



Article

Sea Surface Salinity Estimation and Spatial-Temporal Heterogeneity Analysis in the Gulf of Mexico

Zhiyi Fu ¹, Fangfang Wu ¹, Zhengliang Zhang ¹, Linshu Hu ¹, Feng Zhang ^{1,2,*}, Bifeng Hu ³, Zhenhong Du ^{1,2}, Zhou Shi ⁴ and Renyi Liu ^{1,2,5}

¹ School of Earth Sciences, Zhejiang University, Hangzhou 310027, China; 11838030@zju.edu.cn (Z.F.); 21938035@zju.edu.cn (F.W.); zlzhang@zju.edu.cn (Z.Z.); hulinsu1010@zju.edu.cn (L.H.); duzhenhong@zju.edu.cn (Z.D.); liurenyi@zju.edu.cn (R.L.)

² Zhejiang Provincial Key Laboratory of Geographic Information Science, Hangzhou 310028, China

³ Department of Land Resource Management, School of Tourism and Urban Management, Jiangxi University of Finance and Economics, Nanchang 330013, China; hubifeng@zju.edu.cn

⁴ Institute of Agricultural Remote Sensing and Information Technology Application, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China; shizhou@zju.edu.cn

⁵ Ocean Academy, Zhejiang University, Zhoushan 316021, China

* Correspondence: zfcarnation@zju.edu.cn; Tel.: +86-571-8827-3287

Abstract: As an important parameter to characterize physical and biogeochemical processes, sea surface salinity (SSS) has received extensive attention. Cubist is a data mining model, which can be well-suited to estimate and analyze SSS in the Gulf of Mexico (GOM) because it can reflect the SSS internal heterogeneity in the GOM—overall circular distribution, and the seasonality related to temperature and river discharge changes. Using remote sensing reflectance (Rrs) at 412, 443, 488 (490), 555, and 667 (670) nm and sea surface temperature (SST), a cubist model was developed to estimate SSS with high accuracy with the overall performance demonstrates a root mean square error (RMSE) of 0.27 psu and correlation coefficient of 0.97 of R2. The model divides the GOM area according to model rules into four sub-regions, which include estuary, nearshore, and open sea, reflecting the gradient distribution of SSS. The division of sub-regions and seasonal changes can be explained by the distribution of water bodies, river discharges, and local wind forces since it is obvious that the estuary region reaches the largest low-value area and spreads eastward with the monsoon in the spring when the river flow increases to the highest value. While the east to west wind in the non-summer monsoon period guides the plume westward, and the lowest river discharge in winter corresponds to the smallest low value area. After comparison with other statistical models, the cubist model showed satisfactory results in independent verification of cruise data, proving the estimation capability under different geographical conditions (such as estuaries and open seas) and seasons. Therefore, considering high accuracy and heterogeneity mining, the cubist-based model is an ideal method for coastal SSS estimation and spatial-temporal heterogeneity analysis, and can provide ideas for model construction for coastal areas with similar geographic environments.

Keywords: sea surface salinity; remote sensing; Gulf of Mexico; data mining



Citation: Fu, Z.; Wu, F.; Zhang, Z.; Hu, L.; Zhang, F.; Hu, B.; Du, Z.; Shi, Z.; Liu, R. Sea Surface Salinity Estimation and Spatial-Temporal Heterogeneity Analysis in the Gulf of Mexico. *Remote Sens.* **2021**, *13*, 881. <https://doi.org/10.3390/rs13050881>

Received: 19 January 2021

Accepted: 23 February 2021

Published: 26 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Ocean salinity controls the dynamic and thermodynamic behaviors of seawater, plays a role as a key parameter in oceanic and climate studies, and its distribution provides significant information for studying physical and biochemical marine processes [1,2]. Several processes govern the evolution of salinity, such as evaporation, precipitation, river runoff, formation and melting of sea ice, and internal ocean dynamics such as circulation and mixing of water masses [3]. Thus, changes in salinity can be used to indicate freshwater input to coastal oceans and therefore understand many physical and biogeochemical processes in coastal waters [4,5].

Previous salinity studies were mainly based on measurements from vessels and buoys, and the spatial coverage of these measurements is inadequate to capture complex ocean processes [1]. Satellite-generated sea surface salinity (SSS) maps show a wide range of multi-phase temporal and spatial salinity changes, improving the understanding of ocean circulation and the air–sea interaction and their influence on the global climate [6]. In recent years, optical satellite data and microwave satellite data have been used to estimate SSS in coastal areas [7–9]. Although microwave remote sensing can provide a global SSS map, the low spatial and temporal resolution limits microwave approaches in coastal SSS estimation [10].

Compared with a salinity satellite, an optical satellite possesses higher revisit frequencies and spatial resolution. The use of optical satellite images to estimate SSS in the estuary can be traced back to 1982 [11], and a multiple linear relationship between Landsat MSS and SSS was developed. There are two ways to retrieve SSS, one is using colored dissolved organic material (CDOM) as an intermediate agent due to its significant correlation with SSS [12]. Related studies that use CDOM to track SSS have been conducted in large river estuaries and plume systems [13–17]. The other method directly connects remote sensing reflectance (Rrs) with SSS, since SSS can be expressed directly as a function of remotely sensed ocean color bands because CDOM can be estimated by Rrs [1,7,10]. Both methods are based on the inverse relationship between SSS and CDOM concentration, which performs well in coastal regions [18,19]. A large portion of CDOM in the coastal ocean is terrestrial in origin and is associated with fresh water; thus, it can be used to indicate the water mixing [20,21]. In addition, CDOM primarily absorbs light in the ultraviolet and blue portions of the spectrum and can therefore be retrieved from apparent optical properties [22,23].

Statistical methods are commonly used in SSS estimation; linear models such as multiple linear regression (MLR), multi-source polynomial regression (MPR), and the least squares method have been used to retrieve SSS in previous studies [7,23–28]. Nonlinear methods show great advantages in estimation accuracy owing to the nonlinear relationship of chemical ocean effects. These models performed well in many coastal areas, such as the mid-Atlantic coastal ocean, the Indonesia coastal area, and Chesapeake Bay [10,29–31]. With the population of machine learning methods, SVR (Support Vector Regression), RF (Random Forest), and NN (Neural Network) models have been developed to predict SSS with satisfactory performance [32–35]. However, the interpretability of machine learning methods is poor, and spatial correlation cannot be well reflected, thus complicating the combination analysis between the model result and environmental factors. However, for complex coastal areas, the impact of the surrounding environment is too significant to be ignored.

Diverse habitats and ecosystems, such as barrier islands, mangrove forests, sea grass beds, coral reefs, and river deltas, are located on Gulf shores [36]. The Mississippi–Atchafalaya River system (MARS), the seventh largest freshwater discharge system, dominates the northern Gulf of Mexico (GOM), making coastal lagoons and estuaries typical and critical areas in the Gulf coastal zone [37]. As an abiotic factor, salinity may be highly variable in coastal and estuarine ecosystems due to the unique geographical location [38]. River runoff and human activities have great influence on coastal salinity, making the estimation more complicated than in the open sea [39]. Therefore, the division of different regions was particularly necessary in the GOM. As a decision-tree-based method, the cubist model has been well applied in digital mapping in recent years [40,41]. The cubist model divides datasets according to different rules composed of different environmental variable conditions [42,43], which can mine the internal heterogeneity of data, and thus can be used to spatially partition salinity regions. It can enable spatial-temporal analysis with higher accuracy when estimating SSS.

In this study, we will use the cubist model to predict SSS of the GOM with higher accuracy, using Rrs and sea surface temperature (SST) as input variables. At the same time, it analyzes the temporal and spatial heterogeneity of salinity in coastal areas on the basis of model zoning. The model divides the GOM area based on variable differences rather than geographic locations for the first time. This study can contribute to proposing a model with general applicability to estimate SSS from satellites for coastal areas, and facilitate the selection of variables and forms of SSS model construction for different regions of the GOM.

2. Materials and Methods

2.1. Study Area

The Gulf of Mexico (GOM) possesses an outer shoreline from the Florida Keys to the northwest coast of Cuba, and is the ninth largest body of water in the world [36] (Figure 1). The largest river system in North America (the MARS) comprises a complex estuary in the northern GOM [44]. The GOM is a shallow basin that holds approximately 2.5 million km³ of water. The average water depth is 1615 m, with its deepest point at 4383 m. Approximately 38% of the Gulf is less than 20 m deep—mainly intertidal areas. The continental shelf and slope comprise approximately 42% of the Gulf, and abyssal areas cover approximately 20% [45].

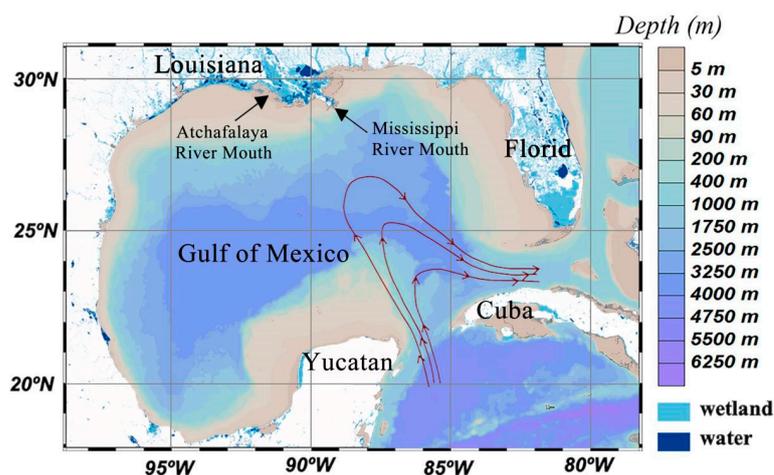


Figure 1. Geographical location of the Gulf of Mexico (GOM); distribution of its river and wetlands; water depth and typical Loop Current path in the GOM.

Both local wind stress and river flow have greater impact on the plume near the estuary. In the spring/summer with high river discharge, the southerly/southeast wind promotes eastward spread of the Mississippi River plume, while in the fall/winter—the low-flow period—the northeast/northern wind transports the river’s fresh water westward [46,47].

2.2. Field Data

The in-situ measured salinity data used in this study were downloaded from Ocean Carbon Data Systems (OCADS, <https://www.nodc.noaa.gov/ocads/>) (access on 1 January 2021). These data were collected by twelve cruises and covered the entire 2018 year with properties, including SSS and SST. The SSS and SST data were obtained ~3 m below the sea surface using a CTD (SBE-38 or SBE-45, Seabird Inc.) integrated in the underway pCO₂ system. Data used for independent verification were distributed in the coastal area, open sea, and west of the Florida Current, guaranteeing that the typical region can be well verified (Table 1).

Table 1. Underway surface salinity measurements used for the development, validation and independent validation of the sea surface salinity (SSS) model. Independent validation cruises were listed in blue font. The cruise GU1609Leg1-3 Fall Pelagic Trawl/Acoustic Survey was listed as GU1609Leg1-3 for convenience. The number of total points used in model development and independent validation was the number obtained after conjugate matching.

Cruise ID	Platforms	Date Range	# of Observations
EQ17	M/V Celebrity Equinox	1/1/2018–1/6/2018	2179
AS17	M/V Allure of the Seas	1/4/2018–1/7/2018	1198
GU1801_Leg1	R/V Gordon Gunter	1/14/2018– 1/22/2018	4178
GU1801_Leg2	R/V Gordon Gunter	1/26/2018–2/9/2018	7421
GU1801_Leg3	R/V Gordon Gunter	2/12/2018– 2/27/2018	5428
GU1801_Leg4	R/V Gordon Gunter	3/1/2018–3/16/2018	7941
GU1802	R/V Gordon Gunter	6/24/2018–7/9/2018	7609
GU1803-transit	R/V Gordon Gunter	7/11/2018– 7/14/2018	1340
GU1803-Leg1	R/V Gordon Gunter	7/20/2018–8/3/2018	7196
GU1803-Leg2	R/V Gordon Gunter	8/6/2018–8/19/2018	4727
GU1804	R/V Gordon Gunter	8/23/2018– 8/31/2018	4445
GU1805-Leg1	R/V Gordon Gunter	9/2/2018–9/9/2018	3563
GU1805-Leg2	R/V Gordon Gunter	9/11/2018– 9/30/2018	9659
EQ18	M/V Celebrity Equinox	1/6/2018– 12/22/2018	872
GU1806	R/V Gordon Gunter	11/10/2018– 12/4/2018	10,127
GM0606	OSV Bold	6/6/2006–6/11/2006	7178
GU1609Leg1-3	R/V Gordon Gunter	9/2/2016–10/1/2016	10,284
EQNX_20190209	M/V Celebrity Equinox	2/9/2019–2/16/2019	2270
Total from all cruises			97,615
Total used in model development and validation			7935
Total used in independent validation			7494

2.3. Satellite Data

Daily standard NASA Moderate Resolution Imaging Spectroradiometer (MODIS/Aqua) level-2 data products were downloaded from NASA Goddard Space Flight Center (<http://oceancolor.gsfc.nasa.gov/>) (access on 1 January 2021), including SST and spectral remote sensing reflectance (Rrs) between 412 and 678 nm. All data products have been validated by in-situ data in the study area to ensure the availability of data in the region. Images with quality level > 1 were discarded from the daily level-2 SST data. Discarding low quality images (quality level >1) from the daily level-2 SST data to guarantee data accuracy. Conjugate matching was processed between the remote sensing images and in situ measurement, and daily images were matched up with in situ underway measurements after discarding low quality data (Figure 2). The spatial resolution of the field and satellite data was reprocessed to approximately 1 km for conjugate matching. As the same consideration of Chen and Hu [34], we chose five visible spectral bands (412, 443, 490, 555, and 670 nm) based on the exponential decay of CDOM absorption from the blue to the red color spectra. SST was used as a model input to capture the difference in temperature between the ocean and rivers.

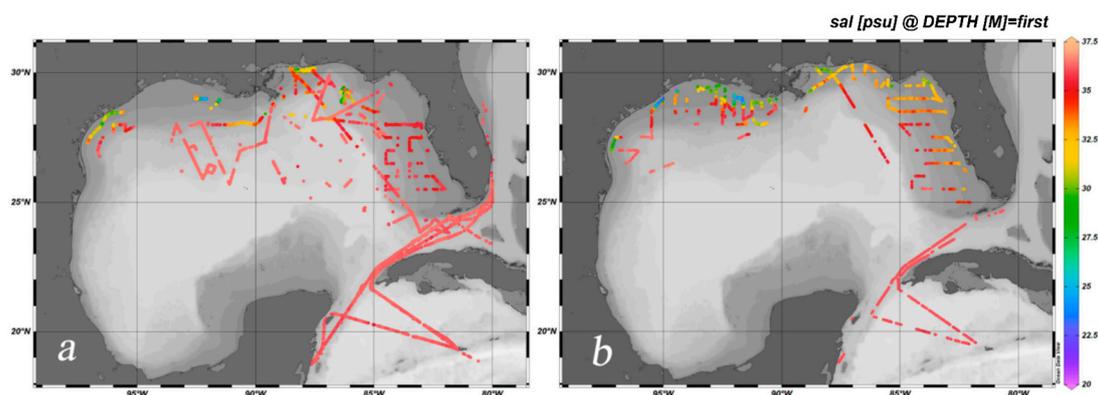


Figure 2. (a) Spatial distribution of conjugate samples of remote sensing and in-situ data in the GOM for SSS model development and validation; (b) spatial distribution of conjugate samples for model independent validation.

2.4. Method

2.4.1. Cubist

Cubist is an algorithm based on M5, which is similar to regression trees and can be used in spatial data mining [48]. As a data partitioning algorithm, it enables exploration of the nonlinear relationship in the observed data [49]. The predicted variable will be fitted by composed linear equations under rules generated by the cubist model, which is different from the CART regression tree model [50] in which its terminal nodes are not predictions [51]. The heterogeneity caused by predictor variables produces different data division conditions, which means “rules”, and will be used to divide the total dataset [52]. For each sub-dataset corresponding to each rule, the cubist model creates linear equations for sub-datasets, respectively, with the form: if {condition} then linear model. By using the linear regression model at each terminal node, a prediction is made and is “smoothed” by considering the prediction in the previous node [53]. The cubist algorithm was implemented in R with the “caret” package in this study.

2.4.2. Other Comparison Methods

Multilinear regression uses linear equations to model the relationship between predictor variables and response variables. It is one of the most widely used methods to express how the response variable depends on multiple independent variables. Similar to MLR, the multiple nonlinear regression (MNR) method uses nonlinear equations, the nonlinear functions include exponential functions, logarithmic functions, power functions, etc., which is also a commonly used empirical method [54]. The MLR and MNR models in this study were implemented in Python with the sklearn package.

The support vector machine (SVM) method is a kernel-based method proposed by Vladimir Vapnik in 1995 [55]. SVM is a general term, possessing two sub-categories support vector classification (SVC) and support vector regression (SVR). SVR uses strips to cover the sample points. The points on the boundary and the points that violate the margin within the two boundaries are regarded as support vectors. The support vector will affect predictions, while the point weight of the non-support vector is zero [56]. In this study, we chose the typical kernel radial basis function (RFB) as the core and implemented the SVM model in Python with the sklearn package.

The multilayer perceptron neural network (MPNN) is a feedforward neural network developed in the Gulf of Mexico and consists of an input layer, a hidden layer and an output layer. The Levenberg–Marquardt optimization and Bayesian regularization algorithm were used for backpropagation. The change of neuron number in the hidden layer will affect the effect of MPNN, because the neuron number of input and output is unchanged [34]. In this study, we used Matlab (R2019a) to implement the construction of MPNN.

3. Results

3.1. Model Performance

When choosing the suitable rule of the cubist model for this study, we used root mean square error (RMSE) and correlation coefficient (R^2) to assess model performance. As seen in Figure 3, the lowest RMSE and highest R^2 appeared when the number of rules was four; the RMSE increased and the R^2 decreased when it exceeded four. This indicates that four rules should be selected owing to the high accuracy and simplicity of the model.

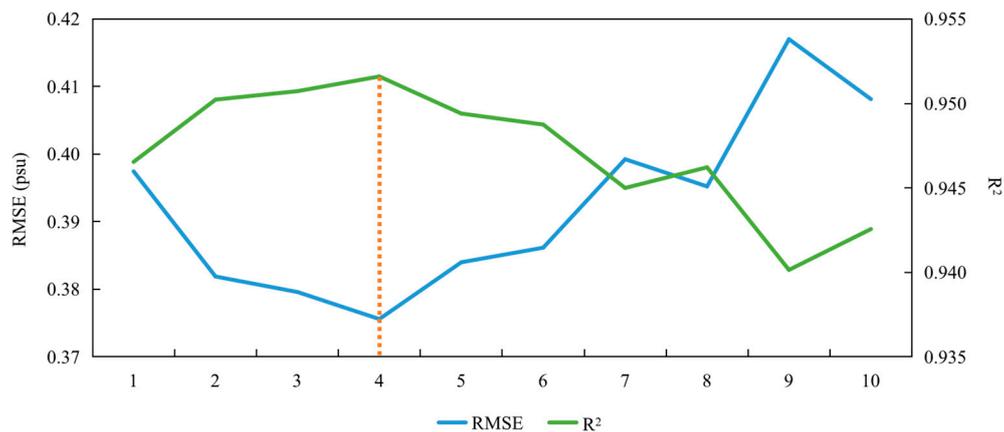


Figure 3. Trend in the RMSE and R^2 when evaluating SSS accuracy and interpretability as rule number increases.

Eighty percent of the dataset was used to build the prediction model, and the rest was used to validate the model. Before dividing, the total dataset was sorted and grouped, and for each group, the data were randomly divided by 8:2. Then two sub-datasets with the same range were created by composing group data. The performance of the cubist model in both the training and validation datasets is shown in Figure 4, colored by data density. R^2 is 0.97 and 0.95 for model development and validation, with RMSE of 0.24 and 0.38 psu, respectively. The other standard statistical measures, mean bias (MB), and mean absolute error (MAE), were also used to compare the cubist model with other methods.

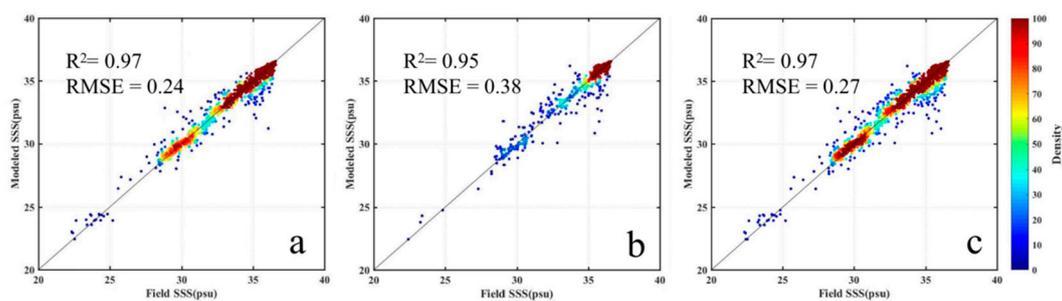


Figure 4. Cubist model performance in estimating SSS in both training (a), validation (b) and total dataset (c); data pairs are color coded by data density.

As presented in Table 2, the performance of training and validation was stable in total, meaning that there was no overfitting in each method. The machine learning approaches were better than the regression methods obviously, since the performance of MLR was the worst with RMSE of 1.00 psu and R^2 of 0.64 for training, and RMSE of 1.04 psu and R^2 of 0.63 for validation. Despite the poor performance of MLR, the correlation coefficient was still higher than 0.6, indicating that a linear relationship can explain part of the SSS, and the regression method can be used as a simple SSS estimation algorithm. As a regression method, the MNR was better owing to the capacity of simulating the inherent nonlinear

relationship, with RMSE of 0.78 psu and R^2 of 0.78, and RMSE of 0.90 psu with R^2 of 0.72 for training and validation, respectively. SVM processes the stable performance showing the little difference between training and validation, the R^2 of 0.85 and RMSE of 0.38 psu for training, and R^2 of 0.84 with RMSE of 0.39 psu for validation. The MPNN model was also developed in the GOM, and our results were similar to those of [34], with R^2 of 0.86 and MB of 0.00 psu for training, and R^2 of 0.85 with MB of 0.00 psu for validation. The complexity of the ocean makes it impossible to directly estimate SSS with high accuracy through linear or simple nonlinear combinations. However, since the use of CDOM to estimate salinity is based on remote sensing mechanisms, simple statistical methods can still fit salinity data. The task of the SVM regression method is to use a fixed-width strip to cover more sample point as possible, making the total error smaller, so there are some limitations for complex data. The internal fitting of the neural network is more complicated, and the learning of data features makes it perform better on large data sets. Comparing all the machine learning methods, the cubist model provided outstanding performance with all indices significantly better than those of other methods. Although the improvement in R^2 was small, the RMSE decreased by nearly half compared with that of the MPNN model.

Table 2. Comparison table of SSS estimation approaches in the GOM, all methods used the same training dataset and validation dataset.

Approach		RMSE (psu)	R^2	MB (psu)	MAE (psu)
MLR [7]	Training	1.00	0.64	0.00	0.61
	Validation	1.04	0.63	0.04	0.63
MNR [31]	Training	0.78	0.78	0.00	0.43
	Validation	0.90	0.72	0.04	0.44
SVM [32]	Training	0.38	0.85	0.00	0.18
	Validation	0.39	0.84	0.02	0.19
Cubist (this study)	Training	0.24	0.97	0.00	0.10
	Validation	0.38	0.95	−0.02	0.16
MPNN [34]	Training	0.62	0.86	0.00	0.33
	Validation	0.67	0.85	0.00	0.35

The year's SSS map estimated by the cubist model is shown in Figure 5. It can be seen that the overall trend of SSS showed a concentric inward value increase. The influence of wind force and GOM dynamics affects the distribution of SSS values [57]. In the northern region, due to the physical mixing of river water with low salinity, SSS was lower throughout the year. In the open sea area, the SSS was usually higher because of minimal river influence.

In summary, the cubist model was superior to other methods in terms of relevance and accuracy. Considering the good interpretability of the cubist model, for it was not a “black box” and can provide the equation of each rule, the cubist model was the most favorable method for estimating SSS within the range 22–38 in the GOM.

3.2. Rule Accuracy Validation

The cubist model divided the dataset into sub-datasets based on conditions and generated an equation for each sub-dataset. Based on the rule partition of the cubist model, it can be seen that the inputs Rrs(412), Rrs(555), and SST, played an important role in dataset division, since the conditions were composed of these three parameters (Table 3). The difference between rule 2 and 4 was the range of SST, while the range of Rrs(555) was the reference to divide rule 1 and 3. Contrary to the rule condition, most model inputs participated in the composition of linear equations of each rule, except Rrs(667). The intercepts of the four equations were very close, but the coefficient of each parameter had great differences.

It is obvious that rule 4 delivered the best performance in each index among the sub-datasets when evaluating the accuracy of each rule-divided dataset, although it possessed the largest amount of data (Table 4). Rule 4 was also the only sub-dataset that did not underestimate SSS. Comparing rules 2 and 4 with rules 1 and 3, it is worth noting that using more data can often make the model more stable. The large difference between training and validation datasets appeared in rule 3; the validation performance was significantly worse than that of the training dataset, with RMSE of 0.69 and R^2 of 0.80.

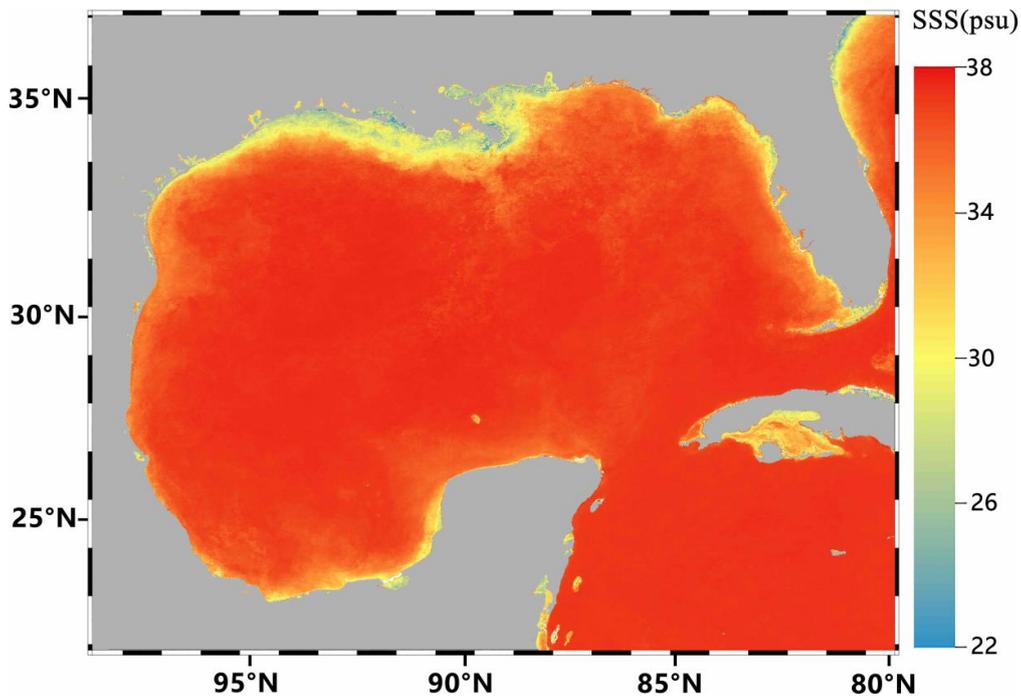


Figure 5. Annual SSS map generated by the cubist model, averaged from daily composed image within the cruise date range.

Table 3. Cubist-generated linear models for each rule, reflecting the usage and weight of each variable.

Rule	Conditions	Data Range						Equation	Count
		Rrs412	Rrs443	Rrs488	Rrs555	Rrs667	SST		
1	Rrs(412) ≤ 0.003746 & Rrs(555) > 0.001552	−0.001744–0.003746	−0.000418–0.00578	0.000846–0.009736	0.001556–0.008024	0.000066–0.003372	18.28–32.15	42.01913 + 2374 * Rrs(412) − 2843 * Rrs(443) + 1107 * Rrs(488) − 948 * Rrs(555) − 0.329 * SST	539
2	Rrs(412) > 0.003746 & SST > 28.98	0.003788–0.016226	0.003156–0.014352	0.002262–0.018128	0–0.011534	−0.000324–0.000924	28.99–32.53	40.31452 + 271 * Rrs(412) − 447 * Rrs(555) − 0.218 * SST − 113 * Rrs(443) + 88 * Rrs(488)	870
3	Rrs(412) ≤ 0.003746 & Rrs(555) ≤ 0.001552	−0.001438–0.003744	−0.000056–0.004428	0.000882–0.004028	0.000112–0.001548	−0.000292–0.000366	20.25–30.75	39.16627 − 501 * Rrs(555) + 240 * Rrs(412) + 348 * Rrs(488) + 261 * Rrs(443) − 0.227 * SST	367
4	Rrs(412) > 0.003746 & SST ≤ 28.98	0.003746–0.030036	0.003124–0.036298	0.003226–0.044100	0.000086–0.028062	−0.000650–0.010122	18.27–28.98	38.57491 − 0.105 * SST − 150 * Rrs(555) + 102 * Rrs(488) − 37 * Rrs(443) + 27 * Rrs(412)	4572

Table 4. Accuracy verification of each rule model for both train and validation, the validation result was listed with blue background.

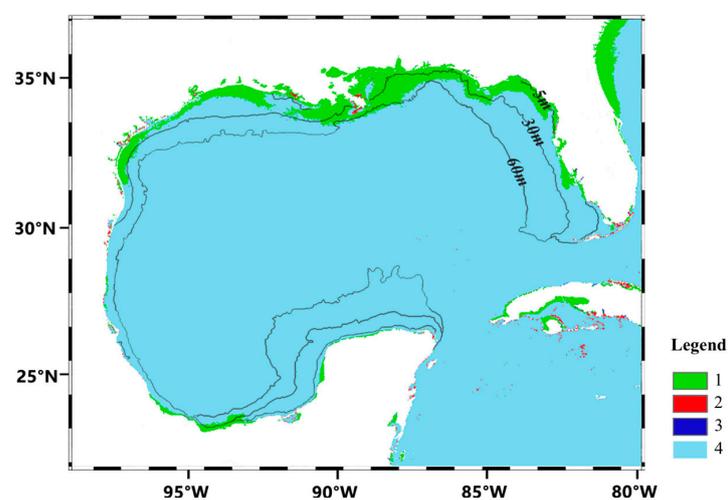
Rule	Count	RMSE (psu)	R ²	MB (psu)	MAE (psu)
1	539	0.51	0.98	−0.03	0.31
	140	0.80	0.97	−0.11	0.48
2	870	0.41	0.88	−0.03	0.21
	207	0.58	0.80	−0.08	0.35
3	367	0.33	0.95	−0.03	0.17
	99	0.69	0.80	−0.04	0.32
4	4572	0.10	0.98	0.00	0.05
	1167	0.14	0.96	0.01	0.07

4. Discussion

4.1. Rule-Based GOM Partition

Based on the rule conditions, the cubist model divided the total GOM area into four parts, which can indicate the spatial heterogeneity of SSS. Incorporating the annual data of model inputs (Rrs(412), Rrs(443), Rrs(488), Rrs(555), Rrs(667), and SST), and conditions applied to region division, the distribution of the four partitions can be distinguished (Figure 6). Due to the strong correlation with the surrounding freshwater environment, the analysis should be combined with the GOM water depth and wetland distribution (Figure 1).

On the whole, region 1 was mainly distributed in the nearshore area, while region 2 and region 3 were scattered in estuaries and nearshore places with water depths less than 5 m. Region 4 occupied the largest area of GOM, covering a large area of open sea. In general, this distribution of the sub-regions corresponded to water depth, the area with depth of less than 30 m comprised most of region 1. The other areas of region 1, such as a small area on the south side of the GOM and west side of Yucatan, were consistent with the distribution of water bodies, which has great influence on the zoning and was also reflected in the distribution of region 2. Most areas of region 2 mainly correspond to the estuary area, where the river water is mixed with seawater, and the salinity was different from that of region 1. Region 4 covered most of the GOM area—mainly open sea areas. Referring to the division rules, we can see that the temperature in region 2 was higher than that in region 4. This may be because a larger amount of data was distributed in summer, and high temperature can also be used to explain that other areas in region 2 are mainly in the Loop Current range, corresponding to warm Florida currents. Region 3 corresponds to very few areas, which may be due to the small number of samples in rule 3.

**Figure 6.** Partitioning according to the model rules; the water depth marked on the figure.

4.2. Seasonal Variations of Surface Salinity

The salinity of the GOM presents a gradient distribution throughout the year, with low values in the nearshore and high values in open seas. The areas near fresh water vary greatly, and sea water with high temperature and high salinity brought by the Loop Current can also be observed (Figure 7).

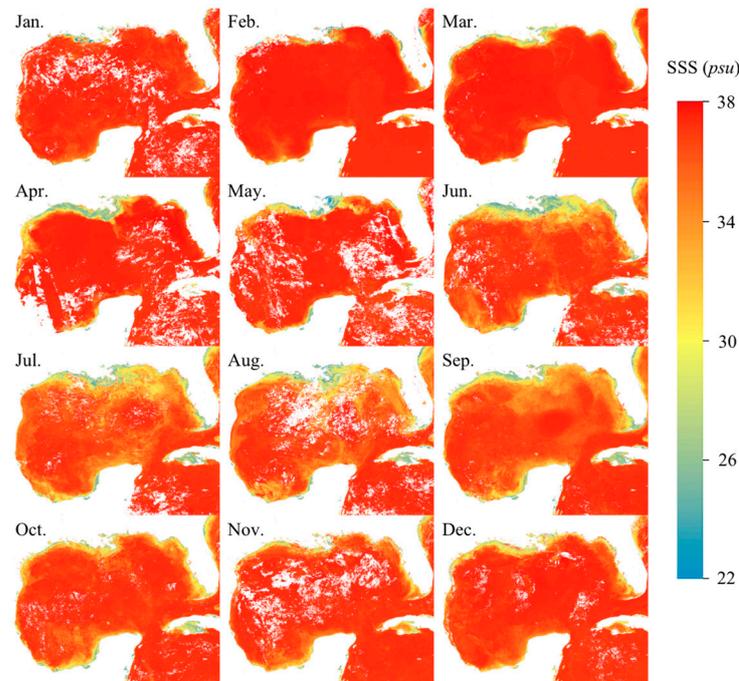


Figure 7. Monthly map of SSS in the GOM, derived from MODIS using the cubist model for 2018.

It can be seen that the most obvious change occurs in the estuary area, reflecting the seasonal cycle, which may be related to the discharge of rivers, since salinity is largely controlled by fluxes of freshwater into and out of the Gulf [58]. The river discharge data were acquired to analyze the influence of river discharge (Figure 8), downloaded from U.S. Army Corps of Engineers website. Mississippi River at Tarbert Landing (Gate ID 01100Q) and Atchafalaya River at Simmesport (Gate ID 03045Q) capturing the water flow from the total Mississippi-Atchafalaya basin. In addition, wind force is an important factor affecting the distribution of salinity. Combining the effects of wind and river flow, the discharge of the Mississippi River increases and the wind plume gradually guides eastward in the spring [59]. Therefore, the SSS values related to the river plume are significantly different than those of the surrounding waters, extending eastward from the mouth of the Mississippi River since March [46,47,60]. Since the monthly image is composed during the cruise period, the data during the maximum flow period in March are not well shown on the monthly map. However, the largest low-value salinity area shown in April indicates the impact of continuous large flow. Although the discharge began to decline in May, the SSS value in the estuary still maintained a low trend. A low SSS plume in the east of the estuary can be clearly observed due to the strong winds out of the south and west during the summer, which drives river plumes eastward [61]. The force of wind directed downcoast from east to west in non-summer periods, and areas with low SSS values were distributed westward due to the effect of wind [62]. At the same time, since the effect of river water is offset by ocean dynamics, the distribution of salinity in autumn is stable with little monthly change [39]. The generally high summer value in open sea areas would have continued throughout the fall, but due to a large amount of precipitation, the value dropped. In addition, as river runoff will reach its lowest annual value, the low-value areas of the estuary contracted in winter.

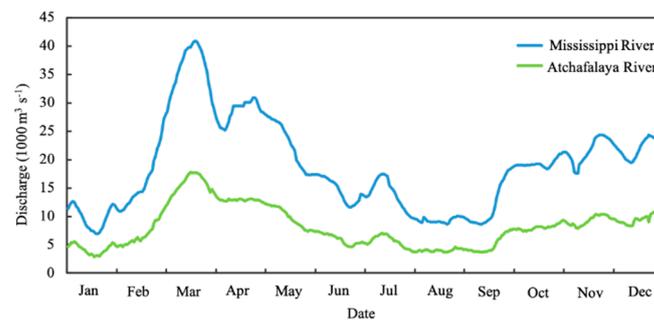


Figure 8. River discharge data in 2018, acquired from Gate ID 01100Q and 03045Q. Blue line presents Mississippi River discharge and green line presents Atchafalaya River discharge.

Along with the estuary area, other surrounding areas with water bodies, including wetlands and small rivers, are also affected by fresh water and present seasonality related to river flux. The seawater entering with the Loop Current is also well reflected, usually bringing high-temperature and high-salinity seawater and flowing out to the east of the GOM.

4.3. Model Evaluation for Various Cases

Cruise data distributed in different areas of the GOM were collected to assess the versatility of the cubist SSS model (Table 5). Since, on the one hand, the chosen independent cruises are located in different geographical locations, reflecting the spatial heterogeneity. On the other hand, the independent cruises covered different seasons, which reflects the temporal heterogeneity. For each cruise, the field-measured dataset was independent from the others, and none of these datasets were used in the model development above. The verification data covered much of the northern GOM and the southeast area near the Loop Current. As seen in the overall results, the model has good generalization ability with an RMSE of 1.64 psu, indicating good predictability in remotely estimating SSS both spatially and temporally.

Table 5. Independent validation of the SSS model, the overall result of each cruise was listed with blue background, the cruise GU1609Leg1-3 Fall Pelagic Trawl/Acoustic Survey was listed as GU1609Leg1-3 for convenience.

Cruise ID		RMSE (psu)	MB (psu)	MR	Count
GM0606	≤30	2.96	2.04	1.07	555
	>30	1.49	−0.02	1.00	2208
		1.88	0.40	1.02	2763
GU1609Leg1-3	≤30	3.01	0.91	1.03	221
	>30	1.53	0.09	1.00	3597
		1.65	0.14	1.00	3818
M2019	>30	0.13	0.04	1.00	914
Total	≤30	2.98	1.72	1.06	776
	>30	1.41	0.05	1.00	6719
		1.64	0.22	1.01	7495

The results based on the underway SSS data collected from cruise GM0606 between June 6 and 11, 2006, are shown in Figure 9. This cruise was collected in the Mississippi-Atchafalaya coastal areas (Figure 9b). For the entire dataset, the RMSE was 1.88, with MB of 0.4 and mean ratio (MR) of 1.02. The variation of satellite SSS along the cruise track agreed well with the field-measured SSS with RMSE of 1.49, MB of −0.02, and MR of 1.0 when the value was higher than 30. However, the model showed higher uncertainties (RMSE = 2.96, MB = 2.04, and MR = 1.07) in the area close to the coastline, especially in

four locations (marked A, B, C, and D in Figure 9a,b). The SSS values in these areas were lower than 30 psu and were overestimated by the cubist model. This may be due to the mixing with Mississippi-Atchafalaya water, and the wind directed eastward in summer, making the low value of SSS in the east of the estuarine area difficult to model. However, except for the cubist method, the RMSE of other methods is greater than 3 psu where the salinity value is lower than 30 psu (Table 6). It indicates that the cubist model has a good estimation potential in river dominated nearshore areas in summer.

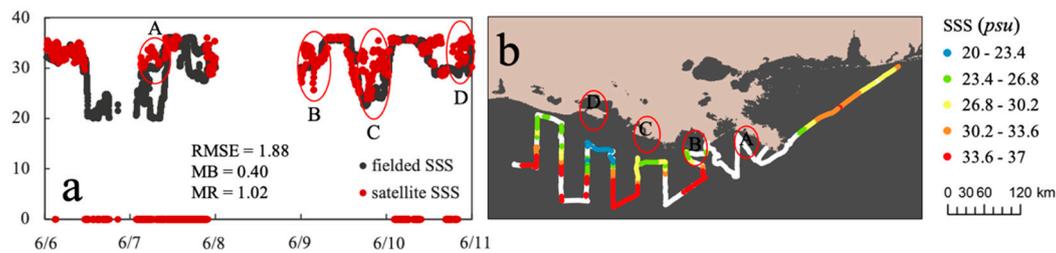


Figure 9. Performance of the SSS model in quantifying SSS in the area around the Mississippi-Atchafalaya estuary in the northern GOM. (a) Comparison between field-measured SSS and MODIS-derived modeled SSS, the red dots on the X-axis indicate that there are no corresponding MODIS-derived SSS; (b) spatial distributions of the MODIS-derived SSS along the cruise track in (a), colored by field-measured value, white color indicates no MODIS data.

Table 6. Model comparison in independent validation of cruise GM0606, the overall result of each cruise was listed with blue background.

Approach		RMSE (psu)	MB (psu)	MR
MLR	≤30	4.13	3.46	1.13
	>30	1.01	−0.12	1.00
		2.06	0.60	1.02
MNR	≤30	4.57	4.25	1.16
	>30	1.53	0.14	1.01
		2.46	0.96	1.04
SVM	≤30	5.13	4.63	1.17
	>30	1.37	0.07	1.00
		2.60	0.98	1.04
MPNN	≤30	3.69	3.02	1.11
	>30	1.19	−0.33	0.99
		1.97	0.34	1.02

The validation result based on one cruise dataset (GU1609 Leg 1-3 Fall Pelagic Trawl/Acoustic Survey) collected in the northeastern GOM between September 2 and October 1, 2016 was shown in Figure 10. A comparison of field and satellite SSS was shown in Figure 10a. Although there were more overestimated values, mainly because a large number of the points were distributed in the open sea, the overall performance was good with an RMSE of 1.65 psu. There was no need for concern with the overestimation of the results with SSS > 30 (RMSE = 1.53, MB = 0.09, and MR = 1.0). The effect of seawater and freshwater mixing was well reflected in the results with overestimated values occurring in the areas near the water bodies (locations marked B and C in Figure 10a,b), and in the coastal area where there is no fresh water influence, underestimated values occurred (locations marked A and D in Figure 10a,b). The advantages of the cubist model in nearshore estimation were still reflected, but a large number of overestimated points in open sea areas makes the overall RMSE results not very prominent. Combined with the results of MB, the estimation ability of the cubist model in the autumn nearshore was still satisfactory (Table 7).

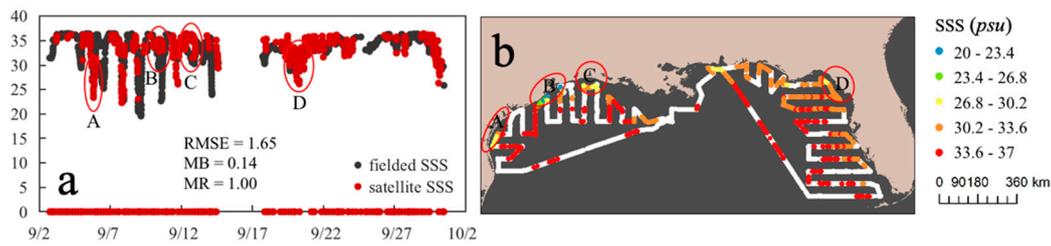


Figure 10. Performance of the SSS model in quantifying SSS in the northern GOM, cruise data distributed in the coastal area and open sea. (a) Comparison between field-measured SSS and MODIS-derived modeled SSS, the red dots on the X-axis indicate that there are no corresponding MODIS-derived SSS; (b) spatial distributions of the MODIS-derived SSS along the cruise track in (a), colored by field-measured value, white color indicates no MODIS data.

The results based on SSS data collected from cruise EQNX_20190209 in the southeastern GOM waters between February 9 and 16, 2019 can be seen in Figure 11. The field-measurement SSS in this area was very stable, because in winter, there was no strong water mass mixing in the open sea areas. The model results were consistent with the previous performance in the open sea area, with primarily overestimated SSS (Table 4), but the most overestimated value did not exceed 0.5 psu. The RMSE of the result is 0.13, which is not significantly different from other machine learning methods results (Table 8), but it still proves the good prediction ability of the model in open sea areas.

Table 7. Model comparison in independent validation of cruise GU1609Leg1-3 Fall Pelagic Trawl/Acoustic Survey, the overall result of each cruise was listed with blue background.

Approach		RMSE (psu)	MB (psu)	MR
MLR	≤30	2.81	−2.74	1.14
	>30	1.34	0.41	1.01
		1.46	0.23	1.02
MNR	≤30	3.27	−2.81	1.14
	>30	1.51	0.20	1.00
		1.67	0.02	1.01
SVM	≤30	5.47	4.80	1.19
	>30	1.06	0.13	1.00
		1.67	0.40	1.02
MPNN	≤30	4.70	3.66	1.15
	>30	1.17	0.21	1.01
		1.60	0.41	1.02

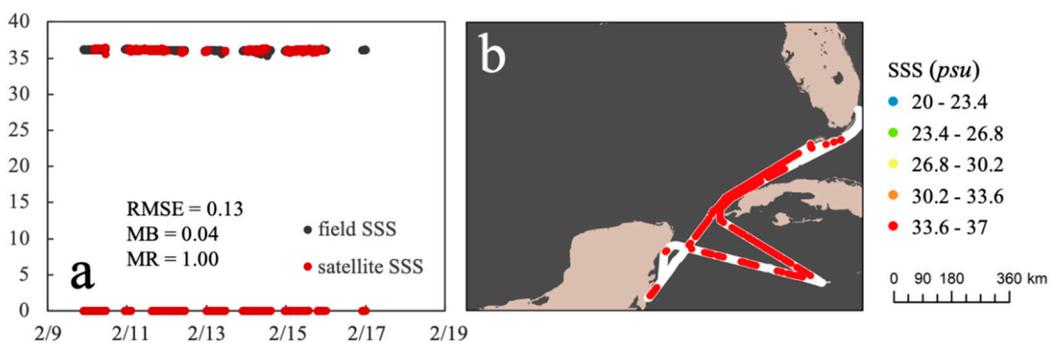


Figure 11. Performance of the SSS model in quantifying SSS in the southeast GOM near the Loop Current. (a) Comparison between field-measured SSS and MODIS-derived modeled SSS, the red dots on the X-axis indicate that there are no corresponding MODIS-derived SSS; (b) spatial distributions of the MODIS-derived SSS along the cruise track in (a), colored by field-measured value, white color indicates no MODIS data.

Table 8. Model comparison in independent validation of cruise M2019.

Approach		RMSE (psu)	MB (psu)	MR
MLR	>30	0.54	−0.15	1.00
MNR	>30	0.29	0.08	1.00
SVM	>30	0.19	0.06	1.00
MPNN	>30	0.18	0.06	1.00

5. Conclusions

In this study, we applied a cubist model to estimate SSS from MODIS images for the GOM area and obtained satisfactory performance with an RMSE of 0.38 psu and R^2 of 0.95. Through accuracy verification for each rule, we found that the linear equations given by the model have good accuracy in each region, which makes the overall salinity prediction more accurate. The model rules divided the GOM into four sub-regions, mainly including the continental shelf, estuary, and open sea area. The influences of water circulation and rivers and distribution of wetlands can be used to explain the rationality of model zoning. Seasonal changes are mainly affected by river discharge and local wind force, reflected in the area and direction of low-salinity-value plumes. Our model performs well in estimating salinity, and the auxiliary verification of other voyages proves the usability of the model under different geographical conditions. Additional estimation and research are needed to locally tune model parameters when extrapolating the model to other areas with similar environmental and geographic conditions. The model provides ideas for model construction in other coastal areas and also provides effective information for explaining the spatial heterogeneity of salinity in coastal areas and exploring seasonal changes in salinity.

Author Contributions: Writing, Z.F.; investigation, Z.F.; methodology, Z.F. and F.W.; software, Z.Z. and L.H.; validation, Z.Z.; data curation, F.W. and Z.Z.; writing—review and editing, F.Z. and B.H.; conceptualization, F.Z. and Z.S.; funding acquisition, Z.D.; supervision, R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2018YFB0505000), National Natural Science Foundation of China (41671391, 41922043, 41871287).

Acknowledgments: Thanks to NOAA, NCEI, and LDEO for providing all the available cruise data and thanks to NASA for providing MODIS satellite data and processing software.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Sun, D.; Su, X.; Qiu, Z.; Wang, S.; Mao, Z.; He, Y. Remote Sensing Estimation of Sea Surface Salinity from GOCI Measurements in the Southern Yellow Sea. *Remote Sens.* **2019**, *11*, 775. [[CrossRef](#)]
- Klemas, V. Remote Sensing of Sea Surface Salinity: An Overview with Case Studies. *J. Coast. Res.* **2011**, *276*, 830–838. [[CrossRef](#)]
- Rao, R.R.; Sivakumar, R. Seasonal variability of sea surface salinity and salt budget of the mixed layer of the north Indian Ocean. *J. Geophys. Res. Ocean.* **2003**, *108*, 9. [[CrossRef](#)]
- Fennel, K.; Hetland, R.D.; Feng, Y.S.; DiMarco, S.F. A coupled physical-biological model of the Northern Gulf of Mexico shelf: Model description, validation and analysis of phytoplankton variability. *Biogeosciences* **2011**, *8*, 1881–1899. [[CrossRef](#)]
- Xue, Z.; He, R.; Fennel, K.; Cai, W.-J.; E Lohrenz, S.; Hopkinson, C.S. Modeling ocean circulation and biogeochemical variability in the Gulf of Mexico. *Biogeosciences* **2013**, *10*, 7219–7234. [[CrossRef](#)]
- Le Vine, D.M.; Kao, M.; Garvine, R.W.; Sanders, T. Remote Sensing of Ocean Salinity: Results from the Delaware Coastal Current Experiment. *J. Atmos. Ocean. Technol.* **1998**, *15*, 1478–1484. [[CrossRef](#)]
- Qing, S.; Zhang, J.; Cui, T.; Bao, Y. Retrieval of sea surface salinity with MERIS and MODIS data in the Bohai Sea. *Remote Sens. Environ.* **2013**, *136*, 117–125. [[CrossRef](#)]
- Font, J.; Camps, A.; Ballabrera-Poy, J. Microwave Aperture Synthesis Radiometry: Paving the Path for Sea Surface Salinity Measurement from Space. In *Remote Sensing of the European Seas*; Springer International Springer: Dordrecht, The Netherlands, 2008; ISBN 9781402067716.

9. Blume, H.-J.C.; Kendall, B.M. Passive Microwave Measurements of Temperature and Salinity in Coastal Zones. *IEEE Trans. Geosci. Remote Sens.* **1982**, *394*–404. [[CrossRef](#)]
10. Urquhart, E.A.; Zaitchik, B.F.; Hoffman, M.J.; Guikema, S.D.; Geiger, E.F. Remotely sensed estimates of surface salinity in the Chesapeake Bay: A statistical approach. *Remote Sens. Environ.* **2012**, *123*, 522–531. [[CrossRef](#)]
11. Khorram, S. Remote sensing of salinity in the San Francisco Bay Delta. *Remote Sens. Environ.* **1982**, *12*, 15–22. [[CrossRef](#)]
12. Owers, D.G.; Harker, G.E.L.; Smith, P.S.D.; Tett, P. Optical Properties of a Region of Freshwater Influence (The Clyde Sea). *Estuarine, Coast. Shelf Sci.* **2000**, *50*, 717–726. [[CrossRef](#)]
13. Palacios, S.L.; Peterson, T.D.; Kudela, R.M. Development of synthetic salinity from remote sensing for the Columbia River plume. *J. Geophys. Res. Ocean.* **2009**, *114*. [[CrossRef](#)]
14. Hu, C.; Montgomery, E.T.; Schmitt, R.W.; Muller-Karger, F.E. The dispersal of the Amazon and Orinoco River water in the tropical Atlantic and Caribbean Sea: Observation from space and S-PALACE floats. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **2004**, *51*, 1151–1171. [[CrossRef](#)]
15. Del Vecchio, R. Influence of the Amazon River on the surface optical properties of the western tropical North Atlantic Ocean. *J. Geophys. Res. Ocean.* **2004**, *109*. [[CrossRef](#)]
16. Del Castillo, C.E.; Miller, R.L. On the use of ocean color remote sensing to measure the transport of dissolved organic carbon by the Mississippi River Plume. *Remote Sens. Environ.* **2008**, *112*, 836–844. [[CrossRef](#)]
17. Binding, C.; Bowers, D. Measuring the salinity of the Clyde Sea from remotely sensed ocean colour. *Estuarine, Coast. Shelf Sci.* **2003**, *57*, 605–611. [[CrossRef](#)]
18. Bricaud, A.; Morel, A.; Prieur, L. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnol. Oceanogr.* **1981**, *26*, 43–53. [[CrossRef](#)]
19. D'Sa, E.J.; Hu, C.; Muller-Karger, F.E.; Carder, K.L. Estimation of colored dissolved organic matter and salinity fields in case 2 waters using SeaWiFS: Examples from Florida Bay and Florida Shelf. *J. Earth Syst. Sci.* **2002**, *111*, 197–207. [[CrossRef](#)]
20. Opsahl, S.; Benner, R. Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature* **1997**, *386*, 480–482. [[CrossRef](#)]
21. Kim, D.-W.; Park, Y.-J.; Jeong, J.-Y.; Jo, Y.-H. Estimation of Hourly Sea Surface Salinity in the East China Sea Using Geostationary Ocean Color Imager Measurements. *Remote Sens.* **2020**, *12*, 755. [[CrossRef](#)]
22. Siegel, D.A.; Michaels, A.F. Quantification of non-algal light attenuation in the Sargasso Sea: Implications for biogeochemistry and remote sensing. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **1996**, *43*, 321–345. [[CrossRef](#)]
23. Marghany, M.; Hashim, M.; Cracknell, A.P. Modelling Sea Surface Salinity from MODIS Satellite Data. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Fukuoka, Japan, 23–26 March 2010; Springer International Springer: Berlin/Heidelberg, Germany, 2010; Volume 6016, pp. 545–556.
24. Wong, M.; Lee, K.; Kim, Y.; Nichol, J.; Li, Z.; Emerson, N. Modeling of suspended solids and sea surface salinity in Hong Kong using Aqua/MODIS satellite images. *Korean J. Remote Sens.* **2007**, *23*, 1–9.
25. Zhao, J.; Temimi, M.; Ghedira, H. Remotely sensed sea surface salinity in the hyper-saline Arabian Gulf: Application to landsat 8 OLI data. *Estuarine, Coast. Shelf Sci.* **2017**, *187*, 168–177. [[CrossRef](#)]
26. Marghany, M.; Hashim, M. A numerical method for retrieving sea surface salinity from MODIS satellite data. *Int. J. Phys. Sci.* **2011**. [[CrossRef](#)]
27. Marghany, M. Least square algorithm for sea surface salinity retrieving from MODIS satellite data. In Proceedings of the 2009 IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia, 18–19 November 2009; pp. 500–503.
28. Marghany, M. Simulation of Tsunami Impact on Sea Surface Salinity along Banda Aceh Coastal Waters, Indonesia. In *Advanced Geoscience Remote Sensing*; IntechOpen: London, UK, 2014.
29. Geiger, E.F.; Grossi, M.D.; Trembanis, A.C.; Kohut, J.T.; Oliver, M.J. Satellite-derived coastal ocean and estuarine salinity in the Mid-Atlantic. *Cont. Shelf Res.* **2013**, *63*, S235–S242. [[CrossRef](#)]
30. Marghany, M.; Hashim, M. Retrieving seasonal sea surface salinity from MODIS satellite data using a Box-Jenkins algorithm. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Sendai, Japan, 1–5 August 2011; pp. 2017–2020. [[CrossRef](#)]
31. Moussa, H.; Benallal, M.A.; Goyet, C.; Lefevre, N.; El Jai, M.C.; Guglielmi, V.; Touratier, F. A comparison of Multiple Non-linear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic ocean based on satellite data. *ESAIM Proc. Surv.* **2015**, *49*, 65–77. [[CrossRef](#)]
32. Rajabi-Kiasari, S.; Hasanlou, M. An efficient model for the prediction of SMAP sea surface salinity using machine learning approaches in the Persian Gulf. *Int. J. Remote Sens.* **2020**, *41*, 3221–3242. [[CrossRef](#)]
33. Liu, M.; Liu, X.; Liu, D.; Ding, C.; Jiang, J. Multivariable integration method for estimating sea surface salinity in coastal waters from in situ data and remotely sensed data using random forest algorithm. *Comput. Geosci.* **2015**, *75*, 44–56. [[CrossRef](#)]
34. Chen, S.; Hu, C. Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements. *Remote Sens. Environ.* **2017**, *201*, 115–132. [[CrossRef](#)]
35. Medina-Lopez, E. Machine Learning and the End of Atmospheric Corrections: A Comparison between High-Resolution Sea Surface Salinity in Coastal Areas from Top and Bottom of Atmosphere Sentinel-2 Imagery. *Remote Sens.* **2020**, *12*, 2924. [[CrossRef](#)]

36. Yáñez-Arancibia, A.; Day, J.W. The Gulf of Mexico: Towards an integration of coastal management with large marine ecosystem management. *Ocean Coast. Manag.* **2004**, *47*, 537–563. [[CrossRef](#)]
37. Deegan, L.A.; Day, J.W.; Gosselink, J.G.; Yáñez-Arancibia, A.; Chávez, G.S.; Sánchez-Gil, P. Relationships Among Physical Characteristics, Vegetation Distribution and Fisheries Yield in Gulf of Mexico Estu-Aries. In *Estuarine Variability*; Elsevier: Amsterdam, Switzerland, 1986.
38. Brokaw, R.J.; Subrahmanyam, B.; Morey, S.L. Loop Current and Eddy-Driven Salinity Variability in the Gulf of Mexico. *Geophys. Res. Lett.* **2019**, *46*, 5978–5986. [[CrossRef](#)]
39. Fournier, S.; Lee, T.; Gierach, M.M. Seasonal and interannual variations of sea surface salinity associated with the Mississippi River plume observed by SMOS and Aquarius. *Remote Sens. Environ.* **2016**, *180*, 431–439. [[CrossRef](#)]
40. Liang, Z.; Chen, S.; Yang, Y.; Zhou, Y.; Shi, Z. High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling. *Sci. Total. Environ.* **2019**, *685*, 480–489. [[CrossRef](#)]
41. Peng, J.; Biswas, A.; Jiang, Q.; Zhao, R.; Hu, J.; Hu, B.; Shi, Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* **2019**, *337*, 1309–1319. [[CrossRef](#)]
42. Yan, F.; Shangguan, W.; Zhang, J.; Hu, B. Depth-to-bedrock map of China at a spatial resolution of 100 meters. *Sci. Data* **2020**, *7*, 1–13. [[CrossRef](#)]
43. Ma, Z.; Zhou, Y.; Hu, B.; Liang, Z.; Shi, Z. Downscaling annual precipitation with TMPA and land surface characteristics in China. *Int. J. Clim.* **2017**, *37*, 5107–5119. [[CrossRef](#)]
44. Dai, A.; Trenberth, K.E. Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *J. Hydrometeorol.* **2002**, *660*–687. [[CrossRef](#)]
45. Ellis, J.T.; Dean, B.J. Gulf of Mexico Processes. *J. Coast. Res.* **2012**, *60*, 6–13. [[CrossRef](#)]
46. Morey, S.L.; Martin, P.J.; O'Brien, J.J.; Wallcraft, A.A.; Zavala-Hidalgo, J. Export pathways for river discharged fresh water in the northern Gulf of Mexico. *J. Geophys. Res. Ocean.* **2003**, *108*. [[CrossRef](#)]
47. Morey, S.L.; Schroeder, W.W.; O'Brien, J.J.; Zavala-Hidalgo, J. The annual cycle of riverine influence in the eastern Gulf of Mexico basin. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
48. Ma, Z.; Shi, Z.; Zhou, Y.; Xu, J.; Yu, W.; Yang, Y. A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai–Tibet Plateau with the effects of systematic anomalies removed. *Remote Sens. Environ.* **2017**, *200*, 378–395. [[CrossRef](#)]
49. Malone, B.P.; Minasny, B.; Odgers, N.P.; McBratney, A.B. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* **2014**, *34*–44. [[CrossRef](#)]
50. Gordon, A.D.; Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. Classification and Regression Trees. *Biometrics* **1984**, *40*, 874. [[CrossRef](#)]
51. Fu, T.; Zhao, R.; Hu, B.; Jia, X.; Wang, Z.; Zhou, L.; Huang, M.; Li, Y.; Shi, Z. Novel framework for modelling the cadmium balance and accumulation in farmland soil in Zhejiang Province, East China: Sensitivity analysis, parameter optimisation, and forecast for 2050. *J. Clean. Prod.* **2021**, *279*, 123674. [[CrossRef](#)]
52. Fu, Z.; Hu, L.; Chen, Z.; Zhang, F.; Shi, Z.; Hu, B.; Du, Z.; Liu, R. Estimating spatial and temporal variation in ocean surface pCO₂ in the Gulf of Mexico using remote sensing and machine learning techniques. *Sci. Total Environ.* **2020**, *745*, 140965. [[CrossRef](#)] [[PubMed](#)]
53. Kuhn, M.; Weston, S.; Keefer, C.; Coulter, N. Cubist Models for Regression; Vignette R Packag. 2016. Available online: <https://mran.microsoft.com/snapshot/2016-09-15/web/packages/Cubist/vignettes/cubist.pdf> (accessed on 25 February 2021).
54. Oosterbaan, R.J. Frequency and Regression Analysis of Hydrologic Data. In *Drainage Principles and Applications*; International Institute for Land Reclamation and Improvement(ILRI): Wageningen, The Netherlands, 1994.
55. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer New York, Inc.: New York, NY, USA, 1995.
56. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [[CrossRef](#)]
57. Vazquez, J.; Gierach, M.M.; Leben, R.R.; Tsontos, V.M. Initial results on the variability of sea surface salinity from Aquarius/SAC-D in the Gulf of Mexico. In Proceedings of the AGU Fall Meeting Abstracts, San Francisco, FL, USA, 3–7 December 2012.
58. Feng, H.; VanDeMark, D.; Wilkin, J. Gulf of Maine salinity variation and its correlation with upstream Scotian Shelf currents at seasonal and interannual time scales. *J. Geophys. Res. Oceans* **2016**, *121*, 8585–8607. [[CrossRef](#)]
59. Salisbury, J.E.; Campbell, J.W.; Linder, E.; David Meeker, L.; Müller-Karger, F.E.; Vörösmarty, C.J. On the seasonal correlation of surface particle fields with wind stress and Mississippi discharge in the northern Gulf of Mexico. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **2004**, *51*, 1187–1203. [[CrossRef](#)]
60. Schiller, R.V.; Kourafalou, V.H.; Hogan, P.; Walker, N.D. The dynamics of the Mississippi River plume: Impact of topography, wind and offshore forcing on the fate of plume waters. *J. Geophys. Res. Ocean.* **2011**, *116*. [[CrossRef](#)]
61. Wiseman, W.; Rabalais, N.; Turner, R.; Dinnel, S.; Macnaughton, A. Seasonal and interannual variability within the Louisiana coastal current: Stratification and hypoxia. *J. Mar. Syst.* **1997**, *12*, 237–248. [[CrossRef](#)]
62. Feng, Y.; DiMarco, S.F.; Jackson, G.A. Relative role of wind forcing and riverine nutrient input on the extent of hypoxia in the northern Gulf of Mexico. *Geophys. Res. Lett.* **2012**, *39*. [[CrossRef](#)]