IEEEAccess Multidisciplinary : Rapid Review : Open Access Journal

Received 21 May 2022, accepted 3 July 2022, date of publication 11 July 2022, date of current version 20 July 2022. Digital Object Identifier 10.1109/ACCESS.2022.3189784

# **RESEARCH ARTICLE**

# Enhanced Sea Surface Salinity Estimates Using Machine-Learning Algorithm With SMAP and High-Resolution Buoy Data

# BALAKRISHNAN KESAVAKUMAR<sup>®1,2</sup>, PALANISAMY SHANMUGAM<sup>®1</sup>, AND RAMASAMY VENKATESAN<sup>2</sup>

<sup>1</sup>Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai 600036, India
<sup>2</sup>National Institute of Ocean Technology, Ministry of Earth Sciences, Chennai 600100, India

Corresponding author: Palanisamy Shanmugam (pshanmugam@iitm.ac.in)

This work was supported in part by the Ministry of Earth Sciences (MoES), Government of India, Ocean Observation Network Program, and in part by the National Geospatial Programme of Department of Science and Technology of the Government of India under Grant OEC1819150DSTXPSHA.

ABSTRACT Despite the recent advances in satellite-based L-band microwave radiometry and retrieval algorithms to provide a unique capability to measure sea surface salinity (SSS) from space and explore its utility for understanding mesoscale dynamics, global ocean circulation, vertical mixing, evaporation rates and climate oscillations, SSS retrieval from L-band microwave radiometric data from the NASA-SMAP (Soil Moisture Active Passive) mission is often biased with systematic errors on larger temporal, spatial scales and random errors on short-time and length scales. To improve the SSS retrieval from SMAP data, we developed a robust algorithm based on a machine-learning approach with high-resolution in-situ data from the Ocean Moored Buoy Network in the Northern Indian Ocean (OMNI), which includes the Arabian Sea (AS) and Bay of Bengal (BoB). The new algorithm was rigorously trained, tested and validated using the in-situ SSS time series measurements from the AS and BoB. Several sensitive variables were examined and used to improve the SSS estimates - such as radiometric and ancillary data from the satellite observations, sea-surface wind and precipitation from the ERA5 data, and SSS/ SST from the OMNI buoy measurements. The OMNI timeseries measurements provided the spatially averaged satellite products to characterize the variability and gross features of SSS in BoB and AS waters on weekly, monthly, seasonal and annual time scales. The systematic validation of SMAP SSS products on a range of spatio-temporal scales showed that the new algorithm improved the SSS estimates by more than 15% in open ocean waters and 25% in river-discharge and precipitation-dominated regions in BoB and AS, when compared to the standard (operational) algorithm. Further analysis demonstrated that the new algorithm reduced clear biases and significant anomalies in SMAP SSS retrievals in the regions of river runoff, surface-freshened ocean and intense tropical cyclones, and captured synoptic/mesoscale SSS features and their seasonal variations in the North Indian Ocean.

**INDEX TERMS** Sea surface salinity, OMNI buoy, SMAP mission, North Indian Ocean, satellite oceanography.

#### **I. INTRODUCTION**

Sea surface salinity (SSS, expressed in PSU) is a key property controlling seawater density and plays a critical role in the

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan<sup>(b)</sup>.

global hydrological cycle, density-driven global ocean circulation, climate dynamics, vertical mixing, air-sea exchange, biogeochemistry, acidification, and ocean heat content [1]–[3]. Changes in the SSS are driven by fresh water flux from the continents, evaporation, precipitation, glacier melts, stratification and vertical mixing. The SSS has been

recognized as a critical parameter of climate variability, which requires its time-series measurements (for instance, the global observation programs such as Global Climate Observing System (GCOS)). The SSS variability on large spatial and temporal scales is also essential for ecological analysis and physical/biological oceanographic studies in coastal environments [4]. Such data are needed for improving the knowledge of ocean circulation and better representing the climate variability in ocean models and associated forecasts (e.g., Indian Ocean Dipole, IOD and the El Niño-Southern Oscillation, ENSO) [5]-[7]. The Climate Variability and Predictability (CLIVAR) mission of the World Climate Research Program (WCRP) and Global Ocean Assimilation Experiment (GODAE) have been specially developed to foster our understanding of the linkages between salinity and climate variability and oceanic dynamics on intraseasonalto-centennial timescales [8]. The SSS measurements are essential to enhance our understanding of the role of haline stratification in dynamics of the mixed-layer and the ocean response to tropical cyclones, because the large freshwater influx from river discharge and precipitation forms barrier layers in the upper ocean and affects the depth of vertical mixing on cyclonic intensification [9]-[11]. Thus, there is a strong justification and necessity for obtaining the SSS measurements in regional seas and global oceans.

In-situ SSS observations were traditionally made from ships, buoys, drifters and moorings within limited ocean regions which resulted in the inadequate distribution of data in the global ocean [7], [12]. Less effort has been made to deploy such observation systems in the nearly unexplored polar oceans. Over the recent decade, with the invention of Argo profiling systems and the increasing number of deployments of surface observation systems, salinity sampling has especially increased in coastal oceans and marginal seas as well as in regions of large-scale dynamics and landsea linkages with strong salinity variations. Although in-situ SSS measurements provide an accurate representation of large-scale salinity variability, such measurements are under sampled on the required temporal, horizontal and vertical scales to resolve finer ocean features such as eddies and fronts [7], [10], [13].

Recently, remote sensing of salinity has become a highly valuable tool for improving many areas of science and applications – such as mesoscale ocean processes, land-sea/coastal-open ocean linkages, estuarine and near-surface mixing, carbon cycle, and air-sea exchange processes [3], [7], [13]. The SSS measurements from space have provided tremendous advantages over the in-situ SSS measurements in terms of covering the global ocean and resolving the space and time scales of major ocean phenomena (e.g., thermohaline circulation on global and ocean basin scales; El Niño-Southern Oscillation on inter-annual timescales). In the context of hydrological and biogeochemical cycles and climate variability, satellite-based SSS measurements are complementary to the in-situ observing systems to study mesoscale salinity variability in the

eddies, fronts [14] and land-sea linkages in the marginal seas and coastal oceans [7], [15]-[17]. To meet the increasing demand for space-borne SSS measurements, three satellite missions were launched to provide the near-global and synoptic views of SSS. These missions are Aquarius/SAC-D mission of NASA/CONAE (June 2011-June 2015), NASA-SMAP (Soil Moisture Active Passive) mission (January 2015-present) [18] and Soil Moisture and Ocean Salinity (SMOS) Mission (2009-present) of ESA [19]. The SMOS and Aquarius missions operated with an overlapping period from mid-2011 to mid-2015, while the SMAP had an overlapping period of only four months with Aquarius (February to June 2015). The SMAP works on active-passive design with an active L-band radar scatterometer, Aquarius uses a passive L-band radiometer, and SMOS is operated with a passive L-band interferometric radiometer. These L-band radiometers measure brightness temperatures (TB) to provide SSS records using a robust inversion scheme. The SSS estimates require accurate corrections for the non-salinity contributions due to extra-terrestrial radiations (both direct and reflected from ocean) from the Sun [10], [20]-[22] and noise contributions from ocean-surface roughness and sea surface temperature (SST) [23]-[27]. There are wellestablished schemes to make these corrections for the direct and ocean-reflected contributions to brightness temperature measurements. For the other contributions, the ocean surface roughness effect on SSS is assessed using ancillary wind data and the roughness parameter derived from the polarized L-band TB measurements and SST measurements (for removing the thermal effects on TB measurements). This requires adequate and accurate ancillary wind and SST data to minimize uncertainties in the SSS retrievals.

To improve the accuracy of satellite SSS measurements, several machine learning algorithms have been developed [28], [29]. These algorithms employ techniques such as random forests and support vector regressions to estimate SSS [30]. The machine-learning algorithms are capable to work with multifaceted data with a large number of predictor variables, and have gained popularity in a wide range of classification and estimation problems (including SSS retrievals from satellite data) [31], [32]. Recent studies reported a low accuracy in SSS retrievals in coastal oceans, estuaries, and outlets of major river systems [33], [34]. The SSS retrievals in such regions could be affected by the lack of in-situ measurements, non-inclusion of sampling differences between satellite and in-situ measurements, and inaccurate corrections applied to TB data for the contributions and contaminations from various sources [10], [35].

The satellite validation studies using in-situ data (collected by stationary and drifter buoys and Argo floats) in open oceans, where the salinity varies little in space and time (i.e., 33-37 PSU with a mean value of approximately 35 PSU) [13], [36], [37], have reported an accuracy of 0.2 PSU for both Aquarius and SMOS SSS. Fewer studies have evaluated satellite SSS observations with in-situ data in regional ocean basins dominated by river runoff, winds, tidal effects

and vertical mixing (that reduces salinity stratification under strong winds and cyclonic storms) [10], [38], [39]. Previous basin-scale studies on evaluating SMOS SSS with in-situ data showed a Root Mean Square Error (RMSE) of 0.52 in the Mediterranean Sea [13] and BoB [7], [16] and 0.52 in the Gulf of Mexico [40]. These studies observed the regional Root Mean Square Deviation (RMSD) to increase from 0.52 PSU in open seas with low river inputs to 2 PSU in coastal seas with high river inputs. This indicates that the regions with high freshwater inputs (from rivers and precipitation) significantly affect the SSS retrievals [3], [10], [41]. Other studies reported that satellite SSS is subject to a bias under high wind speeds; especially, the higher winds associated with the passage of tropical cyclones have greater impacts on the sea surface roughness and brightness temperature [42]. A higher uncertainty in satellite-retrieved SSS originates from inaccurate wind speed data (usually underestimated) for intense cyclones (with surface wind speed often exceeding 15 m s<sup>-1</sup>). Under higher wind speed of more than 20 m s<sup>-1</sup>, the SSS products are severely impacted and discarded during the gridding and generation of satellite Level 3 data. A lower accuracy in SSS retrieval was reported in regions of surface and subsurface freshening due to heavy rainfall and subsequent salinification of the sea surface due to turbulent/vertical mixing [38]. However, no major study has been devoted to evaluate satellite SSS measurements and improve the retrieval algorithm using high-resolution in-situ data for a wide range of oceanic and atmospheric conditions. This addresses a critical gap in areas of validating the satellite-derived SSS products and improving the SSS retrieval algorithm using high-resolution in-situ data, which is part of the crucial demonstration for improving the current capabilities of satellite observing systems and supporting the research community and end users.

There is a challenge with the conventional approaches to improve the accuracy of satellite SSS in tropical and subtropical oceans (especially under the conditions of high winds, waves, sea surface roughness, rains, fresh water inputs, and extreme atmosphere/weather systems). There are other factors that lower the accuracy of SSS retrieval; for example, high uncertainty in the ancillary data, paucity of in-situ measurements, large difference in spatial scales between in-situ measurements and satellite footprints in regions with the subfootprint variability of ocean features (e.g., ~110 km for Aquarius and  $\sim$ 50 km for SMAP SSS), reduced sensitivity of L-band instrument to SSS variations in cold seawater, and interference (RFI) in the vicinity of the continents/ coasts [10], [37], [43], [44]. To overcome such a wide range of estimation problems, we developed a robust algorithm based on machine learning approach with high-resolution insitu data. The new algorithm uses all essential parameters of SSS as the inputs. The SSS estimates from this algorithm are compared with collocated in-situ SSS and results from the existing algorithm for a wide range of oceanic and atmospheric conditions. For this study, in-situ data were collected from a network of buoy systems in the Northern Indian Ocean (including BoB and AS). This study further examines the refined SMAP SSS products using daily, monthly and seasonal data in the BoB and AS regions.

# **II. DATA AND METHODS**

# A. IN-SITU DATA

The in-situ data were obtained from the Ocean Moored Buoy Network in the Northern Indian Ocean (OMNI), which is maintained by the National Institute of Ocean Technology (NIOT) (Fig. 1) [45]. These buoy systems have been deployed to provide ocean observation data over the past two decades. The OMNI buoy systems are equipped with state-of-theart sensors for the measurements of meteorological, surface and subsurface ocean parameters. The surface and subsurface sensors measure conductivity, temperature, current speed and direction at the surface and conductivity and temperature profiles up to 500 m depth. Depending on the activity and application type, the sensor suite is selected and integrated with the Data Acquisition System for further processing and transmission of data via satellite telemetry [45] to the MCC (Mission Control Centre) at NIOT. The moored buoy systems are programmed to transmit data to the MCC at a pre-defined interval.

The OMNI buoy systems are augmented at the predefined locations selected by a group of national experts from the government organizations such as National Hydrographic Office (NHO), Indian Navy, Indian Coast Guard, India Meteorological Department. The buoy data are used to provide real time cyclone updates and present as a critical component for climate models to understand the ocean state [46].

For this study, the in-situ measurement locations in the Bay of Bengal and Arabian Sea are categorized from the Northern Arabian and Bay of Bengal, Middle Arabian and Bay of Bengal, Southern Arabian Sea and Bay of Bengal. Of many types of measurements, conductivity and temperature data were considered in this study. The data from the above locations in different seasons depict a variety of situations such as precipitation, freshwater inputs from rivers, and high saline water intrusion. The buoys were interfaced with a SBE37 MicroCAT sensor (Sea-Bird Scientific, USA) via RS-232 serial port to the DAS (Data Acquisition System). Conductivity and temperature data were stored in the flash memory and DAS. Because conductivity is an intrinsic property of seawater, salinity, and density can be derived from this data [47]. The conductivity sensor used in the SBE 37 has small drifts in the electronics (0.01 PSU/month) that lower power consumption and increase accuracy and resolution [48]. Continuous field observations indicate that the sensor accuracy is within 0.003 mS/month-conductivity in most cases. The conductivity sensor is mounted at the bottom of the moored buoy system (Fig. 2). According to Ohm's Law, resistance is measured from the relation.

$$R' = \frac{\rho L}{A} \tag{1}$$

Conductivity is the multiplicative inverse of resistance (R'), L is the length of the water sample, A is the cross-sectional area of sampled water, and  $\rho$  is the resistivity. The conductivity cell containing three internal platinum electrodes is made up of borosilicate glass. The measurement accuracy is ensured by maintaining the cell geometry and methods for preventing marine growth inside the conductivity cell [49]. Seawater salinity was calculated using the relations (where the calculations of  $R_T$ ,  $r_T$  and coefficient values are not shown for brevity).

$$S = a_0 + a_1 R_T^{1/2} + a_2 R_T + a_3 R_T^{3/2} + a_4 R_T^2 + a_5 R_T^{5/2} + \Delta S$$
(2)

$$\Delta S = \frac{(T-15)}{1+0.0162(T-15)} \times (b_0 + b_1 R_T^{1/2} + b_2 R_T + b_3 R_T^{3/2}$$

$$+b_4 R_T^2 + b_5 R_T^{5/2}) \tag{3}$$

$$R_T = \frac{\kappa}{R_P r_T} \tag{4}$$

$$R = \frac{Conductivity(S, T, P)}{Conductivity(35, 150, 0)}$$
(5)

where  $r_T$  is the temperature coefficient of standard seawater and *R* is the conductivity ratio. Unlike other in-situ observing systems (such as Argo profiler, which measures conductivity and temperatures at 5 m depth below the sea surface), subsurface sensors in the moored OMNI buoy systems record data at 0.45 m depth. This enables a better understanding of air-sea interaction (near-surface) processes and salinity measurements to improve the regional and global SMAP SSS products without being impacted by mixed layer salinity.

#### **B. DESCRIPTION OF THE SATELLITE DATA**

The NASA Soil Moisture Active-Passive (SMAP) mission, launched in January 2015, was put onto a 680 km near-polar sun-synchronous orbit, equator crossings at 6 am and 6 pm local time, with an inclination of about 98.12°, an instrument swath of 1000 km, a spatial resolution of approximately 36 km  $\times$  36 km, and an approximate repeat time of three days [10]. The SMAP mission has collocated radar (active) and L-band microwave radiometer (passive), wherein the L-band microwave radiometer provides global measurements of brightness temperature (TB) at  $\sim$ 1.4 GHz (wavelength = 21 cm) [50], [51]. Despite the mission objective to measure soil moisture, SSS can be estimated from L-band TB data as previously demonstrated from the aircraft-based field campaigns data [52]-[54]. SSS retrievals from L-band TB data require accurate corrections for external contributions (e.g., reflection/emission due to sea surface roughness, temperature and radiation from the Sun and celestial radio sources).

#### C. ECMWF-ERA5 DATA

ECMWF (European Centre for Medium-Range Weather Forecasting) operates the ERA5 reanalysis system to produce the global records of the atmosphere, land surface and ocean state parameters for the period from January 1950 to present. The ERA5 system provides hourly data for several parameters on a 30 km grid throughout the period, together with the uncertainty estimates for these variables at the reduced spatial and temporal resolution from an ensemble (with 3-hourly output at half the horizontal resolution of 63 km grid spacing). The ERA5 data are produced by the Copernicus Climate Change Service (C3S) at ECMWF. This study uses a comprehensive dataset from model simulations and buoy observations. The ECMWF data were obtained from the ECMWF climate change service (https://cds.climate.copernicus.eu/). The ERA5 data corresponding to the in-situ observations were extracted for further analysis. Further, it has the surface dataset consisting of eastward and northward neutral winds at the height of 10 m above the sea surface and precipitation.

### **III. MODEL DESCRIPTION**

#### A. DESCRIPTION OF THE OPERATIONAL ALGORTHM

The brightness temperature is the key parameter measured by microwave radiometers that depends on surface emissivity and SST.



FIGURE 1. Location map of the deep sea moored buoys in the North Indian Ocean (Bay of Bengal and Arabian Sea). The SST and SSS observations from OMNI buoys for the years 2017 and 2018 were used for model development.

The brightness temperature model is used to retrieve surface emissivity (Eq. 7) [55]. The sea surface emissivity for given frequency and viewing geometry is dependent upon the geophysical parameters such as seawater dielectric constant, SSS, SST, surface roughness and foam [54]. The existing model for SSS retrieval incorporates the roughness correction model and implements the Marquardt Levenberg algorithm for minimizing the weighted Sum of Squared Error (SSE) cost function (either linear or nonlinear) over a space of the parameters used in this study. According to the Klein-Swift model [54], the relation of brightness temperature and surface emissivity is given by

$$T_B = T \times e \tag{6}$$

$$T_B = \left(T'_B + T_{SC}\right) / (L_A + T_{UP}) \tag{7}$$



**FIGURE 2.** Schematic of the mooring buoy with the high-resolution SSS and SST sensors (highlighted in red).

where  $T'_B$  is the terrain/surface emission,  $T_{SC}$  is the radiometric temperature of the energy scattered from the surface,  $L_A$  is the atmospheric attenuation, and  $T_{UP}$  is the atmospheric upwelling self-emission. The above relation (Eq. 6) can be written in terms of surface emission and roughness as,

$$T_{B,P}(\Theta, SST, SSS, U_{10}) = e_p(\Theta, SST, SSS) .SST + \Delta T_{B \ rough,p}(\Theta, U_{10})$$
(8)

$$e_p(\Theta, SST, SSS) = 1 - |R_p(\Theta, SST, SSS)|^2 \quad (9)$$

where Rp is the Fresnel reflection coefficient given by

$$R_{\nu}(\Theta, f, SST, SSS) = \frac{\sqrt{(\varepsilon_{sw} - \sin^2 \Theta)} - \varepsilon_{sw} \cos \Theta}{\sqrt{(\varepsilon_{sw} - \sin^2 \Theta)} + \varepsilon_{sw} \cos \Theta} \quad (10)$$

$$R_{h}(\Theta, f, SST, SSS) = \frac{\sqrt{\left(\varepsilon_{sw} - \sin^{2}\Theta\right) - \cos\Theta}}{\sqrt{\left(\varepsilon_{sw} - \sin^{2}\Theta\right)} + \cos\Theta}$$
(11)

for vertical and horizontal polarization respectively. The relation between sea surface salinity and the influencing parameters is given by (Eq. 12)

$$SSS = S_0 (\Theta_i, t_s) + S_1 (\Theta_i, t_s) T'_{BE,V,Sur} + S_2 (\Theta_i, t_s) T'_{BE,V,Sur} + S_3 (\Theta_i, t_s) U_{10}$$
(12)

where SSS is a function of the brightness temperature, SST, viewing geometry ( $\Theta$ , angle of incidence) and roughness parameter and coefficient (S) which is a function of the angle of incidence and temperature. The SSS retrieval algorithm uses brightness temperatures (modelled and measured) by applying the Levenberg-Marquardt least square iterative convergence algorithm. Because the sea surface roughness parameter dominates the errors of SSS retrieval from microwave L-band radiometers, several models were developed based on theoretical and empirical approaches. The small perturbation method (SPM) is one among those models in which RMS (Root Mean Square) of surface wave height is smaller than the wavelength for simulating the surface emissivity. The small-slope approximation method has been used extensively for all ocean wavelengths when the wave slopes are small, because it overcomes the problem of the SPM in the calculations. The two-scale method (TSM) reported in an earlier study [56] has become popular, as it approximates the sea surface in two-scales (small scale and large-scale waves). In other studies, semi-empirical relationships between the brightness temperature, angle of incidence and wind speed were established and tested [50], [53].

# B. ROBUST ALGORITHM BASED ON MACHINE LEARNING APPROACH

In recent years, machine learning has grown out rapidly for developing a wide variety of real-world applications because of its potentiality to intelligently analyse the nature and characteristics of various types of data and enhance the learning algorithm performance. Machine learning algorithms are broadly classified as unsupervised, semi-supervised, supervised, and reinforcement learning type. The machinelearning algorithms are capable of learning from sample data and providing the output results in the form of predictions or classifications based on non-linear statistical models. The ANN(Artificial Neural Network) algorithms can be used to enhance the existing data analysis technique owing to their predictive capabilities [57]-[59]. The number of hidden layers in ANN needs to be increased to achieve a close convergence between the output and measured values, although it increases the complexity of the model. The GRNN (Generalized Regression Neural Network) method, which was proposed for improving SMOS (Soil Moisture and Ocean Salinity) SSS in the South China Sea (SCS), uses salinity data at 5 to 10 m depths from Argo measurements, and explains the significance for correcting SSS using the near surface depth data [41].

Decision trees are based on trees built with nodes and leaves for solving the classification and regression problems. The nodes in the trees indicate the feature from the inputs, the branch indicates decision, and the leaf at the end forms the output. The decision trees are simpler in construction, but its performance reduces when the input data have the noise or outliers. Also, its performance reduces when applied to the new input dataset. In order to minimize the above error, a random forest (RF) regression tool uses multiple decision trees and creates the different sets of trees for the same input values. The collective outputs generated by RF (Fig. 3) present the expected values. The RF based regression was implemented using the 'RandomForestRegressor' function using python. For tuning the random forest regression algorithm, parameters such as the maximum depth of the tree, number of trees in the forest and random state are adjusted for attaining the maximum training score.



FIGURE 3. A flowchart of the random forest architecture.

#### C. RANDOM FOREST REGRESSION

Random forest regression (a supervised machine learning algorithm) uses an ensemble learning technique to combine the output results from multiple learning algorithms in order to produce predictions that are more accurate. The predictions are achieved from several decision trees with leaf and decision nodes constructed during the training process. Leaf nodes are the outcome of the decisions made based on a given dataset. The input data are split into multiple data segments to attain maximum homogeneity and purity based on certain decision criteria and the splitting continues and results in fully-grown trees until the expected output is achieved. The decision trees may suffer from a high degree of variance, lead to over fit the data, and yield poor accuracy on unseen data. This problem can be minimized by increasing the number of trees. Random forest (RF) [60] is a group of decision tree predictors and it works based on a random vector independently sampled with the identical distribution in all the decision trees in the forest. Thus, the generalized error decreases as the size of RF increases. Random selection of input data fields also helps in achieving a better performance accuracy and closeness with the target value. For this study, the RF regression is implemented by forming trees depending on a random vector and a predictor with a numerical value. The predicted variables are based on the average of the predictions from the individual decision tree in the RF method.

# D. MODEL PARAMETERIZATION

The machine-learning approach based on RF was implemented to form a robust algorithm to estimate SSS from satellite microwave L-Band radiometer data. This algorithm uses the input variables from in-situ observations, SMAP, auxiliary inputs and ERA5 data. The brightness temperatures (Tb\_h-horizontally polarized and Tb\_v-vertically polarized) are obtained from SMAP L2 data and the brightness temperature ratio is computed subsequently (Tb\_h/Tb\_v, which helps in stabilizing the sensitivity of Tb at the various ranges of SST) [37]. The ancillary data of SMAP such as NCEP (National Centers for Environmental Prediction) wind speed, significant wave height from NOAA Wave Watch III, NOAA Optimum Interpolation (OI) SST, and reference SSS from HYCOM (HYbrid Coordinate Ocean Model) for each pixel resolution of Tb are obtained from the PODAAC server (podaac-tools.jpl. nasa.gov/drive/files/SalinityDensity/smap). From the ERA5 model, total precipitation, vertical and horizontal components of 10 m neutral wind are obtained. The wind vector provides the direction information of the wind for an improved estimate of SSS.

The in-situ SSS and SST from buoy observations were used to improve the algorithm parameters. The in-situ SSS data were used as the target variable (label data) for training and testing the algorithm performance [39]. Consequently, the new algorithm yields the maximum feature importance score for each in-situ SSS observation (results presented in a later section). The training of the RF algorithm was conducted using in-situ time-series data for the year 2017-2018. Despite the total number of in-situ data from buoy systems being greater than 10,000, quality control application and in-situ/satellite matchup analysis reduced the total number of in-situ data to 3028. For validation, the matchup data were rescaled at the same temporal and spatial resolutions. 70% of these data were used for training and 30% for testing purposes. The results of the algorithm were validated independently using in-situ time-series data for the year 2019.

The significance of the input parameters in the refined model is evaluated using a feature importance bar plot and it depicts that the HYCOM SSS contributes the highest value of 0.65 followed by the subsequent input variables in the descending order (Fig. 4). The quality-controlled input variables are used to train the random forest model (Fig. 5)



FIGURE 4. Feature importance of various input variables.

#### **IV. RESULTS AND DISCUSSION**

This section presents the validation results for the Refined Model (RM) and standard algorithm using in-situ SSS data obtained from the moored OMNI buoy systems. Because the OMNI buoys are located in open sea regions far away from the coast, the in-situ measurements are not contaminated by the land and high radio frequency interference (RFI) effects. The validation matchup data of salinity measurements were obtained under different oceanic and meteorological conditions (including tropical cyclones). In addition to evaluating RM SMAP SSS in comparison with in-situ SSS and standard SMAP SSS data, the RM SMAP SSS products are



FIGURE 5. Flowchart depicting the steps involved for selecting the significant input variables from the various sources and in-situ parameters from buoy data.

further used to study synoptic/mesoscale SSS features and seasonal variations of SSS in BoB and AS.

## A. EVALUATION OF SMAP SSS

To evaluate the satellite-retrieved SSS products, SMAP SSS from the RM and standard algorithm in BoB and AS were compared with OMNI buoy time series in-situ measurement data at 24 locations (14 locations in BoB and 10 locations in AS- Table 1). These measurements show a close consistency between RM SSS and in-situ SSS throughout the observation periods. In contrast, the standard SMAP SSS deviated from in-situ SSS data in regions with strong wind and high waves, river plumes and vertical mixing. The overall RMSD was  $\sim 0.3$  PSU and > 0.5 PSU for the RM and standards algorithm respectively. A significant deviation of the standard algorithm SSS products is owing to inaccurate input data for strong winds, waves, freshwater inputs and vertical mixing/upwelling processes [37]. A slight difference between RM SSS and in-situ SSS may be attributed to the spatiotemporal sampling difference between satellite and in-situ platforms (i.e., the measurement footprints, measurement times, and measurement depths; note that the satellite-based L-band microwave radiometer measures salinities at less than a centimeter depth of the ocean and the buoy systems measure salinities at 0.45 m below the sea surface). These factors often introduce discrepancies in the validation results on a range of spatial and temporal scales [37], [61]. Further, the variation in skin layer SSS caused due to evaporation [62] and the effect of wind in increasing the surface roughness are considered in the correction of geophysical effects [26].

Further comparison based on the daily (continuous) measurements from all OMNI buoy systems in BoB and AS showed a good agreement between RM SSS and in-situ SSS with low bias (Figs. 6 and 7). The RM SSS also depicted

TABLE 1.	Details of the	OMNI buoys fo	r SST and SSS	observations.
----------	----------------	---------------	---------------	---------------

Region	Buoy ID (WMO	Lat (N)	Long	Year
	ID)		(E)	
	BD08 (23091)	17.79	89.21	
	BD09 (23092)	17.46	89.12	
	BD10 (23093)	16.47	87.98	
	BD11 (23094)	13.49	84.16	2017
	BD12 (23095)	10.44	94.02	
	BD13 (23459)	13.97	86.94	
Bay of	BD14 (23460)	6.577	88.23	
Bengal	BD08	17.83	89.22	
	BD09	17.49	89.12	
	BD10	16.47	87.98	
	BD11	13.51	84.18	2018
	BD12	10.50	94.03	
	BD13	14.03	87.00	
	BD14	6.579	88.31	
	AD06 (23456)	18.49	67.46	
	AD07 (23451)	14.96	69.00	
	AD08 (23452)	11.77	68.59	2017
	AD09 (23453)	8.233	73.33	
Arabian	AD10 (23454)	10.32	72.58	
Sea	AD06	18.47	67.45	
	AD07	14.96	68.98	
	AD08	11.78	68.61	2018
	AD09	8.233	73.30	
	AD10	10.32	72.58	

measured salinity variations over the entire observation periods, particularly low-salinity troughs in the northern BoB (due to precipitation and river runoff) during the months of southwest (mainly July and August) and northeast (mainly November-early December) monsoon and highsalinity peaks (nearly month or sub-month) with high fluctuations (due to salt water movements by currents and eddies) during the observation period [37].

A similar performance of the RM was observed in the freshwater-influenced BoB and salty AS with a mean

 TABLE 2. Statistical comparisons of RM SSS and SMAP SSS data vs in-situ

 SSS data for the year 2019.

Cases	SMAP and RM SSS	MRE	MNB	RMSE	SLOPE	R <sup>2</sup>	N
All cases	SMAP SSS	0.006	0.21	0.59	0.97	0.81	479
Case 1	RM SSS SMAP	$0.002 \\ 0.009$	0.07 0.29	0.44 0.75	0.98 0.90	$0.88 \\ 0.77$	
	SSS BM SSS	0.002	0.08	0.68	0.91	0.79	98
Case 2	SMAP	0.002	0.24	0.43	0.98	0.83	80
	RM SSS	0.003	0.11	0.25	0.99	0.92	89
Case 3	SMAP SSS	0.005	0.16	0.39	0.70	0.36	52
Case 4	RM SSS SMAP	0.003 0.003	0.12 0.11	0.32 0.52	0.73 0.82	0.47 0.69	
	SSS RM SSS	0.001	0.02	0.44	0.84	0.75	115
Case 5	SMAP	0.006	0.21	0.80	1.30	0.59	125
	RM SSS	0.002	0.08	0.60	1.33	0.69	123



FIGURE 6. Scatterplots of SMAP SSS vs In-situ SSS (c) and RM SSS vs In-situ SSS (e) in Bay of Bengal waters. Scatterplots of SMAP SSS vs In-situ SSS (d) and RM SSS vs In-situ SSS (f) in Arabian Sea waters. The grey symbol represents SSS data for the years 2017 and 2018. The coloured symbols (red, blue and green) represent the independent validation data of SSS in Bay of Bengal(a) and Arabian Sea(b) for the year 2019.

normalized bias (MNB) of 0.07, a RMSE of 0.44, an intercept of 0.51, a slope of 0.98 and a correlation coefficient of 0.93 (N=479). In contrast, the standard SMAP SSS showed large errors and low slope and correlation coefficient values (MNB 0.21, RMSE 0.59, intercept 0.96, slope 0.97, and



FIGURE 7. (a) Comparison of SSS matchup data from SMAP/In-situ observations and Refined Model for the year 2019 in BoB (17°N/89°E). (b) The corresponding biases (SMAP minus In-situ(black) and RM minus In-situ(red)).

correlation coefficient 0.9). The standard SMAP algorithm showed the similar performance in the middle and southern regions of BoB, but it yielded the relatively high errors and low correlation coefficients in northern BoB waters with high river input (from Ganges and other river systems) due to an underestimation of SSS on both weekly and monthly time scales. In salty AS waters, the standard algorithm severely overestimated SMAP SSS with the larger deviation than the RM. This inconsistency in the performance of the standard algorithm with high errors limit potential applications of satellite salinity observations on local and regional scales.

To examine the source of errors and the performance of the algorithms, SMAP SSS were further compared with in-situ SSS in five different regions on different spatio-temporal scales (Table 2) – Case 1 in the river discharge region, Case 2 in the BoB region, Case 3 in the island region, Case 4 in the deep-sea region of BoB, and Case 5 in the Arabian Sea. These test cases were identified based on certain factors and processes that influence salinity variations and affect the performance of the algorithm. The river discharge region receives a large volume of freshwater from continents and a nearly equal amount of freshwater due to precipitation over the northern BoB [63]. The RM estimated SSS in this region with MNB 0.08, RMSE 0.68, intercept 2.9, slope 0.91 and correlation coefficient 0.88. These statistical values indicate better performance of the RM as compared to the standard algorithm (with MNB 0.29, RMSE 0.75, intercept 3.25, slope 0.9, and correlation coefficient of 0.87). The Case 2 region represents the middle BoB which has lesser RFI effect. For this region, the RM still outperformed the standard algorithm in estimating SSS from SMAP data (MNB 0.11, RMSE 0.25, intercept 0.19, slope 1.0, and correlation coefficient 0.95 for the RM; MNB 0.24, RMSE 0.43, intercept 0.23, slope 0.99, and correlation coefficient 0.91 for the standard algorithm). A similar validation analysis was conducted for the island region (Andaman), where a buoy is moored at a water depth of 3200 m and about 100 nautical miles away from the island coast. This region experiences intense monsoonal rains [64], [65] and high runoff from the continent [66].



FIGURE 8. Scatterplots (first and second row), line plots (third row) and bias plots (fourth row) showing the comparison of SMAP SSS and RM SSS with in-situ measured SSS for the different cases. Case 1 – River discharge region, Case 2 – Central Bay of Bengal region, Case 3 – Near the Island region, Case 4 – Deep sea location in Bay of Bengal, Case 5 – Arabian Sea.

The comparison between RM SSS and in-situ SSS indicated a slightly lower performance of the RM (MNB 0.12, RMSE 0.32, intercept 8.71, slope 0.73, and correlation coefficient 0.68) when compared to its SSS estimates in other regions. However, the RM SSS are more accurate than the standard SMAP SSS products (MNB 0.16, RMSE of 0.39, intercept 9.75, slope 0.70, and correlation coefficient 0.6). Similar validation results were obtained for the Case 4 region of BoB (MNB 0.02, RMSE 0.44, intercept 5.2, slope 0.84, and correlation coefficient 0.86 for the RM; MNB 0.11, RMSE 0.52, intercept 5.8, slope 0.82, and correlation coefficient 0.83 for the standard algorithm). In salty AS water (Case 5), the RM estimated SSS are better consistent with in-situ SSS (MNB 0.08, RMSE 0.60, slope 1.33, and correlation coefficient 0.83) when compared with the standard algorithm (MNB 0.21, RMSE 0.8, slope 1.3, and correlation coefficient 0.76 (Fig. 8(a - t)).

# B. APPLICATION OF THE RM IN THE BoB AND AS

The RM was used to analyse the spatial structure of SSS in BoB and AS for the year 2019. The northern BoB is characterized by low saline water from the discharges of Ganges and Irrawaddy river systems [64]. Precipitation due to southwest (June-August) and northeast (October-December) monsoons further leads to freshening of the bay [60]. In this region, the RM provided accurate SSS estimates over the standard algorithm in coastal and open sea waters.

The RM provided improved SSS estimates over the standard algorithm and reduced the associated deviations due to the dynamic near-surface salinity stratification [13], mismatches between satellite footprint and in-situ point measurement, associated averaging of measurements [67], and RFI [44].

The fresh water fluxes from the rivers and heavy precipitation caused an increase of SST  ${\sim}0.5^\circ$  C in the northern BoB



FIGURE 9. Monthly SSS products derived from the RM for the year 2019. Location in the range between  $3^\circ$  N -  $24^\circ$ N and  $65^\circ$  E- $95^\circ$  E.



2019. Location in the range between 3° N - 24°N and 65° E-95° E.

during summer and a reduction of SST from  $0.5^{\circ}$  to  $1.5^{\circ}$  C in the eastern BoB during summer and winter [68]. This leads to the variations in brightness temperatures (TB) and SSS [10]. Because the RM uses the SST from in-situ data (observed with a sensor with high resolution ~0.0001° and high stability 0.0002°C per month) and ERA5 model data, it improved the SSS estimates over the standard algorithm. As a result, the RM SSS products depicted the spatial coverage of low saline water along the east coast of India, which is fueled by the East India Coastal Current (EICC) [69]. The freshwater tongue like structure is visible in the RM SSS product for the months of September to November [63]

The AS displays high salinity structure due to evaporation, less precipitation and spreading of Persian Gulf waters. As can be seen in Fig. 9, the RM SSS products showed consistently high salinity structure in the AS and in regions of the significant interaction between AS and BoB. The RM SSS products also depicted the southward advection of high saline water from AS to mid BoB during July-November (Fig. 9).



**FIGURE 11.** Plots showing the variables/ parameters for the month of May 2019. a) Sea surface salinity (SSS in PSU) from the in-situ observations RM and standard SMAP algorithm, b) Sea surface atmospheric pressure (hPa), c) Short wave radiation  $(Wm^{-2})$ , d) Long wave radiation  $(Wm^{-2})$ , e) Sea surface temperature (°C), f) Surface water current speed  $(cms^{-1})$ , g) Wind speed and wind gust  $(ms^{-1})$ , h) Wind direction (degrees), and i) Surface water current direction (degrees).



**FIGURE 12.** Plots showing the variables/parameters for the month of November 2019. a) Sea surface salinity (SSS in PSU) from in-situ observations, and RM and standard SMAP algorithm, b) Sea level Pressure (hPa), c) Short wave radiation (Wm<sup>-2</sup>), d) Long wave radiation (Wm<sup>-2</sup>), e) Precipitation (mm), f) Surface water current speed (ms<sup>-1</sup>), g) Wind speed and wind gust (ms<sup>-1</sup>), h) Wind direction (degrees), i) Surface water current direction (degrees) during the extreme conditions of Bulbul on 8 Nov 2019.

The EICC contributes to the southward movement of fresh water along the east coast of India (thus giving rise to the West India Coastal Current, WICC) and to the intrusion of fresh water from BoB to AS [70], [71].

The intrusion of BoB fresh water into AS is clearly observed during the months of January-April and this exchange is reversed during the months of June-October (Fig. 9).



**FIGURE 13.** Plots showing the variables/parameters measured/estimated on 8 November 2019. a) SMAP SSS (PSU), b) RM SSS (PSU), c) Sea level pressure (hPa), d) Sea surface temperature (°C), e) U component of neutral wind at 10 m (ms<sup>-1</sup>), f) V component of neutral wind at 10 m (ms<sup>-1</sup>), and g) Total precipitation (mm), and h) Quality flag of SMAP.

The climatological products of the RM demonstrated the spatial structures of salinity in this region (Fig. 10).

The coincidence of two tropical cyclones with satellite passes enabled a better assessment of the algorithm performance during the extreme cyclonic conditions. For this analysis, the tropical cyclone Fani of May 2019 (Fig. 11) and Bulbul of November 2019 were considered. The standard SMAP SSS products overestimated the salinities due to high surface winds and the highly varying local parameters. The quality flag of SMAP data during the cyclones indicated a high uncertainty associated with the brightness temperature measurements.

In general, vertical mixing, Ekman pumping, rainfall and runoff control the salinities during the cyclonic events [9]. The upwelling process significantly contributed to the observed salinity variability during the cyclonic events. During the Bulbul cyclone, the minimum distance between the eye of the cyclone and the in-situ buoy location was 72 km on 8 November 2019 when the cyclone turned into a very severe cyclonic storm (VSCS). Fig. 12 (a) shows the performance of RM under severe cyclonic conditions, wherein it estimated the salinity variation of 0.3 PSU which was not detected by the standard algorithm. High precipitation occurred due to the Bulbul cyclone freshened the surface layer on the subsequent days [63].



FIGURE 14. Plots showing the variables/parameters measured/estimated during an extreme event on 2 May 2019(TC-Fani). a) SMAP SSS (PSU), b) RM SSS (PSU), c) Sea level pressure (hPa), d) Sea surface temperature (°C), e) U component of neutral wind at 10 m (ms<sup>-1</sup>), f) V component of neutral wind at 10 m (ms<sup>-1</sup>), g) Total precipitation (mm), and h) Quality flag of SMAP.

The freshening of the surface layer was dependent on the degree of cyclonic intensity and cyclone category [38]. Our analysis showed that passive microwave brightness temperatures were significantly affected in high winds (>15 m s<sup>-1</sup>), gusts and shifts in wind direction during the Bulbul and Fani. Other parameters such as surface currents, sea surface wind speed, short wave radiation, and long wave radiation indicated an extreme condition on 8 November 2019. Overall, the RM gave satisfactory results when compared to the standard algorithm. Fig. 13 shows the spatial SSS products generated using ERA 5 model data and SMAP radiometric products on 8 November 2019. The track of SMAP (L2) matched with the cyclone track and the quality flag obtained from SMAP data indicated the large uncertainty with the SMAP SSS products near the eye of the cyclone.

Fig. 13 (b) shows the RM SSS product generated during the cyclonic event.

Our analysis based on the sea level pressure, U and V components of wind, precipitation (Fig. 14) and standard SMAP SSS products showed that the signatures of the Fani cyclone were only partially captured in this region and the quality flag indicated the high values along the SMAP track. In contrast, the RM SSS products displayed spatially consistent SSS structures with the available input data. The remotely sensed input variables are affected due to the extreme conditions and this bias is fed to the model which results in a similar bias in RM SSS estimates.

# **V. CONCLUSION**

A machine-learning based refined algorithm for SSS retrievals using high-resolution in-situ data has been presented in this study. The total number of quality-controlled training data collected over the period 2017-2018 and used for this work, was around 3028. This dataset was critical for random forest (RF) learning and algorithm refinement.

The RF learning database consisted of SMAP observations, ERA 5 data, in-situ measurements of SSS, SST and other ancillary parameters. To conduct an independent and systematic validation of the RM, the in-situ time series measurements of SSS (from OMNI buoys) and the corresponding SMAP data were used to generate a matchup dataset for the year 2019 (N=479).

Results showed that the refined model (RM) estimated SSS with an improvement of 15% in open sea waters and 25% in river-discharge and precipitation-dominated regions (with a root mean square error less than  $\sim$ 0.68 PSU, a correlation coefficient greater than 0.88 and a slope value greater than 0.91). In contrast, the standard SMAP algorithm yielded significantly high errors in regions of river runoff, freshened ocean surface (due to precipitation) and intense tropical cyclones. The standard SMAP algorithm showed the poor prediction accuracy or missing data in regions with large salinity changes. This limits the potential application of SMAP SSS in the North Indian Ocean. The reduced accuracy could be caused by the deficient or inadequate/inaccurate input data and the algorithm parameters.

This study also demonstrated the applicability of RM SSS to study SSS variability in AS and BoB waters. The monthly and seasonal RM SSS products showed large, visible salinity changes as the result of the freshwater inputs from Ganges and other river systems in the northern BoB throughout the southwest (June-August) and northeast (October-December) monsoons. The spatial coverage of the freshening signature further extending to areas through middle of the BoB and along the coasts of Myanmar clearly reflected the differences and contrasts over the river runoff areas and shallow shelf seas [72]. The RM SSS products captured a high SSS structure in the AS throughout the year. This persistent SSS structure was manifested and maintained due to spreading of the intermediate Persian Gulf water and surface Arabian sea high-salinity water. The high SSS water was advected southward in the AS and the seasonality of this structure was found to be dependent on coastal currents and eddies during the fall and winter months [71].

Further, the spatial SMAP SSS maps during the passage of tropical cyclone (TC, Fani) provided an opportunity to investigate the influence of TC intensity and wind speeds on the SSS signatures and to make a robust assessment of SMAP SSS in complex oceanic conditions. The SMAP SSS maps exhibited a larger surface salinification along the path and on the left-hand side of the cyclonic storm. The movement VOLUME 10, 2022

of this cyclone resulted in strong vertical mixing processes and surface freshening (high precipitation) on the forward and right side of the storm during its passage. Results of the RM and standard algorithms showed that because of the high surface winds and inaccurate input wind data, the standard algorithm produced biased SSS along the storm trajectory and its adjoining high-wind regions. In regions surrounded by the eyewall of the cyclone, the standard SMAP SSS products were heavily biased and removed consequently. In contrast, the RM exhibited reduced SSS anomaly due to the storm motion in BoB.

Although the SSS retrieval from satellite microwave L-band radiometer data is challenging due to the current limitations/methodological complexities as well as the uncertainties associated with many factors and input datasets, the present study demonstrated significant improvements in SMAP SSS measurements, which will lead to a better understanding of the large-scale SSS variability in open-ocean regions and high-frequency small-scale SSS variability in shallow shelf regions.

#### ACKNOWLEDGMENT

The authors would like to thank the Director of the National Institute of Ocean Technology (NIOT) for extending the support and encouragement toward this work with the Indian Institute of Technology Madras (IITM) and also would like to thank the 'Group Head' and staff of the Ocean Observation Systems (OOS) Group and Vessel Management Cell of the NIOT and ship staff for their excellent help and support. The National Aeronautics and Space Administration (NASA) for providing measurement data through the PODAAC server, and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing model data.

#### REFERENCES

- [1] R. Venkatesan, N. Vedachalam, M. A. Muthiah, B. Kesavakumar, R. Sundar, and M. A. Atmanand, "Evolution of reliable and cost-effective power systems for buoys used in monitoring Indian seas," *Mar. Technol. Soc. J.*, vol. 49, no. 1, pp. 71–87, Jan. 2015, doi: 10.4031/MTSJ.49.1.8.
- [2] L. Yu, "A global relationship between the ocean water cycle and nearsurface salinity," *J. Geophys. Res.*, vol. 116, no. C10, pp. 1–17, 2011, doi: 10.1029/2010JC006937.
- [3] V. P. Akhil, M. Lengaigne, F. Durand, J. Vialard, A. V. S. Chaitanya, M. G. Keerthi, V. V. Gopalakrishna, J. Boutin, and C. de Boyer Montégut, "Assessment of seasonal and year-to-year surface salinity signals retrieved from SMOS and aquarius missions in the bay of Bengal," *Int. J. Remote Sens.*, vol. 37, no. 5, pp. 1089–1114, Mar. 2016, doi: 10.1080/01431161.2016.1145362.
- [4] E. A. Urquhart, B. F. Zaitchik, M. J. Hoffman, S. D. Guikema, and E. F. Geiger, "Remotely sensed estimates of surface salinity in the Chesa-peake bay: A statistical approach," *Remote Sens. Environ.*, vol. 123, pp. 522–531, Aug. 2012, doi: 10.1016/j.rse.2012.04.008.
  [5] C. de Boyer Montégut, "Mixed layer depth over the global ocean: An
- [5] C. de Boyer Montégut, "Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology," *J. Geophys. Res.*, vol. 109, no. C12, pp. 1–20, 2004, doi: 10.1029/2004JC002378.
- [6] G. Grunseich, B. Subrahmanyam, V. S. N. Murty, and B. S. Giese, "Sea surface salinity variability during the Indian ocean dipole and ENSO events in the tropical Indian ocean," *J. Geophys. Res., Oceans*, vol. 116, no. C11, pp. 1–43, Nov. 2011, doi: 10.1029/2011JC007456.
- [7] Ñ. Vinogradova, T. Lee, J. Boutin, K. Drushka, S. Fournier, R. Sabia, D. Stammer, E. Bayler, N. Reul, A. Gordon, O. Melnichenko, L. Li, E. Hackert, M. Martin, N. Kolodziejczyk, A. Hasson, S. Brown, S. Misra, and E. Lindstrom, "Satellite salinity observing system: Recent discoveries and the way forward," *Frontiers Mar. Sci.*, vol. 6, pp. 1–23, May 2019, doi: 10.3389/fmars.2019.00243.

- [8] F. M. Bingham and S. D. Howden, "Sea surface salinity measurements in the historical database," *J. Geophys. Res.*, vol. 107, pp. 1–10, Dec. 2002, doi: 10.1029/2000JC000767.
- [9] K. Balaguru, G. R. Foltz, L. R. Leung, J. Kaplan, W. Xu, N. Reul, and B. Chapron, "Pronounced impact of salinity on rapidly intensifying tropical cyclones," *Bull. Amer. Meteorol. Soc.*, vol. 101, no. 9, pp. E1497–E1511, Sep. 2020, doi: 10.1175/BAMS-D-19-0303.1.
- [10] N. Reul, "Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010– 2019)," *Remote Sens. Environ.*, vol. 242, Jun. 2020, Art. no. 111769, doi: 10.1016/j.rse.2020.111769.
- [11] P. N. Vinayachandran, V. S. N. Murty, and V. R. Babu, "Observations of barrier layer formation in the Bay of Bengal during summer monsoon," *J. Geophys. Res. Ocean.*, vol. 107, no. 12, pp. 1–9, 2002, doi: 10.1029/2001jc000831.
- [12] B. Kesavakumar, M. ArulMuthiah, S. Elango, D. Gowthaman, P. Kaliyaperumal, P. Senthilkumar, R. Sridharan, and R. Venkatesan, "Design of optimal power source for NIOT offshore moored buoy system," in *Proc. Ocean Electron. (SYMPOL)*, Oct. 2013, pp. 189–195, doi: 10.1109/SYMPOL.2013.6701929.
- [13] J. Boutin, "Satellite and *in situ* salinity: Understanding nearsurface stratification and subfootprint variability," *Bull. Amer. Meteorol. Soc.*, vol. 97, no. 8, pp. 1391–1407, Aug. 2016, doi: 10.1175/BAMS-D-15-00032.1.
- [14] O. Melnichenko, A. Amores, N. Maximenko, P. Hacker, and J. Potemra, "Signature of mesoscale eddies in satellite sea surface salinity data," *J. Geophys. Res., Oceans*, vol. 122, no. 2, pp. 1416–1424, Feb. 2017, doi: 10.1002/2016JC012420.
- [15] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.
- [16] S. Fournier, J. Vialard, M. Lengaigne, T. Lee, M. M. Gierach, and A. V. S. Chaitanya, "Modulation of the Ganges-Brahmaputra river plume by the Indian ocean dipole and eddies inferred from satellite observations," *J. Geophys. Res., Oceans*, vol. 122, no. 12, pp. 9591–9604, Dec. 2017, doi: 10.1002/2017JC013333.
- [17] C. Maes, N. Reul, D. Behringer, and T. O'Kane, "The salinity signature of the equatorial Pacific cold tongue as revealed by the satellite SMOS mission," *Geosci. Lett.*, vol. 1, no. 1, pp. 1–7, Dec. 2014, doi: 10.1186/s40562-014-0017-5.
- [18] R. Venkatesan, S. A. Sannasiraj, M. V. Ramanamurthy, P. Senthilkumar, and G. Dhinesh, "Development and performance validation of a cylindrical buoy for deep-ocean tsunami monitoring," *IEEE J. Ocean. Eng.*, vol. 44, no. 2, pp. 415–423, Apr. 2019, doi: 10.1109/JOE.2018.2819238.
- [19] N. Reul, J. Tenerelli, J. Boutin, B. Chapron, F. Paul, E. Brion, F. Gaillard, and O. Archer, "Overview of the first SMOS sea surface salinity products. Part I: Quality assessment for the second half of 2010," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1636–1647, May 2012, doi: 10.1109/TGRS.2012.2188408.
- [20] D. M. Le Vine, S. Abraham, F. Wentz, and G. S. E. Lagerloef, "Impact of the sun on remote sensing of sea surface salinity from space," *Int. Geosci. Remote Sens. Symp.*, vol. 1, no. 1, pp. 288–291, 2005, doi: 10.1109/IGARSS.2005.1526164.
- [21] N. Reul, J. E. Tenerelli, N. Floury, and B. Chapron, "Earth-viewing L-band radiometer sensing of sea surface scattered celestial sky radiation—Part II: Application to SMOS," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 3, pp. 675–688, Mar. 2008, doi: 10.1109/TGRS.2007.914804.
- [22] N. Reul, J. Tenerelli, B. Chapron, and P. Waldteufel, "Modeling sun glitter at L-band for sea surface salinity remote sensing with SMOS," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2073–2087, Jul. 2007, doi: 10.1109/TGRS.2006.890421.
- [23] T. Meissner, F. J. Wentz, and D. M. Le Vine, "The salinity retrieval algorithms for the NASA aquarius version 5 and SMAP version 3 releases," *Remote Sens.*, vol. 10, no. 7, pp. 1–26, 2018, doi: 10.3390/rs10071121.
- [24] T. Meissner and F. J. Wentz, "The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3004–3026, Aug. 2012, doi: 10.1109/TGRS.2011.2179662.
- [25] S. H. Yueh, W. Tang, A. G. Fore, and A. Hayashi, "Impact of ocean wave height on L-band passive and active microwave observation of sea surfaces," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 12, pp. 5491–5499, Dec. 2015, doi: 10.1109/JSTARS.2015.2432134.

- [26] S. H. Yueh, "L-band passive and active microwave geophysical model functions of ocean surface winds and applications to aquarius retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4619–4632, Sep. 2013, doi: 10.1109/TGRS.2013.2266915.
- [27] S. H. Yueh, S. J. Dinardo, A. G. Fore, and F. K. Li, "Passive and active L-band microwave observations and modeling of ocean surface winds," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3087–3100, Aug. 2010, doi: 10.1109/TGRS.2010.2045002.
- [28] S. Chen and C. Hu, "Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements," *Remote Sens. Environ.*, vol. 201, pp. 115–132, Nov. 2017, doi: 10.1016/j.rse.2017. 09.004.
- [29] H. Moussa, M. A. Benallal, C. Goyet, N. Lefevre, M. C. E. L. Jai, V. Guglielmi, and F. Touratier, "A comparison of multiple nonlinear regression and neural network techniques for sea surface salinity estimation in the tropical Atlantic Ocean based on satellite data," *ESAIM, Proc. Surv.*, vol. 49, pp. 65–77, Feb. 2015, doi: 10.1051/proc/ 201549006.
- [30] D. Rains, R. Sabia, D. Fernandez-Prieto, M. Marconcini, and T. Katagis, "Extended analysis of SMOS salinity retrieval by using support vector regression (SVR)," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 2265–2268, doi: 10.1109/IGARSS.2014.6946921.
- [31] A. Ammar, S. Labroue, E. Obligis, M. Crépon, and S. Thiria, "Building a learning database for the neural network retrieval of sea surface salinity from SMOS brightness temperatures," 2016, arXiv:1601.04296.
- [32] S. Rajabi-Kiasari and M. Hasanlou, "An efficient model for the prediction of SMAP sea surface salinity using machine learning approaches in the Persian Gulf," *Int. J. Remote Sens.*, vol. 41, no. 8, pp. 3221–3242, Apr. 2020, doi: 10.1080/01431161.2019.1701212.
- [33] S. Bao, H. Wang, R. Zhang, H. Yan, and J. Chen, "Comparison of satellitederived sea surface salinity products from SMOS, aquarius, and SMAP," *J. Geophys. Res., Oceans*, vol. 124, no. 3, pp. 1932–1944, Mar. 2019, doi: 10.1029/2019JC014937.
- [34] N. Vinogradova, T. Lee, J. Boutin, K. Drushka, S. Fournier, R. Sabia, D. Stammer, E. Bayler, N. Reul, A. Gordon, O. Melnichenko, L. Li, E. Hackert, M. Martin, N. Kolodziejczyk, A. Hasson, S. Brown, S. Misra, and E. Lindstrom, "Satellite salinity observing system: Recent discoveries and the way forward," *Frontiers Mar. Sci.*, vol. 6, p. 243, May 2019, doi: 10.3389/fmars.2019.00243.
- [35] E. P. Dinnat, D. M. Le Vine, J. Boutin, and T. Meissner, "Sea surface salinity: Inter-comparison of satellite products, *in situ* measurements, and impact of differences in retrieval algorithm," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 2943–2946, doi: 10.1109/IGARSS.2017.8127615.
- [36] T. Lee, "Consistency of Aquarius sea surface salinity with Argo products on various spatial and temporal scales," *Geophys. Res. Lett.*, vol. 43, no. 8, pp. 3857–3864, Apr. 2016, doi: 10.1002/2016GL068822.
- [37] W. Tang, A. Fore, S. Yueh, T. Lee, A. Hayashi, A. Sanchez-Franks, J. Martinez, B. King, and D. Baranowski, "Validating SMAP SSS with *in situ* measurements," *Remote Sens. Environ.*, vol. 200, pp. 326–340, Oct. 2017, doi: 10.1016/j.rse.2017.08.021.
- [38] J. Sun, G. Vecchi, and B. Soden, "Sea surface salinity response to tropical cyclones based on satellite observations," *Remote Sens.*, vol. 13, no. 3, p. 420, Jan. 2021, doi: 10.3390/rs13030420.
- [39] E. Jang, Y. J. Kim, J. Im, and Y.-G. Park, "Improvement of SMAP sea surface salinity in river-dominated oceans using machine learning approaches," *GISci. Remote Sens.*, vol. 58, no. 1, pp. 138–160, Jan. 2021, doi: 10.1080/15481603.2021.1872228.
- [40] S. Fournier, T. Lee, and M. M. Gierach, "Seasonal and interannual variations of sea surface salinity associated with the Mississippi river plume observed by SMOS and aquarius," *Remote Sens. Environ.*, vol. 180, pp. 431–439, Jul. 2016, doi: 10.1016/j.rse.2016.02.050.
- [41] Z. Mu, W. Zhang, P. Wang, H. Wang, and X. Yang, "Assimilation of SMOS sea surface salinity in the regional ocean model for south China sea," *Remote Sens.*, vol. 11, no. 8, p. 919, Apr. 2019, doi: 10.3390/rs11080919.
- [42] T. Meissner, L. Ricciardulli, and A. Manaster, "Tropical cyclone wind speeds from WindSat, AMSR and SMAP: Algorithm development and testing," *Remote Sens.*, vol. 13, no. 9, p. 1641, Apr. 2021, doi: 10.3390/rs13091641.
- [43] J. Boutin, N. Martin, X. Yin, J. Font, N. Reul, and P. Spurgeon, "First assessment of SMOS data over open ocean: Part II—Sea surface salinity," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1662–1675, May 2012, doi: 10.1109/TGRS.2012.2184546.

- [44] R. Oliva, E. Daganzo, P. Richaume, Y. Kerr, F. Cabot, Y. Soldo, E. Anterrieu, N. Reul, A. Gutierrez, J. Barbosa, and G. Lopes, "Status of radio frequency interference (RFI) in the 1400–1427 MHz passive band based on six years of SMOS mission," *Remote Sens. Environ.*, vol. 180, pp. 64–75, Jul. 2016, doi: 10.1016/j.rse.2016.01.013.
- [45] R. Venkatesan, "In situ ocean subsurface time-series measurements from OMNI buoy network in the Bay of Bengal," Curr. Sci., vol. 104, no. 9, pp. 1166–1177, 2013.
- [46] H. Lobeto, M. Menendez, and I. J. Losada, "Future behavior of wind wave extremes due to climate change," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, Dec. 2021, doi: 10.1038/s41598-021-86524-4.
- [47] G. C. Johnson, J. M. Toole, and N. G. Larson, "Sensor corrections for seabird SBE-41CP and SBE-41 CTDs," *J. Atmos. Ocean. Technol.*, vol. 24, no. 6, pp. 1117–1130, Jun. 2007, doi: 10.1175/JTECH2016.1.
- [48] K. Drushka, W. E. Asher, J. Sprintall, S. T. Gille, and C. Hoang, "Global patterns of submesoscale surface salinity variability," *J. Phys. Oceanogr.*, vol. 49, no. 7, pp. 1669–1685, Jul. 2019, doi: 10.1175/JPO-D-19-0018.1.
- [49] R. Venkatesan, K. Ramesh, M. A. Muthiah, K. Thirumurugan, and M. A. Atmanand, "Analysis of drift characteristic in conductivity and temperature sensors used in moored buoy system," *Ocean Eng.*, vol. 171, pp. 151–156, Jan. 2019, doi: 10.1016/j.oceaneng.2018.10.033.
- [50] C. Gabarró, M. Vall-Llossera, J. Font, and A. Camps, "Determination of sea surface salinity and wind speed by L-band microwave radiometry from a fixed platform," *Int. J. Remote Sens.*, vol. 25, no. 1, pp. 111–128, Jan. 2004, doi: 10.1080/0143116031000115175.
- [51] G. Lagerloef, C. Swift, and D. Le Vine, "Sea surface salinity: The next remote sensing challenge," *Oceanography*, vol. 8, no. 2, pp. 44–50, 1995, doi: 10.5670/oceanog.1995.17.
- [52] J. D. Droppleman, R. A. Mennella, and D. E. Evans, "An airborne measurement of the salinity variations of the Mississippi river outflow," *J. Geophys. Res.*, vol. 75, no. 30, pp. 5909–5913, Oct. 1970, doi: 10.1029/JC075i030p05909.
- [53] R. M. Lerner and J. P. Hollinger, "Analysis of 1.4 GHz Radiometric measurements from Skylab," *Remote Sens. Environ.*, vol. 6, no. 4, pp. 251–269, 1977, doi: 10.1016/0034-4257(77)90047-5.
- [54] L. Klein and C. T. Swift, "An improved model for the dielectric constant of sea water at microwave frequencies," *IEEE J. Ocean. Eng.*, vol. OE-2, no. 1, pp. 104–111, Jan. 1977, doi: 10.1109/JOE.1977.1145319.
- [55] F. W. LEBERL, "A review of: 'Microwave remote sensing-active and passive," *Int. J. Remote Sens.*, vol. 5, no. 2, p. 463, Mar. 1984, doi: 10.1080/01431168408948820.
- [56] S. H. Yueh, "Modeling of wind direction signals in polarimetric sea surface brightness temperatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 6, pp. 1400–1418, Nov. 1997, doi: 10.1109/36.649793.
- [57] A. Ammar, S. Labroue, E. Obligis, C. E. Mejia, M. Crepon, and S. Thiria, "Sea surface salinity retrieval for the SMOS mission using neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 3, pp. 754–764, Mar. 2008, doi: 10.1109/TGRS.2008.915547.
- [58] P. Jain and M. C. Deo, "Neural networks in ocean engineering," *Ships Offshore Struct.*, vol. 1, no. 1, pp. 25–35, Jul. 2006, doi: 10.1533/saos.2004.0005.
- [59] S. Thiria, C. Mejia, F. Badran, and M. Crepon, "A neural network approach for modeling nonlinear transfer functions: Application for wind retrieval from spaceborne scatterometer data," *J. Geophys. Res.*, vol. 98, no. C12, p. 22827, 1993, doi: 10.1029/93JC01815.
- [60] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata J., Promoting Commun. Statist. Stata*, vol. 20, no. 1, pp. 3–29, Mar. 2020, doi: 10.1177/1536867X20909688.
- [61] H.-Y. Kao, G. Lagerloef, T. Lee, O. Melnichenko, T. Meissner, and P. Hacker, "Assessment of aquarius sea surface salinity," *Remote Sens.*, vol. 10, no. 9, p. 1341, Aug. 2018, doi: 10.3390/rs10091341.
- [62] L. Yu, "On sea surface salinity skin effect induced by evaporation and implications for remote sensing of ocean salinity," J. Phys. Oceanogr., vol. 40, no. 1, pp. 85–102, Jan. 2010, doi: 10.1175/2009JPO4168.1.
- [63] A. V. S. Chaitanya, M. Lengaigne, J. Vialard, V. V. Gopalakrishna, F. Durand, C. Kranthikumar, S. Amritash, V. Suneel, F. Papa, and M. Ravichandran, "Salinity measurements collected by fishermen reveal a 'river in the Se' flowing along the eastern coast of India," *Bull. Amer. Meteorol. Soc.*, vol. 95, no. 12, pp. 1897–1908, Dec. 2014, doi: 10.1175/BAMS-D-12-00243.1.
- [64] K. Ashin, M. S. Girishkumar, K. Suprit, and V. P. Thangaprakash, "Observed upper ocean seasonal and intraseasonal variability in the Andaman sea," *J. Geophys. Res., Oceans*, vol. 124, no. 10, pp. 6760–6786, Oct. 2019, doi: 10.1029/2019JC014938.

- [65] C. D. Hoyos and P. J. Webster, "The role of intraseasonal variability in the nature of Asian monsoon precipitation," *J. Climate*, vol. 20, no. 17, pp. 4402–4424, Sep. 2007, doi: 10.1175/JCLI4252.1.
- [66] T. Furuichi, Z. Win, and R. J. Wasson, "Discharge and suspended sediment transport in the Ayeyarwady River, Myanmar: Centennial and decadal changes," *Hydrol. Process.*, vol. 23, pp. 1631–1641, 2009, doi: 10.1002/hyp.7295.
- [67] N. T. Vinogradova and R. M. Ponte, "Small-scale variability in sea surface salinity and implications for satellite-derived measurements," *J. Atmos. Ocean. Technol.*, vol. 30, no. 11, pp. 2689–2694, Nov. 2013, doi: 10.1175/JTECH-D-13-00110.1.
- [68] A. Behara and P. N. Vinayachandran, "An OGCM study of the impact of rain and river water forcing on the Bay of Bengal," J. Geophys. Res. Ocean., vol. 121, no. 4, pp. 1–14, 2016, doi: 10.1002/2015JC011325.
- [69] S. R. Shetye, A. D. Gouveia, D. Shankar, S. S. C. Shenoi, P. N. Vinayachandran, D. Sundar, G. S. Michael, and G. Nampoothiri, "Hydrography and circulation in the western bay of Bengal during the northeast monsoon," *J. Geophys. Res., Oceans*, vol. 101, no. C6, pp. 14011–14025, Jun. 1996, doi: 10.1029/95JC03307.
- [70] S. P. Kumar, "Intrusion of the bay of Bengal water into the Arabian sea during winter monsoon and associated chemical and biological response," *Geophys. Res. Lett.*, vol. 31, no. 15, pp. 4–7, 2004, doi: 10.1029/2004GL020247.
- [71] P. V. H. Kumar, B. Mathew, M. R. R. Kumar, A. R. Rao, P. S. V. Jagadeesh, K. G. Radhakrishnan, and T. N. Shyni, "Thermohaline front' off the east coast of India and its generating mechanism," *Ocean Dyn.*, vol. 63, nos. 11–12, pp. 1175–1180, Dec. 2013, doi: 10.1007/s10236-013-0652-y.
- [72] Y. Li, W. Han, M. Ravichandran, W. Wang, and T. Shinoda, "Bay of Bengal salinity stratification and Indian summer monsoon intraseasonal oscillation: 1. Intraseasonal variability and causes," *Nature*, vol. 175, no. 4449, p. 238, 2017, doi: 10.1038/175238c0.



**BALAKRISHNAN KESAVAKUMAR** received the B.E. degree in electrical and electronics engineering (EEE) from Anna University, Tamil Nadu, India, in 2010. He is currently pursuing the M.S. degree with the Department of Ocean Engineering, IIT Madras. He is also working as a Scientist with the National Institute of Ocean Technology. He is involved in deployment of different types of ocean observation platforms in Indian Ocean and Arctic Region. His current research interests include d development of machine learning-based models

satellite oceanography and development of machine learning-based models.



**PALANISAMY SHANMUGAM** received the Ph.D. degree in optical/microwave remote sensing techniques from Anna University, Chennai, India, in 2002.

He is currently a Professor with the Department of Ocean Engineering, IIT Madras, Chennai. His research interests include ocean optics and imaging, satellite oceanography, radiative transfer in the ocean, algorithm development, and underwater wireless optical communication studies.

Dr. Shanmugam has been the Principal Investigator of several projects funded by the Government of India.



**RAMASAMY VENKATESAN** received the Ph.D. degree from the Indian Institute of Science Bengaluru, India, in 2001. He is currently the Heads of the Ocean Observation Group, National Institute of Ocean Technology. He has received 20 awards at national and international level and to his credit 150 papers, technology transfer of seven products to industry, and four patents awarded. His research interests include ocean observation methods and ocean policy and management. He is also the Chair

of the GOOS-Regional Alliances of IOC-UNESCO and the Vice Chair of WMO SG-OOIS.

...