57	0.94	0.97	0.94
	C8	0.32	0.65	-0.12	0.87	0.89	0.80
	C9	37.08	-0.88	-14866.34	0.85	0.88	0.77
	C10	0.97	0.68	-9.18	0.89	0.92	0.85
	C11	0.34	-0.60	-0.22	0.93	0.96	0.92
	C12	0.34	0.41	-0.23	0.89	0.92	0.84
Salinity	Sn1	180.48	-0.16	-32638.62	0.24	0.24	0.06
	Sn2	180.47	0.16	-32635.55	0.24	0.24	0.06
Turbidity	T1	5.18	0.12	-26.00	0.12	0.12	0.01
	T2	1.65	0.12	-1.75	0.12	0.12	0.01
	T3	0.99	0.12	0.01	0.12	0.12	0.01
	T4	1.30	0.07	-0.70	0.07	0.07	0.00
	T5	3.85	0.06	-13.90	0.06	0.06	0.00
	T6	1.37	0.05	-0.90	0.05	0.05	0.00
	T7	1.50	0.02	-1.25	0.02	0.02	0.00
	T8	4.12	0.01	-16.05	0.06	0.06	0.00
	T9	1.53	0.19	-1.34	0.25	0.25	0.06
	T10	1.41	0.10	-1.01	0.06	0.06	0.00
	T11	1.41	0.06	-1.01	0.07	0.07	0.00
	T12	1.31	0.09	-0.72	0.29	0.29	0.08
	T13	85.15	0.13	-7292.28	0.25	0.25	0.06
	T14	7.22	0.13	-51.39	0.25	0.25	0.06
	T15	1.56	-0.19	-1.45	0.25	0.25	0.06
	T16	1.63	0.03	-1.68	0.09	0.09	0.01
	T17	1.74	0.16	-2.06	0.25	0.25	0.06
	T18	1.09	-0.01	-0.18	0.09	0.09	0.01
	T19	1.18	0.09	-0.39	0.25	0.25	0.06
	T20	1.26	0.08	-0.60	0.28	0.07	0.08
	T21	1.32	0.09	-0.77	0.09	0.09	0.01
	T22	1.25	0.11	-0.58	0.28	0.07	0.08

The performance of uncalibrated turbidity models varied with NRMSE values ranging from 0.99 to 85.15, *r* from -0.19 to 0.19, and NSE from -7292.28 to 0.01 (Figure 4c; Table 4). Among the 22 models, uncalibrated model T3 demonstrated the highest performance (NRMSE: 0.99, *r*: 0.12, NSE: 0.01), while uncalibrated model T13 had the least favorable performance (NRMSE: 85.15, *r*: -0.13, NSE: -7292.28) (Figure 4c; Table 4). None of the uncalibrated empirical turbidity models yield significant performance metrics (NRMSE: 5.86 ± 17.78, *r*: 0.07 ± 0.08, NSE: -337.09 ± 1553.51). The calibrated empirical turbidity models displayed a relatively improved performance (NRMSE: 0.15 ± 0.10, *r*: 0.13 ± 0.09, NSE: 0.03 ± 0.03). The calibrated models showed NRMSE values ranging from 0.02 to 0.29, *r* values from 0.00 to 0.29, and NSE values from 0.02 to 0.08 (Figure 4c; Table 4). Among the models, calibrated model T12 showed the highest performance with NRMSE, *r*, and NSE values of 0.29, 0.29, and 0.08, respectively. However, the scatterplot depicting the calibrated turbidity model T12 shows inadequate performance (Figure 5c). Calibrated model T7 exhibited the lowest performance, with NRMSE, *r*, and NSE values of 0.02, 0.02, and 0.00, respectively.

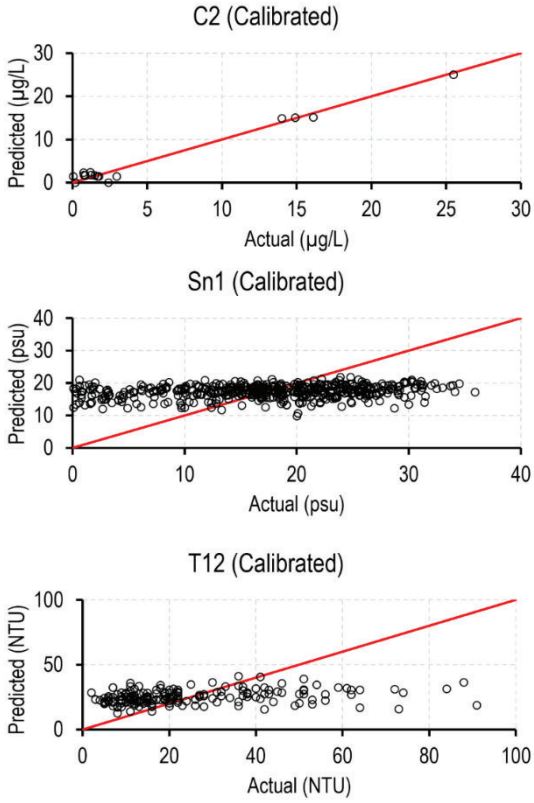

Figure 5: Observed and modeled (a) chlorophyll-a, (b) salinity, and (c) turbidity values generated form the calibrated empirical models over Matagorda Bay. Red lines indicate a 1:1 relationship.

1.6.2. ML Models

Figure 6 and Table 5 present the statistical measures for the optimal salinity models (selected based on testing NSE value) generated from various ML families during training and testing phases. Among the diverse ML families, a trend of closely competitive performance emerges. However, DNN stands out as the most effective family for salinity simulations. The optimal DNN achieved a training performance of NRMSE: 0.90 ± 0.08, *r*: 0.45 ± 0.10, NSE: 0.19 ± 0.13, and a testing performance of NRMSE: 0.87 ± 0.06, *r*: 0.49 ± 0.09, NSE: 0.23 ± 0.12. The most significant input variables for this model were B3, B4, B6, B7, and B1. The relative importance for these inputs were 22%, 15%, 14%, 13%, and 13%, respectively.

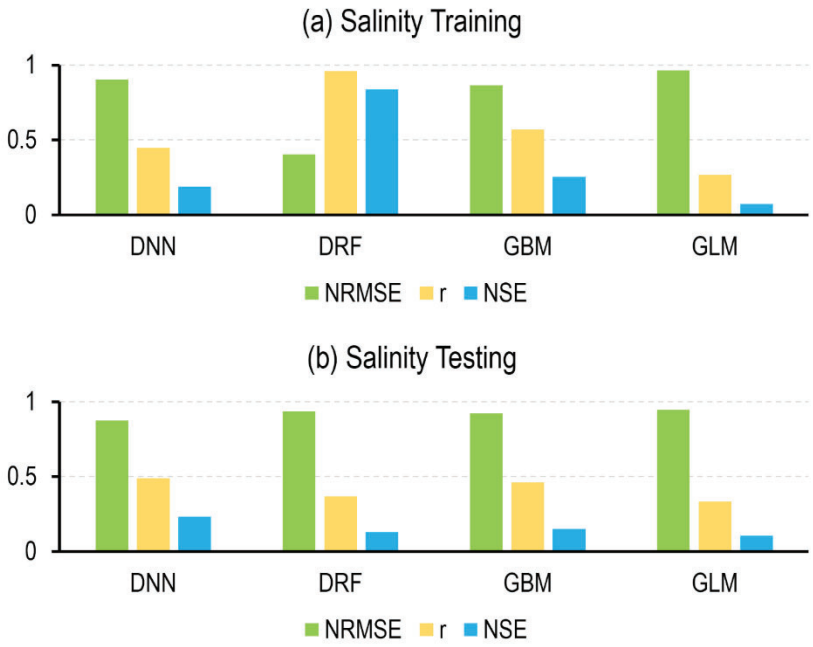

Figure 6: Performance metrics for optimal salinity models derived from different ML model families (e.g., DNN, DRF, GBM, GLM). Metrics are displayed for the (a) training and (b) testing phases.

Table 5: Performance metrics for the optimal ML model generated for salinity and turbidity during training and testing phases.

WQI	ML Family	Training			Testing		
		NRMSE	***r***	**NSE**	**NRMSE**	***r***	**NSE**
Salinity	DNN	0.90 ± 0.08	0.45 ± 0.10	0.19 ± 0.13	0.87 ± 0.06	0.49 ± 0.09	0.23 ± 0.12
	DRF	0.40 ± 0.01	0.96 ± 0.00	0.84 ± 0.01	0.93 ± 0.04	0.37 ± 0.08	0.13 ± 0.08
	GBM	0.86 ± 0.04	0.57 ± 0.07	0.25 ± 0.07	0.92 ± 0.03	0.46 ± 0.10	0.15 ± 0.07
	GLM	0.96 ± 0.01	0.27 ± 0.02	0.07 ± 0.01	0.95 ± 0.03	0.33 ± 0.08	0.10 ± 0.05
Turbidity	DNN	0.59 ± 0.13	0.81 ± 0.06	0.65 ± 0.11	0.63 ± 0.11	0.79 ± 0.11	0.60 ± 0.20
	DRF	0.36 ± 0.02	0.95 ± 0.01	0.87 ± 0.01	0.76 ± 0.10	0.65 ± 0.11	0.42 ± 0.18
	GBM	0.66 ± 0.17	0.79 ± 0.07	0.56 ± 0.14	0.73 ± 0.10	0.73 ± 0.13	0.47 ± 0.20
	GLM	0.93 ± 0.03	0.36 ± 0.08	0.13 ± 0.05	0.86 ± 0.07	0.63 ± 0.16	0.25 ± 0.14

Following DNN closely in performance, GBM ranks as the second-best performing family. The optimal GBM model, showcased training performance metrics of NRMSE: 0.86 ± 0.04, *r*: 0.57 ± 0.07, NSE: 0.25 ± 0.07, and testing performance metrics of NRMSE: 0.92 ± 0.03, *r*: 0.46 ± 0.10, NSE: 0.15 ± 0.07. DRF presents a slightly weaker performance, being the second-least effective family. The optimal DRF model exhibited a training performance of NRMSE: 0.40 ± 0.01, *r*: 0.96 ± 0.00, NSE: 0.84 ± 0.01, and a testing performance of NRMSE: 0.93 ± 0.04, *r*: 0.37 ± 0.08, NSE: 0.13 ± 0.08. Finally, GLM ranks as the least proficient model. The optimal GLM model resulted in a training performance of NRMSE: 0.96 ± 0.01, *r*: 0.27 ± 0.02, NSE: 0.07 ± 0.01, and a testing performance of NRMSE: 0.95 ± 0.03, *r*: 0.33 ± 0.08, NSE: 0.10 ± 0.05. Overall, these results underscore the superiority of DNN, GBM, and DRF for salinity simulation, while highlighting the limitations of GLM (Figure 6; Table 5). Figure 7 presents scatterplots illustrating observed and modeled salinity values generated for the optimal model in each ML family during both training and testing phases. As depicted in Figure 7, the DRF model is performing well during training compared to other models.

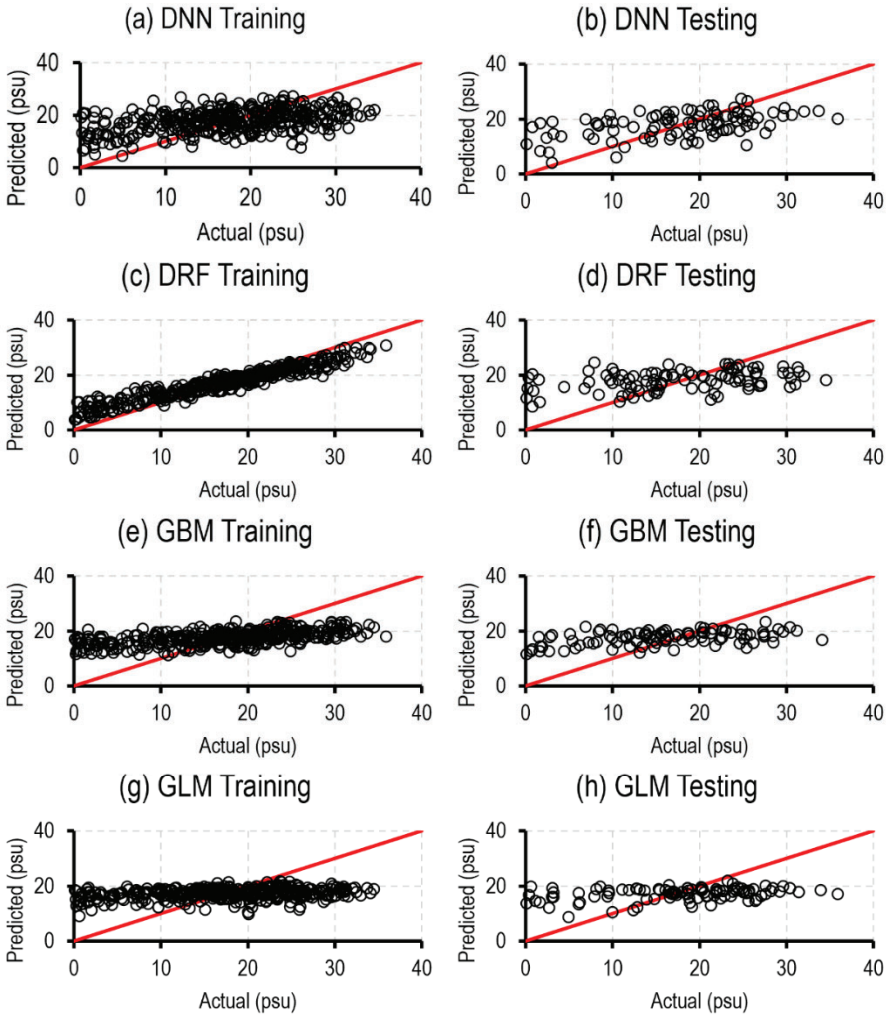

Figure 7: Observed and modeled salinity values generated for the optimal model in each ML family during the training and testing phases. Red lines indicate a 1:1 relationship.

Figure 8 and Table 5 present the optimal turbidity models from the various ML families. All DNN, GBM, and DRF turbidity models demonstrated closely competitive performances. The optimal DNN model achieved a training performance of NRMSE: 0.59 ± 0.13, *r*: 0.81 ± 0.06, NSE: 0.65 ± 0.11, and a testing performance of NRMSE: 0.63 ± 0.11, *r*: 0.79 ± 0.11, NSE: 0.60 ± 0.20 (Figure 8; Table 5). The most significant input variables for this model were B4, B5, B3, B7, and B1. The relative importance for these inputs were 20%, 16%, 14%, 14%, and 14%, respectively. The GBM family provided the second-best performance. The optimal GBM model has a training performance of NRMSE: 0.66 ± 0.17, *r*: 0.79 ± 0.07,

NSE: 0.56 ± 0.14, and a testing performance of NRMSE: 0.73 ± 0.10, *r*: 0.73 ± 0.13, NSE: 0.47 ± 0.20. The performance of the DRF was slightly weaker than the DNN and GBM families. The optimal DRF model produced training performances of NRMSE: 0.36 ± 0.02, *r*: 0.95 ± 0.01, NSE: 0.87 ± 0.01, and testing performances of NRMSE: 0.76 ± 0.10, *r*: 0.65 ± 0.11, NSE: 0.42 ± 0.18. Once more, the GLM family produced models with the lowest performance. The optimal GLM model produced training performance metrics of NRMSE: 0.93 ± 0.03, *r*: 0.36 ± 0.08, NSE: 0.13 ± 0.05, and testing performance of NRMSE: 0.86 ± 0.07, *r*: 0.63 ± 0.16, NSE: 0.25 ± 0.14 (Figure 8; Table 5). Figure 9 depicts scatterplots that illustrate the relationship between observed and modeled turbidity values produced by the optimal model from each ML family during both training and testing phases. All model families tend to overpredict turbidity values during the training phase. The performance of DNN, GBM, and DRF models is close for both training and testing.

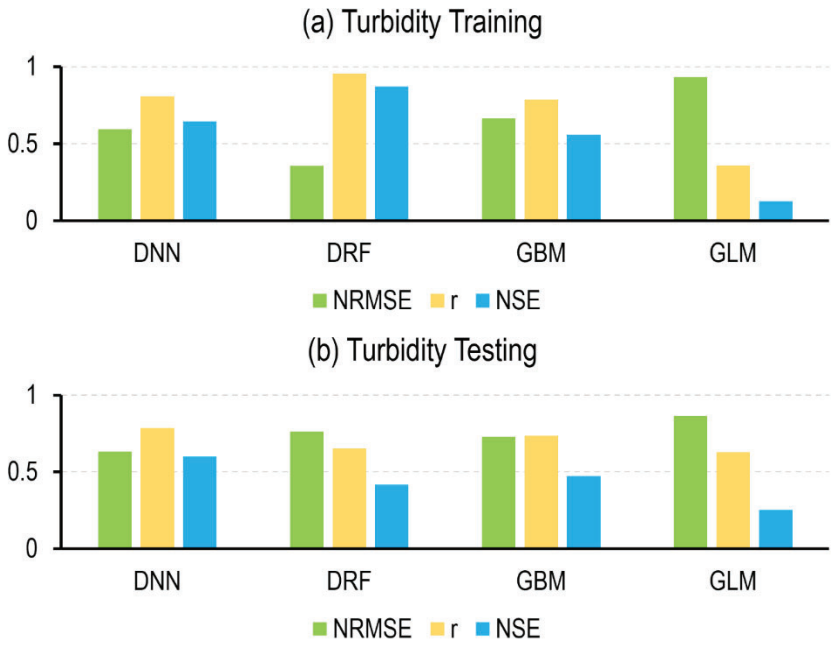

Figure 8: Performance metrics for different turbidity ML model families (e.g., DNN, DRF, GBM, GLM). Metrics are displayed for the (a) training and (b) testing phases.

1.6.3. Applications of Optimal Models

The optimal model for chlorophyll-a was determined to be the calibrated empirical C2 model. However, for salinity and turbidity, the optimal models were identified as DNN models. These models were then used to retrieve WQIs across all of Matagorda Bay. Figure 10 provides examples of the retrieved products for August and November of 2018.

Spatial and temporal variations in WQIs are outside the scope of this study. However, turbidity reflects the clarity and suspended particle levels in the water (Davies-Colley & Smith, 2001), chlorophyll-a is a key indicator of phytoplankton and algae levels (Boyer et al., 2009; Kasprzak et al., 2008; Rakocevic-Nedovic & Hollert, 2005), and salinity is crucial for understanding the bay's conditions and the organisms it can support (Aguilar, 2017; Marshall et al., 2019; Onabule et al., 2020).

The high spatial resolution of WQI data, such as those produced in this study, provides opportunities to analyze and monitor trends in the health of aquatic systems, significantly contributing to previous conservation and pollution assessment efforts (El-Zeiny & El-Kafrawy, 2017; Snyder et al., 2017). Specifically, these data could be used to (1) expand in-situ observational footprints and improve current efforts to conserve, enhance, and restore important aquatic habitats; (2) define areas of "anomalous" WQIs that could need further future investigation; (3) enable a demonstration and use of satellite observations to improve understanding of factors controlling spatial and temporal variability of WQ across complex aquatic systems; and (4) help realize an opportunity and process to enable standard WQ monitoring protocols to be expanded and complemented by free satellite products.

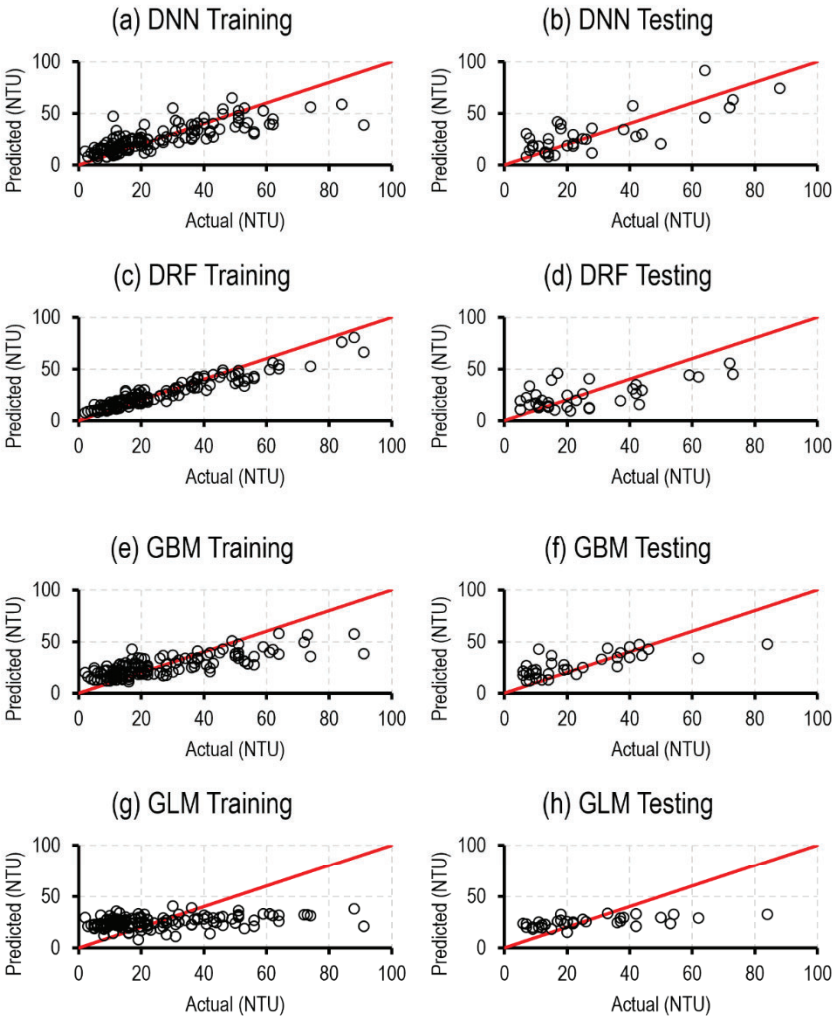

Figure 9: Observed and modeled turbidity values generated for the optimal model in each ML family during the training and testing phases. Red lines indicate a 1:1 relationship.

1.7. Discussion

Performance of the implemented empirical and ML models are influenced by the quality of input datasets. Landsat-8 surface reflectance data showed reasonable quality over Matagorda Bay. Relevant studies indicate a significant correlation (r: 0.70) between Landsat-8-derived and in-situ-derived surface reflectance data (Schild et al., 2017). In addition, we removed pixels over cloud-contaminated regions and these with negative surface reflectance. Atmospheric conditions can significantly affect the quality of Landsat-8 imagery. Aerosols, clouds, and water vapor in the atmosphere can scatter and absorb light, leading to errors in the measured surface reflectance values (Kaufman et al., 1997; Vermote et al., 2016). However, Landsat-8 level 2 data used in this study is atmospherically corrected (Danbara, 2014; Lim & Choi, 2015). Landsat 8, like any satellite, has a calibration process to ensure the accuracy of its radiometric measurements (Danbara, 2014; Lim & Choi, 2015). The angle of the sun and the presence of sun glint (sunlight reflecting directly off the water's surface) can impact the reflectance measurements (Misra et al., 2018; Pahlevan et al., 2017; Wei et al., 2018). To reduce these effects, we removed these pixels from our modelling exercise.

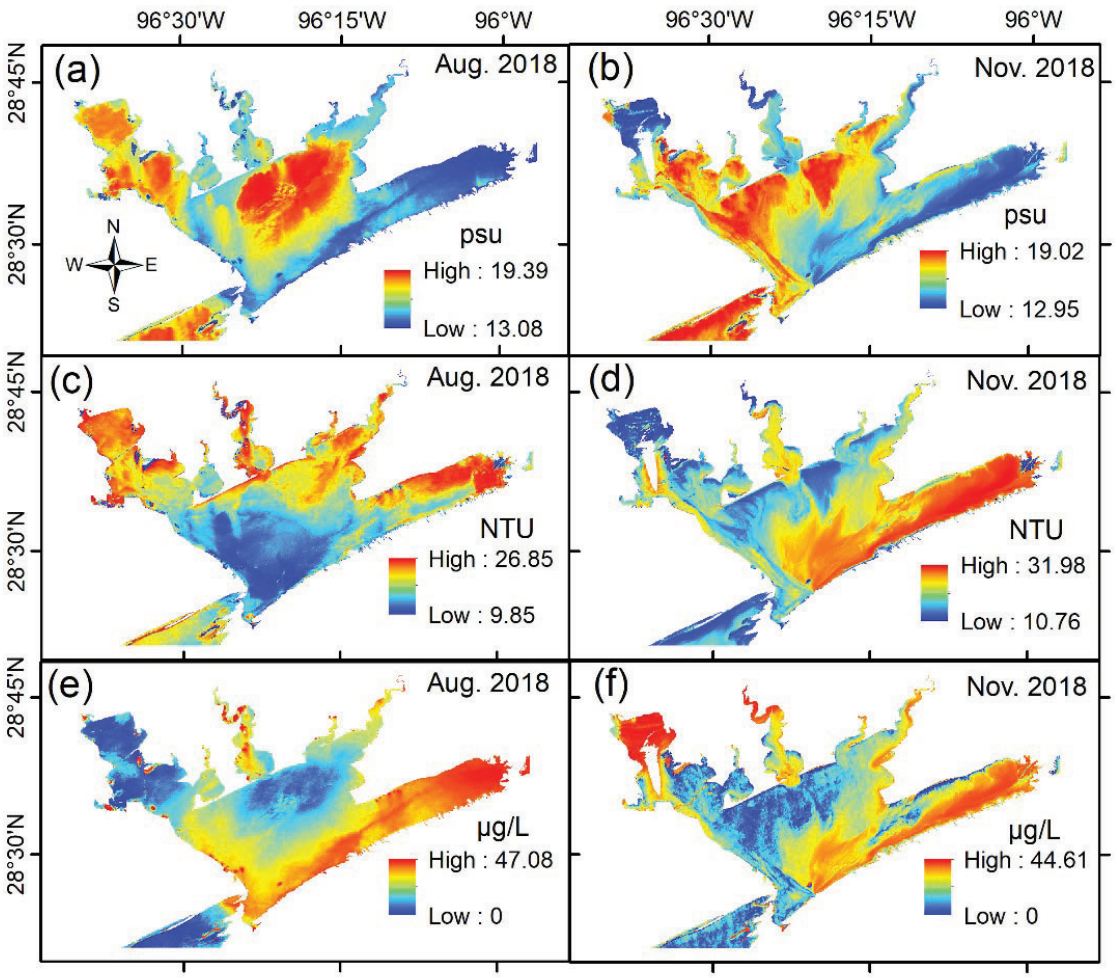

Figure 10: Modeled WQI data produced using optimal salinity, turbidity, and chlorophyll-a models over Landsat-8 images acquired on (a, c, e) August 22, 2018, and (b, d, f) November 26, 2018.

Matagorda Bay is influenced by four major rivers: the Colorado River, the Lavaca River, the Carancahua River, and the Tres Palacios River. Suspended solids and dissolved organic matter from these rivers can directly affect water turbidity, salinity, and reflectance in the area (F. Wang & Xu, 2008). Colored dissolve organic matter, a yellow substance, contains inherently optical properties that affect surface reflectance (Binding & Bowers, 2003; Bowers & Brett, 2008). In addition, previous studies have reported significant negative correlations between salinity and total suspended solids (Fang et al., 2010). In this study, correlations of 0.05 ± 0.04 was estimated between salinity and turbidity. Turbidity, dissolved organic matters, and suspended solids directly influence water color and spectral reflectance. This indirect relationship between salinity and spectral bands motivated the use of Landsat 8 data to estimate salinity, a non-optical WQI. Similar principles have been applied in several studies utilizing Landsat-5 (Khorram,

1985; Lavery et al., 1993; Nazeer & Bilal, 2018; Vuille & Baumgartner, 1993; Xie et al., 2013; Zhang et al., 2012). MODIS (Urquhart et al., 2012) , and Landsat-8 (Ansari & Akhoondzadeh, 2020; Nguyen et al., 2018; Zhao et al., 2017) data for salinity retrieval. However, the relatively lower correlation between Landsat-8-derived spectral reflectance and salinity might explain the significantly lower performance of empirical models and the relatively lower performance of the ML over Matagorda Bay.

1.7.1. Empirical Models

Most uncalibrated empirical models produced poor performance measures. The inadequacies of these models are more likely due to their application in water bodies, for which they were not originally designed. These models were initially developed across geographic regions encompassing lakes, rivers, and estuaries, each with distinct surface areas, hydrological conditions, and surface reflectance signatures. These discrepancies between the model's training conditions and the environmental characteristics of Matagorda Bay contribute to their poor performance.

The empirical models, however, demonstrated improved performance when calibrated over Matagorda Bay. The calibrated chlorophyll-a models showed the most significant improvement. The average ± standard deviation of the NRMSE, *r*, and NSE measures increased from 126.91 ± 424.70, 0.42 ± 0.67, and -1961890.27 ± 6791336.75 for uncalibrated models to 0.91 ± 0.03, 0.94 ± 0.04, and 0.89 ± 0.06 for calibrated models, respectively. The highest performance of the calibrated models was reported for Model C2, which used B1, B3, B5, B6, B7. These bands were all highly correlated ($r > 0.80$) to in-situ chlorophyll-a values, which likely made it easier for this model to generalize a linear relationship between surface reflectance and chlorophyll-a measurements. The metrics of the calibrated C2 model outperform metrics published in previous studies (0.74 ± 0.12) (Elangovan & Murali, 2020).

It is notable that the number of data points utilized in creating the empirical models for chlorophyll-a (total: 17) may raise concerns. However, these measurements were the only ones available that coincided with Landsat-8 overpasses over Matagorda Bay. They represent the sole data points temporally synchronized with the dates of Landsat-8 acquisition throughout the 10-year study period (2014–2023). Some of these in-situ chlorophyll-a data were gathered in the field when weather conditions allowed, while others were obtained from historical records (e.g., TCEQ, TWPD, LCRA). Results of our chlorophyll-a models should be used with caution. We acknowledge that a larger sample size would undoubtedly enhance the robustness of the empirical models, and we are actively pursuing efforts to acquire additional data to expand the sample size in future studies. Despite the limited sample size, our primary objective was to evaluate the feasibility of empirical models in estimating chlorophyll-a using the available data. Empirical salinity models also experienced improvements after calibration over Matagorda Bay. The RMSE, *r*, and

NSE measures increased from 180.47 ± 0.01, 0.00 ± 0.22, and -32637.09 ± 2.18 for uncalibrated models to 0.24, 0.24, and 0.06 for the calibrated model, respectively. The calibrated salinity models, Sn1 and Sn2, used B1, B2, B3, and B4. These bands indicated a weakly inverse relationship (*r*: -0.16 to -0.09) with in-situ salinity measurements. This weak relationship likely hinders the empirical models in establishing meaningful linear relationships between in-situ salinity observations and surface reflectance. In addition, the performance of calibrated salinity models was significantly lower than the performance reported for the original study. In its original applications, the *r* values for Sn1 and Sn2 were 0.84 (Zhao & Temimi, 2016) and 0.77 (Zhao & Temimi, 2016), respectively. However, the original studies only used 33 data points, which might have led to overfitting and produced misleading performance measures.

The calibrated turbidity models demonstrated some improvement in performance. The NRMSE, *r*, and NSE measures increased from 5.86 ± 17.78, 0.07 ± 0.08, and -337.09 ± 1553.51 for uncalibrated models to 0.15 ± 0.10, 0.13 ± 0.09, and 0.03 ± 0.03 for calibrated models, respectively. The highest performing calibrated empirical turbidity model is T12, which used B2, B3, B4, and B5. These bands had a positive, but low, correlation to in-situ turbidity (*r*: 0.05 - 0.12). Model T12 had higher performances reported in its original studies (*r*: 0.85 ± 0.18) compared to that calibrated to Matagorda Bay (Liu & Wang, 2019). This might be due to the complex turbidity patterns over Matagorda Bay.

1.7.2. ML Models

ML algorithms performed better for both salinity and turbidity compared to empirical models. ML algorithms are much more complex, and equipped to ascertain intricate relationships between inputs (surface reflectance) and target (salinity and turbidity) datasets (Zhu et al., 2022). In addition, ML algorithms are better able to map nonlinear relationships, such as those between salinity and turbidity and surface reflectance (correlation between salinity and surface reflectance: -0.10 ± 0.04; correlation between turbidity and surface reflectance 0.04 ± 0.04).

The DNN, GBM, and DRF families examined in this study demonstrated acceptable performance when used to retrieve turbidity and salinity over Matagorda Bay. Notably, DNN performed better than the other families when simulating both salinity and turbidity. DNN is widely recognized for its efficacy in modeling high-dimensional and complex datasets, particularly those characterized by intricate nonlinear associations between input and target variables (Guo et al., 2022; Wei et al., 2022). DNNs excel in extracting hierarchical features from input and target data, endowing them with stronger modeling capabilities relative to alternative ML families (Guo et al., 2022; Mathew et al., 2021; Peterson et al., 2020). The GBM family consistently secured the second-highest model performance when simulating salinity and turbidity. However, no large differences were found in the performance of both DRF and GBM. Remarkably, the GLM family consistently produced the poorest-performing models for both salinity and

turbidity simulations. GLM models adopt a straightforward linear regression approach and lack the intricacy required to dissect the multifaceted relationships inherent in the multidimensional and nonlinear datasets used in this study. Consequently, the limitations of GLM models prevent them from effectively capturing the complexities in the data.

Most of the ML models exhibited a slight decrease in performance during the testing phase compared to the training phase. This relatively modest reduction in testing performance can be attributed to the limited number of training samples (Fallatah et al., 2022; Sun et al., 2021). For instance, the salinity models were trained with only 382 datapoints, while the turbidity models had a training set of 138 datapoints. Another potential explanation for the variation in performance between the training and testing sets may stem from differences in the input/target patterns within the model inputs in the testing set compared to those in the training set. Salinity models were tested with 96 datapoints, while turbidity was tested with 35. Attempts to model chlorophyll-a using ML models were unsuccessful, likely due to a lack of in-situ data. Chlorophyll-a measurements have only 17 datapoints. It is likely that this dataset did not provide large enough target time series to allow the ML algorithms to adequately analyze the complex relationship between surface reflectance and the provided chlorophyll-a observations.

1.8. Conclusion

The integration of public-domain remote-sensing data-modeling techniques has significantly enhanced WQI monitoring capabilities. Landsat-8 offers a unique opportunity to observe spatiotemporal trends of WQ within expansive aquatic ecosystems. With advances in modeling techniques, numerous inquiries have arisen regarding the most effective approaches to model WQIs. This study examines the capabilities of both empirical and ML models to determine optimal methods for retrieving salinity, turbidity, and chlorophyll-a from Landsat-8 data over Matagorda Bay, Texas.

The uncalibrated empirical models exhibited poor performance when applied to Matagorda Bay. The unique environmental characteristics of Matagorda Bay contributed to the failure of these uncalibrated models to produce WQIs. However, when calibrated over Matagorda Bay, model performance improved when used to retrieve chlorophyll-a data. ML models, on the other hand, yielded meaningful results for both salinity and turbidity. Among the implemented ML families, DNN demonstrated the highest performance. This model was able to successfully map complex nonlinear relationships between Landsat-8 reflectance datasets and salinity and turbidity.

Our methodology offers a point of reference, a structured framework, and valuable insights for employing empirical and ML models on Landsat-8 data to characterize WQIs. This approach encourages broader, enhanced use of remote-sensing datasets by the scientific community, end-users, and decision-makers. Given the higher performance of calibrated C2 empirical models for chlorophyll-a and DNN

models for both salinity and turbidity, the modeled WQIs could be used to expand the footprint of in-situ observations and improve current efforts to conserve, enhance, and restore important habitats in aquatic ecosystems. The developed approach serves as a guide to enhance monitoring procedures for aquatic ecosystems.

Section (II)

Oyster Habitat Suitability Indices for Matagorda Bay, Texas: A Comparative Analysis

2.1. Abstract

Oysters in Matagorda Bay have experienced population declines over the past century. Oyster Habitat Suitability Index (OHSI) models evaluate habitat quality and may greatly aid restoration projects. Environmental and water quality indicators are used to generate these OHSIs. This study used remote-sensing-derived water quality datasets from 2018 to 2023 to produce five OHSI models for Matagorda Bay on both monthly and annual scales. Each of these models used five physical parameters (model inputs): salinity, turbidity, water temperature, depth, and water velocity. OHSI1 is a classification-based model, classifying habitat quality in each model input based on specific thresholds that define unsuitable, moderate, and optimal habitat conditions. OHSI2 is a regression-based model that regressed these inputs against total live oyster count to depict habitat quality. Both unweighted and weighted versions of OHSI1 and OHSI2 were implemented. OHSI3 is a machine learning (ML)–based model that uses total live oyster count as a target. OHSI scores from all models generally demonstrated spatial and temporal variability in habitat quality. Across all models, optimal habitat was generally found in shallower regions, near sources of freshwater, and in areas with high water flow. Unsuitable habitat was generally identified in the primary bay, where water depths are greater and water flow is slow. OHSI1 indicated that spring and fall were optimal for oysters, while winter was unsuitable. OHSI2 identified winter months as optimal and summer as unsuitable. Conversely, OHSI3 found the most suitable conditions in summer and determined that spring and late fall were unsuitable. This study evaluated performance variations among the various OHSI modeling approaches and is intended to serve as a framework for future OHSI modeling efforts. The OHSI models generated from this study provide valuable insights into long-term habitat restoration activities over Matagorda Bay.

2.2. Introduction

Texas is the second largest supplier of the Eastern oyster in the nation. However, Matagorda Bay supplies less than 5% of the Texan harvest (Haby, 2012). Historical oyster reefs in Matagorda Bay were once economically and ecologically significant, but these reefs have experienced drastic population declines over the past century (Fisheries & Moore, 1907). Declining populations in the Gulf of Mexico are largely due to exploitive harvesting and dredging practices as well as poor water quality (Aguilar, 2017; Pollack et al., 2012).

The removal of oysters from local ecosystems leads to many adverse effects because oysters play a fundamental role in maintaining and improving marine ecosystems (Grabowski & Peterson, 2011). When oysters are removed from estuaries, water quality worsens, and habitat many marine species is significantly affected (Lenihan, 1999; Plutchak et al., 2010; Sharma et al., 2016). Oysters are filter feeders that improve water quality by removing nutrients and sediments which can be harmful to marine organisms and vegetation (Aguilar, 2017; Culbertson, 2008; TinHan et al., 2018). Without removing excess nutrients by oysters, there is an increased risk of algal blooms and anoxic conditions (Culbertson, 2008; De Santiago et al., 2019). Oysters also serve as habitat for many fishery species (Peyre et al., 2014; TinHan et al., 2018), providing interstitial living space and foraging grounds for predators (Stanley et al., 2024). The removal of oysters decreases the abundance, diversity, and biomass of the marine species that use these habitats (Stanley et al., 2024).

Oyster restoration activities are increasingly critical in replenishing oyster populations and improving environmental quality (De Santiago et al., 2019). In Matagorda Bay, oyster restoration projects aim to stimulate oyster production and provide new habitat for reefs to expand (De Santiago et al., 2019). Oyster restoration specifically targets historically productive reefs (Puckett et al., 2018), because viable substrate is critical for successful restoration (Chowdhury et al., 2019). However, many other factors influence habitat quality for oysters, such as salinity and food availability (Chowdhury et al., 2019; Theuerkauf & Lipcius, 2016). Restoration efforts are put at risk when sites are selected without considering the other physical and biological factors that influence oyster's ability to survive (Pollack et al., 2012).

Restoration site selection may be aided through tools such as Habitat Suitability Indices (HSIs) (Pollack et al., 2012). Oyster Habitat Suitability Index (OHSI) models evaluate and score marine habitats based on their ability to support oyster life (Pollack et al., 2012; Theuerkauf & Lipcius, 2016). OHSIs quantify the physical requirements of oysters (Brown & Hartwick, 1988; Puckett et al., 2018; Theuerkauf & Lipcius, 2016), and score habitat accordingly. This provides exact locations of habitat ranging from unsuitable to optimal (Brown & Hartwick, 1988; Cho et al., 2012).

OHSI modeling has been performed for oyster populations across the world. Table 6 presents the input variables and modeling techniques previously used to generate OHSI models. OHSI are generally produced through weighted (Pollack et al., 2012; Snyder et al., 2017) or unweighted (Cake, 1983; Cho et al., 2012; Soniat & Brody, 1988; Theuerkauf & Lipcius, 2016) overlay models. Overlay models are analytical tools used in Geographic Information Systems to combine multiple layers of spatial data and thereby identify areas where different datasets overlap or intersect (Reddy, 2018). The MaxEnt model has also been used to evaluate oyster habitat suitability (Linhoss & Mickle, 2022). This model is more complex

than the overlay approach; it provides information about habitat quality through probability of oyster presence based on select physical parameters (Linhoss & Mickle, 2022).

OHSIs are most effective when they consider the physical parameters that best suit the needs of a local population. The most frequently implemented model parameters were salinity (Cake, 1983; Cho et al., 2012; Pollack et al., 2012; Soniat & Brody, 1988; Theuerkauf & Lipcius, 2016), substrate (Cake, 1983; Soniat & Brody, 1988; Theuerkauf & Lipcius, 2016), and temperature (Cho et al., 2012; Pollack et al., 2012; Snyder et al., 2017). Other commonly used parameters include turbidity (Pollack et al., 2012; Snyder et al., 2017), water depth (Pollack et al., 2012; Theuerkauf & Lipcius, 2016), and chlorophyll-a (Cho et al., 2012; Snyder et al., 2017). Over specific regions of Matagorda Bay (Figure 11), Turner et al. (2016) generated an OHSI for Tres Palacios and Carancahua Bays using a regression-based model. This OHSI model used mean salinity, minimum salinity, and salinity during spawning as variables to explain habitat suitability.

OHSIs generated from overlay models have faced criticism because they are biased and inherently simple (Theuerkauf & Lipcius, 2016). Unweighted models do not consider varying significance of physical parameters (Theuerkauf & Lipcius, 2016), while weighted models assume relationships between physical conditions and oyster populations (Cho et al., 2012). OHSI models like these depend heavily on theory (Lavallin, 2021). Model validation is critical to ensure the model can effectively predict suitable habitat (Puckett et al., 2018; Soniat & Brody, 1988; Theuerkauf & Lipcius, 2016). However, many OHSI overlay models are unvalidated (Snyder et al., 2017). Models may be validated by correlating OHSI scores to a biologic response from the oyster, such as population density (Puckett et al., 2018; Soniat & Brody, 1988; Theuerkauf & Lipcius, 2016) and biomass (Smith et al., 2022).

ML provides a new approach to modeling OHSI (Garcia-Quintas et al., 2023; Santoso et al., 2023). ML algorithms could be used to conceptualize complex relationships between biophysical conditions (e.g., model inputs) and oysters. These models are data driven, meaning they independently form a relationship between physical parameters and the response from the oyster population (Lavallin, 2021). These models are also innovative because they rely less on theory (Lavallin, 2021) and have been tested with independent data. ML models have produced HSIs for species such as the shortfin mako shark (Garrison, 2023), seabirds (Garcia-Quintas et al., 2023), grizzled leaf monkey (Santoso et al., 2023), and the Asian tiger mosquito (Georgiades et al., 2023); however, this approach is not commonly used for OHSIs.

Table 6: Input variables and model types used to generate OHSIs, globally.

		Source							
		(Cake, 1983)	(Soniat & Brody, 1988)	(Theuerkauf & Lipcius, 2016)	(Pollack et al., 2012)	(Cho et al., 2012)	(Snyder et al., 2017)	(Linhoss & Mickle, 2022)	(Turner et al., 2016)
Input physical parameters	Salinity	X	X	X	X	X		X	X
	Turbidity				X		X		
	Temperature				X	X	X	X	
	Depth			X	X				
	Chlorophyll-a					X	X		
	Substrate	X	X	X					
	Suspended particulate matter					X			
	Dissolved oxygen				X	X		X	
	Clutch coverage, suitable clutch	X	X						
	Frequency of killing floods	X	X						
	Disease/predator	X	X						
	Larval transportation								
	Abundance	X	X						
	Hydrodynamics					X			
	Bathymetry							X	
	Dermo/disease/ parasites								
Model type	Weighted overlay model				X		X		
	Unweighted overlay model	X	X	X		X			X
	MaxEnt model							X	

This study aims to generate different OHSI models over Matagorda Bay, primarily focusing on water quality, on both a monthly and an annual basis. These models considered five physical parameters: salinity, turbidity, temperature, depth, and water velocity (Figure 12). This study produces OHSIs through three modeling approaches: classification-based models (OHSI1), regression-based models (OHSI2), and ML models (OHSI3; Figure 33). Unweighted and weighted versions were implemented for OHSI1 and OHSI2. OHSI scores produced by OHSI1 and OHSI2 models were validated using the Texas Parks and Wildlife Department (TPWD) oyster dredge dataset. We intended to analyze the variability and performance of OHSI scores produced through various modeling approaches. OHSI maps produced from this study could be used to indicate exact locations with high and low capability to sustain oyster populations in Matagorda Bay.

2.3. Study Area

Matagorda Bay, also known as the Lavaca-Colorado Estuary, is the second-largest estuary system in Texas (Figure 11). It is comprised of a primary bay, Matagorda Bay, and secondary bays, including Lavaca Bay, Tres Palacios Bay, and Carancahua Bay (Figure 11). The estuary is primarily fed by the Colorado River, Lavaca River, and Tres Palacios River. These rivers are critical to water quality in Matagorda Bay because freshwater input regulates salinity, transports sediments, and supplies nutrients that stimulate primary production (Wilber & Bass, 1998).

Matagorda Bay is home to many oyster reefs, which are generally concentrated near sources of freshwater (Figure 11). Half Moon Reef is large, historically productive reef that covered 200 ha in the early 1900s, when oyster production was at its peak (Pollack et al., 2012). Frequently harvested areas of Half Moon Reef were declared void in a 1907 study (Fisheries & Moore, 1907), before becoming completely barren of oysters by the 1930s (Aguilar, 2017).

In addition to overharvesting, dredging also depleted populations. In the early 1900s, many Texan estuaries were dredged for oyster shell, which was used as construction material (Pollack et al., 2012). Unregulated dredging resulted in major habitat loss throughout Matagorda Bay, effectively removing and depleting local oyster populations (De Santiago et al., 2019). Permitted dredging led to further depletion of oysters in Matagorda Bay; permits approved between 1961 and 1982 granted permission to dredge approximately 213 acres of oyster reef in total (Luckenbach et al., 1999).

Agencies such as TPWD and the U.S. Army Corps of Engineers (USACE) have pursued oyster reef restoration efforts in Matagorda Bay (Culbertson, 2008). Such efforts aim to restore substrate to provide greater opportunity for spat recruitment and inspire population growth (De Santiago et al., 2019; Peyre et al., 2014). Half Moon Reef is one of the largest reef restoration projects in the United States (De Santiago et al., 2019). The reef was restored in 2014 using 27 parallel rows of limestone and concrete substrate (De Santiago et al., 2019). Restoration caused notable increases in spat recruitment and increased oyster population density at the reef (De Santiago et al., 2019).

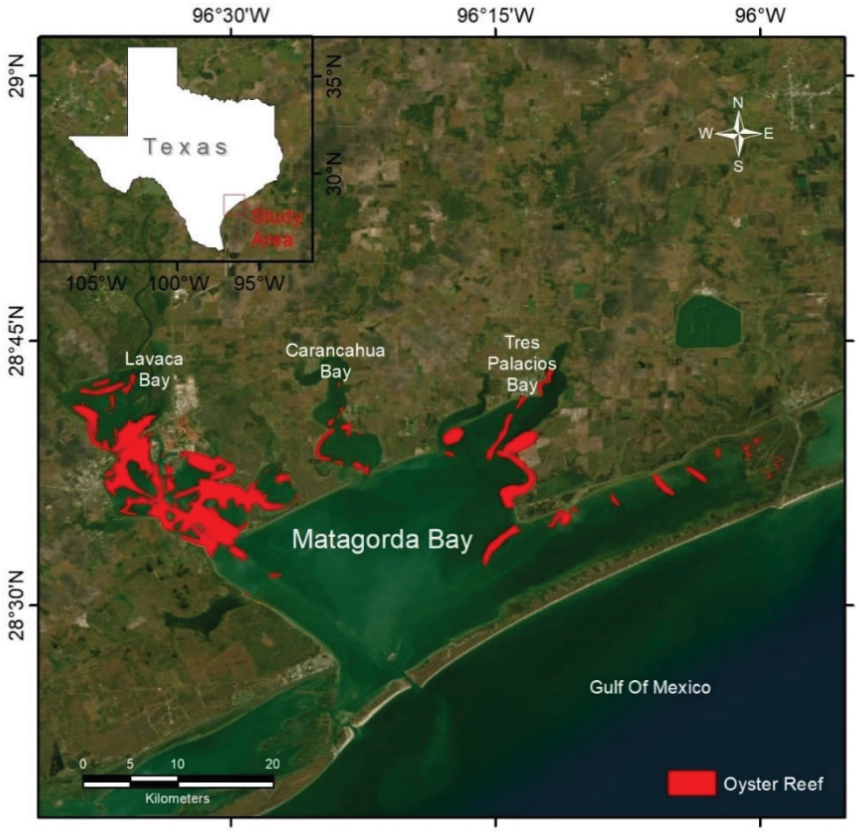

Figure 11: Matagorda Bay, including the primary (Matagorda) and secondary bays (Lavaca Bay, Carancahua Bay, and Tres Palacios Bay). Red polygons indicate regions of oyster reefs. These locations were obtained from the NOAA Gulf of Mexico Data Atlas and include surveys conducted between 2005 and 2007 (Anson et al., 2011).

2.4. Data

OSHIs were produced over Matagorda Bay using five types of physical data: including salinity, turbidity, temperature, depth, and water velocity. These variables were selected because they are significant to oysters, as indicated by previous studies over Matagorda Bay and other Texan estuaries. Salinity, for example, is a key factor that contributes to oyster abundance, biomass, and spat recruitment (LaPeyre et al., 2015; Marshall et al., 2021; Pollack et al., 2011). Oysters can tolerate a large range of salinity (Du et al., 2021; Peyre et al., 2019); however, prolonged events of critically low or high salinities result in mortality (Du et al., 2021). Low salinity reduces spat settlement yield (Marshall et al., 2019; Pollack et al., 2011),

while high concentrations are associated with oyster mortality due to increased disease and parasites (Culbertson, 2008; Du et al., 2021).

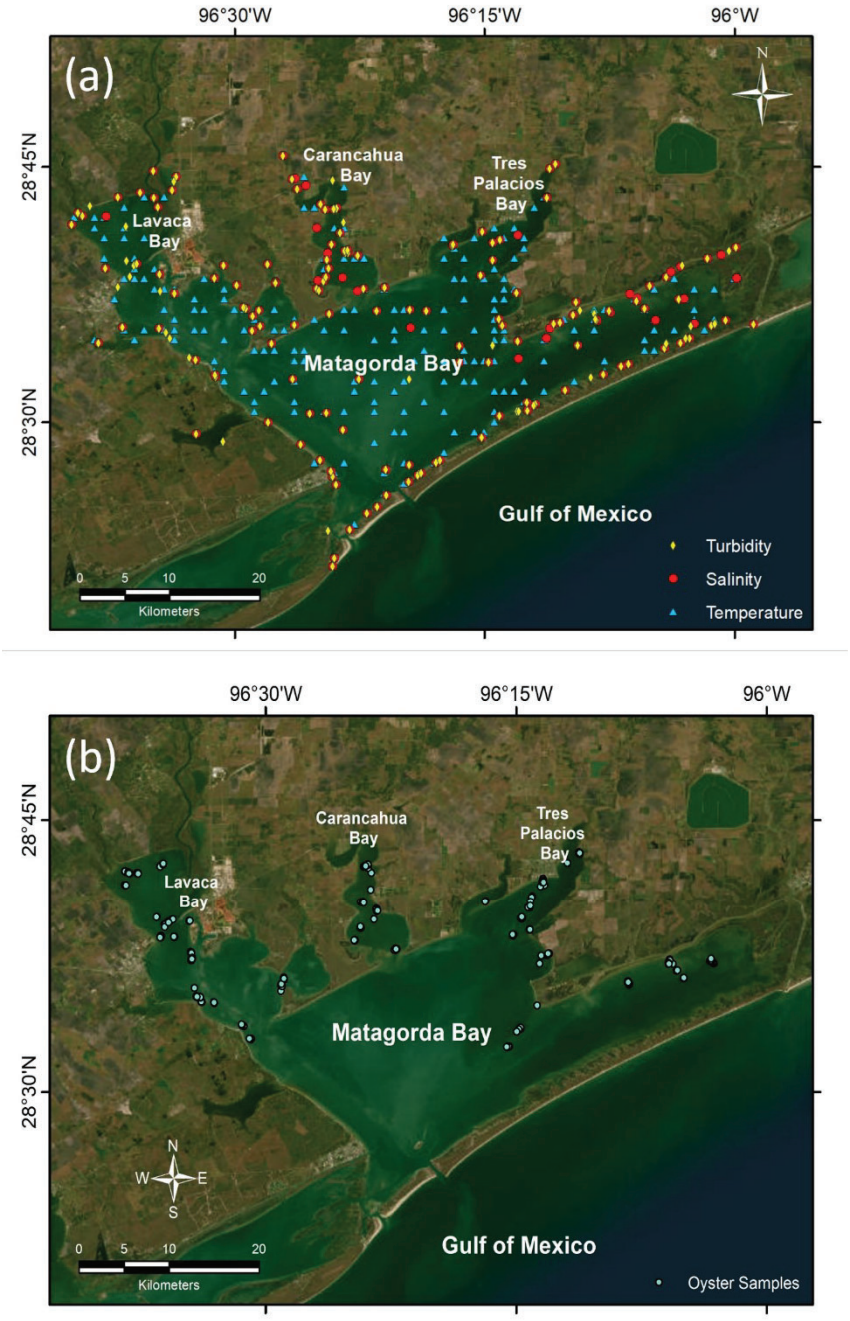

Figure 12: (a) Locations of in-situ salinity, turbidity, and temperature datasets. Salinity and turbidity data were used to produce ML models to simulate salinity and turbidity from Landsat -8 images over Matagorda Bay. Temperature data were used to validate the ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS)-derived water temperature. (b) Locations of TPWD oyster samples. This dataset was used to produce OHSI2 and OHSI3 models and to validate OHSI1 and OHSI2.

Turbidity has been found to have an inverse relationship with species richness and diversity (Aguilar, 2017). Turbidity influences filtration capabilities and sedimentation rates (Zimmerman et al., 2007). Low turbidity has been linked to improved filtration capabilities (Ehrich & Harris, 2015). Elevated turbidity inhibits filtration capabilities (Ehrich & Harris, 2015; Wang et al., 2008) and leads to a reduction in primary production, decreased growth rates, and burial of biomass (Southwell et al., 2017; Wilber & Clarke, 2010).

Temperature has a significant impact on oyster biomass (Aguilar, 2017), spawning capabilities (Casas et al., 2018), and filtration rates (Snyder et al., 2017). Low temperatures reduce the oysters' filtration rates, and therefore inhibit feeding (Bagenda et al., 2019). Spawning in regions such as Matagorda Bay generally begins when temperatures reach 25°C (Casas et al., 2018). However, higher temperatures promote disease and inhibit filtration (Ehrich & Harris, 2015; Motes et al., 1998).

Water velocity influences recruitment, settlement rates, and the delivery of nutrients to oyster beds (Snyder et al., 2017). Low water velocities increase the risk for sedimentation, which requires oysters to expend greater energy to filter the water column; it may clog gills, resulting in mortality (Stanley et al., 2024). Greater water velocities improve water quality, increase food supply, and increase oyster growth (Stanley et al., 2024).

Water depth is a factor that controls light penetration, which influences water temperature and phytoplankton growth in the lower water column, where oysters reside (Porter et al., 2004). Healthy oyster beds are generally found in shallower regions (Chapman et al., 2021). This study used live oyster count as an indicator of suitable habitat and oyster health. Live count is a measure used to observe the natural settlement (Wilkie et al., 2013) of oysters. Relatively high live oyster count indicates optimal physical conditions and lower count indicates unsuitable habitat quality.

In this study, turbidity and salinity were retrieved from Landsat-8 images using ML models (Section I). These products were generated over Matagorda Bay during the period from 2018 to 2023 at 100-m spatial resolution. Figure 12a shows the spatial distribution of the in-situ measurements used to generate these models. Water temperature was generated from the ECOSTRESS images. The ECOSTRESS instrument has been collecting data on the International Space Station since June 2018 (Fisher et al., 2020; Weidberg et al., 2023). The mission of ECOSTRESS was to observe evapotranspiration, which is analyzed through surface temperature and emissivity (Fisher et al., 2020). This instrument records surface temperature through five thermal infrared radiation bands (Fisher et al., 2020). In-situ temperature observations (n: 893) were obtained from Texas Commission on Environmental Quality (TCEQ, 2023), Lower Colorado River Association (LCRA, 2023), and TWPD (Figure 12a). Sea surface temperature was extracted from

ECOSTRESS images over dates and locations of in-situ samples during the period from 2018 to 2023. ECOSTRESS-derived temperature images were resampled to 100 m.

Water depth across Matagorda Bay was obtained from the National Oceanic and Atmospheric Agency (NOAA) Bathymetry Digital Elevation Model (BDEM). This dataset was produced from 23 bathymetric surveys collected during the period from 1934 to 1992 (National Centers for Environmental, 1998). Mean monthly tide provided by NOAA's BDEM was used to generate monthly depth datasets. The monthly depth images were resampled to a 100-m spatial resolution.

Water velocity was obtained from the TxBLEND model, a hydrodynamic and salinity transport model that simulates water velocity. This model uses generalized wave continuity and momentum equations to simulate water velocity (Matsumoto et al., 2014; Schoenbaechler et al., 2011). Net water velocity was computed using average circulation throughout the bay over several tidal cycles. The velocity product was generated from a representative 28-day period (August 11–September 7, 1997) with median salinity (23.87 psu) that is closest to the long-term average. This dataset has a spatial resolution of 100 m.

Oyster dredging datasets were obtained from TPWD for samples collected between 2018 and 2023. Figure 12b illustrates the spatial variability of these sample locations. Dredge samples were collected during a 30-second tow. The dredge sample was used to calculate total live count, total dead count, spat count, and the average length of shells dredged. The TPWD dataset was comprised of 568 samples, and provided water quality data at the locations of oyster samples, including temperature, salinity, turbidity, and dissolved oxygen.

Input datasets for the OHSI2 and OHSI3 models were extracted at the locations and dates coinciding with TPWD oyster samples (Figure 12b). Because not many TPWD sample collection and acquisition dates coincided with both Landsat-8 and ECOSTRESS acquisition dates, temperature, salinity, and turbidity data was extracted from images within ±14 days of the TPWD sample date.

2.5. **Methods**

Three approaches were used to generate OHSI models over Matagorda Bay, including classification-based models (OHSI1), regression-based models (OHSI2), and ML-based models (OHSI3; Figure 13). OHSI1 uses classified (class 1, class 2, and class 3) datasets as model inputs and scores habitats as unsuitable (class 1), moderate (class 2), or optimal habitat (class 3). OHSI2 uses suitability indices (SIs) computed for each model input variable. SIs are generated via regression analysis, quantifying the complex relationship between a biologic response—live oyster count, in this case—to an individual physical parameter (e.g., salinity, turbidity, velocity, temperature, depth). OHSI2 models produce habitat scores ranging from 0 to 1, depicting a range of suitability from unsuitable (score 0) to optimal (score 1). OHSI1

and OHSI2 were implemented using both weighted and unweighted versions (Figure 13). Weights of each input parameter were assigned based on based on these reported in previous studies (Pollack et al., 2012). However, due to the wide range of inputs used to generate OHSI models, no previous OHSI model has implemented the same combination inputs as the ones used in this study. Therefore, the weights of our model inputs were manipulated to reflect those of previous studies but not replicating them. The weights selected for model inputs are presented in Table 7. Salinity was assigned the greatest weight (40%) and depth was also assigned significant weight (25%). Weights assigned to the other model parameters include 10% for turbidity, 15% for temperature, and 10% for velocity. Lastly, OHSI3 generated from ML models that predict total live counts using the five model inputs. OHSI3 provides predicted total live oyster count, where higher oyster counts indicate optimal habitat quality and lower indicate unsuitable habitat.

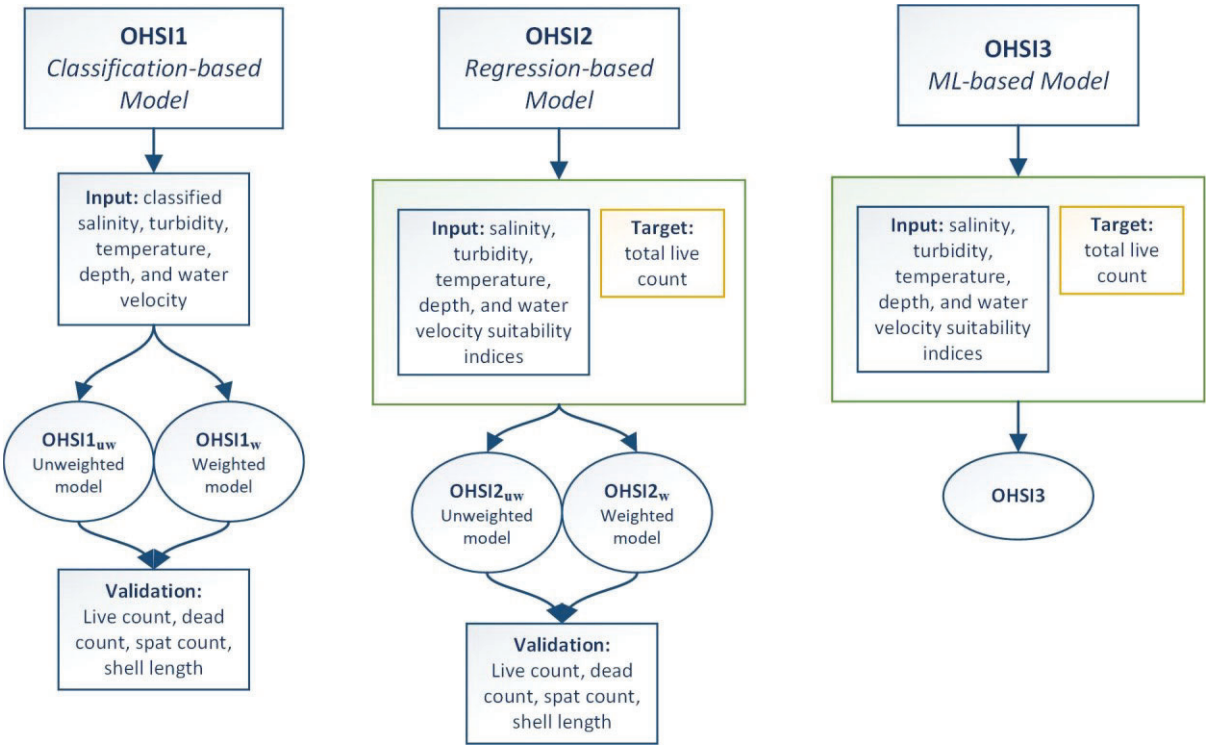

Figure 13: Overview of the modeling approaches for OHSI over Matagorda Bay.

Table 7: Weights assigned to each model input used for OHSI1w and OHSI2w models.

Physical Parameter	Weight
Salinity	40%
Turbidity	10%
Temperature	15%
Depth	25%
Water Velocity	10%

2.5.1. **Classification-Based Models (OHSI1)**

Many studies generated OHSI models based on the classification of inputs into unsuitable, moderate, and optimal classes (Pogoda et al., 2023). In this study, we generated OHSI1 models by classifying model inputs into three classes, depicting unsuitable (class 1), moderate (class 2), and optimal (class 3) conditions. Thresholds for these conditions were established from recent literature published over Texan and Floridian estuaries, as outlined in Table 8.

For salinity, the optimal conditions threshold was between 12 and 24 psu. Moderate conditions ranged from 7 to 12 psu, while unsuitable conditions were identified as those with salinity less than 7 psu or greater than 24 psu. Optimal turbidity concentrations were classified as concentrations less than 20 NTU, while unsuitable conditions were greater than 20 NTU. Temperature is optimal between 20 and 30°C, moderate from 15 to 20°C and from 30 to 35°C, and unsuitable when temperature is less than 15°C or greater than 35°C. Depth is optimal between 1 and 2 m, moderate from 0.5 to 1 m and 2 to 3.5 m, and unsuitable depth is less than 0.5 m or more than 3 m. Velocity is optimal with values greater than 0.01 m/s and unsuitable with values lower than 0.01 m/s.

Table 8: Optimal, moderate, and unsuitable conditions for each input used to generate OHSI1 models.

Input Parameter	Unit	Unsuitable Conditions (Class 1)	Moderate Conditions (Class 2)	Optimal Conditions (Class 3)	Sources
Salinity	psu	<7, >24	7-12	12-24	(Barnes et al., 2007; Dekshenieks et al., 1993; Lebreton et al., 2021; Pollack et al., 2012; H. Wang et al., 2008)
Turbidity	NTU	>20	-	<20	(Aguilar, 2017; Dekshenieks et al., 1993)
Temperature	ºC	<15, >35	15-20, 30-35	20-30	(Barnes et al., 2007; Chávez-Villalba et al., 2010; Dekshenieks et al., 1993; Peyre et al., 2021; Pollack et al., 2012)
Depth	m	<0.5, >3.0	0.5-1, 2-3.5	1-2	(Cassis et al., 2011; Pollack et al., 2012)
Water Velocity	m/s	<0.01	-	>0.01	(Campbell & Hall, 2019)

Two versions of the OHSI1 model were generated: unweighted ($OHSI1_{uw}$) (Eq. 1) and weighted ($OHSI1_{w}$) (Eq. 2):

$$OHSI1_{uw} = {(S + Tu + Te + D + V)}/{5} \quad (5)$$

$$OHSI1_{w} = (w_S * S) + (w_{Tu} * Tu) + (w_{Te} * Te) + (w_D * D) + (w_V * V) \quad (6)$$

where *S, Tu, Te, D,* and *V* are the classified salinity, turbidity, temperature, depth, and velocity datasets, respectively; *w* values are the weights assigned to salinity (w_s), turbidity (w_{tu}), temperature (w_{te}), depth (w_d), and velocity (w_v). OHSI habitat scores produced by OHSI1 models range from 1 to 3, indicating unsuitable (1) to optimal (3) habitat conditions. Outputs from these models were validated using total live counts, total dead counts, and total spat counts, all provided by the TPWD. These indicators are indicative to oyster abundance (Pollack et al., 2012).

2.5.2. **Regression-Based Models (OHSI2)**

OHSI2 used SIs as model inputs. SIs are values assigned to concentrations of a parameter, where higher values signify positive effects and lower values indicate negative effects. SIs were produced through regression analysis using the Max-Bin approach advanced in (Turner & Montagna, 2016). The Max-Bin approach is a method to analyze the relationship between a biologic response and a physical parameter. The primary purpose of these models is to evaluate peak responses of an individual species and describe ideal conditions of a specific physical (input) parameter (Turner & Montagna, 2016). Biological responses include species observations, such as abundance, shell length, or biomass, whereas, physical parameters include salinity and depth (Turner & Montagna, 2016). This model assumes a relationship between the physical parameter and the biological response; further, there is a specific range of the physical parameter that peaks the biologic response. In addition, the biologic response is assumed to decline in values of the physical response before and after the peak biologic response (Turner & Montagna, 2016). This study used live oyster count as the biologic response. The Max-Bin regression model was implemented to evaluate the relationship between the biologic response and salinity, turbidity, temperature, depth, and velocity. The following logarithmic model was implemented:

$$Y = a \times \exp\left(-0.5 \times \left(\ln\left(\frac{X/c}{b}\right)^2\right)\right) \quad (7)$$

where *Y* is the biological response, oyster total live count in this case, and *X* is the ecological driver (e.g., salinity, turbidity, temperature, depth, and velocity). Model parameters *a, b,* and *c* describe the shape of the distribution curve, where *a* is the peak biological response, *b* is the skewness of the biological response, and *c* is the value of the biophysical parameter when the biological response is at its peak (Turner & Montagna, 2016).

Once the model parameters (*a, b,* and *c)* were generated using Eq. 3, they were used to generate SIs for monthly average salinity, turbidity, temperature, depth, and velocity datasets. SIs were then used as model inputs for OHSI2 models according to these equations:

$$OHSI2_{uw} = {(SI_s + SI_{tu} + SI_{te} + SI_d + SI_v)}/{5} \quad (8)$$

$$OHSI2_w = (w_s * SI_s) + (w_{tu} * SI_{tu}) + (w_{te} * SI_{te}) + (w_d * SI_d) + (w_v * SI_v) \quad (9)$$

where *SI* is the suitability index for salinity (SI_s), turbidity (SI_{tu}), temperature (SI_{te}), depth (SI_d), and velocity (SI_v), and *w* values are the weights assigned to salinity (w_s), turbidity (w_{tu}), temperature (w_{te}), depth (w_d), and velocity (w_v). OHSI2 habitat scores produced range from 0 to 1, depicting unsuitable to optimal conditions. A quantile classification approach was employed to categorize the scores based on arbitrary thresholds. Scores greater than 0.388 were deemed optimal habitat, those ranging from 0.313 to 0.388 were classified as moderate habitat, and scores below 0.313 were considered unsuitable habitat. Outputs from these models were validated using total live counts, total dead counts, and total spat counts, provided by the TPWD.

2.5.3. **ML-Based Models (OHSI3)**

The OHSI3 model was generated using ML algorithms. Three ML algorithms were implemented, including deep neural network (DNN), distributed random forest (DRF), and gradient boosting machine (GBM). The open-source H2O-AML (automated ML) platform was employed (accessible at: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html). H2O-AML provides user-friendly, fully automated supervised learning algorithms, catering to both those well-versed in the field and those without expertise (Truong et al., 2019). DNN is a feedforward network that uses multiple hidden layers composed of neurons to analyze complex relationships between inputs and target features (Kamilaris & Prenafeta-Boldú, 2018; Mathew et al., 2017; Oyebisi & Alomayri, 2023; Tang et al., 2020). DRF combines multiple weak decision trees to produce a strong ensemble forest (Asgari et al., 2022; Shrivastav & Kumar, 2022). The GBM family generates an ensemble model using parallel regression trees (Natekin & Knoll, 2013; Shrivastav & Kumar, 2022). Comprehensive descriptions regarding structures and hyperparameters for H2O-AML families can be found on the H2O-AML website (https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html).

The input data for the OHSI3 model was comprised of live count as a model target and the salinity, turbidity, temperature, depth, and velocity as model inputs. Model inputs were normalized within the range

of 0 to 1 to guarantee equal consideration of all input variables. Input and target data were randomly split into training (64%), validation (16%), and testing (20%) frames. Each ML algorithm was simulated over 50 runs, and the model with highest testing performance was reported for each algorithm. To prevent overfitting in each model, early stopping criteria were enforced, employing the mean-squared error as the stopping metric. This involved setting a stopping round value of 5 and a stopping tolerance of 0.0001. Model performance was evaluated through the Nash-Sutcliffe efficiency (NSE) coefficient, correlation coefficient (r), and normalized root mean square error (NRMSE).

The predicted total live counts were used to generate the habitat scores. The assumption is that live oyster count is indicative of habitat suitability. Therefore, greater live count predicted by the OHSI3 model indicates more suitable habitat. A quantile classification approach was employed to categorize the predicted total live counts based on arbitrary thresholds. Total live counts greater than 9 were deemed optimal habitat, those ranging from 9 to 8 were classified as moderate habitat, and those below 8 were considered unsuitable habitat.

2.6. Results

2.6.1. Model Inputs

Figure 14a illustrates the relationship between in-situ and ECOSTRESS-derived temperature data extracted over Matagorda Bay. Temperature extracted from ECOSTRESS images displayed high correlation (r: 0.84) to in-situ data. Relatively lower correlations were estimated between salinity and turbidity datasets retrieved from Landsat-8 data and in-situ observations (salinity r: 0.49; turbidity r: 0.79; Figure 14b and 4c).

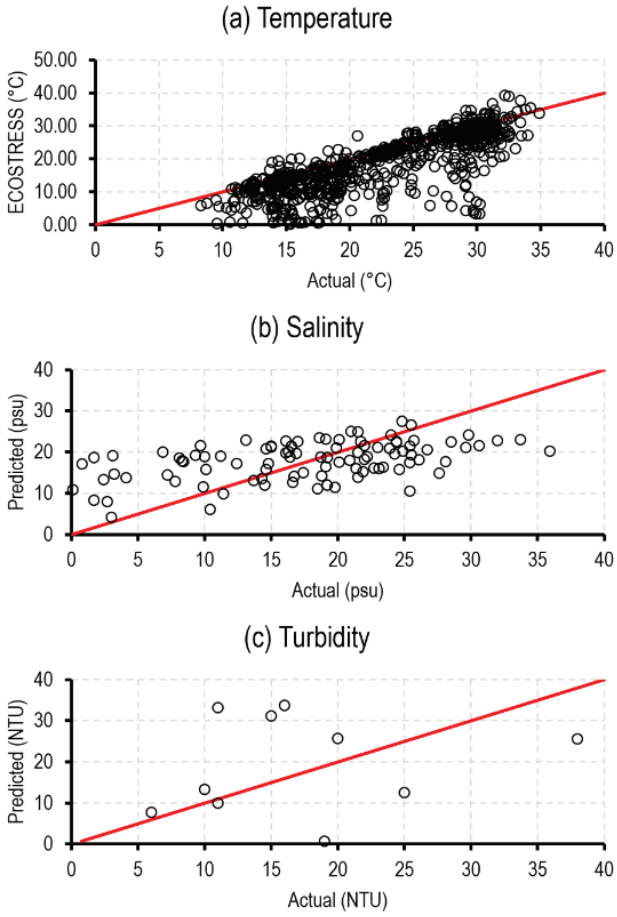

Figure 14: Correlation between (a) in-situ and ECOSTRESS-derived temperature, (b) in-situ and Landsat-8-retrived salinity, and (c) in-situ and Landsat-8-retrived turbidity. Data shown in both (b) and (c) represent testing outputs for ML models used to retrieve salinity and turbidity from Landsat-8 data. Red lines indicate a 1:1 relationship.

2.6.2. Inputs for OHSI1

Statistics for monthly and annual OHSI1 inputs are presented in Table 4. Salinity inputs have no spatial or temporal variability (Figure 15). The monthly and annual salinity datasets (Figure 15a) were classified over the entire bay (>99%) as optimal habitat. The turbidity inputs (Figure 15b) also had no spatial or temporal variability and were classified almost entirely (>90%) as unsuitable. The annual temperature dataset (Figure 15c) had no spatial variability and classified the entire bay (98%) as optimal habitat. However, the monthly temperature inputs had significant temporal and some spatial variability. Winter temperatures (December, January, February) were classified entirely (>88%) as unsuitable habitat. March was entirely (99%) classified as moderate habitat as temperatures began to rise. April and May were entirely (>99%) optimal habitat. June was found to be less optimal habitat as temperatures began to rise in the summer. July was classified as entirely (99%) moderate as temperatures exceeded 30°C. August, September, and October temperatures were entirely (99%) classified as optimal habitat. November showed a decline in optimal habitat coverage as temperatures began to decline into winter. Depth had the greatest spatial variability (Figure 15d), with optimal habitat primarily in the secondary bays and unsuitable habitat in the deeper, primary bay. The velocity datasets (Figure 15e) also illustrated some spatial variability.

2.6.3. Inputs for OHSI2

Statistics for monthly and annual OHSI2 inputs are presented in Table 4. Over the entire bay, the salinity SIs had no spatial or temporal variability. Each salinity SI had a score of 0.61 ± 0.00 (average ± standard deviation; Figure 16a). Turbidity SIs varied spatially and temporally. The lowest average values were found in November (0.15 ± 0.02; Figure 16b), whereas the summer months (June, July, August) had the highest averages (0.20 ± 0.02). The annual dataset had an average SI of 0.18 ± 0.02. The temperature SIs had spatial and temporal variability (Figure 16c). Monthly average SIs were lowest in the summer and highest in winter. July had the lowest monthly average SI value of 0.24 ± 0.01, and February had the highest monthly SI of 0.63 ± 0.01. The annual dataset had an average SI of 0.39 ± 0.00. The depth dataset had substantial spatial variability, but little temporal variability (Figure 16d). October had the lowest monthly average SI (0.06 ± 0.09) and January had the highest monthly average SI (0.10 ± 0.18). The average SI of the annual depth dataset was 0.08 ± 0.14. The average SI of the velocity dataset was 0.52 ± 0.35. Velocity over Matagorda Bay demonstrated high spatial variability (Figure 16e).

2.6.4. Inputs for OHSI3

Statistics for monthly and annual OHSI3 inputs are presented in Table 9. Over the entire bay, the salinity inputs had spatial variability (Figure 17a). Salinity concentrations were lowest near freshwater sources and increased toward the primary bay. These datasets had low temporal variability. The lowest monthly average was from May (17.57 ± 0.39 psu) and the highest monthly average was from October (18.58 ± 0.29 psu). The annual dataset had an average of 18.13 ± 0.24 psu. The turbidity inputs had greater spatial and temporal variability (Figure 17b). Turbidity was lowest in the primary bay and highest near freshwater sources, as the river inputs are generally much more turbulent than the bay. Monthly averages were lowest in the summer months, specifically in June (21.56 ± 1.04 NTU) and greatest in November (25.09 ± 1.37 NTU). The annual average turbidity was 22.09 ± 0.26 NTU. The temperature datasets had the greatest temporal variability (Figure 17c). Water temperatures were lowest in the winter months (December, January, February) and highest in the summer months (June, July, August). February had lowest monthly average (10.72 ± 0.40°C) and July had the highest monthly average (31.07 ± 0.50°C). Annual average water temperature is 22.52 ± 0.85°C. The depth datasets had significant spatial variability (Figure 17d). Temporal variability represents monthly variation of tidal influence. Tidal influence was lowest in January and greatest in October. The annual average depth over Matagorda Bay is -2.78 ± 1.37 m. The water velocity average was low (0.01 ± 0.01 m/s), but illustrated high spatial variability (Figure 17e).

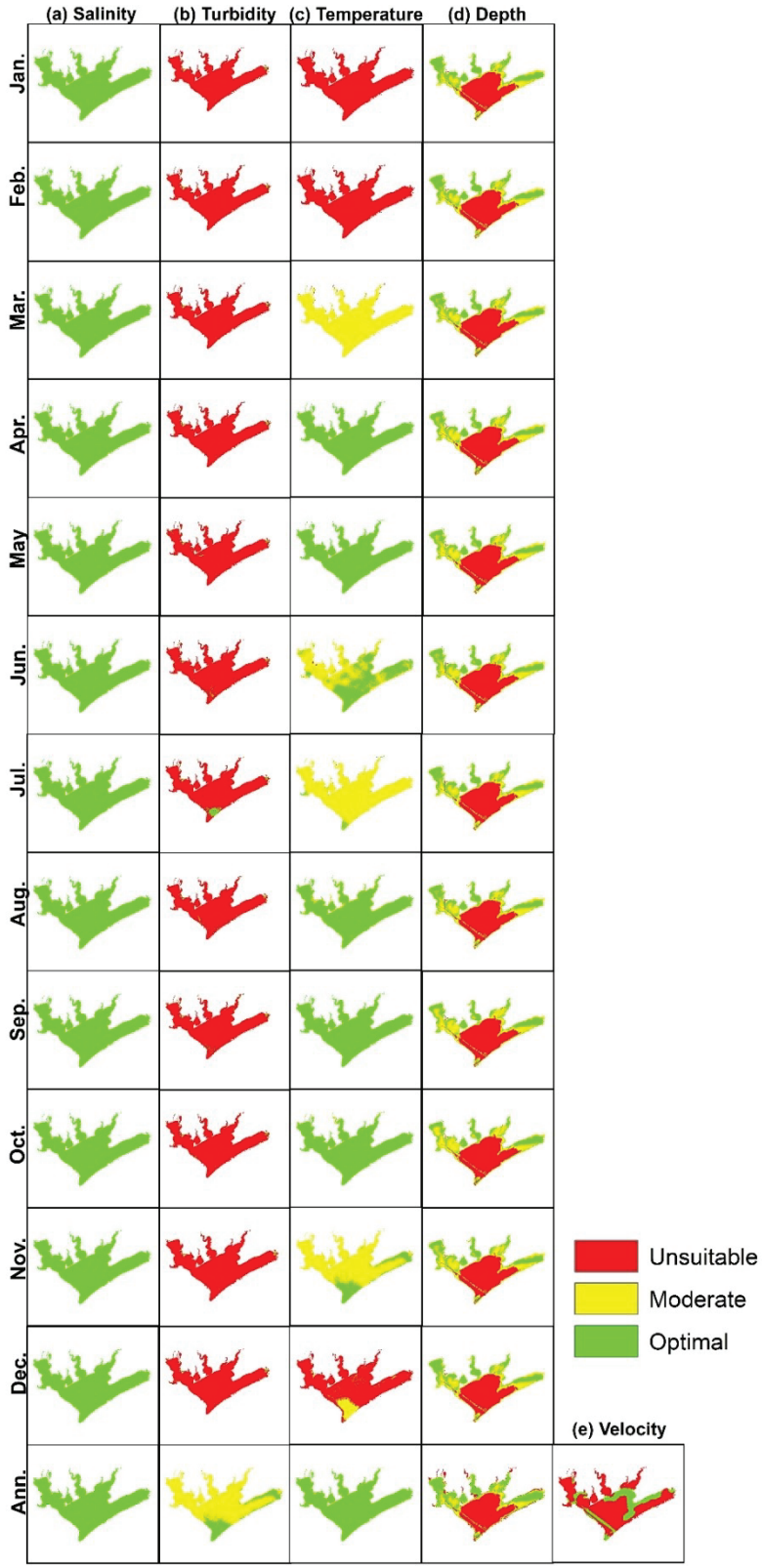

Figure 15: Annual and monthly inputs for (a) salinity, (b) turbidity, (c) temperature, (d) depth, and (e) water velocity for OHSI1uw and OHSI1w models.

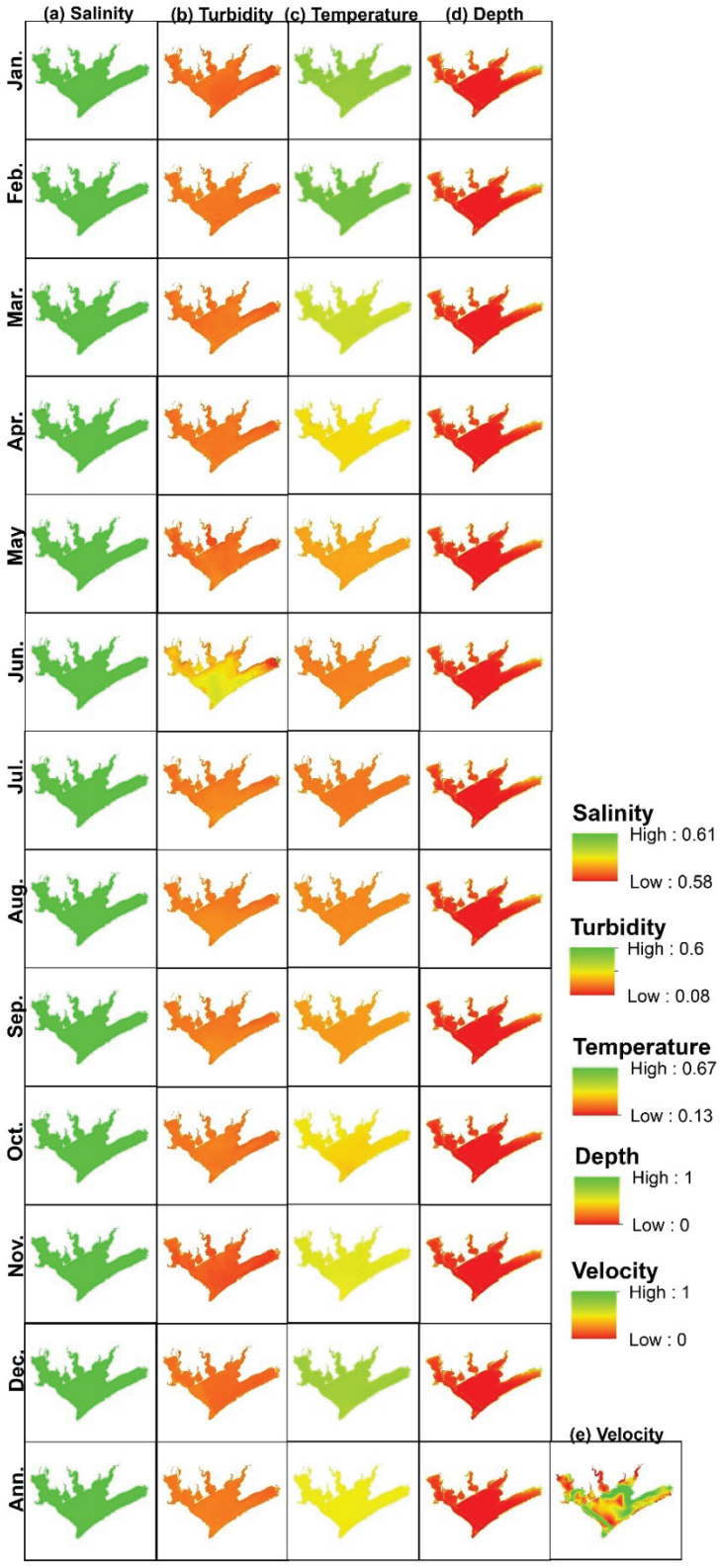

Figure 16: Annual and monthly inputs for (a) salinity, (b) turbidity, (c) temperature, (d) depth, and (e) water velocity for OHSI2uw and OHSI2w models.

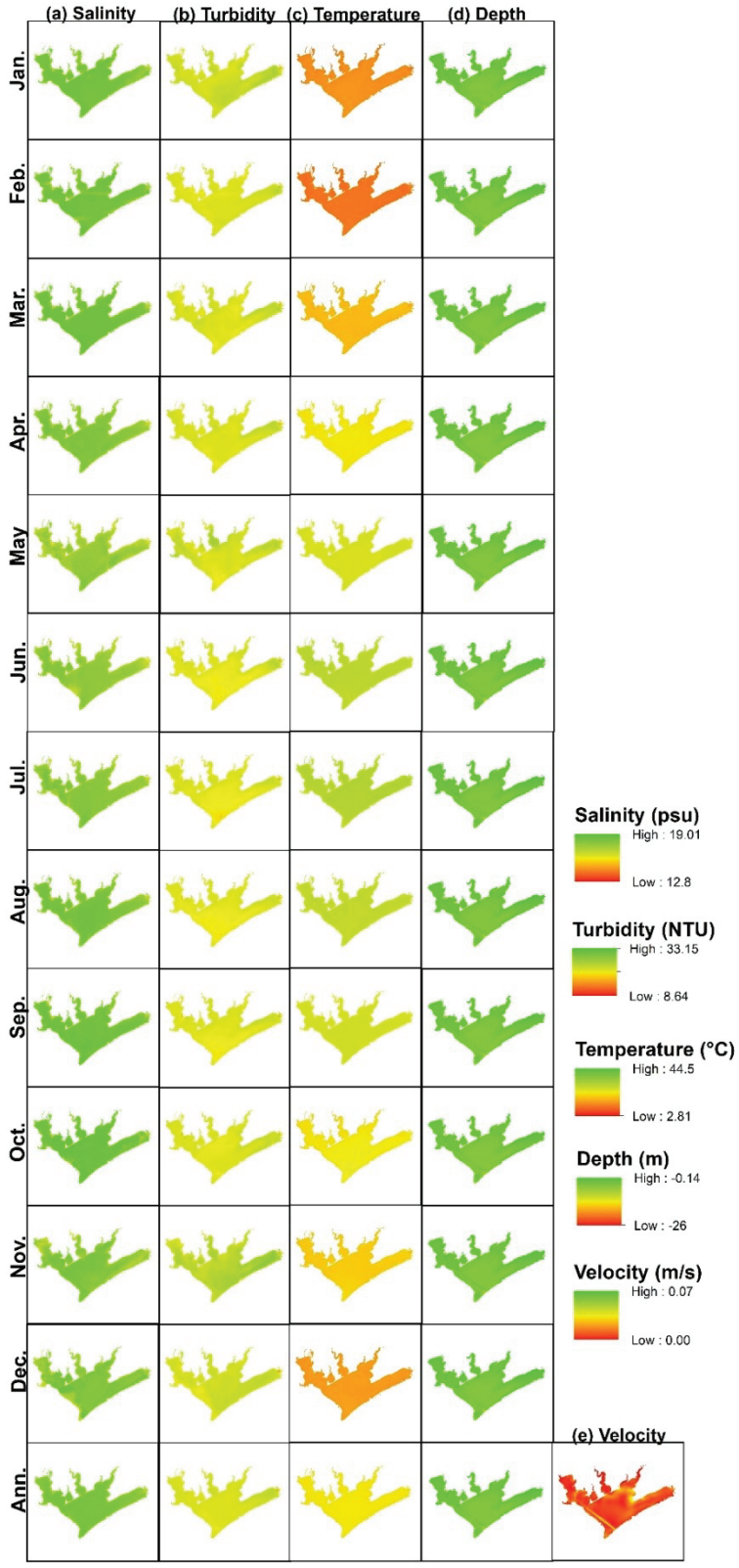

Figure 17: Annual and monthly inputs for (a) salinity, (b) turbidity, (c) temperature, (d) depth, and (e) water velocity for OHSI3 models.

Table 9: Statistics for inputs used for OHSI1, OHSI2, and OHSI3 models generated over the entire Matagorda Bay. Statistical measurements include percent coverage of optimal (Opt. (%)), moderate (Mod. (%)), and unsuitable (Unsuit. (%)) conditions for the OHSI1 inputs. Statistics for OHSI2 and OHSI3 inputs were quantified through average and standard deviation (Avg. ± Std. Dev.), and minimum (Min.) and maximum (Max.) values for SI, and actual inputs, respectively.

Model Parameter	Month	OHSI1 (classes)			OHSI2 (SI values)			OHSI3 (actual values)		
		Opt. (%)	Mod. (%)	Unsuit. (%)	Avg. ± Std.Dev.	Min.	Max.	Avg. ± Std.Dev.	Min.	Max.
Salinity	1	100.00	0.00	0.00	0.61 ± 0.00	0.60	0.61	18.37 ± 0.24	15.40	18.75
	2	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	18.40 ± 0.25	14.80	18.79
	3	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	18.49 ± 0.21	14.92	18.84
	4	100.00	0.00	0.00	0.61 ± 0.00	0.60	0.61	18.06 ± 0.22	15.05	18.63
	5	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	17.57 ± 0.39	14.47	18.68
	6	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	17.75 ± 0.32	13.46	18.58
	7	100.00	0.00	0.00	0.61 ± 0.00	0.60	0.61	17.85 ± 0.27	14.45	18.64
	8	100.00	0.00	0.00	0.61 ± 0.00	0.60	0.61	18.30 ± 0.26	14.83	18.66
	9	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	18.47 ± 0.26	14.80	18.92
	10	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	18.58 ± 0.29	15.01	19.01
	11	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	18.01 ± 0.38	14.84	18.65
	12	100.00	0.00	0.00	0.61 ± 0.00	0.59	0.61	18.05 ± 0.36	14.93	18.97
	Ann.	99.41	0.00	0.59	0.61 ± 0.00	0.60	0.61	18.13 ± 0.24	14.59	18.71
Turbidity	1	1.05	0.00	98.95	0.17 ± 0.02	0.11	0.42	23.37 ± 1.01	13.38	27.66
	2	1.06	0.00	98.94	0.19 ± 0.01	0.13	0.42	22.24 ± 0.78	13.60	26.96
	3	1.10	0.00	98.90	0.19 ± 0.02	0.12	0.48	22.35 ± 0.98	13.33	26.83
	4	1.10	0.00	98.90	0.19 ± 0.02	0.13	0.53	22.36 ± 0.83	11.77	26.16
	5	1.57	0.00	98.43	0.18 ± 0.02	0.09	0.49	22.81 ± 1.35	12.36	28.27
	6	1.94	0.00	98.06	0.20 ± 0.02	0.13	0.60	21.56 ± 1.04	10.32	26.01
	7	4.20	0.00	95.80	0.20 ± 0.02	0.14	0.48	21.64 ± 1.00	12.82	25.84
	8	1.79	0.00	98.21	0.20 ± 0.02	0.14	0.50	21.51 ± 0.88	12.68	25.06
	9	1.96	0.00	98.04	0.19 ± 0.02	0.13	0.56	21.85 ± 0.91	11.13	26.23
	10	1.27	0.00	98.73	0.19 ± 0.02	0.11	0.49	22.21 ± 0.89	11.50	27.48
	11	0.84	0.00	99.16	0.15 ± 0.02	0.11	0.44	25.09 ± 1.37	12.68	28.97
	12	1.29	0.00	98.71	0.18 ± 0.02	0.10	0.53	23.13 ± 0.96	13.04	27.01
	Ann.	0.93	6.86	92.21	0.18 ± 0.01	0.14	0.46	22.09 ± 0.26	16.80	26.21
Temperature	1	0.00	0.06	99.94	0.56 ± 0.02	0.38	0.67	13.52 ± 0.61	9.26	14.84
	2	0.00	0.03	99.96	0.63 ± 0.01	0.42	0.67	10.72 ± 0.40	9.32	14.14
	3	0.25	99.06	0.69	0.46 ± 0.01	0.32	0.66	17.81 ± 0.39	13.17	23.56
	4	99.97	0.03	0.00	0.37 ± 0.01	0.21	0.50	22.18 ± 0.39	19.18	28.06
	5	99.79	0.00	0.21	0.30 ± 0.01	0.26	0.43	26.60 ± 0.39	22.74	28.27
	6	45.41	54.22	0.37	0.25 ± 0.01	0.14	0.32	30.04 ± 0.54	26.10	38.58
	7	2.23	97.59	0.18	0.24 ± 0.01	0.14	0.29	31.07 ± 0.50	27.93	35.31

	8	94.76	5.22	0.02	0.26 ± 0.01	0.18	0.43	29.28 ± 0.44	26.35	34.67
	9	99.95	0.05	0.00	0.29 ± 0.01	0.19	0.42	27.28 ± 0.34	24.29	29.12
	10	98.89	1.11	0.01	0.37 ± 0.01	0.25	0.66	22.42 ± 0.70	18.85	26.26
	11	16.49	83.49	0.02	0.42 ± 0.01	0.35	0.58	19.71 ± 0.36	16.54	22.77
	12	0.01	11.50	88.49	0.53 ± 0.01	0.42	0.66	14.67 ± 0.36	12.63	16.26
	Ann.	98.15	1.82	0.03	0.39 ± 0.00	0.32	0.45	22.52 ± 0.85	12.01	26.52
Depth	1	34.87	24.44	40.69	0.10 ± 0.18	0.00	1.01	-2.65 ± 1.37	-20.97	-0.14
	2	34.50	24.51	40.99	0.09 ± 0.16	0.00	1.01	-2.70 ± 1.37	-21.02	-0.19
	3	33.20	25.62	41.17	0.08 ± 0.15	0.00	0.96	-2.76 ± 1.37	-21.08	-0.25
	4	32.37	26.01	41.62	0.08 ± 0.13	0.00	0.88	-2.82 ± 1.37	-21.14	-0.31
	5	32.23	26.10	41.66	0.08 ± 0.13	0.00	0.86	-2.83 ± 1.37	-21.15	-0.32
	6	32.72	25.93	41.35	0.08 ± 0.14	0.00	0.90	-2.80 ± 1.37	-21.12	-0.29
	7	33.70	25.29	41.01	0.09 ± 0.16	0.00	1.00	-2.72 ± 1.37	-21.04	-0.21
	8	33.20	25.62	41.17	0.08 ± 0.15	0.00	0.96	-2.76 ± 1.37	-21.08	-0.25
	9	32.51	25.46	42.02	0.07 ± 0.11	0.00	0.73	-2.92 ± 1.37	-21.24	-0.41
	10	31.89	25.78	42.33	0.06 ± 0.09	0.00	0.62	-3.00 ± 1.37	-21.32	-0.49
	11	31.90	26.30	41.79	0.07 ± 0.12	0.00	0.81	-2.86 ± 1.37	-21.18	-0.35
	12	33.82	25.23	40.95	0.09 ± 0.16	0.00	1.00	-2.71 ± 1.37	-21.03	-0.20
	Ann.	31.31	19.95	48.74	0.08 ± 0.14	0.00	0.89	-2.78 ± 1.37	-25.09	-0.28
Velocity		24.70	72.37	2.93	0.52 ± 0.35	0.00	1.20	0.01 ± 0.01	0.00	0.07

OHSI2 and OHSI3 models were constructed using the oyster total live count dataset as model targets. Inputs and target datasets were extracted at the TPWD sample locations shown in Figure 12b. Table 10 displays statistics for the input and target data at sample locations. Oyster samples generally occurred over locations of existing reefs. Live count varied significantly. No live oysters were found at locations throughout the primary bay. All samples taken at Half Moon Reef had zero total live count. The highest live counts (>100 oysters) were sampled at reefs in Lavaca Bay and Tres Palacios Bay. Oyster samples were taken at depths <3 m, where average salinity is moderate (17 psu), the average turbidity is low (22 NTU), temperature is warm (21°C), and water velocity is slow (0.01 m/s).

Table 10: Statistics for model OHSI 2 and OHSI3 inputs and targets (e.g., total live counts) extracted at the TPWD sample locations shown in Figure 12b. Statistical measures include average (Avg.), standard deviation (Std. Dev.), minimum value (Min.), and maximum value (Max.).

Statistical measure	Live Count	Salinity	Turbidity	Temperature	Velocity	Depth
Average	9.77	17.03	22.52	20.95	0.01	1.31
Std. Dev.	19.34	1.61	3.22	7.98	0.01	0.48
Min.	0.00	11.79	11.35	0.07	0.00	0.14
Max.	213.00	19.27	37.40	36.76	0.03	3.03

2.6.5. **OHSI1 Outputs**

Table 11 shows statistics for habitat percent coverage generated by OHSI1 models over Matagorda Bay, on both a monthly and an annual basis. Figure 18 shows the spatial distributions of theses habitats. The $OHSI1_{uw}$ model classified few regions (<1%) as optimal. Optimal habitat is identified in regions near freshwater sources where water is shallow, and velocity is high. Spatial variability in $OHSI1_{uw}$ scores follow the trajectory of freshwater flow, specifically from the Colorado River in the eastern arm, classifying these regions as moderate. Unsuitable habitats are found in secondary bays in the winter. Despite optimal conditions of other parameters, such as depth, water velocity, and salinity, the $OHSI1_{uw}$ model classifies these regions as unsuitable due to the cold water temperatures. The $OHSI1_{uw}$ model classified the primary bay as unsuitable habitat in the monthly and annual scores. The $OHSI1_{w}$ classified most habitat as optimal or moderate. Late spring (April and May) and early fall (September and October) had the greatest optimal habitat coverage (>90%). The other months had high optimal habitat coverage (>50%), which was generally concentrated in secondary bays and the eastern arm. The primary bay was classified as moderate habitat in the months with spatial variability. This model had low (<2%) unsuitable habitat coverage.

.

Table 11: Habitat percent coverage for the monthly and annual datasets, produced by the OHSI1uw and the OHSI1w models.

Month	$OHSI1_{uw}$			$OHSI1_{w}$		
	Opt. (%)	Mod. (%)	Unsuit. (%)	Opt. (%)	Mod. (%)	Unsuit. (%)
1	0.00	13.76	86.24	51.77	49.97	0.60
2	0.00	13.50	86.50	49.00	50.34	0.66
3	0.00	43.01	56.99	59.55	39.81	0.64
4	0.08	60.63	39.28	93.80	5.63	0.57
5	0.11	60.86	39.04	93.79	5.56	0.65
6	0.01	50.01	49.97	79.24	20.14	0.62
7	0.00	45.90	54.09	61.66	37.76	0.58
8	0.02	60.06	39.91	93.24	6.26	0.49
9	0.13	60.82	39.05	93.90	5.47	0.63
10	0.16	60.22	39.63	93.94	5.50	0.57
11	0.00	44.36	55.63	65.44	32.96	1.60
12	0.00	16.42	83.58	51.65	47.92	0.43
Annual	0.12	65.36	34.52	98.33	1.67	0.00

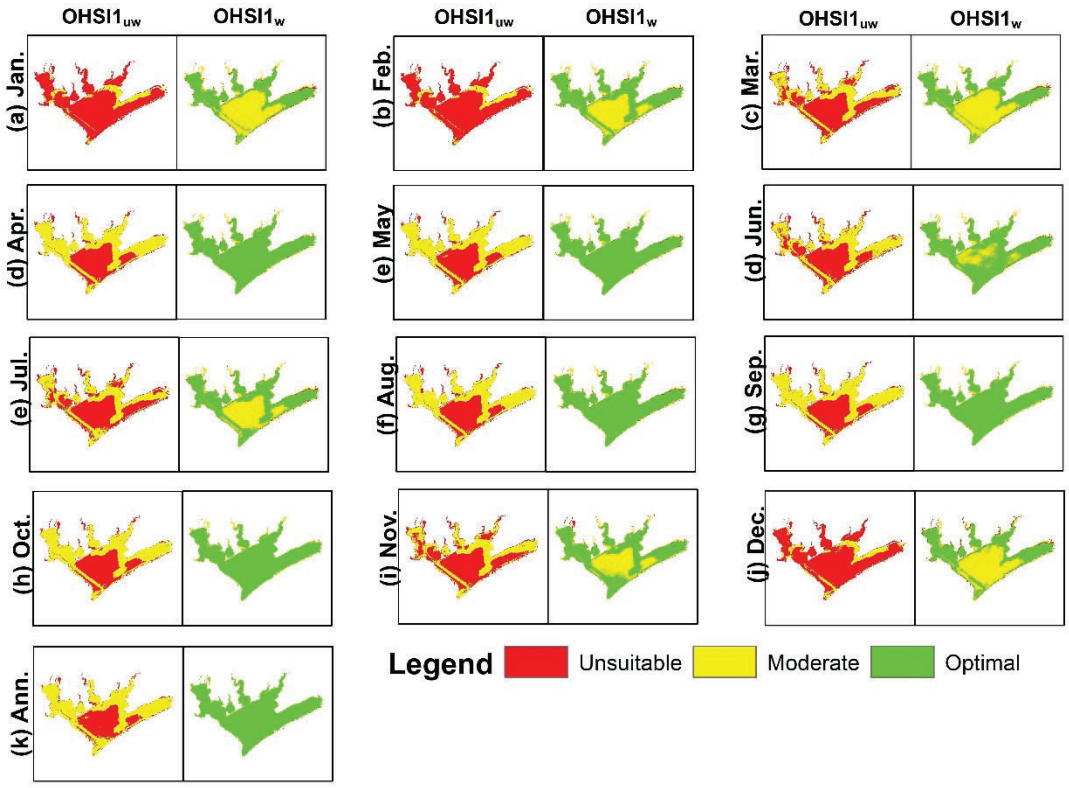

Figure 18: Monthly (a)-(j) and annual (k) scores produced by the $OHSI1_{uw}$ and $OHSI1_{w}$ models. Red indicates unsuitable habitat, yellow moderate habitat, and green optimal habitat.

Figure 19 illustrates the spatial correlation between the total live count (Figure 19a, 19d), total spat count (Figure 19b, 19e), and total dead count (Figure 19c, 19f), as well as the annual $OHSI1_{uw}$ and $OHSI1_{w}$ scores. All samples, including total live, dead, and spat were taken in locations identified as moderate and optimal habitat by $OHSI1_{uw}$ and $OHSI1_{w}$. Live and spat count were greatest in secondary bays, specifically in Lavaca and Tres Palacios Bay. Year-round moderate conditions are identified by $OHSI1_{uw}$ at the explicit locations with high live and spat count. Lowest live and spat count in these secondary bays are in locations identified as unsuitable during the winter months. Total dead count was highest in Lavaca Bay, where live and spat counts are high.

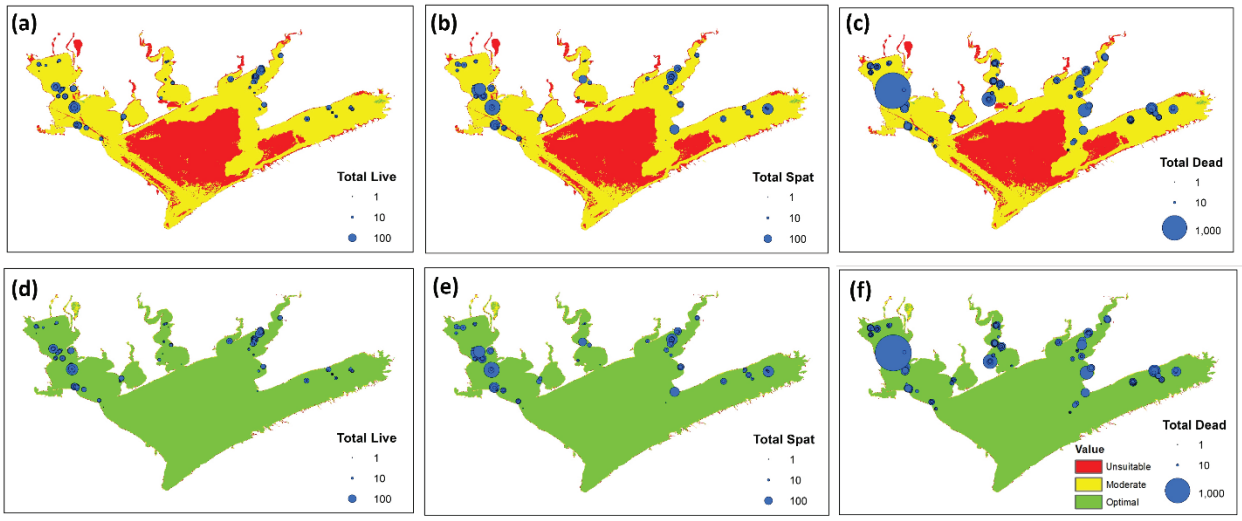

Figure 19: Oyster samples (blue), including total live (a), (d), total spat (b), (e), and total dead (c), (f) over the annual score maps produced by $OHSI1_{uw}$ (a)–(c) and $OHSI1_{w}$ (d)–(f), where red indicates unsuitable habitat, yellow is moderate habitat, and green is unsuitable habitat.

2.6.6. **OHSI2 Outputs**

Coefficients for the Max-Bin regression analysis are displayed in Table 12. Figure 20 shows visualizations of each fitted Max-Bin regression model. Coefficients extracted for the salinity model were a = 0.90, b = 1.08, and c = 13.43 psu. The optimal salinity concentration was ~20 psu. This model found that total live declined at salinities less than 20 psu and greater than 20 psu; however, lower salinities (<10 psu) had more detrimental effects on the live counts. Coefficients extracted for the turbidity model were a = 0.79, b = 0.19, and c = 21.33 NTU. This model found the optimal turbidity concentration to be ~20 NTU, where low (<15 NTU) and high (>32 NTU) turbidities have similar effects on live counts. Coefficients extracted for the temperature model were a = 0.67, b = 0.46, and c = 16.00°C. The optimal temperature depicted by this model was 10°C. The relationship between normalized live and temperature steeply declines in temperatures <7°C. Temperatures greater than 10°C experience only a slight decline in normalized total live. Coefficients extracted for the water depth were a = 1.00, b = 0.10, and c= -1.85 m. The optimal depth was ~1.5 m, and total live experienced significant drops in depths >3 m and <1 m. Coefficients generated for the velocity model were a = 1.20, b = 0.84, and c = 0.03 m/s. Total live increased when water velocity increased.

Table 12: Model coefficients for the Max-Bin regression models between input parameters and total live count.

Parameter	Coefficient		
	a	*b*	*c*
Salinity	0.90	1.08	13.43
Turbidity	0.79	0.19	21.33
Temperature	0.67	0.46	16.00
Depth	1.00	0.10	-1.85
Velocity	1.20	0.84	0.03

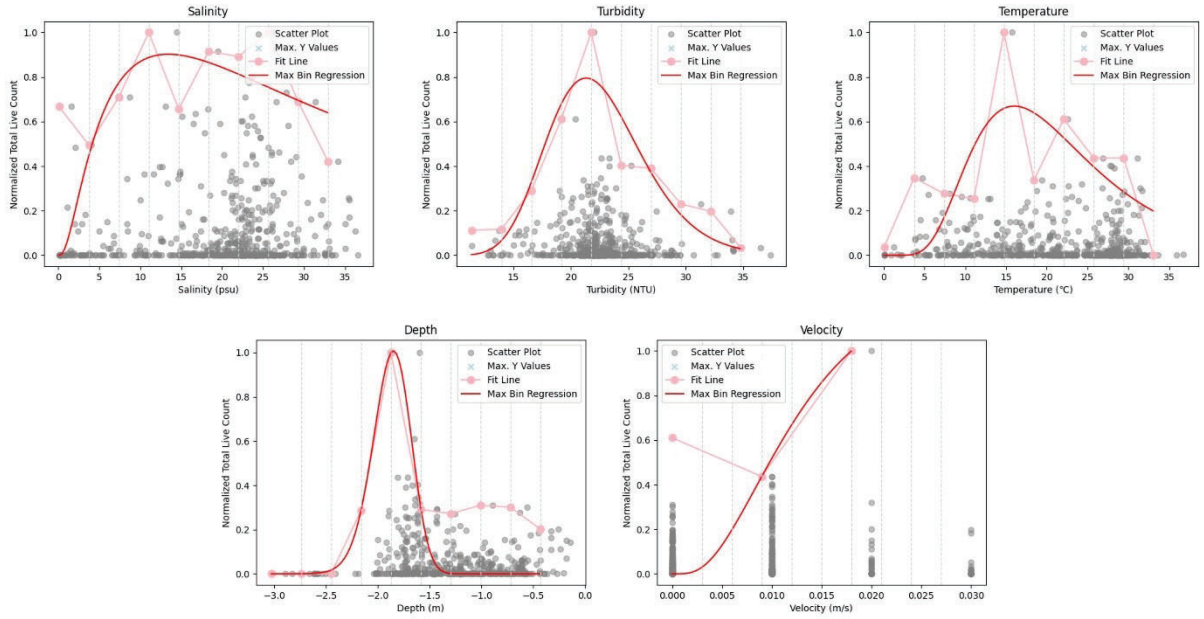

Figure 20: Fitted Max-Bin regression models, depicting the relationship between physical parameters (salinity, turbidity, temperature, depth, and velocity) and normalized total live counts.

Table 13 shows statistics for habitat percent coverage generated by OHSI2 models over Matagorda Bay, on both a monthly and an annual basis. These models identified spatially diverse habitat scores and identified similar spatial and temporal trends (Figure 21). $OHSI2_{uw}$ identified the most optimal habitat in the winter. The summer months and early fall had the greatest unsuitable habitat coverage. Optimal habitat was generally identified near freshwater sources, along the trajectory of flow. These regions are generally shallow in depth, with moderate salinity, higher turbidity, and high flow. Unsuitable habitat is located in the

primary bay, where water is deeper, flow is slower, and salinity is higher. Unsuitable habitat is also located in the secondary bays when temperatures are warm (spring, summer).

$OHSI2_w$ identified greater coverage of optimal habitat than $OHSI2_{uw}$. $OHSI2_w$ identified the greatest coverage of optimal habitat in the winter months, specifically in secondary bays and along the boundaries of the primary bay. These areas identified as optimal are generally shallower regions, where flow is high, and salinity is low due to influence from river discharges. Unsuitable habitat coverage was greatest during the summer months and was primarily found throughout the secondary bays and in the center of the primary bay. The secondary bays are identified as unsuitable during the summer as temperatures rise beyond oyster tolerance, according to the Max-Bin model. The primary bay is unsuitable as this area is deeper, has less flow, and experiences temperature rises during the summer.

Table 13: Monthly and annual habitat coverage produced by the OHSI2 models (OHSI2uw and OHSI2w), which identified optimal, moderate, and unsuitable habitat.

Month	$OHSI2_{uw}$			$OHSI2_w$		
	Opt. (%)	Mod. (%)	Unsuit. (%)	Opt. (%)	Mod. (%)	Unsuit. (%)
1	44.45	41.67	13.88	71.71	28.29	0.00
2	51.58	42.63	5.79	85.30	14.70	0.00
3	37.34	37.04	25.62	54.50	43.07	2.43
4	30.98	32.49	36.53	43.41	44.78	11.80
5	26.74	27.03	46.23	35.59	39.73	24.67
6	25.64	26.13	48.23	33.54	36.01	30.45
7	25.64	25.55	48.81	33.68	34.63	31.69
8	26.44	26.15	47.40	34.61	37.23	28.16
9	26.02	27.54	46.45	34.57	38.83	26.60
10	29.22	32.32	38.46	40.50	45.40	14.10
11	31.48	33.75	34.77	45.67	45.64	8.69
12	41.91	40.91	17.18	65.63	33.98	0.39
Ann.	32.27	33.32	34.42	46.05	53.95	0.00

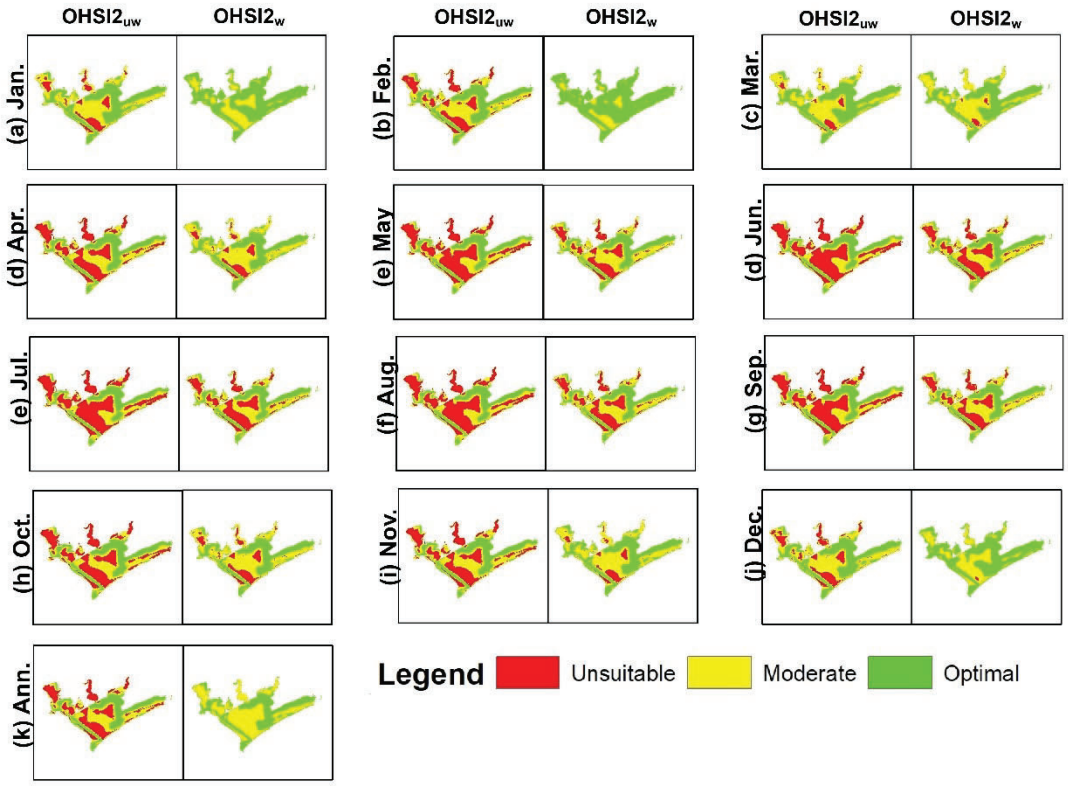

Figure 21: Monthly (a) – (j) and annual (k) scores produced by the OHSI2uw and OHSI2w models. Red indicates unsuitable habitat, yellow indicates moderate, and green is optimal habitat.

Figure 22 illustrates the spatial correlation between the total live count (Figure 22a, 22d), total spat count (Figure 22b, 22e), and total dead count (Figure 22c, 22f) and the annual $OHSI2_{uw}$ and $OHSI2_{w}$ scores. Oyster samples were collected over a range of habitat qualities, as identified by these models. Live and spat counts over optimal habitat identified by $OHSI2_{uw}$ were generally higher than those located over moderate and unsuitable habitat. Dead count was greatest in Lavaca Bay where monthly $OHSI2_{uw}$ maps illustrate year-round unsuitable conditions. For the $OHSI2_{w}$, live and spat count were generally greatest in locations that are year-round optimal habitat. However, both live and spat count had samples of high count and poor habitat quality while some locations of low oyster count were present in optimal habitat quality. Dead count is greatest in locations where monthly maps illustrate unsuitable conditions, specifically in the summer and fall.

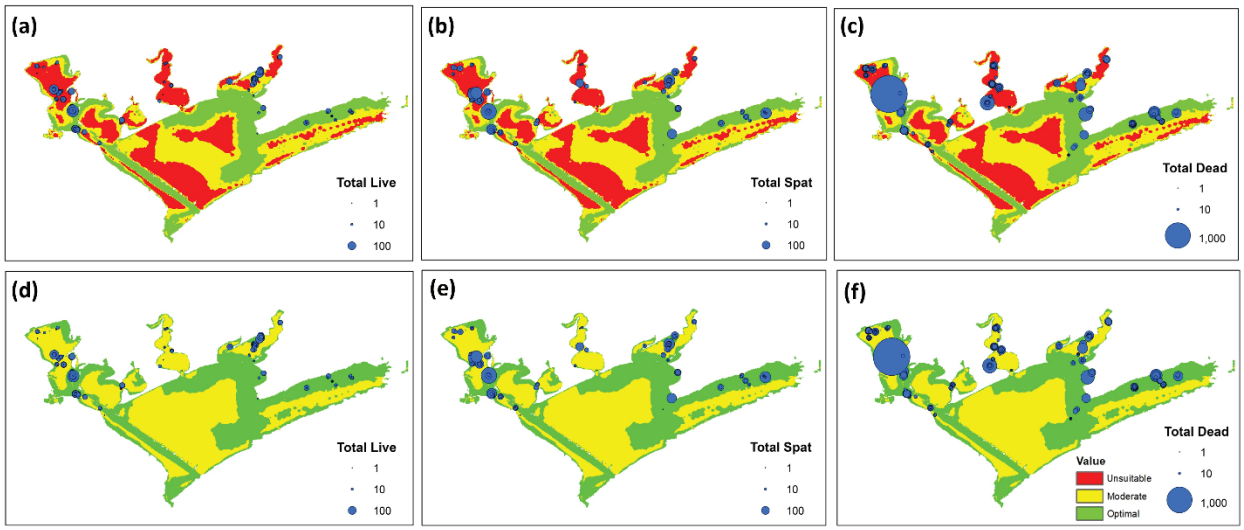

Figure 22: Oyster samples (blue), including total live (a), (d), total spat (b), ©, and total dead (c), (f) over the annual score maps produced by $OHSI2_{uw}$ (a)–(c) and $OHSI2_{w}$ (d)–(f), where red indicates unsuitable habitat, yellow is moderate habitat, and green is unsuitable habitat.

2.6.7. **OHSI3 Outputs**

Figure 23 and Table 14 present the statistical measures for the optimal ML models used to generate the OHSI3. Among the diverse ML families, a trend of closely competitive performance emerges. However, DRF stands out as the most effective family for salinity simulations. The optimal DRF achieved a training performance of NRMSE: 0.53 ± 0.04, *r*: 0.92 ± 0.03, NSE: 0.72 ± 0.05, and a testing performance of NRMSE: 0.85 ± 0.06, *r*: 0.56 ± 0.10, NSE: 0.27 ± 0.12. The relative importance for input variables were 37% for depth, 23% for salinity, 17% for turbidity, 16% for temperature, and 7% for velocity. Following DRF closely in performance, GBM ranks as the second-best performing family. The optimal GBM model, showcased training performance metrics of NRMSE: 0.68 ± 0.19, *r*: 0.79 ± 0.19, NSE: 0.53 ± 0.24, and testing performance of NRMSE: 0.87 ± 0.05, *r*: 0.54 ± 0.10. DNN presents a slightly weaker performance, being the second-least effective family. The optimal DNN model exhibited a training performance of NRMSE: 0.84 ± 0.17, *r*: 0.58 ± 0.23, NSE: 0.29 ± 0.26, and testing performance of NRMSE: 0.88 ± 0.09, *r*: 0.48 ± 0.12, NSE: 0.22 ± 0.21. Figure 24 presents scatterplots illustrating observed and modeled total live counts generated for the optimal model in each ML family during both training and testing phases. As depicted in Figure 24, the DRF model is performing well during training compared to other models.

The optimal DRF model was further implemented to predict live count over Matagorda Bay. Table 15 presents statistics measures of the monthly and annual live counts as well as habitat coverage depicted for

the monthly and annual datasets. Average total live count predicted by OHSI3 was highest in July (13.01 ± 3.65). This month had that highest optimal habitat coverage (Figure 25e). The low average total live count was 7.18 ± 2.20 in November. This month also had the lowest optimal habitat and highest unsuitable habitat coverage (Figure 25i). The OHSI3-generated maps (Figure 25) were more spatially and temporally complex than OHSI1 and OHSI2 maps. This model identified optimal regions in many locations throughout the bay. Optimal habitat was concentrated in secondary bays and in scattered regions throughout the primary bay. Regions of bay in contact with river mouths were identified as unsuitable. The deep shipping channel was generally identified as unsuitable habitat. Unsuitable habitat dominated (>50%) April (Figure 25d), May (Figure 25e), and November (Figure 25i) suitability maps, as the primary bay becomes unsuitable.

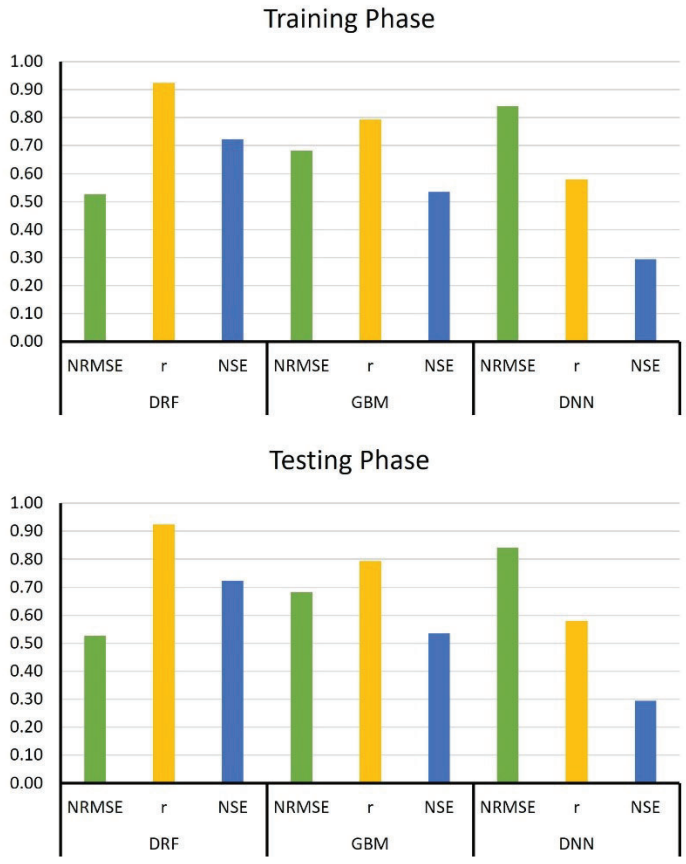

Figure 23: Performance measures for the optimal models produced from various ML model families (DRF, GBM, DNN) for the (a) training and (b) testing phases.

Table 14: Training and testing performance (average ± standard deviation) for the optimal model in each ML-based algorithm.

ML Algorithm	Statistical Measure	Training Phase	Testing Phase
DRF	NRMSE	0.53 ± 0.04	0.85 ± 0.06
	r	0.92 ± 0.03	0.56 ± 0.10
	NSE	0.72 ± 0.05	0.27 ± 0.12
GBM	NRMSE	0.68 ± 0.19	0.87 ± 0.05
	r	0.79 ± 0.19	0.54 ± 0.10
	NSE	0.53 ± 0.24	0.24 ± 0.11
DNN	NRMSE	0.84 ± 0.17	0.88 ± 0.09
	r	0.58 ± 0.23	0.48 ± 0.12
	NSE	0.29 ± 0.26	0.22 ± 0.21

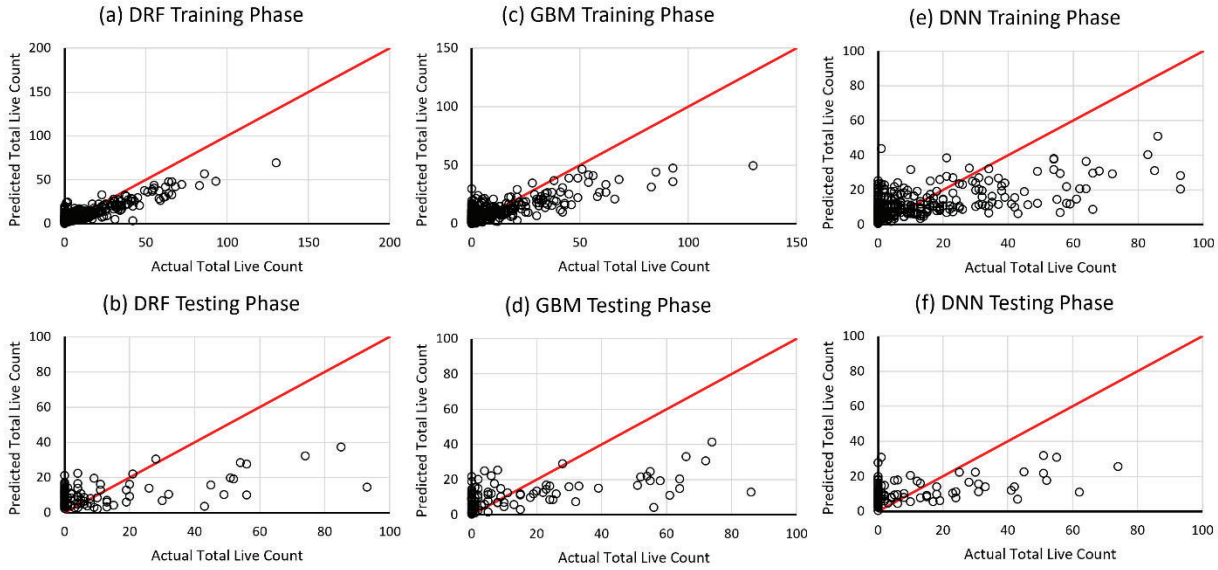

Figure 24: Predicted and actual total live count produced by the optimal models displayed in Table 14. Red lines indicate a 1:1 relationship.

Table 15: Statistics for monthly and annual total live counts and scores predicted by the OHSI3 over the entire bay.

Month	Avg. ± Std. Dev.	Min.	Max.	Opt. (%)	Mod. (%)	Unsuit. (%)
1	9.58 ± 3.52	3.28	29.27	48.93	13.82	37.24
2	9.34 ± 3.06	3.59	28.47	45.64	13.95	40.40
3	9.56 ± 3.11	3.25	34.77	53.31	13.80	32.89
4	8.12 ± 2.63	2.84	25.60	26.83	14.36	58.81
5	8.22 ± 2.69	2.39	26.37	29.13	12.32	58.55
6	10.11 ± 2.61	2.75	28.28	70.47	11.43	18.09
7	13.01 ± 3.65	2.79	29.12	85.42	2.90	11.68
8	9.48 ± 2.63	2.86	26.52	54.71	15.11	30.17
9	9.63 ± 3.36	2.97	32.53	47.27	19.07	33.67
10	8.98 ± 3.08	3.00	36.46	33.00	23.24	43.76
11	7.18 ± 2.20	2.97	28.99	13.89	6.34	79.76
12	9.40 ± 3.63	3.18	43.53	45.05	12.45	42.49
Ann.	8.61	2.78	29.83	31.97	33.28	34.75

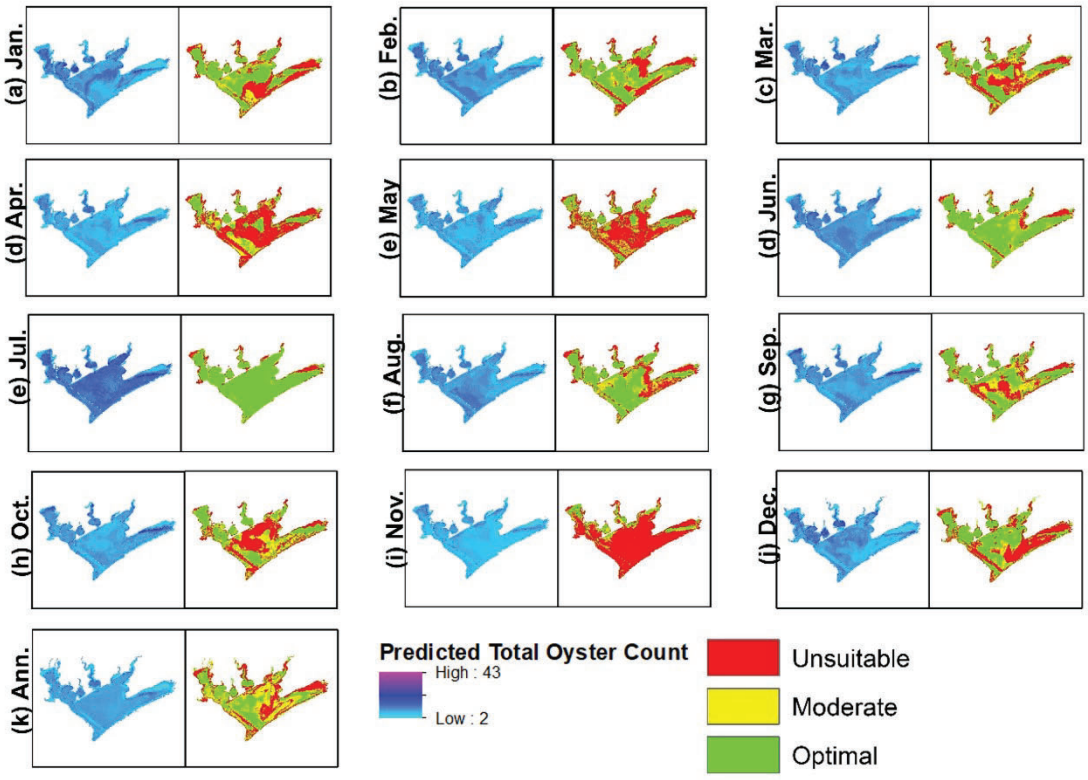

Figure 25: Monthly and annual total live counts and scores produced by the OHSI3 model.

2.7. **Discussion**

The HSI models quantify oyster habitat quality and could assist restoration efforts by providing essential tools for informed site selection. In this study, five OHSI models were developed for Matagorda Bay, primarily focusing on water quality, on both a monthly and an annual basis (2018–2023), using inputs with high spatial resolution (100 m). Input data for these models include salinity, turbidity, temperature, water depth, and velocity. Most OHSI models exhibited spatial and temporal variability. Optimal oyster habitat was generally observed in shallower regions, near freshwater sources with high water velocity. Unsuitable habitat was typically identified in the primary bay, where water depths are greater, water velocity is slower, and the influence from freshwater input is less significant. Each model highlighted a different season with maximum coverage of optimal conditions. OHSI1 indicated that spring and fall months were optimal for oysters, while winter was deemed unsuitable. OHSI2 identified winter months as optimal and deemed summer unsuitable. OHSI3 found the most suitable conditions in summer and determined that spring and late fall were unsuitable.

Remote-sensing datasets have proven to be particularly useful for analyzing spatial and temporal trends of biophysical parameters used to create OHSI models. These datasets are publicly available and exhibit high spatial resolution. However, the quality of the generated OHSI models is significantly controlled by the quality of these datasets. For example, the ECOSTRESS-derived temperature data used in this study demonstrated high correlation to in-situ temperature observations (*r:* 0.84). However, the salinity and turbidity datasets derived from Landsat-8 data show a relatively limited quality (salinity *r:* 0.49, turbidity *r:* 0.79) for several outstanding reasons. The quality of the salinity and turbidity data is function of the ML model performance used to retrieve them from Landsat-8 data. Large in-situ datasets are required to establish a robust ML model and significant and accurate relationships between in-situ salinity and turbidity observations and Landsat-8 surface reflectance data. Generating ML models from small datasets (salinity *n:* 478, turbidity *n:* 173) could lead to obvious bias and low model performance. The optimal ML models for salinity and turbidity show modest performance during the testing phase (salinity: NRMSE: 0.87 ± 0.06, *r*: 0.49 ± 0.09, NSE: 0.23 ± 0.12; turbidity: NRMSE: 0.63 ± 0.11, *r*: 0.79 ± 0.11, NSE: 0.60 ± 0.20). These modest performances affected the model generalization when applied to retrieve monthly salinity and turbidity data over Matagorda Bay. The generated salinity and turbidity encompass a narrow range of spatial variability. For example, the salinity model predicted only small ranges and events of low (<10 psu) and high (>20 psu) salinities could not correctly be predicted by that model. The turbidity model could only identify lower turbidities (<100 NTU). The model successfully failed to predict higher turbidities. Improved estimates of salinity and turbidity could be achieved by increasing the number of in-situ observations to build robust ML models, as well as by refining the model hyperparameters.

The performance of the OHSI models also depends on the number of model input parameters. The generated OHSI models are quite simple given the fact that oyster habitat scores cannot be adequately depicted through a select few physical parameters. Nevertheless, these OHSI models are useful for indicating regions with a likelihood of success based on long-term (2018–2023) water quality conditions. The OHSI models produced in this study focused only on salinity, turbidity, temperature, depth, and water velocity to depict oyster habitat suitability. OHSI produced in other studies included alternative or additional inputs. Substrate (Cake, 1983; Soniat & Brody, 1988; Theuerkauf & Lipcius, 2016) and chlorophyll-a (Cho et al., 2012; Snyder et al., 2017) are frequently used, as substrate and food availability are essential factors that mark habitat as optimal quality. Efforts were made to incorporate these parameters into this study's OHSI models; however, they were not considered due to lack of available data over Matagorda Bay. No substrate data were available. However, depth was selected to serve as a proxy for substrate. In Section I, chlorophyll-a data were retrieved Landsat-8 images using empirical models. However, these models were based on limited number of field observations (n = 17), which hinders their generalization. OHSI model improvements may be made by incorporating additional physical parameters, such as substrate type and chlorophyll-a, because these parameters are critical to oyster habitat suitability.

The performance of the OHSI models also depends on the model structure. OHSI1 is a simple overlay model, which might not be able to capture the complex relationship between the input parameters and the oyster habitat (Theuerkauf & Lipcius, 2016). The performance of this model is also a function of the threshold values used to define optimal, moderate, and unsuitable classes for each model input. These thresholds influence the timing of the maximum coverage of optimal and unsuitable habitat conditions over the entire bay. In addition, assigning weights to different inputs based on previous studies might be subjective and could introduce bias into the final habitat scores. On the other hand, many of the OHSI1 scores demonstrated little to no spatial variability. Small spatial variability could result from the input datasets being entirely classified as a single class. For example, the salinity inputs were nearly exclusively classified as optimal habitat over 99.41–100% of the bay. Turbidity inputs were almost entirely classified as unsuitable habitat over more than 90% of the bay. $OHSI1_{uw}$ and $OHSI_{w}$ scores differed greatly in the habitats they identified. $OHSI1_{uw}$ primarily depicted the bay as unsuitable habitat, with patches of moderate habitat. On the other hand, $OHSI_{w}$ depicted high percentages of optimal habitat, with many $OHSI1_{w}$ maps showing entirely optimal habitat. The nearly constant optimal habitat identified by this model is due to the heavy weight of the salinity inputs, which are classified entirely as optimal. This model only predicted some moderate habitat in half of the monthly images. Despite these differences between $OHSI1_{uw}$ and $OHSI1_{w}$, both models depicted higher quality habitat, either optimal or moderate, in secondary bays, near freshwater sources, and in areas of high flow velocity. Furthermore, both models identified lower quality habitat, either moderate or poor, in the primary bay where salinity is generally higher, water is deeper, and flow is lower.

The OHSI2 is a more complex approach to oyster habitat modeling, as it transformed the values of the input physical parameters (e.g., salinity, turbidity, temperature, water depth, and water velocity) into a SI. SI values established through the Max-Bin regression method reflect those published in previous literature (Turner et al., 2016). The model coefficients (*a*, *b*, and *c*) were generated from annual data and applied to monthly averages; therefore, this model limits monthly variability. Moreover, the model coefficients are function of the number of data points as well as the mathematical fitting function used to generate them. The higher the number of data points ($n = 567$), the more robust these parameters would be. More importantly, the values of these coefficients are significantly influenced by the spatial variability of the locations from which they were extracted (e.g., Figure 12b). Velocity, for example, showed small variability, only depicting velocity values at oyster sample locations through four speeds. The regression model struggled to depict a strong relationship between this velocity and the oyster total live count. The regression model would benefit from the addition of velocity observations that create more spatial variability. In addition, the performance of the OHSI2 model is also a function of the threshold values used to define optimal, moderate, and unsuitable scores. These thresholds influence the timing of the maximum coverage of optimal and unsuitable habitat conditions over the entire bay. OHSI2 models varied in depicted habitat quality. $OHSI2_{uw}$ identified a larger range of unsuitable, moderate, and optimal conditions. A greater coverage of optimal habitat was depicted by the $OHSI2_{w}$ model.

The OHSI3 model generated the most spatially complex habitat scores. These models are inherently less biased, because they do not rely on theories or relationships established in previous literature. They could be used to directly predict oyster abundance indicators such as live, spat, or dead counts. These ML models leverage advanced computational techniques to integrate complex environmental variables, identify suitable habitat areas, and inform targeted conservation strategies. Their ability to handle intricate relationships and adapt to evolving datasets makes them indispensable tools in promoting the sustainability and resilience of oyster populations and coastal ecosystems. However, these models require large datasets to adequately comprehend the intricate relationship between oysters and their environment. This is likely why many studies have not employed a similar ML-based approach to OHSI. ML provides a unique approach to modeling OHSI, as these algorithms can analyze complex, nonlinear relationships. However, model performance from the ML families could be enhanced by optimizing the model hyperparameters, using a larger dataset, and removing extreme outliers. Similar to OHSI2 models, the performance of the OHSI3 model is also a function of the threshold values used to define optimal, moderate, and unsuitable scores. These thresholds influence the timing of the maximum coverage of optimal and unsuitable habitat conditions over the entire bay. The DRF model has been highlighted for their exceptional performance in modeling total live counts. This algorithm is non-parametric, meaning the model does not make unjustified assumptions (Moral-García et al., 2020). Many habitat modeling studies have underscored the DRF as the

best method to depicting habitat quality from select parameters. However, DNN and GBM presented performance competitive as well. GBM is a linear model, however, was the second highest performing family. In addition, this model is non-parametric (Moral-García et al., 2020), which likely contributes to the similar training and testing performance between the optimal DRF and GBM models. The DNN family produced the OHSI model with the poorest training and testing performance. This family is considered much more powerful than other ML families due to its learning capabilities (Christin et al., 2019). The DRF and DNN families have been highlighted for their usefulness in ecological modeling (Christin et al., 2019; Simon et al., 2023).

Oyster habitats are influenced by a wide range of physical parameters. This study contributes to current methodologies for modeling OHSI by analyzing modeling approaches frequently used in OHSI studies and proposing a new method for modeling OHSI based on ML. These OHSI models are useful for identifying regions with a likelihood of success based on the parameters considered. The OHSI scores produced in this study indicate regions of greater and lesser likelihood of restoration success. High OHSI scores at specific locations indicate areas with long-term suitable water quality, while low OHSI scores depict areas with conditions unsuitable for oysters in the long term. The maps produced by this study can be used to identify appropriate sites for restoration and to select suitable sites effectively.

2.8. Conclusion

OHSI models provide insight on habitat quality and may greatly aid oyster restoration efforts. Various approaches have been made to simulate habitat quality for oysters and a wide range of environmental parameters have been used to depict habitat quality. This study generated and validated five OHSI models over Matagorda Bay, utilizing salinity, turbidity, temperature, depth, and water velocity as model parameters. The produced models depicted unsuitable, moderate, and optimal habitat in terms of water quality.

Each OHSI model depicted different seasons as optimal and unsuitable. OHSI1 models identified the greatest optimal habitat coverage in the spring and fall, while winter had the greatest unsuitable habitat coverage. OHSI2 identified winter months as optimal and deemed summer unsuitable. OHSI3 found the highest optimal habitat coverage in summer and determined that spring and late fall were unsuitable. Despite discrepancies between optimal and unsuitable months, models generally depicted shallower regions near freshwater sources with high water velocity as optimal habitat for oysters. The primary bay was generally depicted as unsuitable, as this area is deeper with slower water velocities and less influence from freshwater sources.

Our methodology offers a point of reference, a structured framework, and valuable insights for modeling OHSI. Each model generated in this study provides unique and insightful details about habitat quality in terms of water quality. The developed approach serves as a guide to enhance monitoring procedures for ecological modeling.

3. References

Aguilar, D. (2017). Salinity Disturbance affects Community Structure and Organic Matter on a Restored Crassostrea virginica Oyster Reef in Matagorda Bay, Texas [M.Sc., Texas A&M University - Corpus Christi]. https://www.proquest.com/docview/2003154305/abstract/17A27C0C77294BC8PQ/1

AL-Fahdawi, A., Rabee, A., & Al-Hirmizy, S. (2015). Water quality monitoring of Al-Habbaniyah Lake using remote sensing and in situ measurements. Environmental Monitoring and Assessment, 187(6), 367. https://doi.org/10.1007/s10661-015-4607-2

Allam, M., Yawar Ali Khan, M., & Meng, Q. (2020). Retrieval of Turbidity on a Spatio-Temporal Scale Using Landsat 8 SR: A Case Study of the Ramganga River in the Ganges Basin, India. Applied Sciences, 10(11), Article 11. https://doi.org/10.3390/app10113702

Ansari, M., & Akhoondzadeh, M. (2020). Mapping water salinity using Landsat-8 OLI satellite images (Case study: Karun basin located in Iran). Advances in Space Research, 65(5), 1490–1502. https://doi.org/10.1016/j.asr.2019.12.007

Arıman, S. (2021). Determination of inactive water quality variables by MODIS data: A case study in the Kızılırmak Delta-Balik Lake, Turkey. Estuarine, Coastal and Shelf Science, 260, 107505. https://doi.org/10.1016/j.ecss.2021.107505

Asgari, M., Yang, W., & Farnaghi, M. (2022). Spatiotemporal data partitioning for distributed random forest algorithm: Air quality prediction using imbalanced big spatiotemporal data on spark distributed framework. Environmental Technology & Innovation, 27, 102776. https://doi.org/10.1016/j.eti.2022.102776

Bagenda, D., Nishikawa, S., Kita, H., Kinai, Y., Terai, S., Kato, M., & Kasai, H. (2019). Impact of feeding on oyster depuration efficacy under conditions of high salinity and low temperature. Aquaculture, 500, 135–140. https://doi.org/10.1016/j.aquaculture.2018.10.009

Barnes, T., Volety, A., Chartier, K., Mazzotti, F., & Pearlstine, L. (2007). A Habitat Suitability Index Model for the Eastern Oyster (Crassostrea Virginica), a Tool for Restoration of the Caloosahatchee Estuary, Florida. Journal of Shellfish Research, 26(4), 949–959. https://doi.org/10.2983/0730-8000(2007)26[949:AHSIMF]2.0.CO;2

Bayati, M., & Danesh-Yazdi, M. (2021). Mapping the spatiotemporal variability of salinity in the hypersaline Lake Urmia using Sentinel-2 and Landsat-8 imagery. Journal of Hydrology, 595, 126032. https://doi.org/10.1016/j.jhydrol.2021.126032

Behmel, S., Damour, M., Ludwig, R., & Rodriguez, M. (2016). Water quality monitoring strategies—A review and future perspectives. Science of The Total Environment, 571, 1312–1329. https://doi.org/10.1016/j.scitotenv.2016.06.235

Binding, C., & Bowers, D. (2003). Measuring the salinity of the Clyde Sea from remotely sensed ocean colour. Estuarine, Coastal and Shelf Science, 57(4), 605–611. https://doi.org/10.1016/S0272-7714(02)00399-2

Bormudoi, A., Hinge, G., Nagai, M., Kashyap, M., & Talukdar, R. (2022). Retrieval of Turbidity and TDS of Deepor Beel Lake from Landsat 8 OLI Data by Regression and Artificial Neural Network. Water Conservation Science and Engineering, 7(4), 505–513. https://doi.org/10.1007/s41101-022-00158-9

Bowers, D., & Brett, H. (2008). The relationship between CDOM and salinity in estuaries: An analytical and graphical solution. Journal of Marine Systems, 73, 1–7. https://doi.org/10.1016/j.jmarsys.2007.07.001

Boyer, J., Kelble, C., Ortner, P., & Rudnick, D. (2009). Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. Ecological Indicators, 9(6, Supplement), S56–S67. https://doi.org/10.1016/j.ecolind.2008.11.013

Brody, S., Highfield, W., Arlikatti, S., Bierling, D., Ismailova, R., Lee, L., & Butzler, R. (2004). Conflict on the Coast: Using Geographic Information Systems to Map Potential Environmental Disputes in

Matagorda Bay, Texas. Environmental Management, 34(1), 11–25. https://doi.org/10.1007/s00267-004-0256-x

Brown, & Hartwick, E. (1988). A habitat suitability index model for suspended tray culture of the Pacific oyster, Crassostrea gigas Thunberg. Aquaculture Research, 19(2), 109–126. https://doi.org/10.1111/j.1365-2109.1988.tb00414.x

Buditama, G., Damayanti, A., & Giok Pin, T. (2017). Identifying Distribution of Chlorophyll-a Concentration Using Landsat 8 OLI on Marine Waters Area of Cirebon. IOP Conference Series: Earth and Environmental Science, 98, 012040. https://doi.org/10.1088/1755-1315/98/1/012040

Bugica, K., Sterba-Boatwright, B., & Wetz, M. (2020). Water quality trends in Texas estuaries. Marine Pollution Bulletin, 152, 110903. https://doi.org/10.1016/j.marpolbul.2020.110903

Cabral, P., Santos, J., & Augusto, G. (2011). Monitoring Urban Sprawl and the National Ecological Reserve in Sintra-Cascais, Portugal: Multiple OLS Linear Regression Model Evaluation. Journal of Urban Planning and Development, 137(3), 346–353. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000059

Caillier, J. (2023). An assessment of benthic condition in the Matagorda Bay System using a sediment quality triad approach. Diss. Texas A&M University-Corpus Christi.

Cake. (1983). Habitat Suitability Index Models: Gulf of Mexico American Oyster | U.S. Geological Survey. U.S. Geological Survey. https://www.usgs.gov/publications/habitat-suitability-index-models-gulf-mexico-american-oyster

Campbell, & Hall, S. (2019). Hydrodynamic effects on oyster aquaculture systems: A review. Reviews in Aquaculture, 11(3), 896–906. https://doi.org/10.1111/raq.12271

Carpenter, D., & Carpenter, S. (1983). Modeling inland water quality using Landsat data. Remote Sensing of Environment, 13(4), 345–352. https://doi.org/10.1016/0034-4257(83)90035-4

Casas, S., Lavaud, R., La Peyre, M., Comeau, L., Filgueira, R., & La Peyre, J. (2018). Quantifying salinity and season effects on eastern oyster clearance and oxygen consumption rates. Marine Biology, 165(5), 90. https://doi.org/10.1007/s00227-018-3351-x

Cassis, D., Pearce, C., & Maldonado, M. (2011). Effects of the environment and culture depth on growth and mortality in juvenile Pacific oysters in the Strait of Georgia, British Columbia. Aquaculture Environment Interactions, 1, 259–274. https://doi.org/10.3354/aei00025

Chapman, Rodriguez-Perez, A., Hugh-Jones, T., Bromley, C., James, M., Diele, K., & Sanderson, W. (2021). Optimising recruitment in habitat creation for the native European oyster (Ostrea edulis): Implications of temporal and spatial variability in larval abundance. Marine Pollution Bulletin, 170, 112579. https://doi.org/10.1016/j.marpolbul.2021.112579

Chávez-Villalba, J., Arreola-Lizárraga, A., Burrola-Sánchez, S., & Hoyos-Chairez, F. (2010). Growth, condition, and survival of the Pacific oyster Crassostrea gigas cultivated within and outside a subtropical lagoon. Aquaculture, 300(1), 128–136. https://doi.org/10.1016/j.aquaculture.2010.01.012

Cho, Y., Lee, W., Hong, S., Kim, H., & Kim, J. (2012). GIS-based suitable site selection using habitat suitability index for oyster farms in Geoje-Hansan Bay, Korea. Ocean & Coastal Management, 56, 10–16. https://doi.org/10.1016/j.ocecoaman.2011.10.009

Chowdhury, M., Wijsman, J., Hossain, M., Ysebaert, T., & Smaal, A. (2019). A verified habitat suitability model for the intertidal rock oyster, Saccostrea cucullata. PLOS ONE, 14(6), e0217688. https://doi.org/10.1371/journal.pone.0217688

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. Methods in Ecology and Evolution, 10(10), 1632–1644. https://doi.org/10.1111/2041-210X.13256

Culbertson, J. (2008). Spatial and temporal patterns of eastern oyster (Crassostrea virginica) populations and their relationships to Dermo (Perkinsus marinus) infection and freshwater inflows in West Matagorda Bay, Texas [Ph.D., Texas A&M University]. https://www.proquest.com/docview/304177056/abstract/CF3A06095F2B425BPQ/1

Danbara. (2014). Deriving Water Quality Indicators of Lake Tana, Ethiopia, From Landsat-8. https://doi.org/10.13140/RG.2.1.2998.7282

Davies-Colley, R., & Smith, D. (2001). Turbidity Suspended Sediment, and Water Clarity: A Review1. JAWRA Journal of the American Water Resources Association, 37(5), 1085–1101. https://doi.org/10.1111/j.1752-1688.2001.tb03624.x

De Santiago, K., Palmer, T., Dumesnil, M., & Pollack, J. (2019). Rapid development of a restored oyster reef facilitates habitat provision for estuarine fauna. Restoration Ecology, 27(4), 870–880. https://doi.org/10.1111/rec.12921

Dekshenieks, M., Hofmann, E., & Powell, E. (1993). Environmental Effects on the Growth and Development of Eastern Oyster, Crassostrea virginica (Gmelin, 1791), Larvae: A Modeling Study. Journal of Shellfish Research, 12(2). https://digitalcommons.odu.edu/ccpo_pubs/79

Ding, H., & Elmore, A. (2015). Spatio-temporal patterns in water surface temperature from Landsat time series data in the Chesapeake Bay, U.S.A. Remote Sensing of Environment, 168, 335–348. https://doi.org/10.1016/j.rse.2015.07.009

Du, J., Park, K., Jensen, C., Dellapenna, T., Zhang, W., & Shi, Y. (2021). Massive oyster kill in Galveston Bay caused by prolonged low-salinity exposure after Hurricane Harvey. Science of The Total Environment, 774, 145132. https://doi.org/10.1016/j.scitotenv.2021.145132

Dunne, R. (1999). Spectrophotometric measurement of chlorophyll pigments: A comparison of conventional monochromators and a reverse optic diode array design. Marine Chemistry, 66(3), 245–251. https://doi.org/10.1016/S0304-4203(99)00035-3

Ehrich, & Harris, L. (2015). A review of existing eastern oyster filtration rate models. Ecological Modelling, 297, 201–212. https://doi.org/10.1016/j.ecolmodel.2014.11.023

Elangovan, A., & Murali, V. (2020). Mapping the chlorophyll-a concentrations in hypereutrophic Krishnagiri Reservoir (India) using Landsat 8 Operational Land Imager. Lakes & Reservoirs: Research & Management, 25(4), 377–387. https://doi.org/10.1111/lre.12346

El-Zeiny, A., & El-Kafrawy, S. (2017). Assessment of water pollution induced by human activities in Burullus Lake using Landsat 8 operational land imager and GIS. The Egyptian Journal of Remote Sensing and Space Science, 20, S49–S56. https://doi.org/10.1016/j.ejrs.2016.10.002

Fallatah, O., Ahmed, M., Gyawali, B., & Alhawsawi, A. (2022). Factors controlling groundwater radioactivity in arid environments: An automated machine learning approach. Science of The Total Environment, 830, 154707. https://doi.org/10.1016/j.scitotenv.2022.154707

Fang, L., Chen, S., Wang, H., Qian, J., & Zhang, L. (2010). Detecting marine intrusion into rivers using EO-1 ALI satellite imagery: Modaomen Waterway, Pearl River Estuary, China. International Journal of Remote Sensing, 31, 4125–4146. https://doi.org/10.1080/01431160903229218

Fisher, Lee, B., Purdy, A., Halverson, G., Dohlen, M., Cawse-Nicholson, K., Wang, A., Anderson, R., Aragon, B., Arain, M., Baldocchi, D., Baker, J., Barral, H., Bernacchi, C., Bernhofer, C., Biraud, S., Bohrer, G., Brunsell, N., Cappelaere, B., … Hook, S. (2020). ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resources Research, 56(4), e2019WR026058. https://doi.org/10.1029/2019WR026058

Fisheries, U. S. B. of, & Moore, H. F. (1907). Survey of Oyster Bottoms in Matagorda Bay, Texas. U.S. Government Printing Office.

Garcia-Quintas, A., Roy, A., Barbraud, C., Demarcq, H., Denis, D., & Lanco Bertrand, S. (2023). Machine and deep learning approaches to understand and predict habitat suitability for seabird breeding. Ecology and Evolution, 13(9), e10549. https://doi.org/10.1002/ece3.10549

Garrison. (2023). Home is Where the Habitat Is: Modeling Shortfin Make Habitat Suitability Via Machine Learning [University of Rhode Island]. https://doi.org/10.23860/thesis-garrison-julian-2023

Georgiades, P., Proestos, Y., Lelieveld, J., & Erguler, K. (2023). Machine Learning Modeling of Aedes albopictus Habitat Suitability in the 21st Century. Insects, 14(5), Article 5. https://doi.org/10.3390/insects14050447

Goddijn-Murphy, L., Dailloux, D., White, M., & Bowers, D. (2009). Fundamentals of in Situ Digital Camera Methodology for Water Quality Monitoring of Coast and Ocean. Sensors, 9(7), Article 7. https://doi.org/10.3390/s90705825

González, L., Torres, F., Rodríguez, C., Torregroza, A., & Sandoval, J. (2018). Estimation of water quality parameters using Landsat 8 images: Application to Playa Colorada Bay, Sinaloa, Mexico. Applied Geomatics, 10(2), 147–158. https://doi.org/10.1007/s12518-018-0211-9

González-Márquez, L., Torres-Bejarano, F., Torregroza-Espinosa, A., Hansen-Rodríguez, I., & Rodríguez-Gallegos, H. (2018). Use of LANDSAT 8 images for depth and water quality assessment of El Guájaro reservoir, Colombia. Journal of South American Earth Sciences, 82, 231–238. https://doi.org/10.1016/j.jsames.2018.01.004

Gorokhovich, Y., Cawse-Nicholson, K., Papadopoulos, N., & Oikonomou, D. (2022). Use of ECOSTRESS data for measurements of the surface water temperature: Significance of data filtering in accuracy assessment. Remote Sensing Applications: Society and Environment, 26, 100739. https://doi.org/10.1016/j.rsase.2022.100739

Grabowski, Brumbaugh, R., Conrad, R., Keeler, A., Opaluch, J., Peterson, C., Piehler, M., Powers, S., & Smyth, A. (2012). Economic Valuation of Ecosystem Services Provided by Oyster Reefs. BioScience, 62(10), 900–909. https://doi.org/10.1525/bio.2012.62.10.10

Grabowski, J., & Peterson, C. (2011). Ecosystem Engineers: Plants to Protists. Academic Press.

Guo, H., Huang, J., Chen, B., Guo, X., & Singh, V. (2021). A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery. International Journal of Remote Sensing, 42(5), 1841–1866. https://doi.org/10.1080/01431161.2020.1846222

Guo, H., Tian, S., Jeanne Huang, J., Zhu, X., Wang, B., & Zhang, Z. (2022). Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. ISPRS Journal of Photogrammetry and Remote Sensing, 183, 451–469. https://doi.org/10.1016/j.isprsjprs.2021.11.023

Haby, M. (2012). A review of Palacios shrimp landings, Matagorda Bay oyster resources and statewide economic impacts from the Texas seafood supply chain and saltwater sportfishing. https://repository.library.noaa.gov/view/noaa/43585

He, Y., Jin, S., & Shang, W. (2021). Water Quality Variability and Related Factors along the Yangtze River Using Landsat-8. Remote Sensing, 13(12), Article 12. https://doi.org/10.3390/rs13122241

Hossain, A., Mathias, C., & Blanton, R. (2021). Remote Sensing of Turbidity in the Tennessee River Using Landsat 8 Satellite. Remote Sensing, 13(18), Article 18. https://doi.org/10.3390/rs13183785

Hossen, H., Mahmod, W., Negm, A., & Nakamura, T. (2022). Assessing Water Quality Parameters in Burullus Lake Using Sentinel-2 Satellite Images. Water Resources, 49(2), 321–331. https://doi.org/10.1134/S0097807822020087

Hu, C., Chen, Z., Clayton, T., Swarzenski, P., Brock, J., & Muller–Karger, F. (2004). Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. Remote Sensing of Environment, 93(3), 423–441. https://doi.org/10.1016/j.rse.2004.08.007

Huang, W., Mukherjee, D., & Chen, S. (2011). Assessment of Hurricane Ivan impact on chlorophyll-a in Pensacola Bay by MODIS 250m remote sensing. Marine Pollution Bulletin, 62(3), 490–498. https://doi.org/10.1016/j.marpolbul.2010.12.010

Ighalo, J., & Adeniyi, A. (2020). A comprehensive review of water quality monitoring and assessment in Nigeria. Chemosphere, 260, 127569. https://doi.org/10.1016/j.chemosphere.2020.127569

Jakovljević, G., Govedarica, M., & Álvarez-Taboada, F. (2018). Assessment of biological and physic chemical water quality parameters using Landsat 8 time series. Remote Sensing for Agriculture, Ecosystems, and Hydrology XX, 10783, 349–361. https://doi.org/10.1117/12.2513277

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Kannel, P., Lee, S., Lee, Y., Kanel, S., & Khan, S. (2007). Application of Water Quality Indices and Dissolved Oxygen as Indicators for River Water Classification and Urban Impact Assessment. Environmental Monitoring and Assessment, 132(1), 93–110. https://doi.org/10.1007/s10661-006-9505-1

Kasprzak, P., Padisák, J., Koschel, R., Krienitz, L., & Gervais, F. (2008). Chlorophyll a concentration across a trophic gradient of lakes: An estimator of phytoplankton biomass? Limnologica, 38(3), 327–338. https://doi.org/10.1016/j.limno.2008.07.002

Kaufman, Y., Tanré, D., Gordon, H., Nakajima, T., Lenoble, J., Frouin, R., Grassl, H., Herman, B., King, M., & Teillet, P. (1997). Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect. Journal of Geophysical Research: Atmospheres, 102(D14), 16815–16830. https://doi.org/10.1029/97JD01496

Khorram, S. (1985). Development of water quality models applicable throughout the entire San Francisco Bay and Delta. Photogrammetric Engineering and Remote Sensing. https://www.academia.edu/105928692/Development_of_water_quality_models_applicable_throughout_the_entire_San_Francisco_Bay_and_Delta

Kim, & Montagna, P. (2009). Implications of Colorado river (Texas, USA) freshwater inflow to benthic ecosystem dynamics: A modeling study. Estuarine, Coastal and Shelf Science, 83(4), 491–504. https://doi.org/10.1016/j.ecss.2009.04.033

Kim, Son, S., Kim, Y., Khim, J., Nam, J., Chang, W., Lee, J., Lee, C., & Ryu, J. (2017). Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. Marine Pollution Bulletin, 121(1), 425–434. https://doi.org/10.1016/j.marpolbul.2017.05.026

Kinsey, J., & Montagna, P. (2005). Response of Benthic Organisms to External Conditions in Matagorda Bay. https://doi.org/10.26153/tsw/8998

Krishnaraj, A., & Honnasiddaiah, R. (2022). Remote sensing and machine learning based framework for the assessment of spatio-temporal water quality in the Middle Ganga Basin. Environmental Science and Pollution Research, 29(43), 64939–64958. https://doi.org/10.1007/s11356-022-20386-9

Kucera, C., Faulk, C., & Holt, G. (2002). The effect of spawning salinity on eggs of spotted seatrout (Cynoscion nebulosus, Cuvier) from two bays with historically different salinity regimes. Journal of Experimental Marine Biology and Ecology, 272(2), 147–158. https://doi.org/10.1016/S0022-0981(02)00081-3

Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H., Kampel, M., Richey, J., Stadler, P., Crawford, J., Striegl, R., Vermote, E., Pahlevan, N., & Butman, D. (2019). Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. Remote Sensing of Environment, 224, 104–118. https://doi.org/10.1016/j.rse.2019.01.023

Kumar, A., & Dua, A. (2009). Water quality index for assessment of water quality of river ravi at Madhopur (India). Global Journal of Environmental Sciences, 8(1), Article 1. https://doi.org/10.4314/gjes.v8i1.50824

LaPeyre, M., Serra, K., Joyner, T., & Humphries, A. (2015). Assessing shoreline exposure and oyster habitat suitability maximizes potential success for sustainable shoreline protection using restored oyster reefs. PeerJ, 3. https://doi.org/10.7717/peerj.1317

Lavallin, A. (2021). Machine Learning for Species Habitat Analysis.

Lavery, P., Pattiaratchi, C., Wyllie, A., & Hick, P. (1993). Water quality monitoring in estuarine waters using the landsat thematic mapper. Remote Sensing of Environment, 46(3), 268–280. https://doi.org/10.1016/0034-4257(93)90047-2

LCRA. (2022). 2022 Guidance for Assessing and Reporting Surface Water Quality in Texas.

LCRA. (2023). Water Quality Parameters—LCRA - Energy, Water, Community. https://www.lcra.org/services/els/water-quality-parameters/

Lebreton, B., Pollack, J. B., Blomberg, B., Palmer, T. A., Montagna, P. A., & Nevels, M. (2021). Oyster growth across a salinity gradient in a shallow, subtropical Gulf of Mexico estuary. Experimental Results, 2, e10. https://doi.org/10.1017/exp.2020.72

Lenihan, H. (1999). Physical–Biological Coupling on Oyster Reefs: How Habitat Structure Influences Individual Performance. Ecological Monographs, 69(3), 251–275. https://doi.org/10.1890/0012-9615(1999)069[0251:PBCOOR]2.0.CO;2

Li, N., Ning, Z., Chen, M., Wu, D., Hao, C., Zhang, D., Bai, R., Liu, H., Chen, X., Li, W., Zhang, W., Chen, Y., Li, Q., & Zhang, L. (2022). Satellite and Machine Learning Monitoring of Optically Inactive Water Quality Variability in a Tropical River. Remote Sensing, 14(21), Article 21. https://doi.org/10.3390/rs14215466

Lim, J., & Choi, M. (2015). Assessment of water quality based on Landsat 8 operational land imager associated with human activities in Korea. Environmental Monitoring and Assessment, 187(6), 384. https://doi.org/10.1007/s10661-015-4616-1

Linhoss, A., & Mickle, P. (2022). A Field Validated Model of Temporal Variability in Oyster Habitat Suitability. Frontiers in Marine Science, 9. https://www.frontiersin.org/articles/10.3389/fmars.2022.778936

Liu, D., Yu, S., Cao, Z., Qi, T., & Duan, H. (2021). Process-oriented estimation of column-integrated algal biomass in eutrophic lakes by MODIS/Aqua. International Journal of Applied Earth Observation and Geoinformation, 99, 102321. https://doi.org/10.1016/j.jag.2021.102321

Liu, & Wang. (2019). Modelling Reservoir Turbidity Using Landsat 8 Satellite Imagery by Gene Expression Programming. Water, 11(7), Article 7. https://doi.org/10.3390/w11071479

Luckenbach, M., Mann, R., & Wesson, J. (1999). Oyster Reef Habitat Restoration: A synopsis and synthesis of approaches; proceedings from the symposium, Williamsburg, Virginia, April 1995. Reports. https://doi.org/10.21220/V5NK51

Markogianni, V., Kalivas, D., Petropoulos, G., & Dimitriou, E. (2017). Analysis on the Feasibility of Landsat 8 Imagery for Water Quality Parameters Assessment in an Oligotrophic Mediterranean Lake. International Journal of Geological and Environmental Engineering, 11(9), 906–914.

Markogianni, V., Kalivas, D., Petropoulos, G., & Dimitriou, E. (2018). An Appraisal of the Potential of Landsat 8 in Estimating Chlorophyll-a, Ammonium Concentrations and Other Water Quality Indicators. Remote Sensing, 10(7), Article 7. https://doi.org/10.3390/rs10071018

Marshall, Casas, S., Walton, W., Rikard, F., Palmer, T., Breaux, N., La Peyre, M., Beseres Pollack, J., Kelly, M., & La Peyre, J. (2021). Divergence in salinity tolerance of northern Gulf of Mexico eastern oysters under field and laboratory exposure. Conservation Physiology, 9(1), coab065. https://doi.org/10.1093/conphys/coab065

Marshall, D., Lebreton, B., Palmer, T., De Santiago, K., & Beseres Pollack, J. (2019). Salinity disturbance affects faunal community composition and organic matter on a restored Crassostrea virginica oyster reef. Estuarine, Coastal and Shelf Science, 226, 106267. https://doi.org/10.1016/j.ecss.2019.106267

Masocha, M., Dube, T., Nhiwatiwa, T., & Choruma, D. (2018). Testing utility of Landsat 8 for remote assessment of water quality in two subtropical African reservoirs with contrasting trophic states. Geocarto International, 33(7), 667–680. https://doi.org/10.1080/10106049.2017.1289561

Masocha, M., Mungenge, C., & Nhiwatiwa, T. (2018). Remote sensing of nutrients in a subtropical African reservoir: Testing utility of Landsat 8. Geocarto International, 33(5), 458–469. https://doi.org/10.1080/10106049.2016.1265596

Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep Learning Techniques: An Overview. In A. E. Hassanien, R. Bhatnagar, & A. Darwish (Eds.), Advanced Machine Learning Technologies and Applications (pp. 599–608). Springer. https://doi.org/10.1007/978-981-15-3383-9_54

Mathew, Srinivasa Rao, N., & Mandla, V. (2017). Development of regression equation to study the Total Nitrogen, Total Phosphorus and Suspended Sediment using remote sensing data in Gujarat and Maharashtra coast of India. Journal of Coastal Conservation, 21(6), 917–927. https://doi.org/10.1007/s11852-017-0561-1

Matsumoto, J., Guthrie, C., Crockett, D., & McEwen, T. (2014). TxBLEND Model Extension and Salinity Validation for the Sabine-Neches Estuary: Extending Simulations Through 2013.

McBride, M. R. (2022). Influence of Colorado River Discharge Variability on Phytoplankton Communities in Matagorda Bay, Texas [M.Sc., Texas A&M University - Corpus Christi]. https://www.proquest.com/docview/2723515236/abstract/D6BB98A2751D4A3CPQ/1

Misaghi, F., Delgosha, F., Razzaghmanesh, M., & Myers, B. (2017). Introducing a water quality index for assessing water for irrigation purposes: A case study of the Ghezel Ozan River. Science of The Total Environment, 589, 107–116. https://doi.org/10.1016/j.scitotenv.2017.02.226

Mishra, A., Khali, H., Singh, S., Pande, C., Singh, R., & Chaurasia, S. (2021). An Assessment of In-situ Water Quality Parameters and its variation with Landsat 8 Level 1 Surface Reflectance datasets. International Journal of Environmental Analytical Chemistry, 0(0), 1–23. https://doi.org/10.1080/03067319.2021.1954175

Misra, A., Chapron, B., Nouguier, F., Ramakrishnan, B., & Yurovskaya, M. (2018). Sun-glint imagery of Landsat 8 for ocean surface waves. Remote Sensing of the Open and Coastal Ocean and Inland Waters, 10778, 86–94. https://doi.org/10.1117/12.2324754

Mondejar, J., & Tongco, A. (2019). Near infrared band of Landsat 8 as water index: A case study around Cordova and Lapu-Lapu City, Cebu, Philippines. Sustainable Environment Research, 29(1), 16. https://doi.org/10.1186/s42834-019-0016-5

Montagna, P. (1994). Inflow needs assessment: Effect of the Colorado River diversion on benthic communities. https://doi.org/10.15781/T2H41JS4B

Moral-García, S., Mantas, C., Castellano, J., & Abellán, J. (2020). Non-parametric predictive inference for solving multi-label classification. Applied Soft Computing, 88, 106011. https://doi.org/10.1016/j.asoc.2019.106011

Motes, M., DePaola, A., Cook, D., Veazey, J., Hunsucker, J., Garthright, W., Blodgett, R., & Chirtel, S. (1998). Influence of Water Temperature and Salinity on Vibrio vulnificus in Northern Gulf and Atlantic Coast Oysters (Crassostrea virginica). Applied and Environmental Microbiology, 64(4), 1459–1465.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7. https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021

National Centers for Environmental. (1998). Matagorda Bay, TX (G280) Bathymetric Digital Elevation Model (30 meter resolution) Derived From Source Hydrographic Survey Soundings Collected by NOAA. https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.dem:matagorda_bay_g280_30m

Nazeer, M., & Bilal, M. (2018). Evaluation of Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR) for Water Quality Monitoring: A Case Study for the Estimation of Salinity. Journal of Ocean University of China, 17, 305–310. https://doi.org/10.1007/s11802-018-3380-6

Nguyen, P., Koedsin, W., McNeil, D., & Van, T. (2018). Remote sensing techniques to predict salinity intrusion: Application for a data-poor area of the coastal Mekong Delta, Vietnam. International Journal of Remote Sensing, 39(20), 6676.

Olsen, Z. (2019). Quantifying Nursery Habitat Function: Variation in Habitat Suitability Linked to Mortality and Growth for Juvenile Black Drum in a Hypersaline Estuary. Marine and Coastal Fisheries, 11(1), 86–96. https://doi.org/10.1002/mcf2.10064

Onabule, Mitchell, S., & Couceiro, F. (2020). The effects of freshwater flow and salinity on turbidity and dissolved oxygen in a shallow Macrotidal estuary: A case study of Portsmouth Harbour. Ocean & Coastal Management, 191, 105179. https://doi.org/10.1016/j.ocecoaman.2020.105179

Osawa, T., Mitsuhashi, H., Uematsu, Y., & Ushimaru, A. (2011). Bagging GLM: Improved generalized linear model for the analysis of zero-inflated data. Ecological Informatics, 6(5), 270–275. https://doi.org/10.1016/j.ecoinf.2011.05.003

Oyebisi, S., & Alomayri, T. (2023). Artificial intelligence-based prediction of strengths of slag-ash-based geopolymer concrete using deep neural networks. Construction and Building Materials, 400, 132606. https://doi.org/10.1016/j.conbuildmat.2023.132606

Pahlevan, N., Schott, J., Franz, B., Zibordi, G., Markham, B., Bailey, S., Schaaf, C., Ondrusek, M., Greb, S., & Strait, C. (2017). Landsat 8 remote sensing reflectance (Rrs) products: Evaluations, intercomparisons, and enhancements. Remote Sensing of Environment, 190, 289–301. https://doi.org/10.1016/j.rse.2016.12.030

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M. K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., … Ruiz-Verdù, A. (2022). Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sensing of Environment, 270, 112860. https://doi.org/10.1016/j.rse.2021.112860

Palmer, T., Montagna, P., Pollack, J., Kalke, R., & DeYoe, H. (2011). The role of freshwater inflow in lagoons, rivers, and bays. Hydrobiologia, 667(1), 49–67. https://doi.org/10.1007/s10750-011-0637-0

Pekár, S., & Brabec, M. (2018). Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. Ethology, 124(2), 86–93. https://doi.org/10.1111/eth.12713

Peprah, M., & Mensah, I. (2017). Performance evaluation of the Ordinary Least Square (OLS) and Total Least Square (TLS) in adjusting field data: An empirical study on a DGPS data. South African Journal of Geomatics, 6(1), Article 1. https://doi.org/10.4314/sajg.v6i1.5

Pereira, L., Andes, L., Cox, A., & Ghulam, A. (2017). Measuring Suspended-Sediment Concentration and Turbidity in the Middle Mississippi and Lower Missouri Rivers Using Landsat Data. JAWRA Journal of the American Water Resources Association, 54. https://doi.org/10.1111/1752-1688.12616

Peterson, K., Sagan, V., & Sloan, J. (2020). Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. GIScience & Remote Sensing, 57(4), 510–525. https://doi.org/10.1080/15481603.2020.1738061

Peyre, M., Aguilar Marshall, D., Miller, L., & Humphries, A. (2019). Oyster Reefs in Northern Gulf of Mexico Estuaries Harbor Diverse Fish and Decapod Crustacean Assemblages: A Meta-Synthesis. Frontiers in Marine Science, 6. https://www.frontiersin.org/articles/10.3389/fmars.2019.00666

Peyre, M., Furlong, J., Brown, L., Piazza, B., & Brown, K. (2014). Oyster reef restoration in the northern Gulf of Mexico: Extent, methods and outcomes. Ocean & Coastal Management, 89, 20–28. https://doi.org/10.1016/j.ocecoaman.2013.12.002

Peyre, M., Marshall, D., & Sable, S. (2021). Oyster model inventory: Identifying critical data and modeling approaches to support restoration of oyster reefs in coastal U.S. Gulf of Mexico waters. In Open-File Report (2021–1063). U.S. Geological Survey. https://doi.org/10.3133/ofr20211063

Plutchak, R., Major, K., Cebrian, J., Foster, C., Miller, M., Anton, A., Sheehan, K., Heck, K., & Powers, S. (2010). Impacts of Oyster Reef Restoration on Primary Productivity and Nutrient Dynamics in Tidal Creeks of the North Central Gulf of Mexico. Estuaries and Coasts, 33(6), 1355–1364. https://doi.org/10.1007/s12237-010-9327-9

Pogoda, B., Hausen, T., Rothe, M., Bakker, F., Hauser, S., Colsoul, B., Dureuil, M., Krause, J., Heinicke, K., Pusch, C., Eisenbarth, S., Kreutle, A., Peter, C., & Pesch, R. (2023). Come, tell me how you live: Habitat suitability analysis for Ostrea edulis restoration. Aquatic Conservation: Marine and Freshwater Ecosystems, 33(7), 678–695. https://doi.org/10.1002/aqc.3928

Pollack, Cleveland, A., Palmer, T., Reisinger, A., & Montagna, P. (2012). A Restoration Suitability Index Model for the Eastern Oyster (Crassostrea virginica) in the Mission-Aransas Estuary, TX, USA. PLOS ONE, 7(7), e40839. https://doi.org/10.1371/journal.pone.0040839

Pollack, Kim, H., Morgan, E., & Montagna, P. (2011). Role of Flood Disturbance in Natural Oyster (Crassostrea virginica) Population Maintenance in an Estuary in South Texas, USA. Estuaries and Coasts, 34(1), 187–197. https://doi.org/10.1007/s12237-010-9338-6

Porter, E., Cornwell, J., & Sanford, L. (2004). Effect of oysters Crassostrea virginica and bottom shear velocity on benthic–pelagic coupling and estuarine water quality. Marine Ecology Progress Series, 271, 61–75.

Poshtegal, M., & Mirbagheri, S. (2023). Simulation and modelling of heavy metals and water quality parameters in the river. Scientific Reports, 13(1), Article 1. https://doi.org/10.1038/s41598-023-29878-1

Pu, F., Ding, C., Chao, Z., Yu, Y., & Xu, X. (2019). Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks. Remote Sensing, 11(14), Article 14. https://doi.org/10.3390/rs11141674

Puckett, Theuerkauf, S., Eggleston, D., Guajardo, R., Hardy, C., Gao, J., & Luettich, R. A. (2018). Integrating Larval Dispersal, Permitting, and Logistical Factors Within a Validated Habitat Suitability Index for Oyster Restoration. Frontiers in Marine Science, 5. https://doi.org/10.3389/fmars.2018.00076

Quang, Sasaki, J., Higa, H., & Huan, N. (2017). Spatiotemporal Variation of Turbidity Based on Landsat 8 OLI in Cam Ranh Bay and Thuy Trieu Lagoon, Vietnam. Water, 9(8), Article 8. https://doi.org/10.3390/w9080570

Rakocevic-Nedovic, J., & Hollert, H. (2005). Phytoplankton Community and Chlorophyll a as Trophic State Indices of Lake Skadar (Montenegro, Balkan) (7 pp). Environmental Science and Pollution Research, 12(3), 146–152. https://doi.org/10.1065/espr2005.04.241

Reddy, G. (2018). Spatial Data Management, Analysis, and Modeling in GIS: Principles and Applications | SpringerLink. Geospatial Technologies in Land Resources Mapping, Monitoring and Management. https://link.springer.com/chapter/10.1007/978-3-319-78711-4_7

Renaud, M., & Williams, J. (2023). Movements of Kemp's Ridley (Lepidochelys kempii) and Green (Chelonia mydas) Sea Turtles Using Lavaca Bay and Matagorda Bay.

Ropicki, A., Hanselka, R., Cummins, D., & Balboa, B. (2016). The Economic Impacts of Recreational Fishing in the Matagorda Bay System. https://repository.library.noaa.gov/view/noaa/43595

Santoso, C., Sakti, A., Al Faruqi, I., Adillah, K., Mahezs, F., & Suripto, S. (2023). Modeling habitat suitability for endemic Grizzled leaf monkey (Presbytis comata) using geospatial machine learning approach. Remote Sensing Applications: Society and Environment, 32, 101067. https://doi.org/10.1016/j.rsase.2023.101067

Schaeffer, B., Conmy, R., Duffy, A., Aukamp, J., Yates, D., & Craven, G. (2015). Northern Gulf of Mexico estuarine coloured dissolved organic matter derived from MODIS data. International Journal of Remote Sensing, 36(8), 2219–2237. https://doi.org/10.1080/01431161.2015.1035408

Schild, K., Hawley, R., Chipman, J., & Benn, D. (2017). Quantifying suspended sediment concentration in subglacial sediment plumes discharging from two Svalbard tidewater glaciers using Landsat-8 and in situ measurements. International Journal of Remote Sensing, 38(23), 6865–6881. https://doi.org/10.1080/01431161.2017.1365388

Schoenbaechler, C., Guthrie, C., Matsumoto, J., & Lu, Q. (2011). TxBLEND Model Calibration and Validation For the Laguna Madre Estuary.

Sent, G., Biguino, B., Favareto, L., Cruz, J., Sá, C., Dogliotti, A. I., Palma, C., Brotas, V., & Brito, A. (2021). Deriving Water Quality Parameters Using Sentinel-2 Imagery: A Case Study in the Sado Estuary, Portugal. Remote Sensing, 13(5), Article 5. https://doi.org/10.3390/rs13051043

Sharaf El Din, E. (2020). A novel approach for surface water quality modelling based on Landsat-8 tasselled cap transformation. International Journal of Remote Sensing, 41(18), 7186–7201. https://doi.org/10.1080/01431161.2020.1754497

Shareef. (2015). Estimation and characterization of physical and inorganic chemical indicators of water quality by using SAR images. https://doi.org/10.1117/12.2194503

Shareef, M., Khenchaf, A., & Toumi, A. (2016). Integration of passive and active microwave remote sensing to estimate water quality parameters. 2016 IEEE Radar Conference (RadarConf), 1–4. https://doi.org/10.1109/RADAR.2016.7485127

Sharma, S., Goff, J., Moody, R., Byron, D., Heck Jr., K., Powers, S., Ferraro, C., & Cebrian, J. (2016). Do restored oyster reefs benefit seagrasses? An experimental study in the Northern Gulf of Mexico. Restoration Ecology, 24(3), 306–313. https://doi.org/10.1111/rec.12329

Shi, J., & Hu, C. (2021). Evaluation of ECOSTRESS Thermal Data over South Florida Estuaries. Sensors, 21(13), Article 13. https://doi.org/10.3390/s21134341

Shrivastav, & Kumar, R. (2022). An Ensemble of Random Forest Gradient Boosting Machine and Deep Learning Methods for Stock Price Prediction. Journal of Information Technology Research (JITR), 15(1), 1–19. https://doi.org/10.4018/JITR.2022010102

Silva, G., Campos, D., Brasil, J., Tremblay, M., Mendiondo, E., & Ghiglieno, F. (2022). Advances in Technological Research for Online and In Situ Water Quality Monitoring—A Review. Sustainability, 14(9), Article 9. https://doi.org/10.3390/su14095059

Simon, S., Glaum, P., & Valdovinos, F. (2023). Interpreting random forest analysis of ecological models to move from prediction to explanation. Scientific Reports, 13(1), 3881. https://doi.org/10.1038/s41598-023-30313-8

Singh, A., Jakubowski, A., Chidister, I., & Townsend, P. (2013). A MODIS approach to predicting stream water quality in Wisconsin. Remote Sensing of Environment, 128, 74–86. https://doi.org/10.1016/j.rse.2012.10.001

Smith, R., Hogan, S., Tedford, K., Lusk, B., Reidenbach, M., & Castorani, M. (2022). Long-term data reveal greater intertidal oyster biomass in predicted suitable habitat. Marine Ecology Progress Series, 683, 221–226. https://doi.org/10.3354/meps13949

Snyder, J., Boss, E., Weatherbee, R., Thomas, A., Brady, D., & Newell, C. (2017). Oyster Aquaculture Site Selection Using Landsat 8-Derived Sea Surface Temperature, Turbidity, and Chlorophyll a. Frontiers in Marine Science, 4. https://doi.org/10.3389/fmars.2017.00190

Soniat, & Brody, M. (1988). Field validation of a habitat suitability index model for the American oyster. Estuaries, 11(2), 87–95. https://doi.org/10.2307/1351995

Southwell, Veenstra, J., Adams, C., Scarlett, E., & Payne, K. (2017). Changes in Sediment Characteristics upon Oyster Reef Restoration, NE Florida, USA. Journal of Coastal Zone Management, 20(1), 1–7. https://doi.org/10.4172/2473-3350.1000442

Stanley, Bilskie, M., Woodson, C., & Byers, J. (2024). A model for understanding the effects of flow conditions on oyster reef development and impacts to wave attenuation. Ecological Modelling, 489, 110627. https://doi.org/10.1016/j.ecolmodel.2024.110627

Strobl, R., & Robillard, P. (2008). Network design for water quality monitoring of surface freshwaters: A review. Journal of Environmental Management, 87(4), 639–648. https://doi.org/10.1016/j.jenvman.2007.03.001

Sudheer, K., Chaubey, I., & Garg, V. (2006). Lake Water Quality Assessment from Landsat Thematic Mapper Data Using Neural Network: An Approach to Optimal Band Combination Selection1. JAWRA Journal of the American Water Resources Association, 42(6), 1683–1695. https://doi.org/10.1111/j.1752-1688.2006.tb06029.x

Sun, A., Scanlon, B., Save, H., & Rateb, A. (2021). Reconstruction of GRACE Total Water Storage Through Automated Machine Learning. Water Resources Research, 57. https://doi.org/10.1029/2020WR028666

Tang, C., Luktarhan, N., & Zhao, Y. (2020). SAAE-DNN: Deep Learning Method on Intrusion Detection. Symmetry, 12(10), Article 10. https://doi.org/10.3390/sym12101695

TCEQ. (2012). Surface Water Quality Monitoring Procedures, Volume 1: Physical and Chemical Monitoring Methods.

TCEQ. (2023). Surface Water Quality Viewer. https://tceq.maps.arcgis.com/apps/webappviewer/index.html?id=b0ab6bac411a49189106064b70bbe778

Texas Secretary of State. (2023). Texas Administrative Code. https://texreg.sos.state.tx.us/public/readtac%24ext.TacPage?sl=T&app=9&p_dir=F&p_rloc=183310&p_tloc=29466&p_ploc=14656&pg=3&p_tac=&ti=30&pt=1&ch=290&rl=111

Theuerkauf, S. J., & Lipcius, R. N. (2016). Quantitative Validation of a Habitat Suitability Index for Oyster Restoration. Frontiers in Marine Science, 3. https://www.frontiersin.org/articles/10.3389/fmars.2016.00064

TinHan, T., Mohan, J., Dumesnil, M., DeAngelis, B., & Wells, R. (2018). Linking Habitat Use and Trophic Ecology of Spotted Seatrout (Cynoscion nebulosus) on a Restored Oyster Reef in a Subtropical Estuary. Estuaries and Coasts, 41(6), 1793–1805. https://doi.org/10.1007/s12237-018-0391-x

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. Remote Sensing, 8(8), Article 8. https://doi.org/10.3390/rs8080640

Torres-Bejarano, F., Arteaga-Hernández, F., Rodríguez-Ibarra, D., Mejía-Ávila, D., & González-Márquez, L. (2021). Water quality assessment in a wetland complex using Sentinel 2 satellite images. International Journal of Environmental Science and Technology, 18(8), 2345–2356. https://doi.org/10.1007/s13762-020-02988-3

Trinh, R., Fichot, C., Gierach, M., Holt, B., Malakar, N., Hulley, G., & Smith, J. (2017). Application of Landsat 8 for Monitoring Impacts of Wastewater Discharge on Coastal Water Quality. Frontiers in Marine Science, 4. https://www.frontiersin.org/articles/10.3389/fmars.2017.00329

Truong, A., Walters, A., Goodsitt, J., Hines, K., Bruss, C. B., & Farivar, R. (2019). Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 1471–1479. https://doi.org/10.1109/ICTAI.2019.00209

Turner, & Montagna, P. (2016). The Max Bin Regression Method to Identify Maximum Bioindicator Responses to Ecological Drivers. Elsevier. http://dx.doi.org/10.1016/j.ecoinf.2016.10.007

Turner, Savage, K., Trungale, J., Palmer, T., & Montagna, P. (2016). Effect of Freshwater Inflow on Habitat Suitability Change in Texas Bays.

Urquhart, E., Zaitchik, B., Hoffman, M., Guikema, S., & Geiger, E. (2012). Remotely sensed estimates of surface salinity in the Chesapeake Bay: A statistical approach. Remote Sensing of Environment, 123, 522–531. https://doi.org/10.1016/j.rse.2012.04.008

Vakili, T., & Amanollahi, J. (2020). Determination of optically inactive water quality variables using Landsat 8 data: A case study in Geshlagh reservoir affected by agricultural land use. Journal of Cleaner Production, 247, 119134. https://doi.org/10.1016/j.jclepro.2019.119134

Vargas-Lopez, I. A., Rivera-Monroy, V. H., Day, J. W., Whitbeck, J., Maiti, K., Madden, C. J., & Trasviña-Castro, A. (2021). Assessing Chlorophyll a Spatiotemporal Patterns Combining In Situ Continuous Fluorometry Measurements and Landsat 8/OLI Data across the Barataria Basin (Louisiana, USA). Water, 13(4), Article 4. https://doi.org/10.3390/w13040512

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment, 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Virdis, S., Xue, W., Winijkul, E., Nitivattananon, V., & Punpukdee, P. (2022). Remote sensing of tropical riverine water quality using sentinel-2 MSI and field observations. Ecological Indicators, 144, 109472. https://doi.org/10.1016/j.ecolind.2022.109472

Vuille, M., & Baumgartner, M. F. (1993). Hydrologic investigations in the north Chilean Altiplano using landsat - MSS and - TM data. Geocarto International, 8(3), 35–45. https://doi.org/10.1080/10106049309354418

Wagle, Acharya, T., & Lee, D. (2020). Comprehensive Review on Application of Machine Learning Algorithms for Water Quality Parameter Estimation Using Remote Sensing Data. Sensors and Materials, 32(11), 3879. https://doi.org/10.18494/SAM.2020.2953

Wang, F., & Xu, Y. J. (2008). Development and application of a remote sensing-based salinity prediction model for a large estuarine lake in the US Gulf of Mexico coast. JOURNAL OF HYDROLOGY, 360(1–4), 184–194. https://doi.org/10.1016/j.jhydrol.2008.07.036

Wang, H., Huang, W., Harwell, M. A., Edmiston, L., Johnson, E., Hsieh, P., Milla, K., Christensen, J., Stewart, J., & Liu, X. (2008). Modeling oyster growth rate by coupling oyster population and hydrodynamic models for Apalachicola Bay, Florida, USA. Ecological Modelling, 211(1), 77–89. https://doi.org/10.1016/j.ecolmodel.2007.08.018

Wang, Huang, W., Harwell, M., Edmiston, L., Johnson, E., Hsieh, P., Milla, K., Christensen, J., Stewart, J., & Liu, X. (2008). Modeling oyster growth rate by coupling oyster population and hydrodynamic models for Apalachicola Bay, Florida, USA. Ecological Modelling, 211(1), 77–89. https://doi.org/10.1016/j.ecolmodel.2007.08.018

Ward, & Armstrong, N. (1980). Matagorda Bay, Texas, Its Hydrography, Ecology, and Fishery Resources. Fish and Wildlife Service, U.S. Department of the Interior.

Wei, J., Lee, Z., Garcia, R., Zoffoli, L., Armstrong, R., Shang, Z., Sheldon, P., & Chen, R. (2018). An assessment of Landsat-8 atmospheric correction schemes and remote sensing reflectance products in coral reefs and coastal turbid waters. Remote Sensing of Environment, 215, 18–32. https://doi.org/10.1016/j.rse.2018.05.033

Wei, Wei, L., Yang, H., Wang, Z., Xiao, Z., Li, Z., Yang, Y., & Xu, G. (2022). Water Quality Grade Identification for Lakes in Middle Reaches of Yangtze River Using Landsat-8 Data with Deep Neural Networks (DNN) Model. Remote Sensing, 14(24), Article 24. https://doi.org/10.3390/rs14246238

Wei, Zhang, Y., Huang, C., Wang, Z., Huang, Q., Yin, F., Guo, Y., & Cao, L. (2020). Inland Lakes Mapping for Monitoring Water Quality Using a Detail/Smoothing-Balanced Conditional Random Field Based on Landsat-8/Levels Data. Sensors, 20(5), Article 5. https://doi.org/10.3390/s20051345

Weidberg, N., Chiquillo, L., Román, S., Román, M., Vázquez, E., Olabarria, C., Woodin, S., & Wethey, D. (2023). Assessing high resolution thermal monitoring of complex intertidal environments from space: The case of ECOSTRESS at Rias Baixas, NW Iberia. Remote Sensing Applications: Society and Environment, 32, 101055. https://doi.org/10.1016/j.rsase.2023.101055

Wetzel, G., & Armstrong, N. (1987). Studies Regarding the Distribution and Biomass Densities of, and the Influences of Freshwater Inflow Variations of Finfish Populations in the Matagorda Bay System, Texas. https://hdl.handle.net/2152/118122

Wilber, & Bass. (1998). Effect of the Colorado River Diversion on Matagorda Bay Epifauna. Estuarine, Coastal and Shelf Science, 47(3), 309–318. https://doi.org/10.1006/ecss.1998.0356

Wilber, & Clarke, D. (2010). Dredging Activities and the Potential Impacts of Sediment Resuspension and Sedimentation on Oyster Reefs.

Wilkie, E., Bishop, M., & O'Connor, W. (2013). The density and spatial arrangement of the invasive oyster Crassostrea gigas determines its impact on settlement of native oyster larvae. Ecology and Evolution, 3(15), 4851–4860. https://doi.org/10.1002/ece3.872

Wong, M., Lee, K., Kim, Y., Nichol, J., Li, Z., & Emerson, N. (2007). Modeling of Suspended Solids and Sea Surface Salinity in Hong Kong using Aqua/MODIS Satellite Images. Korean Journal of Remote Sensing, 23(3), 161–169. https://doi.org/10.7780/kjrs.2007.23.3.161

Xie, Z., Zhang, C., & Berry, L. (2013). Geographically weighted modelling of surface salinity in Florida Bay using Landsat TM data. Remote Sensing Letters, 4(1), 75–83. https://doi.org/10.1080/2150704X.2012.693218

Yang, Z., & Anderson, Y. (2016). Estimating Chlorophyll-A Concentration in a Freshwater Lake Using Landsat 8 Imagery.

Yu, X., Yi, H., Liu, X., Wang, Y., Liu, X., & Zhang, H. (2016). Remote-sensing estimation of dissolved inorganic nitrogen concentration in the Bohai Sea using band combinations derived from MODIS data. International Journal of Remote Sensing, 37(2), 327–340. https://doi.org/10.1080/01431161.2015.1125555

Zhang, C., Xie, Z., Roberts, C., Berry, L., & Chen, G. (2012). Salinity Assessment in Northeast Florida Bay Using Landsat TM Data. Southeastern Geographer, 52(3), 267–281.

Zhang, Xue, B., Wang, G., Zhang, X., & Zhang, Q. (2022). Deep Learning-Based Water Quality Retrieval in an Impounded Lake Using Landsat 8 Imagery: An Application in Dongping Lake. Remote Sensing, 14(18), Article 18. https://doi.org/10.3390/rs14184505

Zhao, J., & Temimi, M. (2016). An empirical algorithm for retreiving salinity in the Arabian Gulf: Application to Landsat-8 data. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 4645–4648. https://doi.org/10.1109/IGARSS.2016.7730212

Zhao, J., Temimi, M., & Ghedira, H. (2017). Remotely sensed sea surface salinity in the hyper-saline Arabian Gulf: Application to landsat 8 OLI data. Estuarine, Coastal and Shelf Science, 187, 168–177. https://doi.org/10.1016/j.ecss.2017.01.008

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. Eco-Environment & Health, 1(2), 107–116. https://doi.org/10.1016/j.eehl.2022.06.001

Zimmerman, A., DePaola, A., Bowers, J., Krantz, J., Nordstrom, J., Johnson, C., & Grimes, D. (2007). Variability of Total and Pathogenic Vibrio parahaemolyticus Densities in Northern Gulf of Mexico Water and Oysters. Applied and Environmental Microbiology, 73(23), 7589–7596. https://doi.org/10.1128/AEM.01700-07

Appendix (1):

Photos taken at the AGU conference and during fieldwork

Figure 1: Graduate student Meghan Bygate attending the Fall 2022 American Geophysical Union (AGU) conference.

Figure 2: Graduate student presenting preliminary results at AGU.

Figure 3: Collection of chlorophyll-a samples in Matagorda Bay.

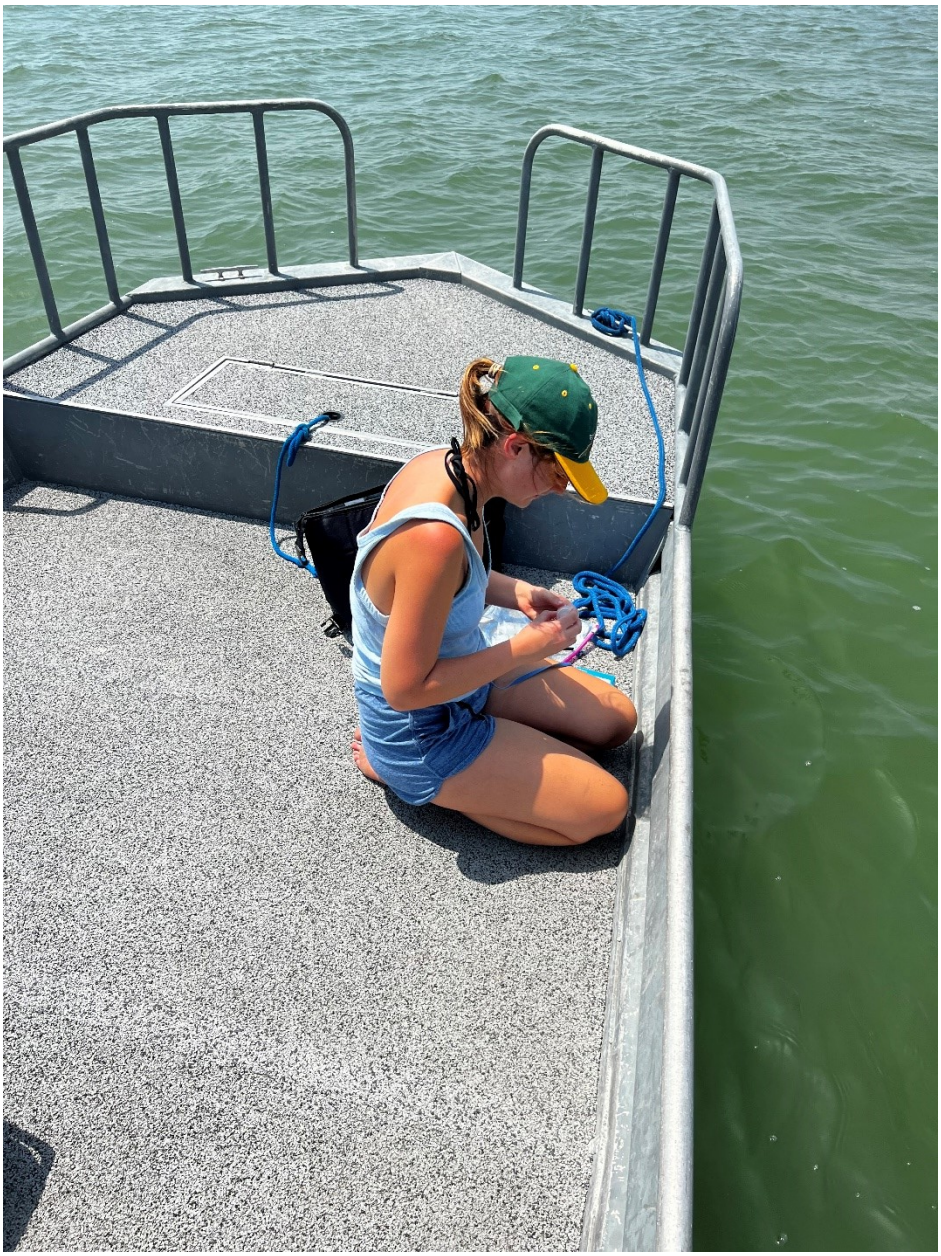

Figure 4: Nutrient sample collection.

Figure 5: Equipment used to collect nutrient samples.

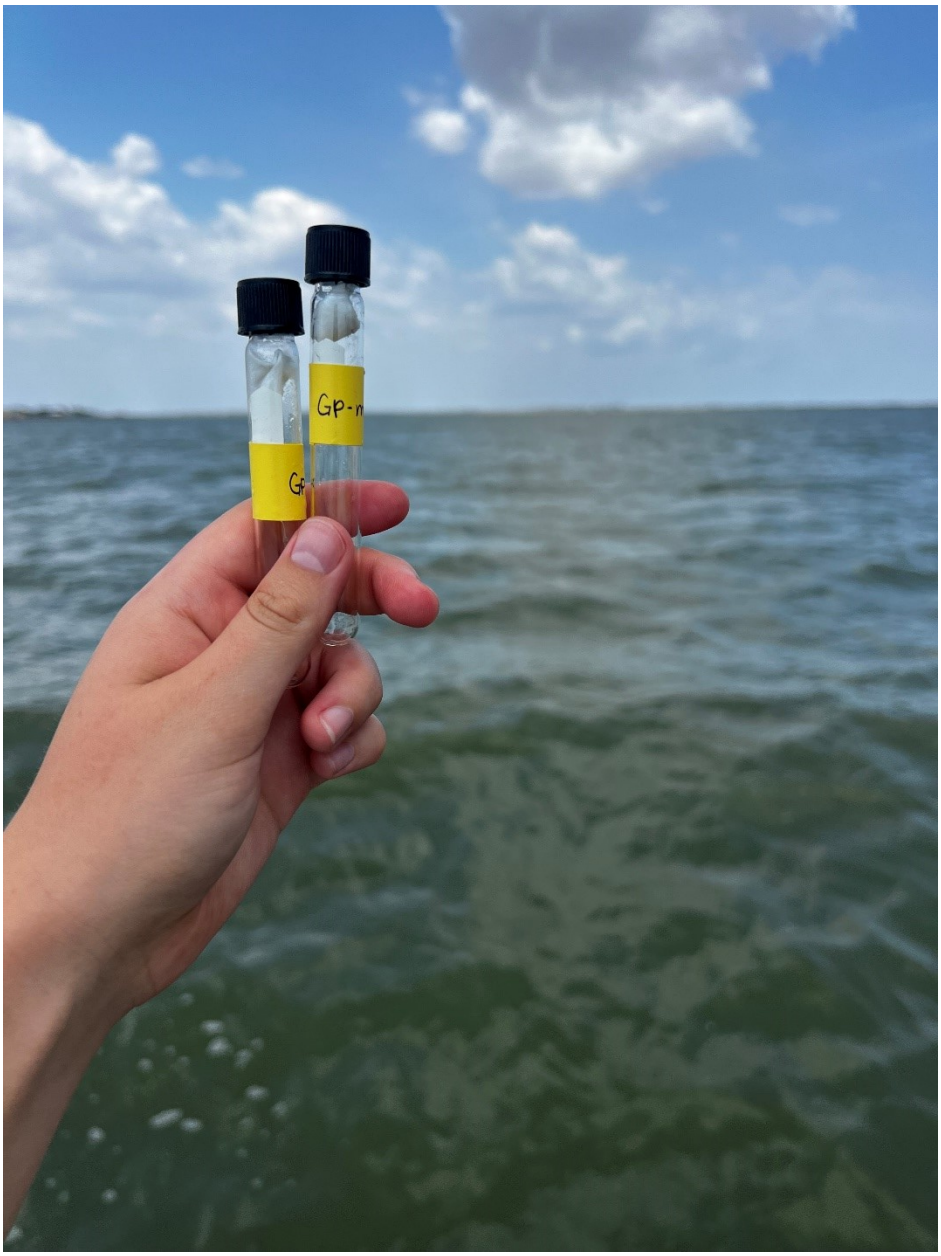

Figure 6: Collected nutrient samples.

Figure 7: Collecting salinity and temperature observations using a Yellow Springs Instrument (YSI).