Article

Sea Surface Salinity Inversion Model for Changjiang Estuary and Adjoining Sea Area with SMAP and MODIS Data Based on Machine Learning and Preliminary Application

Xiaoyu Zhang [1,2,3], Mingfei Wu [1], Wencong Han [1,*], Lei Bi [1], Yongheng Shang [4] and Yingchun Yang [5]

1   School of Earth Sciences, Zhejiang University, Hangzhou 310027, China
2   Hainan Institute of Zhejiang University, Sanya 572000, China
3   Ocean Academy, Zhejiang University, Zhoushan 316000, China
4   The Engineering Center of High Resolution Earth Observation, Zhejiang University, Hangzhou 310027, China
5   College of Computer Science & Technology, Zhejiang University, Hangzhou 310027, China
*   Correspondence: hanwencong0303@163.com

Abstract: Sea surface salinity (SSS) is one of the most important basic parameters for studying the oceanographic processes and is of great significance in identifying oceanic currents. However, for a long time, the salinity observation in the estuary and coastal waters has not been well resolved due to the technology limitation. In this study, the SSS inversion models for the Changjiang Estuary and the adjacent sea waters were established based on machine learning methods, using SMAP (Soil Moisture Active and Passive) salinity data combined with the specific bands and bands ratios of MODIS (Moderate Resolution Imaging Spectroradiometer). The performance of the three machine learning methods (Random Forest, Particle Swarm Optimization Support Vector Regression (PSO-SVR) and Automatic Machine Learning (TPOT)) are compared with accuracy verification by the in-situ measured SSS. Random Forest is proven to be effective for the SSS inversion in flood season, whereas TPOP performs the best for the dry season. The machine learning-based models effectively solve the problem of insufficient time span of SSS observation from salinity satellites. At the same time, an empirical algorithm was established for the SSS inversion for the sea areas with low salinity (<30 psu) where the machine learning based model fails with great errors. The average deviation of the complex SSS inversion models is −0.86 psu, validated with Copernicus Global Ocean Reanalysis Data. The long term series SSS dataset of March and August from 2003 to 2020 was then constructed to observe the salinity distribution characteristics of the flood season and the dry season, respectively. It is indicated that the distribution pattern of CDW can be divided into three categories: northeast-oriented expansion pattern, multi direction isotropic expansion pattern, and a turn pattern of which CDW shows changing direction, namely the northeast-southeast expansion pattern. The pattern of CDW expansion is indicated to be the comprehensive effect of the interaction of different currents. In addition, it is noteworthy that CDW shows increasing expansion with decreasing SSS in the front plume, especially in the flood season. This study not only gives a feasible solution for effective SSS observation, but also provides a dataset of basic oceanographic parameters for studying the coastal biogeochemical processes, evolution of land-sea interaction, and changing trend of material and energy transport by the CDW in the west Pacific boundary.

Keywords: sea surface salinity; satellite remote sensing; inversion models; machine learning methods; Changjiang Estuary and adjacent sea area

1. Introduction

The Changjiang Estuary is China's largest estuary, with an annual runoff of more than 900 billion cubic meters, accounting for about 57% of China's total estuaries [1]. The Changjiang Dilute Water (CDW), carried with terrestrial materials, is of great significance in transporting material to the west Pacific. The interaction of CDW with saline sea water

has a profound impact on the regional biogeochemical and physical processes. The spatial distribution and temporal variation of CDW and corresponding impacting factors are always hot topics but remain completely unresolved [2,3]. The lack of an effective index with long-term observation, sufficient spatial coverage and resolution is one of the main problems. Although different kinds of parameters are used to trace the CDW, such as sea surface temperature (SST), suspended particulate matter (SPM), colored dissolved organic matter (CDOM) or even Chl-a (Chlorophyll a), sea surface salinity (SSS) has been proven to be one of the most effective indexes in providing direct insight into hydrodynamic processes and energy transfer in marginal seas, land–sea interactions, marine ecology variation under pressure of global climate change and human activity, etc. [4–7].

However, the acquirement of SSS needs to be improved greatly to account for the disadvantages of modern observation technology. Generally, technologies monitoring SSS include on-site observation shipboards or buoys and remote sensing by aerospace satellite [7]. Traditionally, marine salinity monitoring relies heavily on voyage surveys combined with fixed station buoy observations. To overcome the disadvantages of on-site observation with low temporal and spatial resolution, limited spatial coverage and the inability of dynamic synchronous observation, observation networks at different levels have been constructed globally and/or regionally [8]. Among them, Argo has become an important means of obtaining global ocean temperature and salinity data [9]. Argo has advantages in automatically collecting, saving online data and transmitting data through communication satellites, but as a device that flows with the currents, Argo is unable to collect data in a fixed station or even area, and the spatial resolution of Argo data is low, most of all, Argo is only deployed in sea areas deeper than 2000 m to observe the vertical profile from surface to deep [10]. Assimilation simulation based on the 3D hydrodynamic model is a promising technology which can simulate the salinity distribution and the variable dynamics of a specific area well [11], but the problem of the scarcity of high-precision historical data remains unresolved.

In recent years, the rapid development of satellite remote sensing technology has provided a new way to monitor SSS synchronously from the aerospace covering a large sea area with an acceptable spatial resolution and temporal frequency. Currently, Soil Moisture and Ocean Salinity (SMOS), Aquarius, Soil Moisture Active and Passive (SMAP) are the three main satellites used to monitor sea surface salinity, and have been successfully launched and are operating normally [12–14]. All three are equipped with L-band microwave radiometers to observe the sea surface brightness temperature. Particularly, Aquarius and SMAP have the ability to perform sea surface roughness correction as they are equipped with scatterometers, which are conducive to obtaining SSS products with higher precision [15–17]. However, their capability in observing coastal sea waters from salinity satellites is greatly limited due to serious errors. Moreover, the spatial resolution of satellite-obtained SSS is too low to identify the subtle interaction of currents in marginal sea. In addition, the narrow temporal span makes it difficult to obtain temporal variation over a long period with impacts from human activity and global changes.

To improve the spatial resolution and temporal span of satellite SSS data, this study aims to establish a complex SSS inversion model based on salinity satellite data combined with ocean color satellite data. Concerning the low spatial resolution of Aquarius\SSS, and the unavailability of SMOS\SSS in the sea areas around Changjiang Estuary, SMAP\SSS data is finally chosen to participate in the model construction. The accuracy of SMAP\SSS in the study area has been verified by its predecessors with the in-situ measured SSS [7]. The Moderate Resolution Imaging Spectroradiometer (MODIS) was selected to provide oceanic color products for this study. At present, the MODIS has been widely used in offshore investigations due to its comprehensive observation capabilities. MODIS has the advantages of a long temporal span, open data access, high temporal and spatial resolution, and plentiful oceanic color products including SST, Chl-a, primary productivity, SPM, CDOM and so on [18]. Previous studies have proved that the incorporation of ocean color data, such as Chl-a, SPM and CDOM, can improve the accuracy of SSS inversion [19]. In this

study, CDOM-related information is utilized for the strong negative relationship between SSS and CDOM, which has been widely observed in coastal waters [20–23]. However, it should be noted that systematic errors can be introduced due to the uncertainty of their internal relationships [24], and the matching sampling points are too few to ensure the temporal and spatial applicability of the constructed model. Meanwhile, since SST has the capability of capturing the temperature differences between riverine freshwater and oceanic saline water, especially in the upwelling region [23], the monthly average product of reflectance combined with MODIS\SST is used as input, and SMAP\SSS product is used as the true value. Concerning the great difference of the Changjiang River runoff in the flood and dry seasons, Random Forest, Particle Swarm Optimization Support Vector Regression (PSO-SVR), and Automatic Machine Learning (TPOT) were used for training in different seasons. The accuracy of the inversion model is verified with the in situ measured test dataset. The model with the best performance is selected to inverse SSS for the study area. Particularly, the coastal area is filled with SSS inversed by empirical models constructed from in situ measured data to compensate for the incapability of salinity satellites in coastal waters with low salinity and high turbidity. The combined inversion models are applied to SSS observations from 2003–2020 for the study area. The spatial expansion patterns and the temporal changing trend of CDW are then discussed. This study provides a way to obtain SSS in the study area with high spatial and temporal resolution. The observation of temporal variation of SSS distribution over a longer period can be achieved, which is conducive to a better understanding of the change of material transport and estuarine geochemical processes of marginal seas in the west Pacific under severe human activities and global warming.

2. Research Area and Data

2.1. Study Area

This paper takes the Changjiang Estuary and its adjacent waters as a study area, including the north of the East China Sea (ECS) and the south of the Yellow Sea. The geographic location is between 26°~36°N, 120°~130°E as shown in Figure 1. The study area has a typical subtropical monsoon climate with an average annual precipitation of more than 1000 mm. The region is controlled alternately by tropical oceanic air masses and polar continental air masses, with distinct seasonal variations in both the climate and precipitation [25].
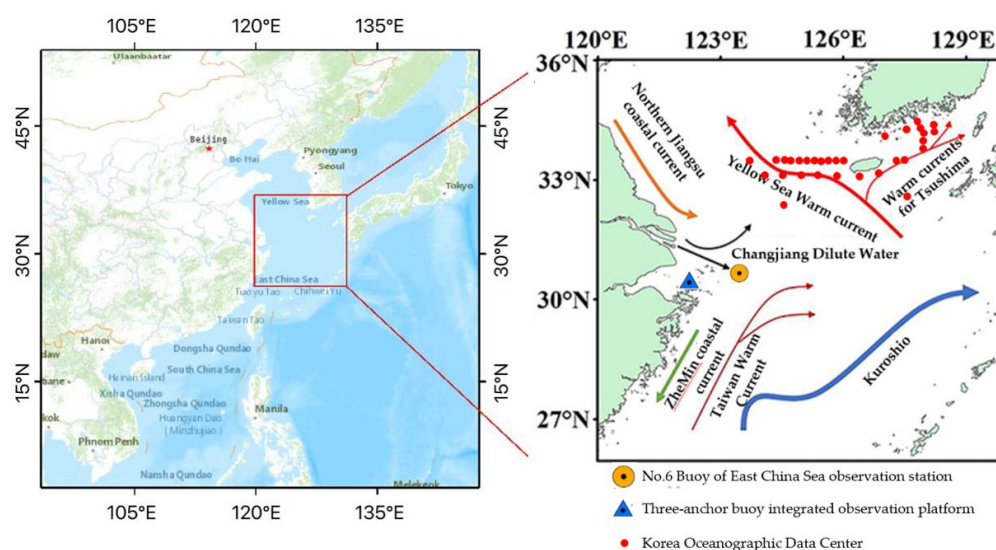

Figure 1. Sketch map of Geographical location, main currents and the stations collecting the in-situ data of the study area.

The hydrodynamic environment in the study area is very complicated with comprehensive impacts from Changjiang Diluted Water (CDW), Taiwan Warm Current (TWC), Yellow Sea Warm Current (YSWC) and ZheMin Coastal Current (ZCC), as shown in Figure 1. The interaction between the fresh water of CDW and different marine currents with saline sea water induces a unique SSS distribution pattern with significant seasonal differences [26]. The monsoon and tide constrain the expansion of CDW further as well [6,27].

TWC is a branch of the Kuroshio and is an important current developed from the South China Sea, flowing northward to the ECS. Usually, TWC exhibits distinct seasonal variation; it is strong in summer with high temperature and low salinity, and weak in winter, showing relatively lower temperature and higher salinity [28]. The TWC has a stable flow path which migrates northward to the mouth of the Changjiang River [29]. In addition, the northward Kuroshio has had a profound impact on the expansion of CDW through its invasion of the continental shelf of ECS [30]. Generally, the front plume of the CDW extends to the northeast due to the southward YSWC with high temperatures and salinity in the dry season. It is generally believed that YSWC is a unique circulation formed from convergence of the TWC, the Tsushima Warm Current and the Yellow Sea rather than derived from the Tsushima Warm Current solely [31,32].

2.2. Satellite Data

Satellite data used in this study includes SMAP salinity products and MODIS ocean color products. The SMAP salinity products was produced by the Remote Sensing Systems (RSS) team and sponsored by the NASA Ocean Salinity Science Team (http://www.remss.com, accessed on 1 September 2021). A monthly average SMAP\SSS L3 grid product with a spatial resolution of 0.25° was selected in this study, and the data is stored in the NetCDF (Network Common Data Form) format [33]. The average accuracy of SMAP\SSS is indicated to be 0.5 psu, which was verified with HYCOM\SSS at a unified resolution, and the mean square error is 0.25 psu, which was evaluated with Argo buoys in the middle and low latitudes. Each monthly file contains information on salinity and wind speed.

MODIS contains 36 bands covering the visible light to the infrared spectrum, of which nine bands are specially designed for oceanographic applications. The one or two time revisits of frequencies per day ensure its capability of dynamic observations, the spatial resolution of the kilometer level is high enough for oceanographic research and has broad application values [34]. In this study, Aqua\MODIS L3 monthly Rrs (remote sensing reflectance) and SST (4 km) products (https://ocean-color.gsfc.nasa.gov/ accessed on 1 September 2021) are used as the input for model building and long time series inversion combined with SMAP\SSS data after spatio-temporal matching. Aqua\MODIS L3 Daily Rrs and SST (4 km) are used for matching with in situ measured SSS data to verify the model. Aqua\MODIS L2 Daily Rrs (1 km) is used to construct the empirical regression model for the near-shore sea waters with low salinity [18]. The specific MODIS products applied in this study are shown in Table 1.

Table 1. List of satellite products used in this study and their usage.

Products	Temporal Resolution	Spatial Resolution	Temporal Span	Function
Aqua\MODIS L3 Rrs and SST	Monthly	4 km	2003–2020/March August October	For SSS inversion model construction; for inversion of sequential SSS
Aqua\MODIS L3 Rrs and SST	Daily	4 km	15–17 August 2020; 28 April 2020	Calibration

Table 1. *Cont.*

Products	Temporal Resolution	Spatial Resolution	Temporal Span	Function
RSS\SMAP L3 SSS	Monthly	40 km	2016–2018/March August; 2020/March August	Ground truth as training data and validation model inversion effect
Aqua\MODIS L2 Rrs	Daily	1 km	2014–2015; 2018–2019, total 40 spatio-temporal matching points	Modeling and Verification of Supplementary Empirical Models for Offshore Sea Areas

2.3. Field Data

The in situ measured SSS used in this paper comes from the Network of Serial Oceanographic Observations (NSO) of the Korea Fisheries Science Center (https://nso.edu, accessed on 10 October 2021), the No. 06 buoy and the three-anchor buoy integrated observation platform of the ECS. The locations of the stations are shown in Figure 1.

NSO is a large-scale offshore observation project that was started in 1961. At present, the network has 25 survey sections with a total of 207 stations, monitoring 17 marine parameters including sea temperature, dissolved oxygen, salinity, zooplankton, phytoplankton and so on. The data can be downloaded from the website of the Korea Oceanographic Data Center (KODC) (https://www.nifs.go.kr/kodc, accessed on 20 October 2021). Since external users can only access data over a time span of up to one year ago, this paper filters the measured NSO data from the archive of 2020 to match the satellite data. All the matching in situ data are measured at depths of 3 m with verified quality [35].

Both the No. 06 buoy and the three-anchor buoy integrated observation platforms in the ECS are equipped with the "Conductivity, Temperature, Depth" detection devices. The No. 06 buoy began to work in October 2014, collecting the temperature, salinity and depth records of four profiles for the following 412 consecutive days by using the anchor chain suspension method. The three-anchor buoy comprehensive observation platform of the ECS works similar to that of the No. 06 buoy in acquiring sea surface parameters, but the underwater data collection is carried out by using a local network, which can remotely control the lifting and lowering of the underwater observation unit and obtain data. Screening in the archives of the surface salinity data at 0 m of No. 06 buoy station from 2014 to 2015 and the three-anchor buoy comprehensive observation platform from 2018 to 2019, a total of 40 sets of measured data were obtained.

2.4. Copernicus Global Ocean Reanalysis Data

The Copernicus Programme is one of the Earth observation programs initiated by the European Union. It aims to provide ocean data for a wide range of researchers with a shared and innovative platform. The Copernicus project provides dynamic observation and reanalysis data in the fields of ocean physics, sea ice and marine biogeochemistry on a global scale [36]. The global $1/12°$ ocean physics reanalysis dataset was produced and distributed by the Mercator Ocean International Team under the Copernicus Project, which developed a complex ocean simulation system (a numerical model) based on ocean observation data of both satellite data and in situ measurements. It can describe, analyze and predict oceanic physical and biogeochemical processes in real time. It produces mapping data such as temperature and salinity at different water depths of the global ocean (containing 50 horizontal layers ranging from 0 to 5000 m) stored and distributed in NC format. This paper uses the gridded monthly mean data of seawater salinity at a depth of

0.5 m as the validation data to evaluate the accuracy of the combined model constructed in this study.

3. Methodology

Figure 2 shows the basic technical flow of this paper. Firstly, an SSS inversion model with machine learning methods is constructed for the study area with salinity >30 psu, the MODIS\Rrs and MODIS\SST L3 monthly products are used as the input, and SMAP\SSS is used as the true value. The inversion SSS by different machine learning models is evaluated with in situ measured SSS from NSO for different seasons, and the model with the best performance is selected. Then, an empirical statistical model for sea areas with a salinity <30 psu is constructed from in situ measured data and validated. With the establishment of an SSS inversion process, the machine learning models combined with empirical models are applied to obtain SSS for March and August as representatives of the dry season and flood seasons from 2003 to 2010. Finally, the long-term analysis of the SSS temporal-spatial variation trend in the study sea area is performed.
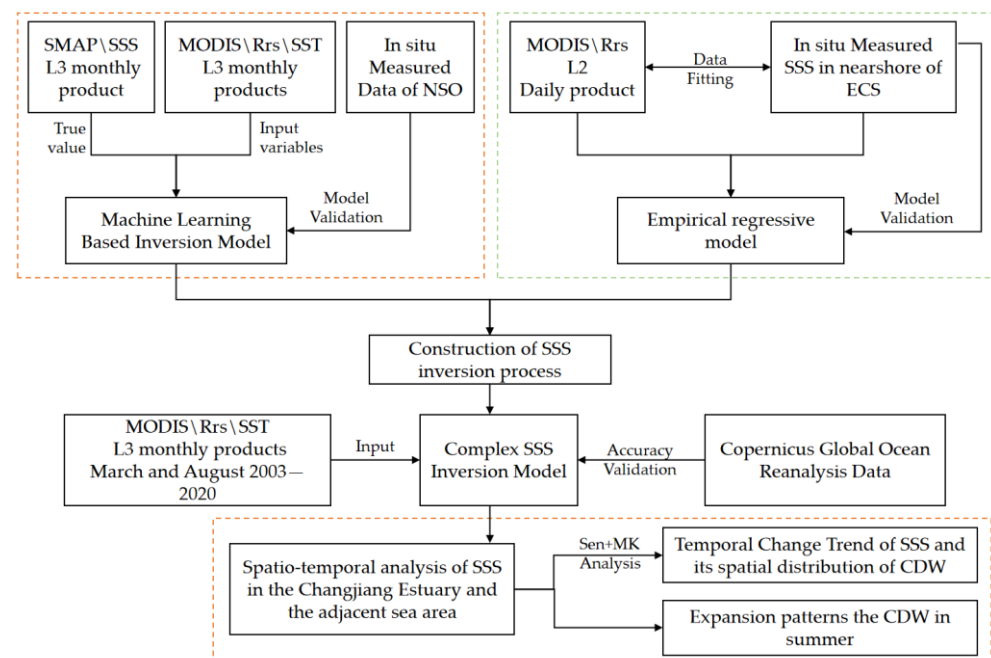

Figure 2. The technology road mapping used in this study.

3.1. Data Preparation

First, MODIS\Rrs and MODIS\SST are resampled using a cubic convolution method to obtain data with the same spatial resolution as SMAP\SSS. This method fits a smooth curve based on the centers of the 16 nearest input cells to determine the new value for that cell. This method was indicated to be appropriate for continuous data. It is geometrically less distorted than the raster achieved by running the nearest neighbor resampling algorithm, although it may result in an output raster containing values outside the range of the input raster [37].

Second, concerning the species and concentration of CDOM will affect the reflectance from the blue to red band of the spectrum, bands in the range of 350–700 nm are widely adopted for the construction of CDOM inversion models [38]. Particularly, five spectral bands (412, 443, 488, 547 and 667 nm) are used as the single band input variables. In addition, regarding the high accuracy of SSS inversion models established previously for

this study sea area with the simple regression method, this paper adopts the band ratio as an input to enhance the accuracy of the model (Equations (1) and (2)) [39].

$$
\begin{aligned}
&The\ Flood\ Season: 34.147e^{-0.285*Rrs555/Rrs443}\\
&the\ Dry\ Season: 98.984e^{-1.056Rrs555/Rrs469}
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&The\ Flood\ Season: 34.147e^{-0.285*Rrs555/Rrs488}\\
&the\ Dry\ Season: 98.984e^{-1.056Rrs555/Rrs488}
\end{aligned}
\tag{2}
$$

$Rrs555/Rrs443$ and $Rrs555/Rrs488$ were selected for the flood season, while $Rrs555/Rrs469$ and $Rrs555/Rrs488$ were selected for the dry season, respectively. Moreover, SST is even more pronounced in indicating the front plume in some cases. Thus, MODIS\SST is adopted as an important variable that can capture the characteristics of CDW expansion. However, it should be kept in mind that SSS and SST do not always change synchronously.

To improve the SSS inversion accuracy, this paper aims to establish different inversion models for the dry and flood seasons, respectively. $Rrs412$, $Rrs443$, $Rrs488$, $Rrs547$, $Rrs667$, $Rrs555/488$, $Rrs555/443$ and SST are inputs in the flood season, whereas Rrs412, $Rrs443$, $Rrs488$, $Rrs547$, $Rrs667$, $Rrs555/488$, $Rrs555/469$ and SST are inputs in the dry season.

Finally, normalization of the input data is performed to reduce the inversion error, which may be induced by the uniformity of the dimensions of each input. The formula for normalization methods was given as follows:

$$
Sstd = \frac{S - Smin}{Smax - Smin}
\tag{3}
$$

S_{std} is the normalized index obtained; S_{min} and S_{max} represent the maximum and minimum of the original data set, respectively.

After comprehensively considering the lack of MODIS reflectivity data in typical months of the flood and dry seasons in the study area, this paper selects March and August as the representatives of the dry season and flood seasons, respectively, from 2016 to 2018 as training data. A total of 2916 valid points were matched in the flood season and 2550 valid points were matched in the dry season. When dividing the training set and the test set, the principle of hierarchical stochastic is followed; that is, the data is classified with a salinity gradient, and the data are randomly selected as the training set and the test set under each classification to prevent errors caused by excessive concentration of the salinity range. The percentage of the training set and the test set is 80% and 20%, respectively.

3.2. Machine Learning Methods

Three conventional machine learning methods, including Random Forest, PSO-SVR and Automatic Machine Learning (TPOT), are introduced to establish the seasonal SSS inversion model (for the flood season and the dry season, respectively). The inversion accuracy of the three methods is compared, and the method with the highest accuracy is finally adopted.

3.2.1. Random Forest

Random forest is a classification/regression method that uses decision trees as the base learner to build a bagging ensemble and further introduces random attribute selection in the training process of decision trees [40]. Random Forest is powerful in modeling nonlinear relationships, which can reduce over-fitting and improve the noise-resisting ability. Regarding regression, the final predicted value of the output is the mean of the predicted values of each tree. Two important parameters, the number of decision trees, namely as Ntree, and the number of features, namely as Mtry, are adjusted, especially when modeling.

3.2.2. PSO-SVR

Support Vector Machine (SVM), first proposed by Cortes and Vapnik in 1995 [41], performs very well in solving small-sample, nonlinear and high-dimensional pattern recognition problems [42]. SVM is based on the Vapnik-Chervonenkis (VC) dimensional theory of statistics and the structural risk minimum (SRM) principle. It seeks the best compromise between model complexity and learning ability based on limited sample information. SVM can be extended to solve regression problems, referred to as Support Vector Regression (SVR). When predicting, SVR maps the input independent variables to the high-dimensional feature space and converts the original nonlinear model into a linear model in the feature space [43]. According to the principle of structural risk minimization [44], the regression fitting process of SVR can be equivalent to the extreme value optimization problem.

For the available kernel functions, this paper tries linear (linear), polynomial (poly) and Gaussian kernel functions (RBF) to select the optimal kernel function. There are three main parameters that may affect the accuracy of the model, namely the selection of the kernel function, the penalty factor C and the kernel function parameter g [45].

Particle Swarm Optimization (PSO) is a swarm intelligent biological heuristic evolution algorithm. It has the advantages of high accuracy and rapid convergence. The solutions to the optimization problem in the PSO algorithm are called particles, and they have a position vector and a velocity vector. Assuming that there are M particles in the population, the initial position of the first particle and its velocity are generated by random selection [46]. In this paper, the mean squared error (MSE) is used as the fitness function, and the PSO algorithm is used to optimize the relevant parameters of SVR. The basic principle of SVR was introduced in the previous section, in which the penalty factor C and the kernel function parameter gamma are the main factors affecting the SVR fitting.

3.2.3. Automatic Machine Learning (TPOT)

Automatic Machine Learning (TPOT) is a new model construction method that automatically selects algorithms and automatically optimizes parameters. TPOT is one of the classic frameworks of automatic machine learning and has the ability to generate any tree structure process. The algorithms of each node in the process can be randomly matched and evolved through genetic programming to optimize the ideal machine learning process [47]. The parameter optimization of TPOT is based on genetic programming. The basic idea of genetic programming draws on the principles of biological evolution and genetic theory in nature. It is a method of automatically and randomly generating search programs. The genetic programming parameter settings in this study are shown in Table 2.

Table 2. Genetic Programming parameters setting.

GP Parameters	Content
Population size	50
generation	50
Individual mutation rate	90%
Individual crossover rate	10%
Filter Methods	10% Elite Retention Strategy, 2 in 3 Tournament Selection, and 1 in 3 based on complexity
Mutation	Replacement, insertion, and deletion accounted for one-third of each type of mutation
Times of repetition	30

3.3. Construction of Empirical Statistical Model for Offshore Sea Waters

An Empirical Statistical Model is constructed for SSS inversion in near-shore sea area with a salinity lower than 30 psu. The in situ SSS is from the two buoy stations of NSO located at the mouth of the Changjiang Estuary and the mouth of Hangzhou Bay, respectively. The salinity measured in the two stations varies from 20–31 psu. To make

up for the lack of spatial distribution with time continuity, the MODIS\Rrs L2 product is used for spatio-temporal matching. A total of 40 valid data points are matched. Due to the limited samples, one empirical model is constructed for both the flood and dry seasons. Among them, 30 samples are used to build the model, and the other 10 samples are used for validity verification.

Especially, the screening in the archives of the surface salinity data at 0 m of the No. 06 buoy station from 2014 to 2015 and the three-anchor buoy comprehensive observation platform from 2018 to 2019, a total of 40 sets of measured data were obtained. The samples are matched at the spatial and temporal scales of ±1 km and ±6 h. A Pearson correlation coefficient is used to analyze the correlation between each single band and band ratio combination with SSS, as shown in Figure 3.
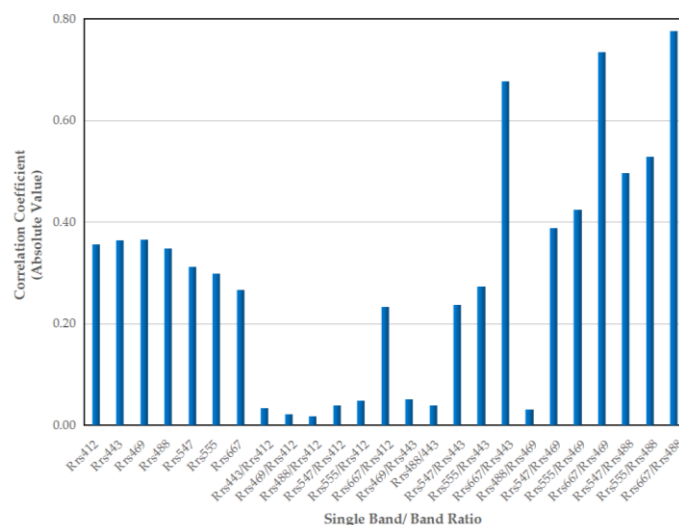

Figure 3. The correlation coefficient between MODIS\Rrs, band ratios and measured SSS.

The band ratios ($Rrs667/Rrs488$) with the most robust correlation coefficient are selected and then fitted with linear, exponential, logarithmic and power mathematical forms. The power regression model with the highest correlation coefficient is used (Table 3).

Table 3. Correlation between band ratio and in situ measured SSS based on different regressive methods.

Bands/Bands Ratio	Fitting Methods	Correlation Coefficient
$Rrs667/Rrs488$	linear	−0.776
	exponential	−0.797
	logarithmic	−0.844
	power	−0.869

3.4. Sen + MK Trend Analysis

The temporal variation trend of SSS is analyzed with Sen's slope estimator combined with the Mann–Kendall method (Sen + MK). Sen's slope estimator is a classic trend calculation method based on non-parametric statistics. Compared with the least squares regression method, the estimated deviation caused by extreme values and outliers has a significant improvement in accuracy and reliability [48]. The formula for calculating the Sen slope is shown in Equation (4):

$$k_{\text{sss}} = \text{Median}\left(\frac{\text{SSS}_j - \text{SSS}_i}{j - i}\right), \ \forall j > i \tag{4}$$

In the formula, k_{sss} is the Sen slope, i and j are the year sequence, SSS_i and SSS_j are the SSS value of a pixel in the year, and Median is the median processing.

The Sen slope method can effectively calculate the variation trend of SSS at each position globally, concerning its inability to determine the significance of the variation trend, the Mann–Kendall method, a non-parametric test method, is introduced, which improves the anti-noise ability greatly. Moreover, with the Sen + MK method, it is not necessary for the sequence data to obey a specific distribution [49].

4. Results and Analysis

4.1. Training Results of Machine Learning Models

With the three different machine learning methods, SSS inversion models are constructed and the training results are compared, respectively. The model with the best accuracy is adopted to inverse SSS.

4.1.1. Random Forest

For the model constructed by Random Forest, the Mean Absolute Error (MAE) and correlation coefficient R^2 are used as an evaluation index, and GridsearchCV is adopted to determine Ntree and Mtry. It is determined that Ntree = 250, Mtry = 3 for August, and Ntree = 175, Mtry = 4 for March. The feature importance of the output of the random forest model established in August and March is shown in Figure 3. With the band ratio input, feature importance in both August and March is improved greatly, implying the significance of incorporation of CDOM optical properties in indicating SSS in the study area. In addition, it is worth noting that SST is more important in March, while $Rrs667$ is more important in August. $Rrs667$ is an important index indicating the suspended solids, the significance of $Rrs667$ can be due to the huge amount of sediments carried by CDW during the flood season; on the contrary, with the decrease of CDW runoff and the increase of SST difference between terrestrial runoff and saline sea water in the dry season, SST gradually replaced the SSS and became the leading role in the SSS inversion. The final accuracy of the test dataset is shown in Figure 4.
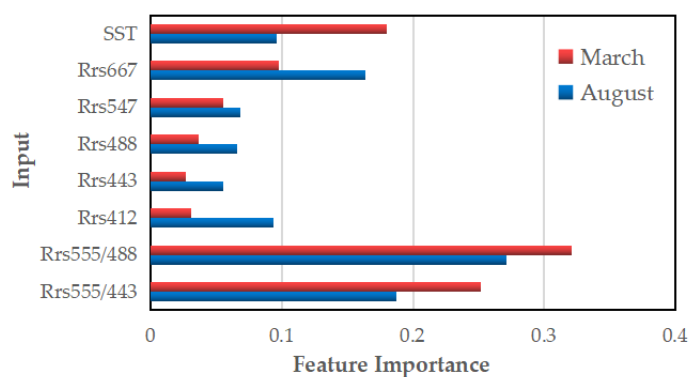

Figure 4. Feature importance of random forest output.

4.1.2. PSO-SVR

The first step involved in the construction of SSS inversion models with the PSO-SVR method is to choose the kernel function. We used a five-fold cross-validation in the training dataset to evaluate the performance of the kernel function (Table 4). The RBF kernel function performs better in both August and March. The RBF kernel function has been proven to be effective in dealing with nonlinear problems. Therefore, only the penalty factor C and the kernel parameter g need to be adjusted. This study uses pop = 40, max_iter = 100, and MSE as the fitness function to set the PSO algorithm to optimize SVR. It is determined that C = 5, g = 0.96 in August, C = 20 and g = 8.38 in March. The convergence rate of optimization in August is obviously slower than that in March.

Table 4. Selection of kernel function.

Seasons	MSE	Kernel		
		RBF	Poly	Linear
August		0.79	0.77	0.73
March		0.74	0.72	0.64

4.1.3. Automatic Machine Learning

MSE is used as the evaluation function, and the optimized machine learning process for August and March was constructed according to the genetic programming parameter settings described in Table 2. This study chooses 50 iterations and the population size is 50. Compared with Random Forest and PSO-SVR, the TPOT runs significantly slower due to its automatic search based on optimized pipelines. The accuracy of the test dataset is shown in Figure 5. In the case of sufficient iterations, the accuracy of the TPOT model is relatively high and stable.
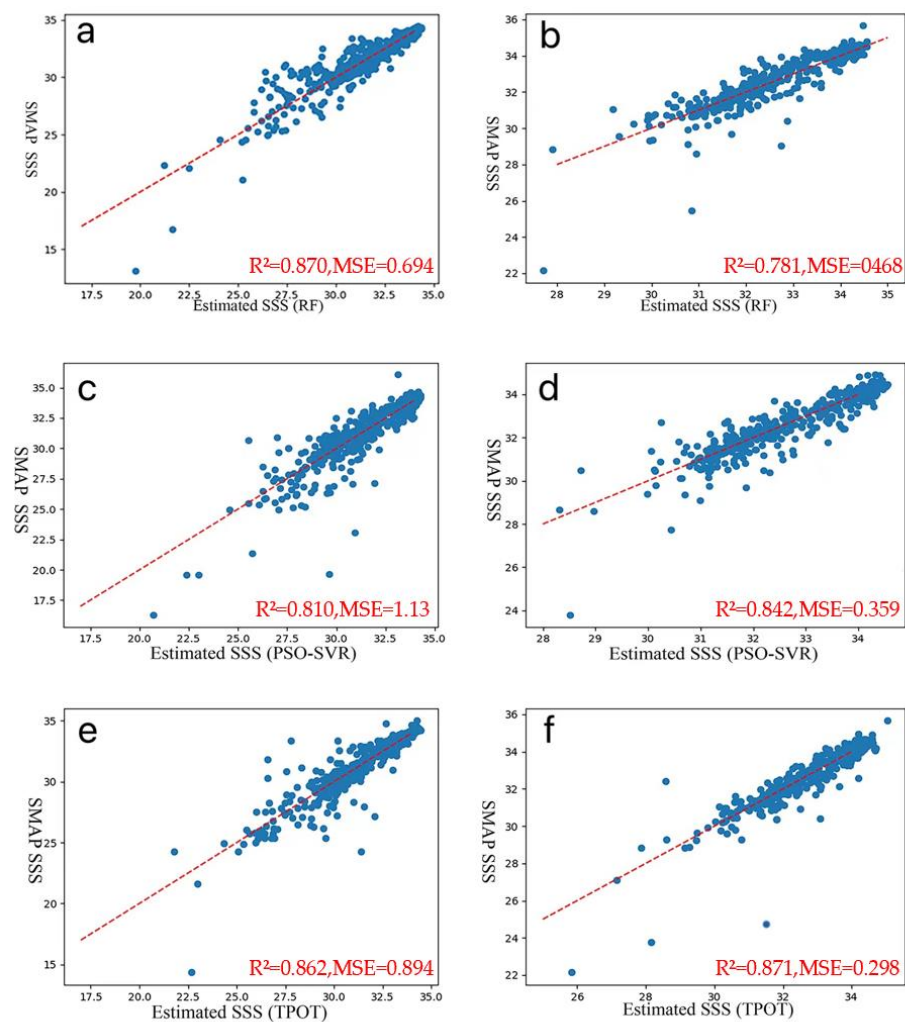
Figure 5. Accuracy validation of Random Forest ((**a**): August; (**b**): March), PSO-SVR ((**c**): August; (**d**): March), TPOT ((**e**): August; (**f**): March). Both horizontal and vertical coordinate units are psu.

4.2. Calibration of Inversion Results of Machine Learning Based Models

According to the verification results of the three different machine learning methods with the test dataset, the Random Forest was selected to invert SSS in the flood season, and the TPOT was utilized to invert SSS in the dry season. However, the inversion results of

the two methods in the region with low SSS (<30 psu) are unsatisfactory. Possible reasons are supposed as follows:

- For the low spatial resolution of SMAP data, few samples are available in coastal regions with low salinity for training.
- The accuracy of SMAP\SSS products in the near-shore waters with low salinity is particularly low.
- The relationships between SSS in the coastal water with the MODIS reflectance and SST are unstable because of the uncertainty of relationships between optical active components such as SPM, CDOM, etc. and SSS.

Since the inversion model constructed in this paper was designed to solve the problem of observing SSS in a longer time span and at larger space scales, the original data involved in the inversion are all monthly average products, which means that each pixel has been averaged on the time scale. Therefore, we speculate that the mapping relationship is still stable at the original monthly average data (4 km) resolution of MODIS.

To verify the model's accuracy, the measured SSS of the National Institute of Fundamental Studies (NIFS) was used for comparison. In order to maintain consistency with the data used in the previous modeling, the daily average Rrs and SST (4 km) of MODIS Aqua L3 products are used as input, which matches the measured SSS with a spatial-temporal resolution of ± 4 km and ± 12 h. A total of 19 points from August 15 to 17 is matched during the flood season, and a total of 13 points on April 28 is matched during the dry season (no SSS is available for March). The comparison between the inversion SSS and the measured SSS is shown in Figure 6. Due to the small fluctuation of the SSS range of the sampling point, MAE was selected to evaluate the inversion effect. The results indicate MAE = 0.50264 in the flood season and MAE = 0.263322 in the dry season, respectively. Although the number of sampling points is relatively small and the SSS varies in a concentrated range, the model shows generalization ability with a high spatial resolution.
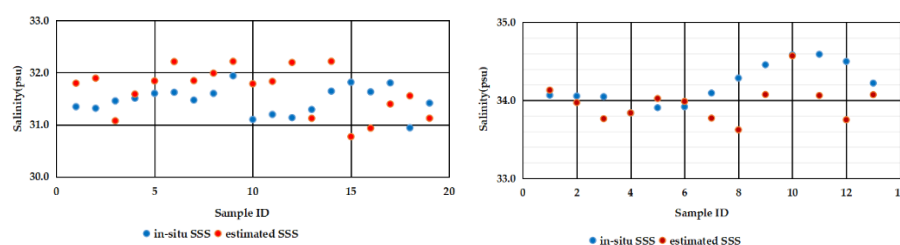

Figure 6. Comparison of in-situ SSS and estimated SSS. The broken lines of August (**left**) and April (**right**) were shown in the chart.

The monthly average products of MODIS in March and August 2020 were used as inputs in the inversion model to obtain SSS. Figure 7 shows the inversion results and the current monthly SMAP\SSS for comparison. The inversion result in August was found to be consistent with the SMAP\SSS, the model results showed finer details around the CDW boundary. By comparing the measured SSS near Jeju Island in 2020 (Figure 6), it is found that the regional salinity fluctuates in the range of 28–30 psu in the SMAP data, and the inversion results show that the salinity is above 31 psu, which is highly consistent with the actual measurement. The inversion results in April are basically consistent with the salinity boundary of SMAP SSS, but some low-value areas in the central part are obviously underestimated. On the whole, the inversion results have a certain accuracy and can be used for long-term research observations such as the observation of the Yangtze River dilute water range.
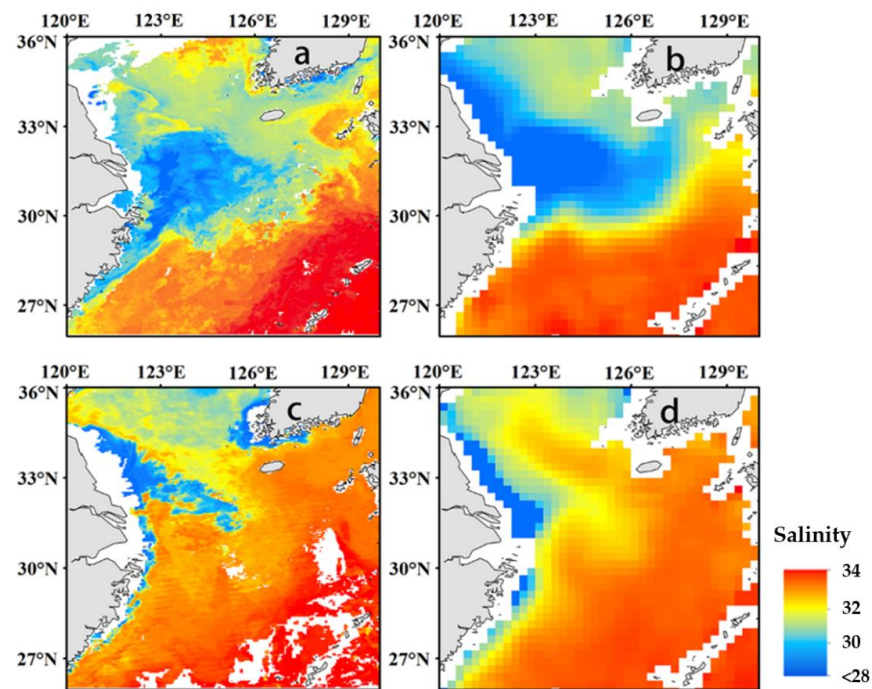
Figure 7. Comparisons between the inversion SSS and monthly SMAP\SSS product ((**a**,**c**) are the inversion SSS for August and March 2020; (**b**,**d**) are SMAP\SSS product in August and March 2020).

4.3. Offshore Empirical Statistical Model Validation

In this paper, the *SSS* inversion model for coastal sea waters with low salinity (<30 psu) is constructed based on the empirical method as mentioned in Section 3.3.

The correlation between each single band and band ratio combination with SSS was analyzed (Figure 3), and it was found that $Rrs(667)/Rrs(488)$ had the strongest correlation with the measured *SSS*. Then, the correlation coefficients between $Rrs(667)/Rrs(488)$ and *SSS* based on the linear, exponential, logarithmic and power mathematical forms were compared (Table 3), and finally, the power form was selected as the *SSS* inversion model.

The formula is shown in Equation (5):

$$SSS = 22.217 \left(\frac{Rrs(667)}{Rrs(488)} \right)^{-0.16} \tag{5}$$

The accuracy of the model is validated by the in situ measured data, as shown in Figure 8. As a model constructed in the context of a limited target area, it shows high accuracy.
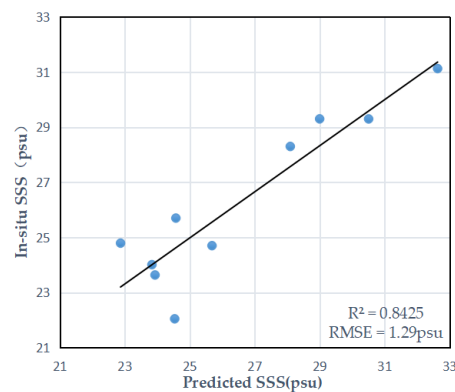

Figure 8. Accuracy evaluation of inversion SSS by empirical model for the near shore sea waters.

4.4. SSS Inversion Process Based on Complex Models

This paper establishes an inversion process based on complex models. The procedure is shown in Figure 9. First, the optimal machine learning based model is used to invert the SSS in August and March. Then, the values >30 psu are retained, whereas the pixels of <30 psu are replaced with new values recalculated by the empirical model. Finally, the new inversion result is output as SSS images for that month.
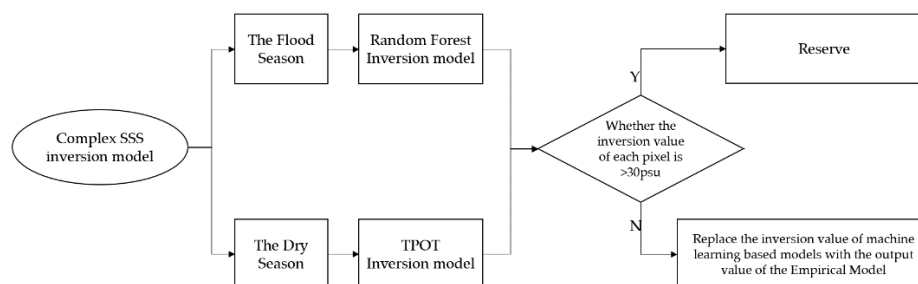

Figure 9. The process of SSS inversion based on the complex models.

To show the effect of SSS inversion with the complex models, SSS inversion in August 2020 is taken as an example. SSS inversion based on the complex models is shown in Figure 10a, whereas SSS inversion with the machine learning model is only shown in Figure 10b. It can be seen that the SSS distribution in the Changjiang Estuary and the adjacent sea area is greatly improved with complex models. Particularly, the front plume of the CDW is highlighted significantly with the accurate delineation of low salinity.
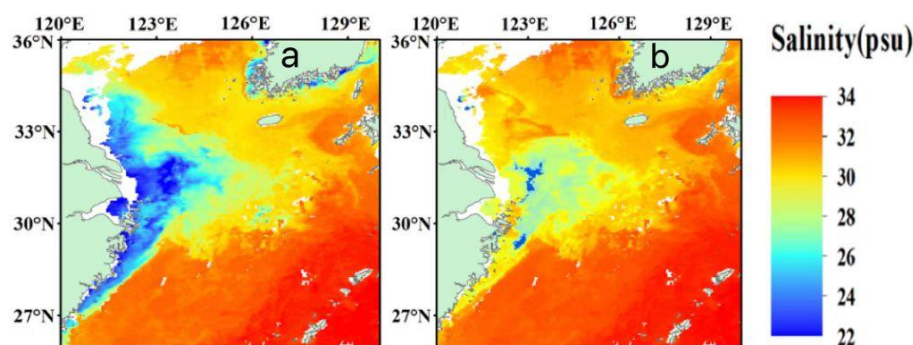

Figure 10. The inversion results in August 2020 based on complex models (**a**) and with machine learning based model only (**b**).

In order to further evaluate the accuracy of the SSS inversion results with the complex models, this paper uses the monthly average data of Copernicus Global Ocean Reanalysis to validate the inversion results. In the correlation analysis between Copernicus reanalysis data and combined inversion results in August 2020, the total number of matching points was 44,899, the correlation coefficient was 0.763, and the average deviation is −0.86 psu. Therefore, the SSS inversion results are in good agreement with the Copernicus reanalysis data. The validation suggests that the complex models constructed in this study are efficient in SSS observations for the study area.

5. Application: Temporal Variation of Spatial Expansion of CDW

With the complex models constructed in this study, the spatial distribution of SSS in the Changjiang Estuary and the adjoining sea area in March and August from 2003 to 2020 is inverted (as shown in Appendix A Figures A1 and A2). The expansion pattern of the CDW in August is summarized, and the temporal variation trend of SSS in August and March is then analyzed, respectively.

5.1. Expansion Pattern of CDW in Summer

Generally, in summer, the salinity of CDW changes dramatically in the range of 22–29 psu, which is the salinity fluctuating range of the plume front of CDW, representing the core area of the CDW [50]. After carefully comparing the inversion SSS images in August from 2003 to 2020 (Figure 11, Appendix A Figures A1 and A2), and referring to previous related studies [51], 27 psu is indicated as the threshold to define the boundary of CDW. The expansion patterns of the CDW in August are then summarized into three categories: the first category is the northeast-oriented expansion pattern, including the northeast expansion subtype, the north-northeast expansion subtype and the northeast-east expansion subtype. The other two are multi-direction isotropic expansion pattern and a turn pattern in which CDW shows changing direction, namely the northeast-southeast expansion pattern. Among them, the northeast-oriented expansion pattern is the dominant pattern of CDW expansion.
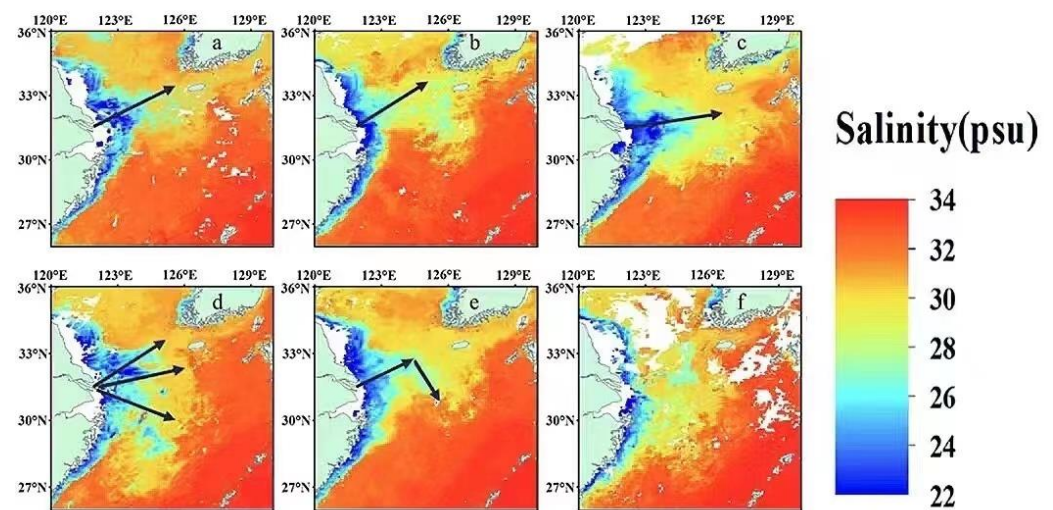

Figure 11. The expansion type of the Yangtze River's diluent water ((**a**) is the northeast expansion type in 2005, (**b**,**e**) are the northeast expansion type in 2004 and 2017, (**c**) is the northeast expansion type in 2020, (**d**,**f**) are the multidirectional expansion type in 2014 and 2015 type).

In most years, the CDW is deflected northeastward by factors such as terrain, runoff, wind speed, circulation and their combined actions. For example, the enhanced TWC or southeasterly wind will make the diversion of the CDW more pronounced and thus expand northernly. The multi-direction expansion pattern and the turn pattern are special cases due to the special marine dynamic conditions of the year. The multi-directional expansion type means that the CDW expands in different directions at the same time. For example, the CDW expands in the northeast and southeast directions at the same time in the SSS inversion image in 2015. The isotropic expansion pattern occurred in 2014 and 2019, the low SSS area was distributed along the coast synchronously without a dominant direction, exhibiting an arc shape. The turn expansion type was found in the SSS inversion results in 2017. The CDW first expanded northeastward, then turned southward at around 125°E, becoming the only expansion pattern of the CDW with a changing direction.
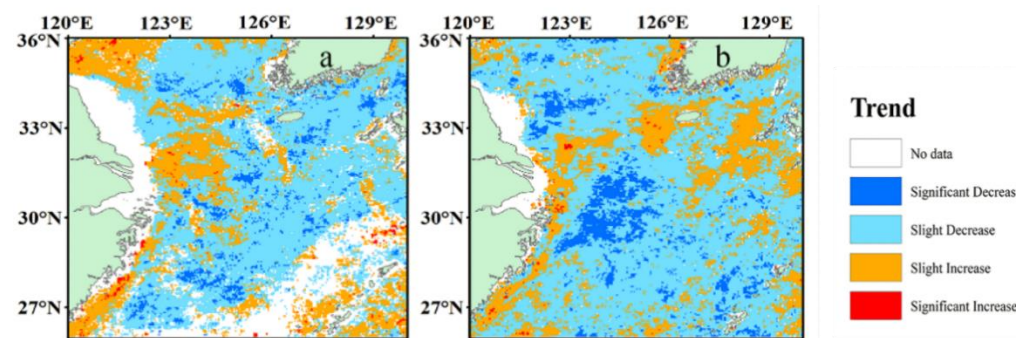
5.2. Changing Trend of SSS from 2003 to 2020

The Sen + MK trend test is used to evaluate the significance of the interannual change of SSS in March and August from 2003 to 2020. The SSS change area is calculated according to the different degrees of confidence, as shown in Table 5.

Table 5. The percentage of areas with changed SSS according to different degree of confidence.

Sen Slope	Z Statistic	Degree of Change	Area Ratio in March	Area Ratio in August
>0	≥1.96	significantly increased	0.91%	0.35%
>0	−1.96~1.96	not significantly elevated	27.33%	22.97%
<0	≤−1.96	significantly lower	6.62%	8.61%
<0	−1.96~1.96	not significantly reduced	65.12%	68.05%

The change rate of SSS calculated pixel by pixel in March and August from 2003 to 2020 is shown in Figure 12. Generally, the study area shows a slight SSS decrease in both March and August, exhibiting an overall downward trend of SSS in recent years. More than 70% of the sea area shows a decreasing trend in SSS in both August and March. The sea area with decreasing SSS in the flood season is larger than that in the dry season. The sea area with a significant and slight decrease in SSS is mainly distributed in the outer parts of the Changjiang Estuary. The area with elevated SSS is located at the plume front of CDW and the activity area of TWC, which intrudes along the submarine valley along the Zhejiang coast. The deceasing trend of SSS in the front plume of CDW may indicate the strength of CDW, particularly in flood season.

Figure 12. Spatial variation trend of SSS in March and August in the waters adjacent to the Yangtze Estuary from 2003 to 2020 ((**a**) is March, (**b**) is August).

On the contrary, the sea area with a slight increase in SSS in both August and March is mainly located in the coastal waters. Particularly, in March, a large area in the northwest of the study area shows slightly increased SSS, in addition, sea areas with significant SSS increases are mainly distributed along the coast of the south of Hangzhou Bay. Whereas, in August, slightly increased SSS occurred near Jeju Island and a very small patch of sea area with significantly elevated SSS was observed in the Changjiang Estuary.

6. Conclusions

This paper takes the Changjiang Estuary and the adjoining sea area (the geographical location between 120°~130°E and 26°~36°N) as the research area. With SMAP\SSS L3 monthly product and MODIS\SST and MODIS\Rrs L3 monthly product as inputs, respective SSS inversion models of the flood season and the dry season are constructed by machine learning methods for the sea area with salinity >30 psu. An empirical method is constructed for the near-shore sea water with a salinity <30 psu without seasonal discrimination, to compensate the low accuracy of the machine learning inversion model in near-shore sea waters with low salinity. With the application of the combined SSS inversion models, SSS in this study area from 2003 to 2020 was performed to observe the change in expansion area and pattern of CDW in the continental shelf of ECS.

The paper reaches the following conclusions:

The combined inversion model achieved acceptable accuracy, particularly for the near shore sea water, SSS inversion accuracy is significantly improved through the combination with empirical model. In sea areas with a salinity between 30–34 psu, where SSS is obtained

with machine learning models, the MAE in the flood season is 0.503 psu, the MAE in the dry season is 0.263 psu. In addition, the experimental verification accuracy of the offshore empirical statistical model is R^2 = 0.8425, RMSE = 1.29 psu. Moreover, the correlation coefficient between the inversion SSS with Copernicus reanalysis SSS reaches 0.764, and the average deviation is −0.86 psu. It is indicated that the accuracy of the constructed combined model in this study satisfies the requirements of large scale dynamic observations. This new model provides a new scheme to obtain acceptable accuracy and enhanced spatial resolution for salinity observation for marginal seas, where the salinity varies widely, and the high turbidity has always been a great challenge for observation from space.

With the application of the combined model, SSS is inversed for the study area. The results show that more than 70% of the sea area has a downward trend in the flood season and the dry season. Particularly, in August, there is a significant decrease in salinity, indicating an expansion of the CDW extension. The increasing expansion of CDW with low salinity may indicate the gradual strengthening of CDW or weakening of TWC but needs more investigation.

From 2003 to 2020, the diluting water expansion patterns of the Yangtze River can be summarized into three categories: the first category is the northeast-directed expansion type, multi-directional expansion type, and the northeast-southeast expansion type. Among them, the northeast-pointing pattern is the most common type of CDW, and the three types occur due to the special marine dynamic conditions of that time.

With the application of a combined SSS inversion model constructed in this study, the time series of seasonal SSS sequence is obtained, the variation of the CDW expansion area and pattern are analyzed from 2003 to 2020. The study suggests an obvious increasing expansion of CDW, which is of great significance for not only a new scheme to observe SSS for coastal water from space, but also suggests a reliable insight into oceanic current variation. More studies are necessary to further explore the internal mechanisms inducing the land–sea interaction process under heavy human activities and global warming.

However, this paper still has the following shortcomings and future work direction:

(1) Consider trying more input parameters. In the future, factors such as chlorophyll and suspended sediment can be considered as inputs to comprehensively consider possible influencing factors related to SSS so as to improve the accuracy of the model.
(2) The inversion model is applicable to nearby sea areas, and its applicability needs to be strengthened.
(3) The missing rate of MODIS original reflectance data is high in winter and during the period of high occurrence of cloud and fog, which leads to partial missing of SSS inversion results. Therefore, in future research work, using appropriate cloud haze detection or data reconstruction methods to recover missing data can greatly improve the feasibility of remote sensing dynamic observation of SSS.
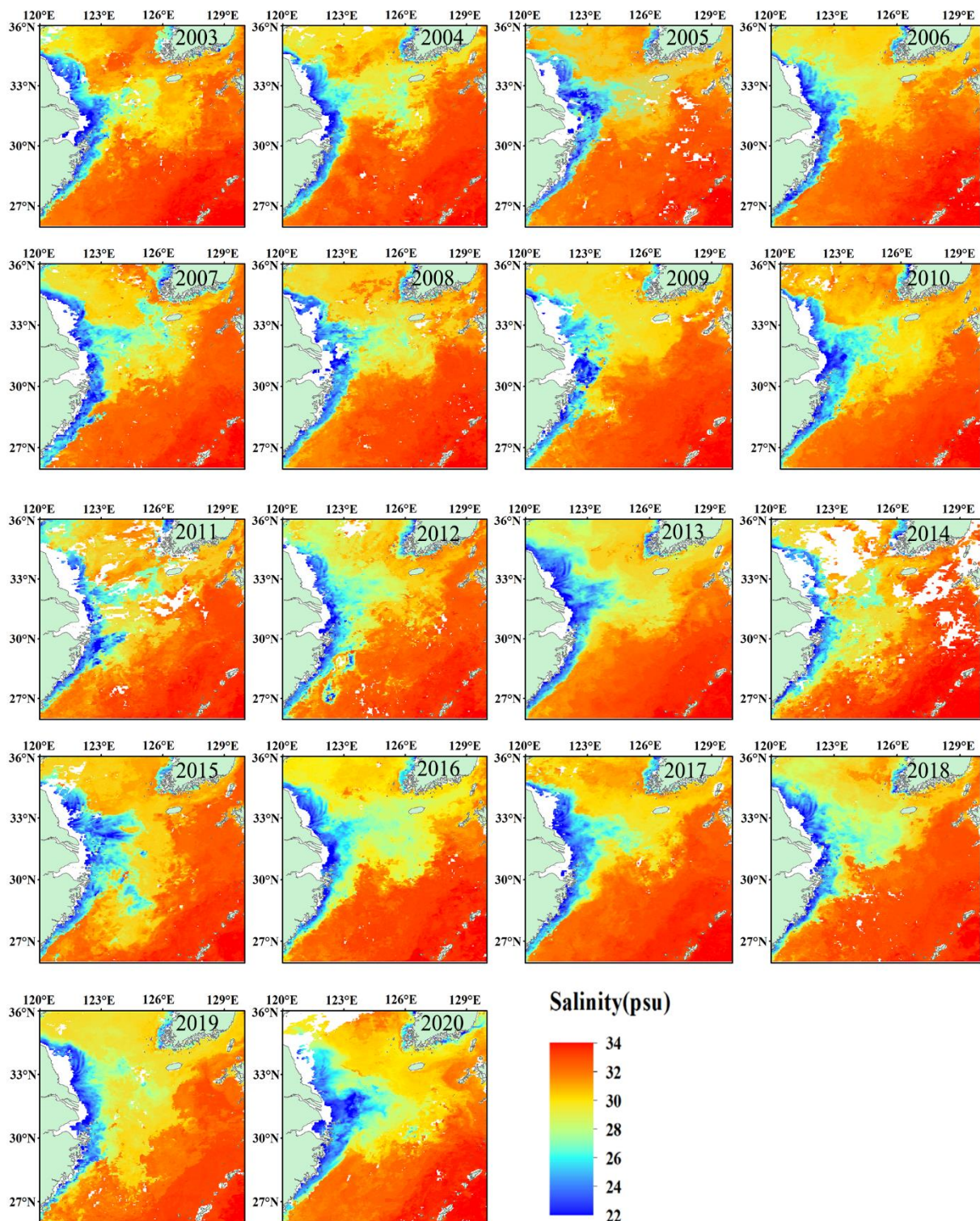
Appendix A

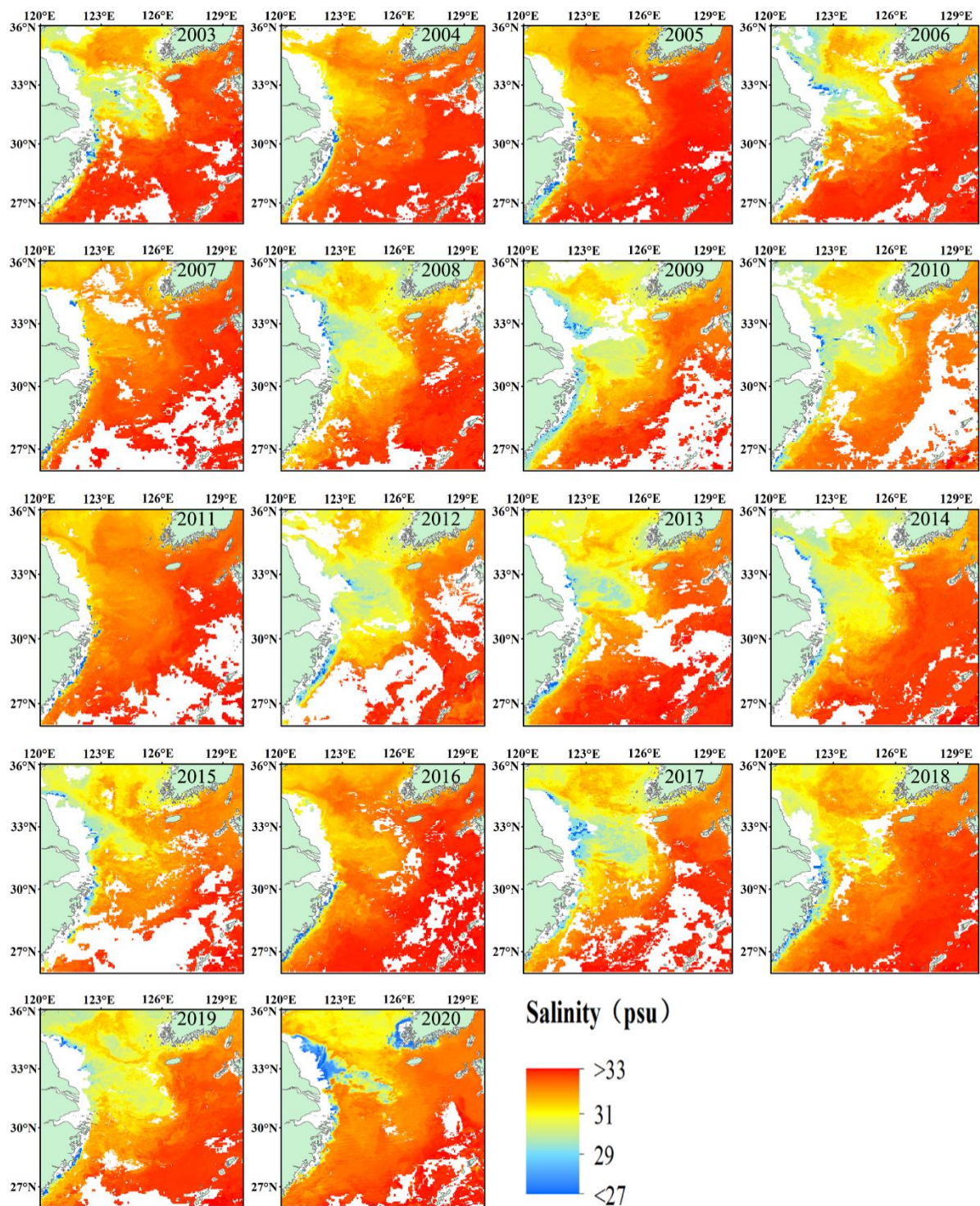**Figure A1.** SSS inversion results for August 2003–2020.

Figure A2. SSS inversion results for March 2003–2020.

References

1. Edmond, J.M.; Spivack, A.; Grant, B.C.; Hu, M.-H.; Chen, Z.; Chen, S.; Zeng, X. Chemical dynamics of the Changjiang estuary. *Cont. Shelf Res.* **1985**, *4*, 17–36. [CrossRef]
2. Bian, C.; Jiang, W.; Quan, Q.; Wang, T.; Greatbatch, R.J.; Li, W. Distributions of suspended sediment concentration in the Yellow Sea and the East China Sea based on field surveys during the four seasons of 2011. *J. Mar. Syst.* **2013**, *121*, 24–35. [CrossRef]
3. Li, G.; Qiao, L.; Dong, P.; Ma, Y.; Xu, J.; Liu, S.; Liu, Y.; Li, J.; Li, P.; Ding, D.; et al. Hydrodynamic condition and suspended sediment diffusion in the Yellow Sea and East China Sea. *J. Geophys. Res. Ocean.* **2016**, *121*, 6204–6222. [CrossRef]

4. Lin, C.; Su, J.; Xu, B.; Tang, Q. Long-term variations of temperature and salinity of the Bohai Sea and their influence on its ecosystem. *Prog. Oceanogr.* **2001**, *49*, 7–19. [CrossRef]

5. Singh, A.; Delcroix, T.; Cravatte, S. Contrasting the flavors of El Niño-Southern Oscillation using sea surface salinity observations. *J. Geophys. Res. Ocean.* **2011**, *116*. [CrossRef]

6. Chang, P.H.; Isobe, A. A numerical study on the Changjiang diluted water in the Yellow and East China Seas. *J. Geophys. Res. Ocean.* **2003**, *108*. [CrossRef]

7. Wu, Q.; Wang, X.; Liang, W.; Zhang, W. Validation and application of soil moisture active passive sea surface salinity observation over the Changjiang River Estuary. *Acta Oceanol. Sin.* **2020**, *39*, 1–8. [CrossRef]

8. Qin, S.; Wang, H.; Zhu, J.; Wan, L.; Zhang, Y.; Wang, H. Validation and correction of sea surface salinity retrieval from SMAP. *Acta Oceanol. Sin.* **2020**, *39*, 148–158. [CrossRef]

9. Abe, H.; Ebuchi, N. Evaluation of sea surface salinity observed by Aquarius. *J. Geophys. Res. Ocean.* **2014**, *119*, 8109–8121. [CrossRef]

10. Yu, F.; Wang, Z.; Liu, S.; Chen, G. Inversion of the three-dimensional temperature structure of mesoscale eddies in the Northwest Pacific based on deep learning. *Acta Oceanol. Sin.* **2021**, *40*, 176–186. [CrossRef]

11. Ahsan, Q.; Blumberg, A.F. Three-Dimensional Hydrothermal Model of Onondaga Lake. *N. Y. J. Hydraul. Eng.* **1999**, *125*, 912–923. [CrossRef]

12. Springel, V.; Wang, J.; Vogelsberger, M.; Ludlow, A.; Jenkins, A.; Helmi, A.; Navarro, J.F.; Frenk, C.S.; White, S.D.M. The Aquarius Project: The subhaloes of galactic haloes. *Mon. Not. R. Astron. Soc.* **2008**, *391*, 1685–1711. [CrossRef]

13. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The soil moisture active passive (SMAP) mission. *Proc. IEEE* **2010**, *98*, 704–716. [CrossRef]

14. Kerr, Y.H.; Waldteufel, P.; Wigneron, J.P.; Delwart, S.; Cabot, F.; Boutin, J.; Escorihuela, M.-J.; Font, J.; Reul, N.; Gruhier, C.; et al. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. *Proc. IEEE* **2010**, *98*, 666–687. [CrossRef]

15. Crow, W.T.; Chan, S.; Entekhabi, D.; Houser, P.R.; Hsu, A.Y.; Jackson, T.J.; Njoku, E.G.; O'Neill, P.E.; Shi, J.; Zhan, X. An observing system simulation experiment for hydros radiometer-only soil moisture products. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 156–160. [CrossRef]

16. Mecklenburg, S.; Drusch, M.; Kerr, Y.H.; Martin-neira, M. ESA's soil moisture and ocean salinity mission: Mission performance and operations. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1354–1366. [CrossRef]

17. Le Vine, D.M.; Lagerloef GS, E.; Colomb, F.R.; Yueh, S. Aquarius: An instrument to monitor sea surface salinity from space. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 2040–2050. [CrossRef]

18. Wang, J.; Deng, Z. Development of a MODIS data based algorithm for retrieving nearshore sea surface salinity along the northern Gulf of Mexico coast. *Int. J. Remote Sens.* **2018**, *39*, 3497–3511. [CrossRef]

19. Esaias, W.E.; Abbott, M.R.; Barton, I. An overview of MODIS capabilities for ocean science observations. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1250–1265. [CrossRef]

20. Hu, C.; Chen, Z.; Clayton, T.D. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sens. Environ.* **2004**, *93*, 423–441. [CrossRef]

21. 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]

22. Maisonet, V.J.; Wesson, J.; Burrage, D.; Howden, S. Measuring coastal sea-surface salinity of the Louisiana shelf from aerially observed ocean color. In Proceedings of the OCEANS 2009, Biloxi, MS, USA, 26–29 October 2009.

23. Palacios, S.L.; Peterson, T.D.; Kudela, R. Development of synthetic salinity from remote sensing for the Columbia River plume. *J. Geophys. Res. Ocean.* **2009**, *114*. [CrossRef]

24. Tietjen, T.; Vähätalo, A.V.; Wetzel, R.G. Effects of clay mineral turbidity on dissolved organic carbon and bacterial production. *Aquat. Sci.* **2005**, *67*, 51–60. [CrossRef]

25. Xie, W.; Yang, J.; Yao, R.; Wang, X. Spatial and Temporal Variability of Soil Salinity in the Yangtze River Estuary Using Electromagnetic Induction. *Remote Sens.* **2021**, *13*, 1875. [CrossRef]

26. Wu, X.D.; Song, J.M.; Li, X.G. Seasonal variation of water mass characteristic and influence area in the Yangtze Estuary and its adjacent waters. *Mar. Sci.* **2014**, *38*, 110–119.

27. Zhou, F.; Xuan, J.L.; Ni, X.B.; Huang, D.J. A preliminary study of variations of the Changjiang Diluted Water between August of 1999 and 2006. *Acta Oceanol. Sin.* **2009**, *28*, 1–11.

28. Qi, J.; Yin, B.; Zhang, Q.; Yang, D.; Xu, Z. Seasonal variation of the Taiwan Warm Current Water and its underlying mechanism. *Chin. J. Oceanol. Limnol.* **2017**, *35*, 1045–1060. [CrossRef]

29. Yang, D.; Yin, B.; Liu, Z.; Feng, X. Numerical study of the ocean circulation on the East China Sea shelf and a Kuroshio bottom branch northeast of Taiwan in summer. *J. Geophys. Res. Ocean.* **2011**, *116*, C05015. [CrossRef]

30. Li, J.; Wei, H.; Zhang, Z.; Lu, Y. A modelling study of inter-annual variation of Kuroshio intrusion on the shelf of East China Sea. *J. Ocean. Univ. China* **2013**, *12*, 537–548. [CrossRef]

31. Niino, H.; Emery, K.O. Sediments of Shallow Portions of East China Sea and South China Sea. *GSA Bull.* **1961**, *72*, 731–762. [CrossRef]

32. Beardsley, R.C.; Limeburner, R.; Yu, H.; Cannon, G.A. Discharge of the Changjiang (Yangtze River) into the East China Sea. *Cont. Shelf Res.* **1985**, *4*, 57–76. [CrossRef]

33. Fore, A.G.; Yueh, S.H.; Tang, W.; Stiles, B.W.; Hayashi, A.K. Combined active/passive retrievals of ocean vector wind and sea surface salinity with SMAP. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7396–7404. [CrossRef]

34. Wang, M.; Tang, J.; Shi, W. MODIS-derived ocean color products along the China east coastal region. *Geophys. Res. Lett.* **2007**, *34*, 1–5. [CrossRef]

35. Jang, E.; Kim, Y.J.; Im, J.; Park, Y.G. Improvement of SMAP sea surface salinity in river-dominated oceans using machine learning approaches. *GIScience Remote Sens.* **2021**, *58*, 138–160. [CrossRef]

36. Thépaut, J.N.; Dee, D.; Engelen, R.; Pinty, B. The Copernicus programme and its climate change service. In Proceedings of the IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1591–1593.

37. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [CrossRef]

38. Brezonik, P.; Menken, K.D.; Bauer, M. Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM). *Lake Reserv. Manag.* **2005**, *21*, 373–382. [CrossRef]

39. Niu, Y. The temporal and spatial differentiation of the surface water salinity of the Yangtze River Estuary based on MODIS. *J. Jilin Univ. (Earth Sci. Ed.)* **2019**, *49*, 1486–1495.

40. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]

41. Cortes, C.; Vapnik, V. Support vector machine. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

42. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]

43. Meng, Q.; Ma, X.; Zhou, Y. Forecasting of coal seam gas content by using support vector regression based on particle swarm optimization. *J. Nat. Gas Sci. Eng.* **2014**, *21*, 71–78. [CrossRef]

44. Vapnik, V.N. *The Nature of Statistical Learningtheory*; Springer: New York, NY, USA, 1995.

45. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]

46. Marini, F.; Walczak, B. Particle swarm optimization (PSO). A tutorial. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 153–165. [CrossRef]

47. Olson, R.S.; Randal, S.; Jason, M.H. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning in Workshop on Automatic Machine Learning. *Autom. Mach. Learn.* **2019**, 151–160.

48. Sen, P.K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **1968**, *63*, 1379–1389. [CrossRef]

49. Kendall, M.G. *Rank Correlation Methods*; Charles Griffin: London, UK, 1975.

50. Halverson, M.J.; Pawlowicz, R. Estuarine forcing of a river plume by river flow and tides. *J. Geophys. Res. Ocean.* **2018**, *113*. [CrossRef]

51. Moon, J.H.; Hirose, N.; PangI, C.; Hyun, K.H. Modeling Offshore Freshwater Dispersal from the Changjiang River and Controlling Factors During Summer. *Terr. Atmos. Ocean. Sci.* **2012**, *23*, 247–260. [CrossRef]