

# A Satellite-Derived Upper-Ocean Stratification Data Set for the Tropical North Atlantic With Potential Applications for Hurricane Intensity Prediction

Nguyen Dac Da<sup>1,2</sup> , Gregory R. Foltz<sup>2</sup> , and Karthik Balaguru<sup>3</sup> 

<sup>1</sup>Cooperative Institute for Marine and Atmospheric Studies (CIMAS), Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA, <sup>2</sup>Atlantic Oceanographic and Meteorological Laboratory (AOML), National Oceanic and Atmospheric Administration, Miami, FL, USA, <sup>3</sup>Marine Sciences Laboratory, Pacific Northwest National Laboratory (PNNL), Seattle, WA, USA

## Key Points:

- Inferring upper-ocean stratification from surface data is feasible with simple regression
- Stratification data set created using satellite data and a statistical model can realistically reproduce large and mesoscale features
- An example of the potential usefulness of the stratification data set is illustrated for tropical cyclone intensity prediction

## Supporting Information:

- Supporting Information S1

## Correspondence to:

G. R. Foltz,  
gregory.foltz@noaa.gov

## Citation:

Da, N. D., Foltz, G. R., & Balaguru, K. (2020). A satellite-derived upper-ocean stratification data set for the tropical North Atlantic with potential applications for hurricane intensity prediction. *Journal of Geophysical Research: Oceans*, 125, e2019JC015980. <https://doi.org/10.1029/2019JC015980>

Received 16 DEC 2019

Accepted 1 SEP 2020

Accepted article online 9 SEP 2020

**Abstract** Upper-ocean stratification strongly impacts vertical mixing and the heat flux between the ocean and atmosphere, especially under extreme conditions of tropical cyclones (TCs). Knowledge of prestorm stratification is important for accurate TC intensity prediction. In situ observations of the tropical ocean have significantly increased in the past decade. However, they are still too sparse to resolve ocean stratification variability in near-real time and on small spatial scales. In this study, based on long-term observations and an ocean reanalysis data set from 2004–2017, we investigate the possibility of retrieving upper-ocean stratification from sea surface temperature (SST), sea surface salinity (SSS), and sea surface height (SSH) using a simple regression method. It is found that more than 90% of the mean seasonal cycle and about 30% to 80% of temperature and salinity stratification anomalies can be reconstructed using surface data from either observations or an ocean reanalysis. Simple regression can be used with satellite observations to create a high-resolution, near-real-time-gridded ocean stratification data set that successfully reproduces both the large and mesoscale variability of ocean stratification. When used in a simple expression for TC-induced SST cooling, the satellite-derived stratification shows improvements over an ocean analysis in terms of variance explained of SST cooling, offering promise as a near-real-time indicator of the ocean's impact on TC intensification.

## 1. Introduction

Over the past several decades, tropical cyclone (TC) track predictions have improved significantly while TC intensity forecasts have seen much smaller advancements (Emanuel, 2018; Rappaport et al., 2012). One of the reasons for the relatively poor prediction skill of TC intensity is the lack of real-time observations of the upper ocean boundary layer beneath the storm (Emanuel, 2018). TC intensity is modulated by the enthalpy flux between the ocean and the TC, which depends on SST and vertical turbulent cooling induced by the TC's wind and modulated by ocean stratification (Bender & Ginis, 2000; D'Asaro et al., 2007; Price, 1981; Vincent, Lengaigne, Madec et al., 2012; Vincent, Lengaigne, Vialard et al., 2012). The maximum potential TC intensity can be predicted if real-time sea surface temperature (SST) under the storm is known (Emanuel, 1999). However, under extreme conditions, real-time SST is difficult to observe. Methods have been proposed to replace this real-time coupled SST by prestorm SST (SST-PI, Emanuel, 1999), temperature averaged to a fixed depth (OC-PI, Lin et al., 2013), or temperature averaged to a variable depth ( $T_{dy}$ , Balaguru et al., 2015).  $T_{dy}$  incorporates a more realistic TC-induced mixing length and therefore gives a closer approximation to real-time coupled SST than SST-PI or OC-PI. For  $T_{dy}$  to be practical, knowledge of prestorm ocean density and temperature stratification is required (Balaguru et al., 2015). In this study, we focus on building an observation-based ocean stratification data set available in high resolution and in near-real time that can be used for TC intensity prediction.

Existing data sets often provide quantities such as mixed layer depth, isothermal layer depth, and barrier layer thickness, which are not directly related to ocean stratification. Mixed layer depth provides information about the upper layer of weak stratification but does not give information about stratification beneath it. Barrier layer thickness only relates to the portion of salinity stratification that occurs within the isothermal layer. However, stratification in the thermocline layer can be important for studying phenomena that are

strong enough to modulate the mixed layer depth and the stratification at the base of the mixed layer, such as tropical cyclones, eddies, Madden Julian Oscillation, and El Niño–Southern Oscillation. Furthermore, the observation-based data sets are available at limited resolutions (i.e., monthly climatologies at 1–2° resolution (de Boyer Montégut et al., 2004, 2007; Kara, 2003; Zeng et al., 2016) or 10-day, 2° resolution for 2001–2009 (Hosoda et al., 2010)). The coarse resolutions of the available data sets do not resolve mesoscale stratification, which can be important for TC intensity changes (Goni et al., 2009). With the availability of satellite data in high resolution and in near-real time, we aim at creating a gridded stratification data set available in near-real time, with 0.25° horizontal and 4-day temporal resolution.

There are long-term reanalysis and analysis products such as the Simple Ocean Data Assimilation (SODA, Carton et al., 2018; [www.soda.umd.edu](http://www.soda.umd.edu)), Global Ocean Data Assimilation System (GODAS, Behringer & Xue, 2004; <http://www.cpc.ncep.noaa.gov>), and HYbrid Coordinate Ocean Model (HYCOM, Cummings, 2005; <http://hycom.org>) that have high resolutions (3 hr to 5 days, 0.5° to 1/12°) and can be used to derive ocean stratification. However, the numerical approach is often costly, and all of the reanalysis data sets mentioned do not assimilate newly available satellite sea surface salinity (SSS) data, which could be important for resolving salinity stratification, a useful indicator of the potential for TCs to undergo rapid intensification in the eastern Caribbean and western tropical Atlantic (Balaguru et al., 2020). Moreover, for intervalidation and applications that require an ensemble approach, such as weather and climate forecasts, it is desirable to have multiple products based on different methods. In this study, we use a statistical approach to predict ocean stratification from satellite surface observations.

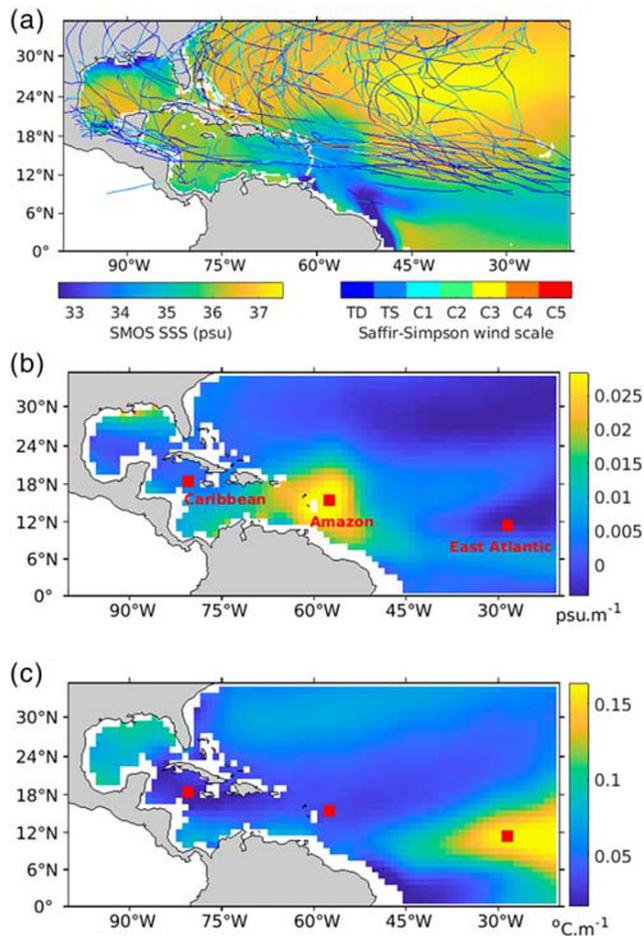
Several studies have explored the use of surface data for predicting subsurface ocean thermal structure (see Klemas & Yan, 2014 for a review). Khedouri et al. (1983) and Mayer et al. (2001) used temperature data along XBT tracks and sea surface height (SSH) from satellite altimetry to show that subsurface temperature variability and SSH variability are highly correlated, implying that SSH observations can be used to retrieve subsurface thermal structure. SST was also found to be a good predictor for temperature reconstruction in the South China Sea (Chu et al., 2000). Fischer (2000), using model output, found that both SST and SSH are important for subsurface temperature inference along the equator. Ali (2004) applied a neural network to moored buoy data in the Arabian Sea from 1994–1995 and found that, on average, about 92% of the variability in thermal structure from the surface to 300 m depth can be explained by the variability of SST, SSH, sea surface wind, net radiation, and net heat flux. These early studies, although for limited regions or time periods, provided a good basis for later large-scale temperature reconstruction studies (Guinehut et al., 2012; Pun et al., 2014, 2016; Su et al., 2015; Su, Huang, et al., 2018; Wu et al., 2012).

Only since the start of the Argo era (~2001) and SMOS (Soil Moisture and Ocean Salinity) satellite salinity measurements (2010) have there been enough salinity data to infer subsurface salinity (Guinehut et al., 2012) or density (Liu et al., 2017) from surface values. Guinehut et al. (2012), using Argo salinity profiles, SSH from altimetry, and simple regression found that only 20–30% of salinity anomalies can be reconstructed from SSH. However, the role of satellite SSS was not mentioned in their study. Liu et al. (2017) applied a simplified dynamic method called isQG (interior + surface quasi-geostrophic balance) to infer subsurface density from satellite SST, SSH, and SSS (SMOS) over the North Atlantic in 2012. They found that SSS is important for inferring density and that SSS from Argo gives better results than SSS from SMOS. Yet their study used only 1 year of SMOS data (2012), which is too short to come to any conclusions regarding the importance of satellite salinity for inferring subsurface density.

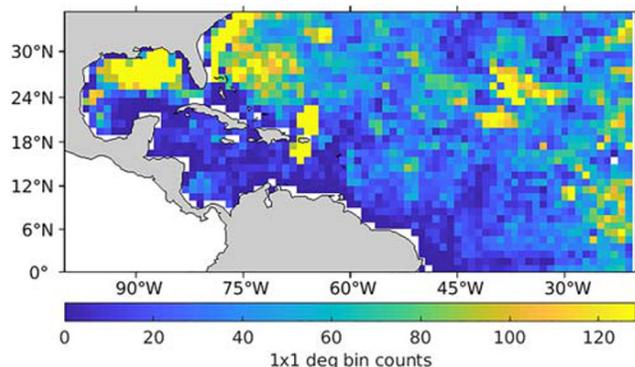
Our goals are first to assess whether we can reliably use surface predictors to infer upper-ocean stratification. We then explore the extent to which high-resolution satellite observations can be used to map upper-ocean stratification fields. With the use of satellite data as predictors, we aim at creating a gridded ocean stratification data set available at a high spatial resolution and that can ultimately be updated in near-real time. Finally, the ability of the satellite-based data set to resolve mesoscale features will be shown, and its potential usefulness for TC intensity prediction will be illustrated.

## 2. Data

In this study, we focus on the tropical North Atlantic (20°–100°W, 0–35°N, Figure 1), an important region for tropical Atlantic climate variability and TCs. It is also strongly impacted by freshwater input from the Mississippi River in the Gulf of Mexico, the Amazon-Orinoco in the western tropical Atlantic, and the



**Figure 1.** The tropical North Atlantic during hurricane season. (a) June–November climatological mean SMOS SSS (color shading) and storm tracks (color coded using Saffir–Simpson wind scale) from HURDAT2 data set (Landsea, & Frank.in, J. L., 2013) over the SMOS period (2010–2017). Corresponding June–November (b) salinity and (c) temperature gradients from the surface to 100 m computed using World Ocean atlas over the period 2005–2017.



**Figure 2.** WOD data coverage after filtering of the period 2004–2017. Some bins have more than 3,000 profiles.

ITCZ. This region thus has strong spatiotemporal variability of temperature and salinity stratification, making it an ideal region for ocean stratification inference. Satellite SSS has much higher spatial resolution than the salinity stratification field derived from the World Ocean Atlas (Figures 1a and 1b), illustrating one of the potential benefits of using satellite data to infer ocean stratification. The study period starts in 2004, when Argo reached global coverage, and ends in 2017.

In situ temperature and salinity profiles are used as predictors (SST and SSS only) and response variables (temperature, salinity, and density stratification) as well as for validation of regression models. We use the World Ocean Database 2018 (WOD), available at <https://www.nodc.noaa.gov/> (Boyer et al., 2018; Garcia et al., 2018), selecting only good-quality profiles (i.e., flag = 0) with at least 10 measurements of temperature and salinity in the upper 100 m and with at least one measurement within 0–10 m. This ensures homogeneous coverage and quality of temperature and salinity stratification as well as the representativeness of surface data. A total of ~92,000 temperature-salinity profiles were found to fit the criteria in the study region during 2004–2017 (Figure 2).

The satellite data consist of SST, SSS, and SSH, which are used as predictors in our regression model. We obtained the blended microwave-infrared SST product, version 5.0, produced by Remote Sensing Systems ([www.remss.com](http://www.remss.com)) with daily and 0.09° resolution. SSS data are from the SMOS mission and are produced by Centre Aval de Traitement des Données SMOS (CATDS, [www.catds.fr](http://www.catds.fr)). We use the Version 3, Level 3, 4-day, 25 km SSS product (Boutin et al., 2018). For SSH, we use gridded multimission delayed-time Level-4 absolute dynamic topography distributed by Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>). The data have daily and 0.25° resolution. In this study, all satellite data sets cover the 2004–2017 period, except SMOS SSS which is available only since 2010.

In addition to observational data sets, we use an ocean reanalysis product to test the feasibility of retrieving ocean stratification from surface data. Compared to observational data, reanalysis temperature, salinity, and SSH have a more consistent resolution and give an independent test of the performance of our regression method. We use reanalysis data from the Global Ocean Forecasting System (GOFS). GOFS uses the Hybrid Coordinate Ocean Model (HYCOM) with the Navy Coupled Ocean Data Assimilation (NCODA) system. The data cover the study period 2004–2017 and are available with 41 vertical layers at 3-hourly frequency and 0.08° spatial resolution (<https://www.hycom.org/>). Due to the large volume of data, we have regridded to daily and 0.25° resolution while keeping the original vertical resolution.

To assess the prediction skill of our regression models for monthly climatologies and higher temporal resolution anomalies, we compare to ocean stratification derived from a near-real-time numerical ocean model analysis, since one of our goals is to produce a satellite-based product in near-real time. We use the GOFS system Version 3.0 with 33 vertical layers, daily and 0.08° resolution, and covering the period 2009–2017. We regrid the data from 0.08° to 0.25° resolution, keeping the original vertical and temporal resolution. Hereafter, these data will be referred to as HYCOM analysis to distinguish it from the HYCOM reanalysis. The fundamental difference is that HYCOM analysis was conducted in nowcast

mode with near-real-time output, whereas HYCOM reanalysis was conducted in hindcast mode. This results in differences in forcing, grid resolution, assimilation schemes, and the amount and quality of the observations assimilated. More information about these data sets can be found at this site (<https://www.hycom.org/>).

To assess the quality of our monthly climatological stratification product, we also compare it to stratification calculated from the World Ocean Atlas 2018 data set (WOA18, Garcia et al., 2019), available at this site (<https://www.nodc.noaa.gov/>). The WOA18 data are available for different decadal periods (1955–1964, 1995–2004, 1981–2010, etc.). Here we select the product that covers 2005–2017, which is closest to our study period (2004–2017). This data set is also used to provide the background stratification during June–November (Figure 1) and to illustrate the computation of our stratification indexes (Figure 4).

Satellite surface chlorophyll *a* concentration (Chl*a*) and TC track data are used in section 5 to illustrate the usefulness of our stratification product by showing the relationships between stratification indices and TC-induced SST cooling and primary production. Chl*a* is obtained from the Ocean Color Climate Change Initiative project, version 3.1. This data set is distributed by European Space Agency and available online at this site (<http://www.esa-oceancolour-cci.org>). Chl*a* data have daily and 4 km resolution. We use the HURDAT2 database (Landsea and Franklin, 2013), which provides TC track data for the North Atlantic basin from 1851 to 2019. This data set is available at this site (<https://www.nhc.noaa.gov/>).

### 3. Methods

#### 3.1. Data Preprocessing

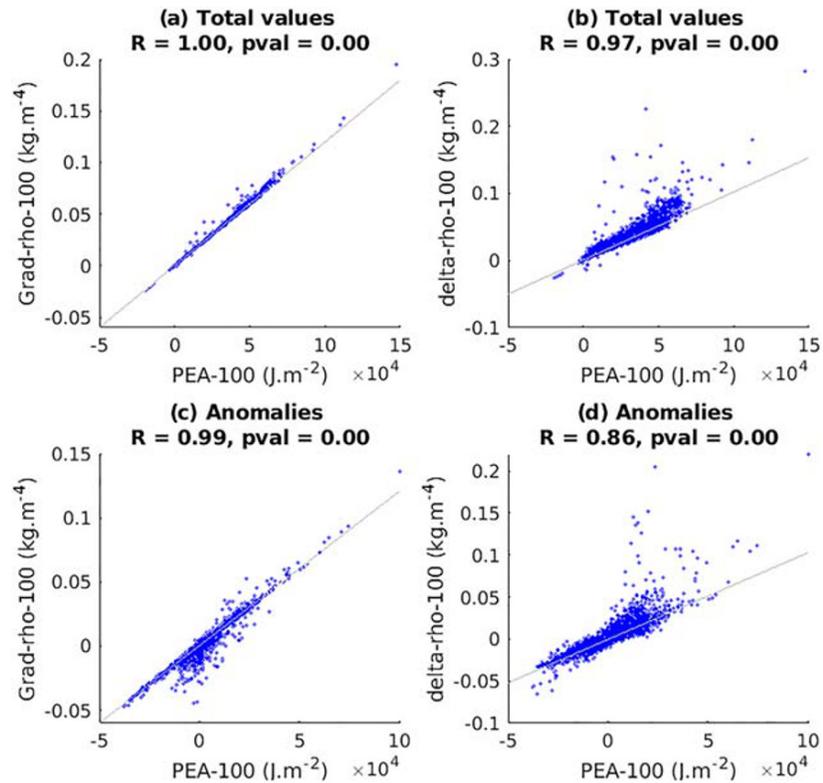
##### 3.1.1. Stratification Index Computation

Some studies use the difference of temperature, salinity, and density between the surface and 100 m (Balaguru et al., 2020) or 200 m (Capotondi et al., 2020; Yamaguchi & Suga, 2019), or between the mixed layer and 50 m below (Balaguru et al., 2015), to characterize upper ocean stratification. However, using two single depths to characterize the stratification of the whole layer may introduce bias to the indexes because it is sensitive to the variations at one of the depths (e.g., due to diurnal heating), whereas a linear fit takes into account the variations of all available data in the water column. Other studies use potential energy anomaly, which represents the energy input required to cause complete vertical mixing of the water column and is hence directly proportional to the strength of stratification (de Boer et al., 2008; MacKenzie & Adamson, 2004; Simpson, 1981; Yamaguchi et al., 2019). In this study, we use vertical gradients of temperature, salinity, and density to characterize stratification of the upper ocean. Stratification defined in this way is more widely applicable than potential energy anomaly and can be used directly in the computation of  $T_{dy}$  (described in section 5.2). Figures 3a and 3c show that potential energy anomaly and density gradients are very similar in the North Atlantic.

In situ, reanalysis, and analysis temperature and salinity profiles are first interpolated onto a vertical grid that extends from the surface to 100 m with 5 m spacing. Data within 5 m of the surface are rare and are extrapolated by assigning with the closest data available within the 0–10 m layer. Potential density at each depth is then computed using the UNESCO 1983 equation of state (Fofonoff & Millard, 1983), which is included in the seawater toolbox provided by the Commonwealth Scientific and Industrial Research Organization (CSIRO, <http://www.cmar.csiro.au/>). We fit all data from each profile to a simple linear regression model  $y = ax + b$  where  $y$  is either temperature, salinity, or density of the profile,  $x$  is the corresponding depths,  $b$  is a constant, and  $a$  represents the averaged rate of change between the surface and a given depth  $d$ . These values of  $a$  represent the gradients of temperature (Grad-T), salinity (Grad-S), and density (Grad-Rho).

For the gradient method, it is important to determine an appropriate value of  $d$  for the calculation. Price (2009) and Balaguru et al. (2020) found that a lower depth of 100 m is relevant for tropical cyclone studies. This depth is also deeper than the 20–80 m range of mixed layer depths found in the tropical ocean (de Boyer Montégut et al., 2004). Therefore, we compute our stratification indexes from the surface to  $d = 100$  m to include both the weak stratification of the mixed layer and strong stratification at the base of mixed layer. Over the North Atlantic, the density gradient (Grad-rho) agrees better with potential energy anomaly than the density difference between two depths ( $\Delta$ -rho, Figure 3).

Figure 4 illustrates the computation of temperature, salinity, and density gradients at the three locations indicated in Figure 1. The eastern Atlantic location is characterized by strong temperature stratification

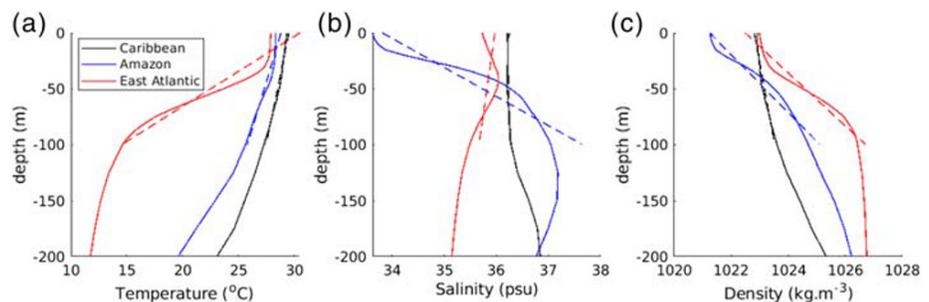


**Figure 3.** Comparisons of (a) Grad-rho and (b) Delta-rho stratification indexes to PEA. (c) and (d) Same as (a) and (b) except comparisons are shown for anomalies from the seasonal cycle. All calculations are performed over the tropical North Atlantic using WOD18 profiles during 2004–2017.

but weak salinity stratification. In contrast, the Amazon location has strong salinity stratification but weak temperature stratification. As a result, the magnitudes of density stratification are similar in the two regions. In the Caribbean, however, both temperature and salinity stratification are weak, leading to the weakest density stratification of the three locations.

### 3.1.2. Computation of Climatology and Anomalies

By definition, the annual cycle of ocean stratification and its predictors are invariant from year to year. We separate the strong, known annual signal from the total signal in order to emphasize the generally smaller interannual anomaly signal that is more difficult to infer. This improves the accuracy of the regression model for interannual anomalies and thus the total signal. Many studies cited in section 1 (Guinehut et al., 2012; Liu et al., 2017; Su, Li, et al., 2018) have used this method.



**Figure 4.** Climatological mean July (a) temperature, (b) salinity, and (c) density profiles from WOA18 at the locations marked by the red squares in Figure 1. Dashed lines show linear fits to the data (i.e., vertical gradients), between the surface and 100 m.

For WOD data, the stratification metrics from individual profiles are binned into monthly,  $1^\circ \times 1^\circ$  maps for climatology computation. For a given month, bins having less than three profiles are discarded, leaving about 41% of the total 2086  $1^\circ \times 1^\circ$  water bins on average, excluding the eastern Pacific. Monthly climatologies of stratification are then interpolated to daily climatologies from which anomalies are obtained by subtracting daily climatology values from individual profile values.

For the satellite data, we first compute monthly climatologies from each data set over 2004–2017, except 2010–2017 for SMOS SSS. The data are then regridded to  $1^\circ \times 1^\circ$  and linearly interpolated to daily climatologies to match the WOD data. To calculate anomalies, satellite data are interpolated to the WOD profiles' times and locations and then the gridded climatological values are subtracted. For satellite SST and SSS, when climatological values differ by more than  $\pm 1^\circ\text{C}$  or  $\pm 1$  psu, respectively, from WOD climatological values, the satellite climatological and associated anomaly values are discarded.

For the HYCOM reanalysis and analysis, monthly climatologies of stratification are first computed over the 2004–2017 and 2009–2017 periods, respectively. These monthly mean climatologies are then interpolated to daily means before computing daily anomalies to match the WOD data. For comparison with stratification predicted from the regression model, HYCOM analysis stratification anomalies are interpolated to the WOD profiles' times and locations.

### 3.2. Model Design

In this study, we choose to predict ocean stratification directly from surface data. An alternative is to map the three-dimensional structures of temperature and salinity first and then compute ocean stratification. This approach would not only require more resources but would lead to inconsistencies in the stratification fields due to spatial and temporal mismatches in the profiles used to create the gridded temperature and salinity fields. Existing observational mixed layer depth data sets (de Boyer Montégut et al., 2004; Hosoda et al., 2010) were also created by first computing the mixed layer depth from each profile and then gridding.

The feasibility of using SST, SSS, and SSH to predict upper-ocean stratification is tested using two different data sets: in situ-based observations with sparse and irregular spatial and temporal distributions, and a reanalysis with consistent and higher spatial and temporal resolutions. We design different regression models for each data set. One model (called “RegWOD”) uses purely observations, with WOD SST and SSS and satellite SSH as the predictors and stratification computed from WOD profiles as the responses. SSH from satellite altimetry is used in RegWOD due to the lack of global in situ observations of SSH. The other model (called “RegHYCOM”) uses predictors (SST, SSS, and SSH) and responses (Grad-T, Grad-S, and Grad-Rho) from the HYCOM reanalysis. Both models are applied to monthly mean climatologies and daily anomalies, resulting in separate estimates of the climatological and anomalous stratification derived from surface data. The performances of RegWOD and RegHYCOM are assessed to determine the skill with which ocean stratification can be retrieved using only surface data.

To investigate the possibility of using high-resolution satellite observations to produce a high-resolution stratification product, one possible approach is to replace in situ SST and SSS with satellite data in RegWOD and test its prediction skill. Another approach is to build a regression model, called “RegSAT,” which uses satellite MW-IR SST, SMOS SSS, and altimetry SSH as predictors and co-located WOD-based stratification as the response. However, RegSAT can only be trained over the period 2010–2017 due to limitations of the SMOS SSS data. Comparison of the performances of RegWOD using satellite data and RegSAT will give a comprehensive assessment of the ability to use high-resolution satellite data for creating an upper-ocean stratification data set. The sources of data for predictors and responses and study periods used in each model are summarized in Table 1.

### 3.3. Model Training and Skill Assessment

The regression models RegWOD, RegHYCOM, and RegSAT have the following form:

$$Y = a_1X_1 + a_2X_2 + \dots + a_kX_k + b \quad (1)$$

where  $k$  represents the number of predictors used in the model,  $Y$  represents the responses (Grad-T, Grad-S, or Grad-Rho) at a given location,  $X$  represents predictors in the model (SST, SSS, and SSH),  $a$  represents the regression coefficients to be determined, and  $b$  is a constant term. Note that  $b$  is constant only in time and can vary with location.

**Table 1**  
Specifications of Data Sources for Predictors and Responses and the Periods Used for the Different Statistical Models

Model	Predictors			Responses (grad-T, grad-S, grad-rho)	Period
	SST	SSS	SSH		
RegWOD	WOD	WOD	altimetry	WOD	2004–2017
RegHYCOM	HYCOM Reanalysis				2004–2017
RegSAT	MW-IR	SMOS	altimetry	WOD	2010–2017
RegWOD for final product	MW-IR	SMOS	altimetry		2010–2017

### 3.3.1. Data Sampling

In this study, we adopt a supervised learning approach to build our regression model. Therefore, it is important to separate a portion of the data as an independent test set that can be used to assess the prediction skill of the regression model after training. We randomly select 80% of the (X,Y) data for model training of RegWOD, RegHYCOM, and RegSAT. The remaining 20% are used as an independent test set in each regression model. The training and test data sets must be representative of the original population in terms of range and variability. We check the representativeness using a two-sample Komogorov-Smirnov test. Figure 5 shows an example of the distributions of the original, training, and test data sets. At the 95% confidence level, the test failed to reject the null hypothesis that the samples have the same variability and are drawn from the same population.

### 3.3.2. Model Training

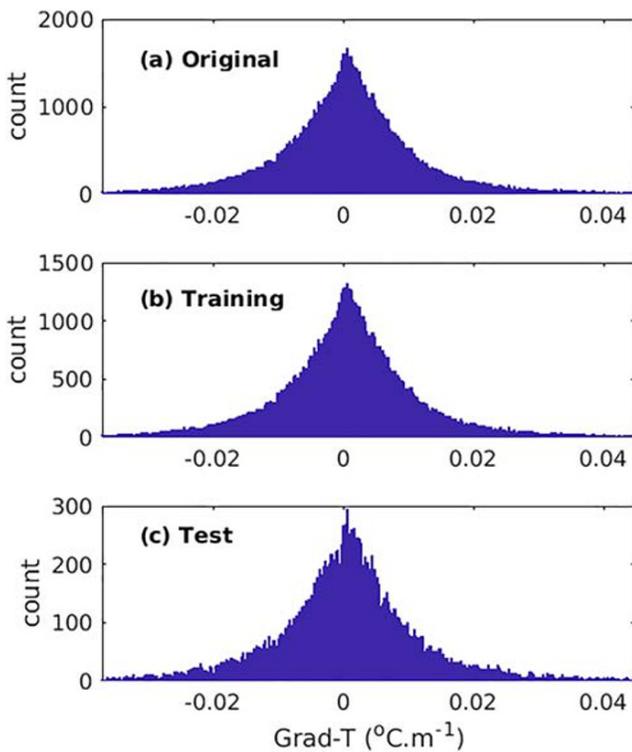
The goal of model training is to determine, for a given location, the optimum model size  $k$ , the best predictors to use, and the associated regression coefficients. This process is divided into two stages: training to obtain optimum model form (predictor selection and model size selection) and training to obtain optimum regression coefficients associated with the optimum model form (James et al., 2013).

The model size can vary from one to four (SST, SSS, SSH, and constant term). To choose the best combination of predictors for a given model size, we first create a model for each combination of predictors (e.g., there are six possible combinations for the two-predictor model: constant-SST, constant-SSS, constant-SSH, SST-SSS, SST-SSH, and SSS-SSH). To avoid overfitting, we apply the so-called fivefold cross-validation method (James et al., 2013) by randomly dividing the training data, which is 80% of total data, into fivefolds (parts). For each model we choose fourfolds of the training data to train the model and the remaining onefold to test the model prediction skill. This process is repeated five times, each time with a new fold as the test set. Residual standard error (RSE), which represents how far on average the prediction value deviates from the observations, is commonly used to estimate the prediction skill of statistical models (James et al., 2013):

$$RSE = \sqrt{\frac{\sum_{i=1}^N (Y_i^{predicted} - Y_i^{test})^2}{N - 2}} \quad (2)$$

where  $Y_i^{predicted}$  is the predicted stratification for the test set by the regression model,  $Y_i^{test}$  is the ocean stratification in the test set, and  $N$  is the size of the test set. The combination of predictors with the smallest RSE, averaged over all five testing folds, is chosen as the best predictor combination for that model size. For example, taking temperature stratification as the response, the best one-predictor model could be SST, the best two-predictor model might be {constant term, SSH}, and so on.

After the best predictor combinations were calculated for model sizes of one to four predictors, we determined the best model size using the one standard error rule (James et al., 2013). We selected the simplest



**Figure 5.** Example of representativeness of (b) training and (c) testing set against (a) original data set. X axis shows temperature stratification in RegWOD data set and Y axis shows count number.

possible model with insignificantly different performance compared to the minimum RSE model. Any model with an averaged RSE within one standard error of the minimum RSE model fits this criterion. The idea is that if a more complicated model does not perform significantly better than a simpler model, the simpler model is preferable. More details about cross validation, feature selection, and model size selection procedures can be found in sections A1–A3.

The final step in the training process is to calculate the best regression coefficients associated with the best model form. During the training process, only 80% (fourfold) of the training data were used to train the model. After the best model form was selected, we recalculated the regression coefficients using all data in the training set instead of only 80%. The refined regression coefficients capture the variability in all of the training data, improving the prediction skill over the independent test set.

### 3.3.3. Model Testing

The next step is to use the final model for prediction on the independent test set. This test set was not used in the training process and thus provides an unbiased estimate of the model's prediction skill. We assess the skill using the coefficient of determination (R2), which calculates the fraction of the total variance that is explained by the model:

$$R2 = 1 - \frac{\sum_{i=1}^N (Y_i^{\text{predicted}} - Y_i^{\text{test}})^2}{\sum_{i=1}^N (Y_i^{\text{test}} - \bar{Y}^{\text{test}})^2} \approx 1 - \frac{RSE^2}{\text{Variance}(Y^{\text{test}})} \quad (3)$$

$$\text{Variance}(Y^{\text{test}}) = \frac{\sum_{i=1}^N (Y_i^{\text{test}} - \bar{Y}^{\text{test}})^2}{N - 1} \quad (4)$$

where  $\text{Variance}(Y^{\text{test}})$  represents the total variability of observations in the independent test set, and the ratio  $\frac{RSE^2}{\text{Variance}(Y^{\text{test}})}$  represents the variance not explained by the regression model. RSE and R2 are thus interdependent, and for short, we will only show R2 as a metric of model skill.

### 3.3.4. Method Adjustments

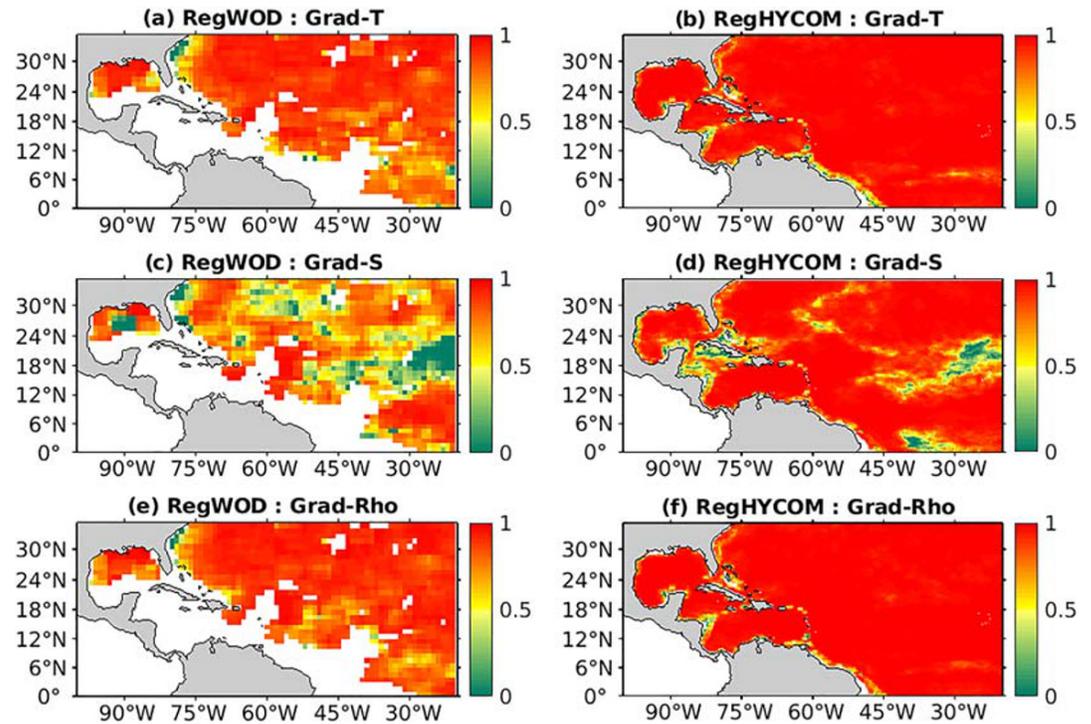
We apply regression locally in order to take into account the geographical variations in the relationship between surface predictors and ocean stratification. For gridded data such as the HYCOM reanalysis, at each grid point the temporal coverage is great enough to give a stable regression. However, the irregularity and scarcity of WOD data requires a different approach. We perform the regression over a  $1^\circ \times 1^\circ$  grid and apply a search window of  $5^\circ$  longitude  $\times$   $3^\circ$  latitude. For each location and associated search window, we perform the regression only when there are more than 30 measurements for the whole period. The same method is used to calculate model prediction skill.

## 3.4. Final High-Resolution Data Set Creation

The best climatology and anomaly regression models at each location are then applied to high-resolution satellite data to obtain the high-resolution stratification data sets. Our final stratification data set has the resolution of the coarsest satellite product, which is SMOS SSS at  $0.25^\circ$  and 4 days.

Over the study region, there are locations without enough data for stable regression coefficients, even when the  $5^\circ \times 3^\circ$  window has been applied. These missing locations are filled using the nearest neighbors method. If the missing bin is over the ocean and at least three of eight neighboring bins have data, the regression coefficients of the neighboring bins are averaged and assigned to the missing bin. Note that different locations may have different model sizes (i.e., different number of predictors used). In this case, the regression coefficients of the predictors that are not used in the final model are assigned to zero before averaging and assigning to the missing bin. These steps are repeated until all missing ocean bins are filled with an averaged regression model of neighboring bins.

Due to the irregularly distributed observations, the regression models of two neighboring bins may strongly differ, resulting in noisiness in the predicted stratification field. We overcome this problem by applying a  $9^\circ \times 9^\circ$  moving average window. The smoothed regression coefficients are then resampled to the same resolution as the predictors ( $0.25^\circ$ ). This averaging is only applied when creating the regression models so that the stratification data set retains the high resolution of the original satellite data.



**Figure 6.** Performance ( $R^2$  maps over independent test set) of RegWOD (left column: a, c, and e) and RegHYCOM (right column: b, d, and f) on monthly climatology data for temperature (top row: a, b), salinity (middle row: c, d) and density (bottom row: e, f) stratification.

## 4. Results

### 4.1. Subsurface Data Inference Feasibility

Before calculating stratification from satellite data, we establish the feasibility of inferring stratification from surface data. Figure 6 shows the performance of RegWOD (left column) and RegHYCOM (right column) over the independent test set (20% of total data) for the monthly mean climatology. The prediction skill of RegWOD for temperature stratification is high over most areas, with  $R^2$  between 0.6 and 1 (Figure 6a). The skill of RegHYCOM is higher, generally ranging from 0.9 to 1 (Figure 6b). This is due to more consistent spatial–temporal resolution in the reanalysis compared to observations. Using a neural network and data from a buoy over the period 1994–1995, Ali (2004) also found an average  $R^2$  of 0.92 for the annual cycle of the thermal structure from the surface to 300 m. The white areas in RegWOD maps (Figures 6a, 6c, and 6e) are where data are missing or the sample size is too small for a reliable calculation.

For the monthly climatology, the overall performances of RegWOD and RegHYCOM are very high ( $R^2$  of 0.94 and 0.99, respectively, calculated across all grid points; Table 2). For salinity stratification, the performance of RegWOD is high over most areas (Figure 6c). The locations of the regions of poor performance are similar between RegWOD and RegHYCOM, but the areas are larger in RegWOD. The overall performances of RegWOD and RegHYCOM for salinity stratification are also very high at 0.91 and 0.97, respectively (Table 2). Annual variability of density stratification is strongly impacted by temperature stratification so that its  $R^2$  map and overall  $R^2$  are very similar to those of temperature stratification (Figures 6e and 6f; Table 2).

The prediction skills of RegWOD and RegHYCOM are lower when applied to daily anomalies (Figure 7). There are noticeable differences between RegWOD and RegHYCOM for temperature and density stratification (Figures 7a, 7b, 7e, and 7f), though the agreement is better for salinity stratification. RegWOD and RegHYCOM both have high skill for salinity stratification in areas where salinity stratification is strong: over the Mississippi River plume in the northern Gulf of Mexico, near the outflow from the Amazon-Orinoco River system, and in the 5°–10°N band of the ITCZ (Figure 1b). There is no clear relationship between temperature stratification magnitude and the models' prediction skill. While RegWOD has much weaker overall

**Table 2**  
Overall Performance (*R*-Squared) of RegWOD and RegHYCOM Over the Independent Test Set

	Climatology		Anomaly	
	RegWOD	RegHYCOM	RegWOD	RegHYCOM
Grad-T	0.94	0.99	0.30	0.55
Grad-S	0.91	0.97	0.70	0.76
Grad-Rho	0.94	0.98	0.51	0.69

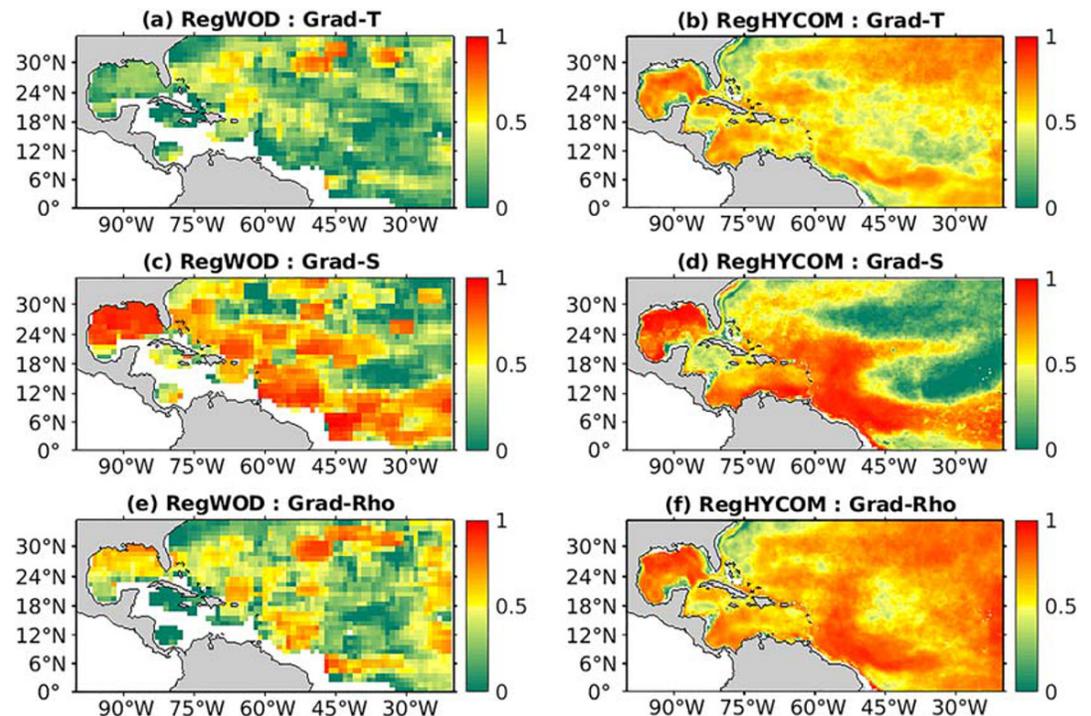
performance for temperature stratification (0.30 vs. 0.55 for RegHYCOM), its overall performance for salinity stratification is similar (0.70 vs. 0.76). The poorer performance of RegWOD for temperature stratification is connected to its poorer performance for density stratification compared to RegHYCOM (*R*<sup>2</sup> of 0.51 and 0.69, respectively). These results show that simple regression is more skillful for salinity stratification than temperature stratification in the tropical North Atlantic.

The differences in model skill for temperature and salinity stratification may be due in part to differences in the magnitudes of climatologies and anomalies. The ratios of the standard deviations of anomalies to standard deviations of climatologies for temperature, salinity, and density stratification are 31%, 52%, and 36%, respectively for WOD data and 32%, 47%, and 36%, respectively, for HYCOM reanalysis (Figure S1). The weaker variability of anomalies compared to climatologies may explain the reduction in model skill for anomalies compared to climatologies. Similarly, the weaker variability of temperature stratification anomalies compared to salinity stratification anomalies (31% vs. 52% in WOD data and 32% vs. 47% in HYCOM reanalysis) may explain the larger drop in model skill for temperature stratification.

In summary, using a simple regression method with surface data from observations or reanalysis, we can reconstruct about 90–100% of the monthly climatological variations of upper ocean stratification and about 30–80% of the anomaly variance. These results are encouraging for the prospect of inferring stratification from satellite surface data.

#### 4.2. Feasibility of Creating High Resolution Stratification Data Set Based on Satellite Surface Predictors

In this section, we investigate the possibility of using high-resolution satellite data to create a stratification data set. For reference, first we compare the skill of RegWOD using in situ SST and SSS with the skill of the HYCOM analysis (Table 3, Column 2 vs. Column 1). The HYCOM analysis is chosen because of its availability in near-real time, which is the ultimate goal for our satellite-based product. For temperature



**Figure 7.** Similar to Figure 6 but for daily anomaly data. Performance (*R*<sup>2</sup> maps) of RegWOD (left panels: a, c, and e) and RegHYCOM (right panels: b, d, and f) on their independent test set of daily anomaly data.

**Table 3**  
Variance in Observed Stratification Explained ( $R^2$ ) by HYCOM Analysis, RegSAT, and RegWOD Using Different Sources of SST and SSS Data

Variable	HYCOM analysis (1)	Prediction using RegWOD					Prediction using RegSAT
		WOD SST, WOD SSS (2)	MW-IR SST, WOD SSS (3)	WOD SST, SMOS SSS (4)	MW-IR SST, SMOS SSS (5)	MW-IR SST, Corr-SSS (6)	MW-IR SST, SMOS SSS (7)
Grad-T-clim (1)	0.94	0.94	0.94	0.93	0.94	0.94	0.94
Grad-S-clim (2)	0.71	<b>0.90</b>	<b>0.90</b>	<b>0.78</b>	<b>0.78</b>	<b>0.81</b>	<b>0.78</b>
Grad-Rho-clim (3)	0.92	0.94	0.94	0.92	0.92	0.93	0.93
Grad-T-anom (4)	0.25	0.25	<u>0.19</u>	0.24	<u>0.18</u>	<u>0.18</u>	<u>0.19</u>
Grad-S-anom (5)	0.17	<b>0.68</b>	<b>0.68</b>	0.20	0.20	<b>0.27</b>	<b>0.27</b>
Grad-Rho-anom (6)	0.37	<b>0.45</b>	0.41	<u>0.28</u>	<u>0.24</u>	<u>0.25</u>	<u>0.26</u>

Note. Corr-SSS stands for corrected SSS. Italics indicate reference values (HYCOM Analysis), regular text means the magnitude of the difference relative to the reference is  $<0.05$ , underline means the reference is larger by at least 0.05, and bold means the reference is smaller by at least 0.05.

stratification, the performance of RegWOD is similar to that of HYCOM, with  $R^2$  of 0.94 for climatological data and 0.25 for anomalies. For salinity stratification, RegWOD outperforms HYCOM analysis for the climatology (0.90 vs. 0.71) and especially anomalies (0.68 vs. 0.17). For density stratification, RegWOD has comparable skill for the climatology (0.94 vs. 0.92) but better skill for anomalies (0.45 vs. 0.37). These results suggest that RegWOD, if based on in situ surface data, can provide better stratification fields in near-real time compared to the HYCOM analysis and much of the improvement is salinity related.

As mentioned in section 3.2, satellite SST and SSS can be used as predictors in RegWOD, which was built based on in situ SST and SSS training data, or the satellite data can be used to build a satellite-only version of the regression model (RegSAT). We will consider both of these options. The HYCOM analysis is chosen as the reference to assess the skill of the regression model with satellite predictors (section 2). Note that for anomaly data, RegWOD was trained over the period 2004–2017 and RegSAT was trained over the period 2010–2017 due to the limitation of SMOS SSS data. For RegSAT, we used the same time locations as in the test set of RegWOD. For stratification anomalies, the comparisons of skill between RegWOD (with satellite predictors) and RegSAT are conducted over the test set of RegSAT (~14,000 data points) and are summarized in Table 3.

Comparing Columns 2 and 3 in Table 3, we see that replacing WOD SST with MW-IR SST does not significantly change the performance of RegWOD, except for a decrease in  $R^2$  from 0.25 to 0.19 for anomaly temperature stratification. This indicates that the quality of MW-IR SST is comparable to WOD SST. However, when we replace WOD SSS with SMOS SSS (Columns 2 and 4 in Table 3), the performance of RegWOD decreases for climatological salinity stratification ( $R^2$  goes from 0.90 to 0.78), anomalous salinity stratification (0.68 to 0.20), and anomalous density stratification (0.45 to 0.28), in agreement with the findings of Liu et al. (2017). Similarly, when both WOD SST and SSS are replaced with their satellite counterparts (Columns 2 and 5 in Table 3), the decreases in performance are consistent with the linear superposition of decreases due to satellite SST and SSS. These results suggest that the cause of the decrease in skill may be the resolution and/or quality of the satellite SSS data.

For all climatologies and temperature and density stratification anomalies, RegSAT, trained directly with satellite SST and SSS, performs similarly to satellite RegWOD (Columns 5 and 7 in Table 3) and has much worse performance than in situ RegWOD for anomalies (Columns 2 and 7 in Table 3). RegSAT has a small improvement compared to satellite RegWOD (columns 5 and 7) for anomalies, especially salinity stratification (0.27 for RegSAT vs. 0.20 for RegWOD). To test the sensitivity to the training period, we trained RegWOD over the RegSAT training period (2010–2017) and found that the performance is similar

compared to training over the longer period. Hence, for salinity stratification the improvement of RegSAT compared to satellite-RegWOD is not due to the difference in their training period and instead is likely the result of better adaptation of RegSAT to SMOS SSS anomalies (i.e., the same satellite data are used to train and predict with RegSAT, but this is not the case with satellite-RegWOD). Overall, we conclude that the regression model based on satellite surface data provides a useful source of ocean stratification information that is independent of the HYCOM analysis and comparable in quality. Although RegSAT is slightly better than satellite-RegWOD for salinity stratification prediction over the 2010–2017 common period, we will use RegWOD to create our final product because it was trained with more data over a longer period (2004–2017) and thus is more stable.

### 4.3. Refinements to the Satellite Regression Model

SST and SSS are the most important predictors for temperature and salinity stratification, respectively (see Text S1 and Figure S3 in the supporting information). Results in section 4.1 showed that RegWOD is much more skillful for anomalous salinity than anomalous temperature stratification prediction when using in situ SST and SSS. Results in section 4.2 showed that replacing in situ SSS with SMOS SSS caused much worse performance of RegWOD than replacing in situ SST by MW-IR SST. This led to the conclusion that the prediction skill of RegWOD for salinity and density stratification anomalies can potentially be increased by improving the quality of satellite SSS. In this section, we investigate the reasons why satellite SSS may be worse than SST and the potential to improve satellite SSS for predicting stratification.

To assess the quality of satellite SST and SSS, we compare them to WOD SST and SSS, respectively (Figure 8). The comparison periods are 2004–2017 for SST and 2010–2017 for SSS. Figures 8a–8c show that MW-IR SST agrees well with WOD SST, with  $R^2$  of 0.99, 0.99, and 0.78 for total, climatology, and anomaly values, respectively. However, the agreement between SMOS and WOD SSS is much worse, especially for anomalies, with  $R^2$  of 0.91, 0.91, and 0.31 for total, climatology, and anomaly values, respectively. These results explain the much larger degradation in prediction skill of RegWOD when replacing WOD SSS with SMOS SSS compared to replacing WOD SST with MW-IR SST.

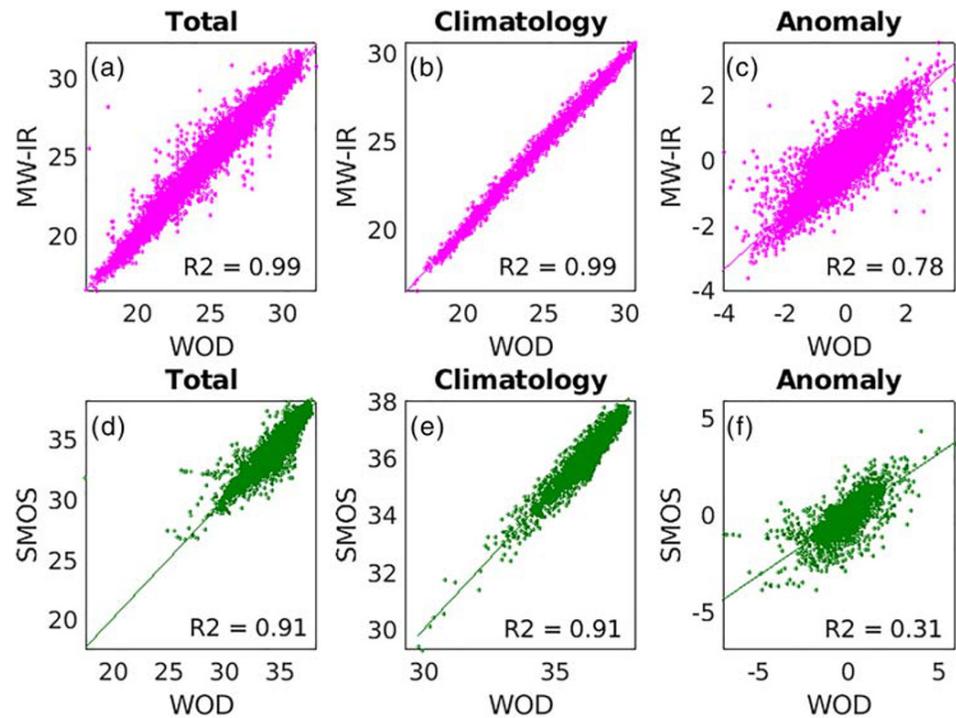
The poorer performance of SMOS SSS could be due to its resolution and/or quality. MW-IR SST has a higher resolution than SMOS SSS (0.09° vs. 0.25° and daily vs. 10 day), and WOD in situ SST and SSS are from individual profiles on scales of meters and hours. Thus, the interpolation of satellite data to WOD times and locations would create better match-ups for satellite SST than SSS. To test this, we re-gridded satellite SST to the same spatiotemporal resolution as SMOS SSS before interpolating to the WOD profiles' time-locations. Then we recalculated  $R^2$  for both climatology and anomaly data over the SMOS period (2010–2017). The performance of this re-gridded SST data is the same as the performance of the original SST (figure not shown). Hence, poorer data quality, not coarser resolution, is likely the reason for the worse performance of SMOS SSS. Boutin et al. (2018) reported that besides being strongly influenced by land sea contamination and civil radio frequency interference, SMOS SSS has low skill for low-salinity water.

To overcome this partially, we build a statistical model to correct SMOS SSS toward WOD SSS using simple regression, with WOD SSS as response, and SMOS SSS and a constant term as predictors. For consistency, the model is trained using the same method described in section 3.3. Its performance is tested on an independent test set (20% of the total data) using time locations that are the same as the independent test set used in RegWOD. Over the test set, corrected SSS has better agreement with WOD SSS than the original SMOS SSS, with  $R^2$  higher by 0.06 and 0.18 for climatology and anomaly data, respectively. Using the corrected SSS,  $R^2$  of RegWOD for salinity stratification increases by 3% for the climatology and 7% for anomalies (Table 3, Columns 5 and 6) compared to using the original SSS. These increases are significant at the 5% level using a bootstrap method. The resultant overall performance of RegWOD using corrected SSS is comparable to RegSAT and the HYCOM analysis (Table 3, Columns 1, 5, and 6). Therefore, we will use RegWOD with satellite data (corrected SSS) to create our final high-resolution stratification data set.

## 5. Final Product Performance

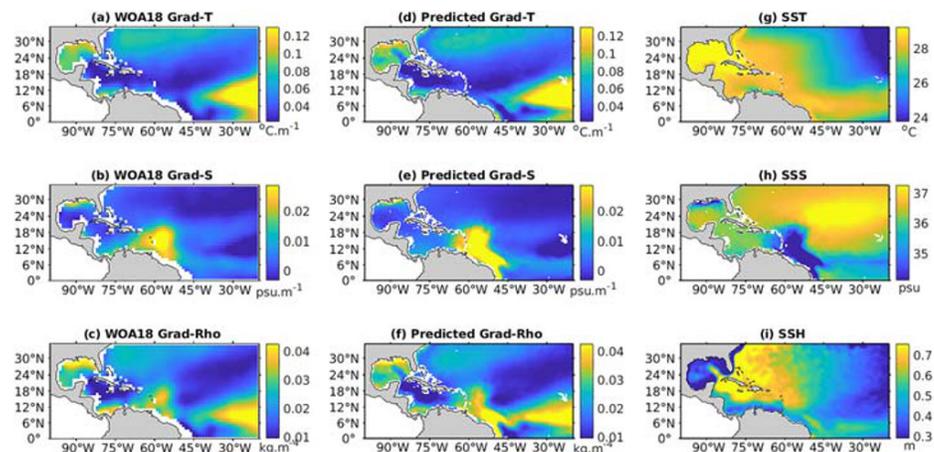
### 5.1. Mesoscale Resolving

Figure 9 gives an example of high-resolution climatological satellite data mapped to upper-ocean stratification fields using the statistical model RegWOD. The predicted stratification fields (Figures 9d–9f) reproduce

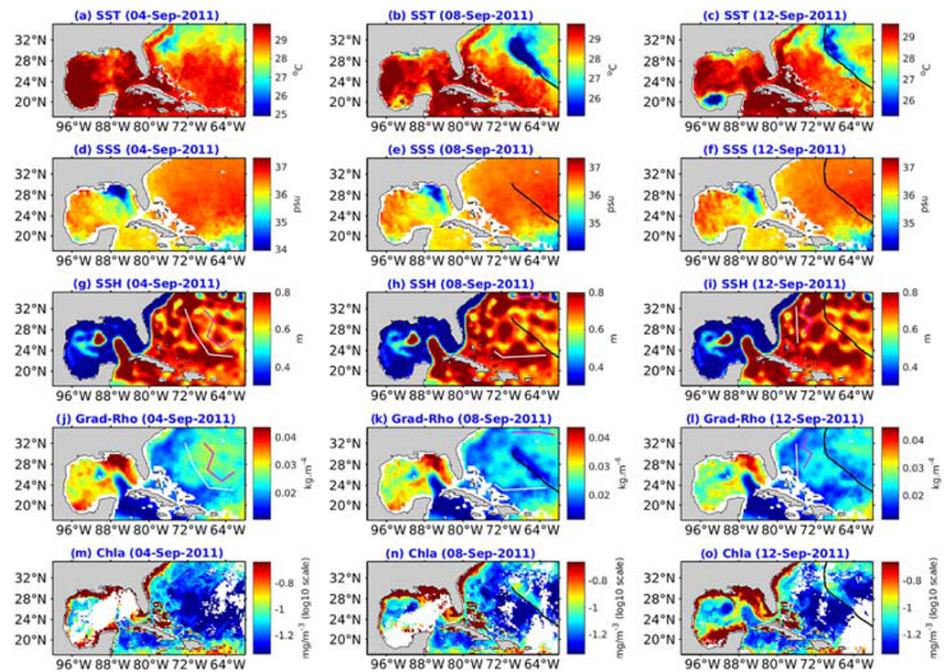


**Figure 8.** Comparison of MW-IR versus WOD SST (upper panels: a–c) and SMOS versus WOD SSS (lower panels: d–f): Total values (left: a, d), monthly  $1^\circ \times 1^\circ$  climatology (middle: b, e), and anomaly (right: c, f). Data for which climatology pair values differ by more than  $1^\circ\text{C}$  or 1 psu were removed.

well the large-scale variations of stratification derived from the WOA18 data set (Figures 9a–9c). The predicted fields additionally have higher horizontal resolution that allows small-scale features to be resolved in greater detail. In the Gulf of Mexico and Caribbean Sea, temperature and density stratification vary out of phase with SSH. The region of high SSH in the Gulf of Mexico is associated with the well-known Loop Current (Figure 9i). This feature is visible in the predicted temperature and density stratification fields (Figures 9d and 9f), with weak stratification where SSH is high. The feature is absent in the WOA18-derived



**Figure 9.** July climatologies of (a)–(c) stratification from WOA18, (d)–(f) stratification predicted using RegWOD, and corresponding satellite surface data: (g) MW-IR SST, (h) corrected version of SMOS SSS, and (i) altimetry SSH. Note that the sign convention is such that positive values represent stable stratification (temperature, salinity, and density decreasing, increasing, and increasing with depth, respectively).



**Figure 10.** Daily maps of (a)–(i) satellite predictors, (j)–(l) density stratification, and (m)–(o) chlorophyll *a*. The magenta and white solid lines connect the warm and cold eddies respectively in (g)–(i) and their corresponding density stratification in (j)–(l). Black solid lines show the track of hurricane Katia.

maps (Figures 9a and 9c), probably due to the lack of observations in this area for the month. Also, the shape and size of the Amazon plume, characterized by low SSS (Figure 9h), are much more realistically reproduced in the predicted stratification fields (Figures 9e and 9f) than WOA18 derived fields (Figures 9b and 9c). The strong horizontal gradient of density stratification near 45°W to 6°N (Figure 9f) associated with the retroflexion of the North Brazilian Current (Lumpkin and Garzoli, 2005) is poorly represented in the WOA18 derived density stratification map (Figure 9c). Moreover, the stratification in coastal regions in WOA18 derived maps (Figures 9a–9c) is missing due to a lack of in situ observations. This has been overcome by using satellite data to infer stratification (Figures 9d–9f). Therefore, there is additional value in our product compared to existing observational products.

Figure 10 shows daily maps of satellite SSH (Figures 10g–10i) and satellite-derived density stratification (Figures 10j–10l) in September 2011, focusing on active eddy regions. In addition to the permanent feature of the Loop Current in the Gulf of Mexico, we also see that warm and cold eddies, connected by the white and magenta solid lines in Figures 10g–10i, are associated with weak and strong stratification, respectively (Figures 10j–10l). There are exceptions, such as the low stratification values along the black lines in Figures 10h and 10i that do not resemble the eddy field. The black lines illustrate the track of Hurricane Katia, which caused strong mixing and SST cooling (Figures 10b and 10c) that resulted in low stratification (Figures 10k and 10l) and high Chla concentration (Figures 10n and 10o). The hurricane impact is too small to be visible in the SSH field because SSH represents steric height integrated from depths of thousands of meters to the surface. Another exception is the abnormally high stratification along the western coast of Florida due to the impacts of low-salinity water coming from the Mississippi (Figures 10d–10f).

In summary, our stratification products can realistically reproduce both the large-scale and mesoscale variations of stratification over the North Atlantic, with physical connections to tropical cyclones and primary production.

## 5.2. Applications

One of the main objectives of this study is to create a stratification data set for the prediction of TC-induced SST cooling and TC intensity change using the coupled SST index ( $T_{cp}$ , Balaguru et al., 2015). In this section we illustrate how the data set created can be used for this objective.

The TC-induced mixing length is computed as follows (Balaguru et al., 2015):

$$L = h + \left( \frac{2\rho_0 u_*^3 t}{\kappa g \alpha} \right)^{\frac{1}{3}} \quad (5)$$

where  $h$  is the initial mixed layer depth,  $\rho_0$  is the sea water density,  $u_*$  is the friction velocity,  $t$  is the time period of mixing,  $\kappa$  is the von Kármán constant,  $g$  is the acceleration due to gravity, and  $\alpha$  is density stratification beneath the mixed layer.

$T_{dy}$  is then determined based on an initial temperature profile and the mixing length with the assumption that the TC's wind mixes the whole water column from the surface to  $L$ :

$$T_{dy} = \frac{1}{L} \int_{-L}^0 T(z) dz \quad (6)$$

In order to compute  $T_{dy}$ , in addition to the storm state, we need to determine three unknown ocean variables: initial mixed layer depth ( $h$ ), density stratification at the base of the mixed layer, and prestorm temperature from the surface to  $L$ . These variables can be derived from existing high-resolution data sets such as HYCOM analysis. Here we show how to use the stratification data set in this study to compute the unknown variables. We then show a comparison between  $T_{dy}$  obtained from HYCOM analysis and our data set that is based on satellite observations.

Initial mixed layer depth and ocean stratification are two main ocean variables required for the computation of  $T_{dy}$  (Equation 5). We put our first effort on the production of the stratification data set because the derivation of mixed layer depth is more complicated and will require another study. We investigated the possibility of retrieving mixed layer depth from satellite data by adopting the same method and satellite predictors used in RegWOD for stratification. The skill of this mixed layer model is significantly lower than RegWOD for stratification. The spatially averaged decorrelation time scale of density stratification ( $\sim 12$  days) is twice that of mixed layer depth ( $\sim 6$  days) over the tropical North Atlantic computed using HYCOM reanalysis data (see Figure S4 in the supporting information). This means that the ocean has more memory for stratification than mixed layer depth. In this application, we therefore approximate along-TC-tracks prestorm mixed layer depth from WOA climatological mixed layer depth for simplicity.

Knowing mixed layer depth  $h$ , stratification at the base of the mixed layer can be computed as

$$\alpha_{(T,rho)} = \frac{Grad_{(T,rho)} * (h + 50)}{50} \quad (7)$$

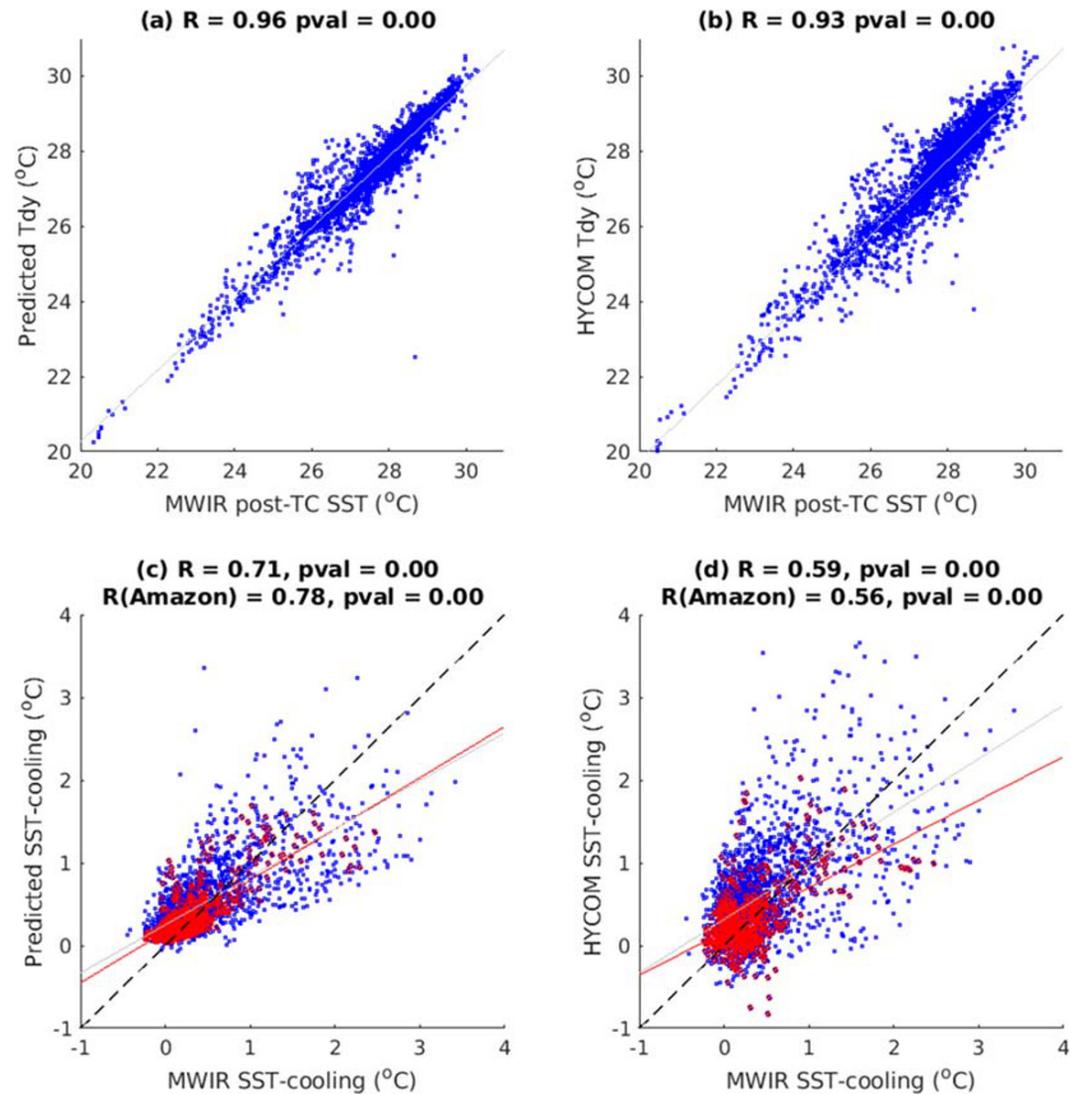
where  $\alpha_{(T,rho)}$  are, respectively, temperature and density stratification at the base of the mixed layer.

Knowing  $h$ , SST, and  $\alpha_T$  we can reconstruct a temperature profile:

$$\begin{aligned} T(z) &= \text{SST} \quad \text{for } z \leq h \\ T(z) &= \text{SST} + \alpha_T * (z - h) \quad \text{for } z > h \end{aligned} \quad (8)$$

Figures 11a and 11b show a comparison between post-TC SST and  $T_{dy}$  obtained from the RegWOD and HYCOM analysis data sets. We can see that, even with a rough approximation of mixed layer depth using WOA climatology,  $T_{dy}$  obtained from the data set of this study has a very high correlation with observations (0.96). The performance of  $T_{dy}$  obtained from our data set is slightly better than  $T_{dy}$  obtained from HYCOM analysis (0.93), where mixed layer depth, stratification at the base of the mixed layer, and prestorm temperature profiles are from HYCOM analysis.

The high correlation between  $T_{dy}$  and post-TC satellite SST is probably largely due to the strong spatiotemporal variation of SST itself. In order to remove this spatiotemporal variation, we show in Figures 11c and 11d the comparison of SST cooling, which is the difference between pre-TC (3–5 days ahead) and post-TC (1 day after) SST. In terms of correlation with satellite SST cooling, our data set shows better performance than HYCOM analysis (0.71 vs. 0.59) over the North Atlantic and especially over the Amazon plume (0.78 vs. 0.56). Moreover, the SST cooling bias from our data set, taking satellite SST cooling as reference, is much smaller over the North Atlantic (0.06 vs. 0.15°C) and slightly higher over the Amazon plume (0.06 vs. 0.04°C).



**Figure 11.** Comparison between post-TC (1 day after TC) satellite SST and  $T_{dy}$  obtained from RegWOD data set (predicted  $T_{dy}$ ) and from HYCOM analysis data set (a, b) and their respective SST cooling (c, d) over the period 2010–2017. SST cooling is the difference between pre-TC (3–5 days ahead of TC) and post-TC SST. The red scatter points show data within the Amazon plume (40–70°W, 10–25°N).

$T_{dy}$  computed from RegWOD and from HYCOM Analysis have the same significant correlations with TC intensification rate over the North Atlantic (0.20) and the Amazon plume (0.24). TC intensification rate is computed as the trend of TC intensity over a 24 consecutive hours period. However, compared to HYCOM, RegWOD  $T_{dy}$  can better separate intensifying and weakening TCs over the North Atlantic: the Student's  $t$  value representing the difference of means between intensifying and weakening TCs computed from the RegWOD  $T_{dy}$  is 16.00 vs. 11.91 computed from HYCOM  $T_{dy}$ . Over the Amazon plume, their  $t$  values are similar: 3.44 from RegWOD  $T_{dy}$  and 3.53 from HYCOM  $T_{dy}$ .

In summary, even with a rough approximation of mixed layer depth,  $T_{dy}$  computed from the RegWOD data set shows very good agreement with satellite observations and provides good prediction skill for TC intensification rate. The comparison with HYCOM analysis also illustrates the potential value of our data set in near-real time. Besides this application, the usefulness of the data set is further demonstrated in two other applications in the supporting information (Texts S2 and S3 and Figures S5 and S6). One application reveals an interesting geographically dependent relationship between ocean stratification and SST cooling (Text S2

and Figure S5). The other application reveals opposite relationships between ocean stratification and primary production: positive over river influenced areas and negative over the other region of the North Atlantic (Text S3 and Figure S6).

## 6. Summary and Conclusions

We investigated the possibility of using a linear statistical method to retrieve upper-ocean stratification from surface data (SST, SSS, and SSH) in the tropical North Atlantic, a region with strong spatiotemporal variability of temperature and salinity stratification. Using observations (RegWOD) or reanalysis data (RegHYCOM), our regression models were able to reconstruct more than 90% of the mean seasonal cycle and about 30% to 80% of temperature and salinity stratification daily anomalies. Simple regression therefore is a viable approach to reconstruct upper-ocean stratification. Statistical models for climatological and anomalous data agree that SST and SSS are the most important predictors for temperature and salinity stratification, respectively. We found that SSH is a secondary but useful predictor.

We assessed the feasibility of creating a high-resolution, near-real-time stratification data set based on satellite sea surface data by comparing the prediction skill of our regression model with the HYCOM analysis. We found that the prediction skills of the in situ (RegWOD) and satellite-based regression models (RegSAT) are comparable to the skill of the HYCOM analysis. The final stratification data set can reproduce both large- and small-scale features of the temperature and salinity stratification fields, providing good prediction skill for TC-induced SST cooling and TC intensification tendency and revealing interesting insights into the relationships between stratification and TC-induced SST cooling and primary productivity in the tropical North Atlantic. This data set will be available online and in near-real time at this site (<https://www.aoml.noaa.gov/>).

Despite the value of satellite data for retrieving upper-ocean stratification, we found that the skill of the statistical model and the final data set can potentially be improved with higher quality of satellite SSS data. It would be useful to test our methods using satellite SSS data from the more recent Soil Moisture Active Passive (SMAP) mission when the data record is long enough. The stratification data set produced is capable of resolving mesoscale variability, though in the scope of this study we have not been able to investigate the usefulness of the high resolution for a particular application. This will be left to our future studies or to the potential users of this data set. In addition, the quality of SSH data and how it impacts satellite-based stratification inference is still an open question. The performance of RegHYCOM proved that simple regression is a powerful method capable of reproducing close to 100% of the annual variations and 60–80% of anomaly variations of stratification. For nonhomogenous observational data, other methods such as support vector machines or deep neural networks may be worth trying.

Our results offer potential for real-time satellite-based monitoring of upper-ocean stratification for improved tropical cyclone intensity guidance. The methods we have developed are also likely to be applicable in other tropical cyclone basins and may be useful for analyses of other phenomena such as the Madden-Julian Oscillation (MJO) and monsoons, which are impacted by ocean stratification (Li et al., 2014, 2017).

## Appendix A.

### A.1 Cross Validation

Cross-validation is used to estimate the averaged prediction skill of a model over a training set by dividing the training data into multiple folds (fivefolds in this study). Each time one new fold is left out for testing and the rest is used to fit the model. After fitting on the training set and prediction on the test set are done, the residual standard error (RSE) of the test fold is computed (between predicted and test values) and recorded. This process is repeated until all folds are used for testing. After that, two important statistics are computed: fivefold mean RSE and its associated standard error. The fivefold mean RSE provides a measure of prediction skill of the model over the whole training set which can be used to determine the best model, that is, model associated with minimum fivefold mean RSE at a given size. The standard error can be used later to judge whether the prediction skill of two models at different sizes are significantly different over the same training set. This process is summarized in this diagram:

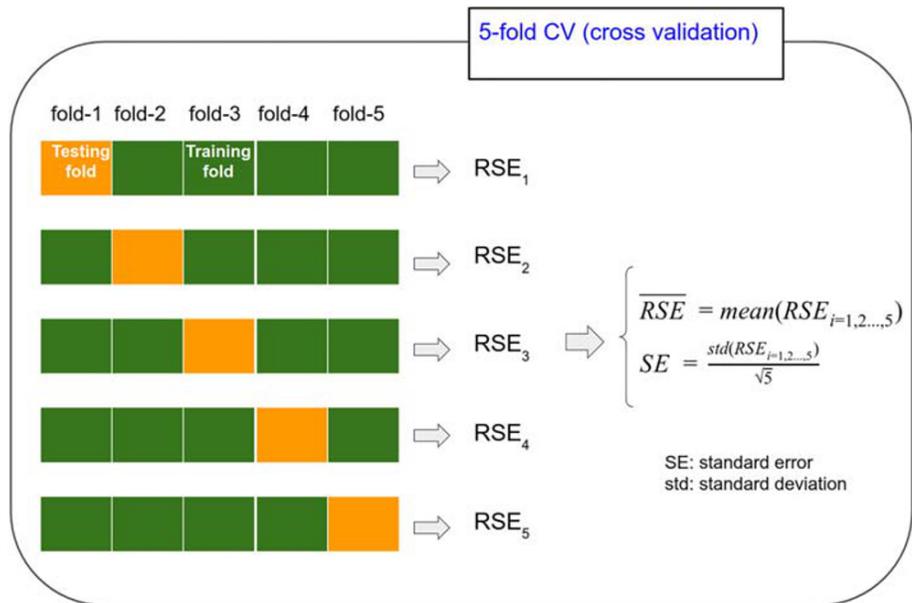


Figure A1. Schematic of the cross validation technique.

### A.2 Feature Selection

For a given location and response, Grad-T for instance, the optimum number of predictors (model size) can vary from one to four, and predictors can be any combination of SST, SSS, SSH, and constant term. To choose

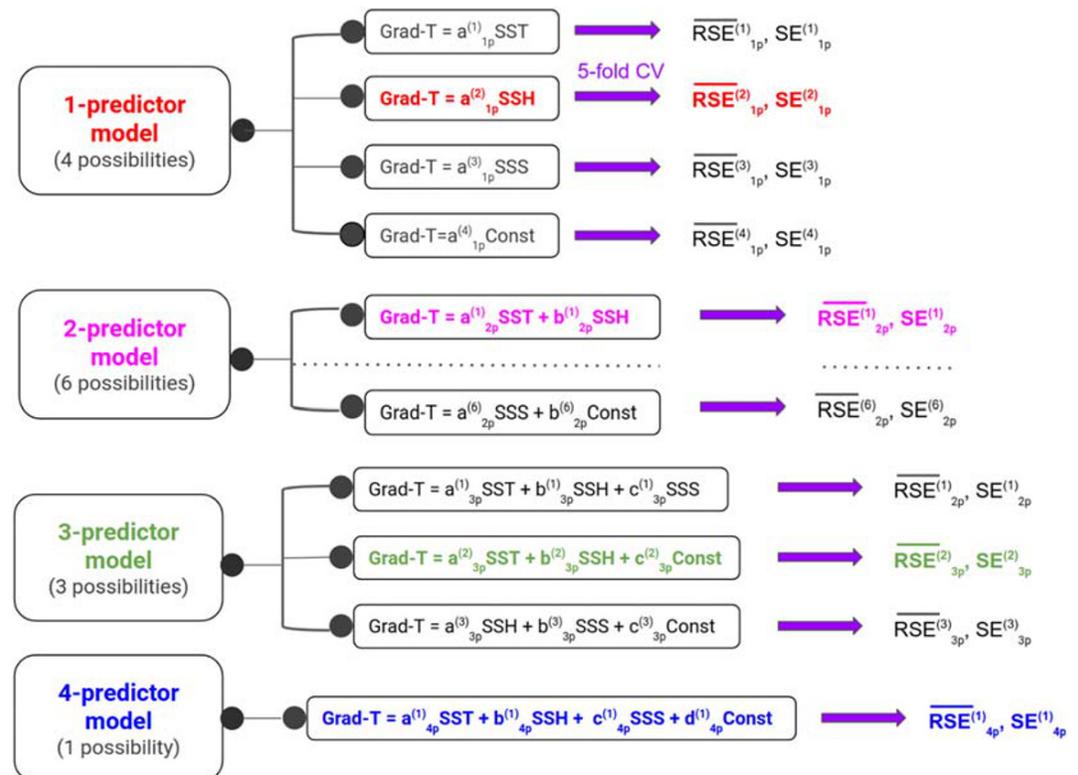
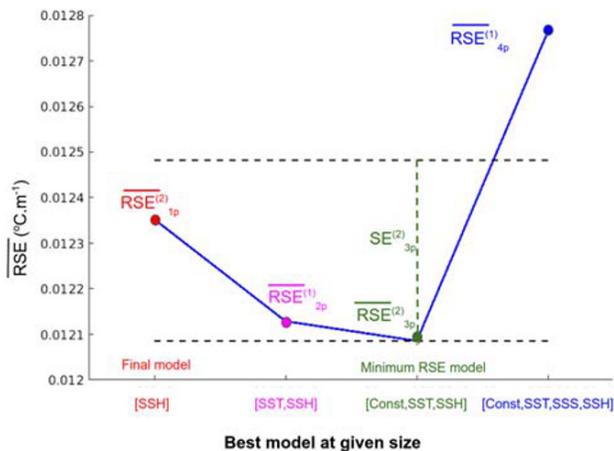


Figure A2. Schematic of the method used to determine the best model form.



**Figure A3.** Schematic illustrating the one standard error rule used to determine the final model.

the best model form (combination of predictors) for a given model size, fivefold cross validation is applied to all possible model forms associated with the model size. As an illustration, the best model form of a given size that has the minimum  $RSE$  is shown in the same color as the model size (left column) in the following diagram:

### A.3 Model Selection Using 1 Standard Error Rule

To choose the final model, local minimum  $RSE$  values for different model sizes are plotted to find the model size with the minimum  $RSE$  value (three-predictor model in this case). The black dashed lines show the one standard error interval associated with the three-predictor model. The simplest model whose  $RSE$  lies within  $RSE_{3p}^{(2)}$  and  $RSE_{3p}^{(2)} + SE_{3p}^{(2)}$  is the one-predictor model. The 1-p model is much simpler than the 3-p model yet has similar prediction skill. Simple models are more robust than complicated models in terms of preventing overfitting. Therefore, the 1-p model is chosen as the final model for Grad-T at this location.

### Data Availability Statement

All data used in this study are available publicly from the websites given in section 2. The final data sets and computer programs used to create them are available for download at this site ([ftp.aoml.noaa.gov/pub/phod/foltz/strat\\_dataset/](ftp.aoml.noaa.gov/pub/phod/foltz/strat_dataset/)).

### Acknowledgments

N. D. D. and G. F. were supported by the Climate Observations and Monitoring Division of NOAA's Climate Program Office and base funds to NOAA/AOML's Physical Oceanography Division. K. B. acknowledges support from NOAA's Climate Program Office, Climate Monitoring Program (Award NA17OAR4310155). K. B. is also supported by the Office of Science (BER) of the U.S. Department of Energy as part of the Regional and Global Modeling and Analysis (RGMA) Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

### References

- Ali, M. M. (2004). Estimation of ocean subsurface thermal structure from surface parameters: A neural network approach. *Geophysical Research Letters*, *31*, L20308. <https://doi.org/10.1029/2004GL021192>
- Balaguru, K., Foltz, G. R., Leung, L. R., Asaro, E. D., Emanuel, K. A., Liu, H., & Zedler, S. E. (2015). Dynamic potential intensity: An improved representation of the ocean's impact on tropical cyclones. *Geophysical Research Letters*, *42*, 6739–6746. <https://doi.org/10.1002/2015GL064822>
- Balaguru, K., Foltz, G. R., Leung, L. R., Kaplan, J., Xu, W., Reul, N., & Chapron, B. (2020). Pronounced impact of salinity on rapidly intensifying tropical cyclones. *Bulletin of the American Meteorological Society*, *101*(9), E1497–E1511. <https://doi.org/10.1175/BAMS-D-19-0303.1>
- Behringer, D. W., & Xue, Y. (2004). Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. In *Eighth symposium on integrated observing and assimilation systems for atmosphere, oceans, and land surface, AMS 84th Annual Meeting, Washington state convention and trade center, Seattle, Washington* (pp. 11–15). Boston, MA USA: American Meteorological Society.
- Bender, M. A., & Ginis, I. (2000). Real-case simulations of hurricane-ocean interaction using a high-resolution coupled model: Effects on hurricane intensity. *Monthly Weather Review*, *128*(4), 917–946. [https://doi.org/10.1175/1520-0493\(2000\)128<0917:RCOHO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0917:RCOHO>2.0.CO;2)
- Boutin, J., Vergely, J. L., Marchand, S., D'Amico, F., Hasson, A., Kolodziejczyk, N., & Vialard, J. (2018). New SMOS sea surface salinity with reduced systematic errors and improved variability. *Remote Sensing of Environment*, *214*, 115–134. <https://doi.org/10.1016/j.rse.2018.05.022>
- Boyer, T. P., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Locarnini, R. A., et al. (2018). World ocean database 2018. In A. V. Mishonov (Ed.), *Technical Editor, NOAA Atlas NESDIS 87* (Chap. 3, 6, 15). Silver Spring, MD USA: U.S. Department of Commerce.
- Carton, J. A., Chepurin, G. A., & Chen, L. (2018). SODA3: A new ocean climate reanalysis. *Journal of Climate*, *31*(17), 6967–6983. <https://doi.org/10.1175/JCLI-D-18-0149.1>
- Chu, P. C., Fan, C., & Liu, W. T. (2000). Determination of vertical thermal structure from sea surface temperature. *Journal of Atmospheric and Oceanic Technology*, *17*, 9. [https://doi.org/10.1175/1520-0426\(2000\)017<0971:DOVTSF>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<0971:DOVTSF>2.0.CO;2)
- Cummings, J. A. (2005). Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society Part C*, *131*, 3583–3604. <https://doi.org/10.1256/qj.05.105>
- D'Asaro, E. A., Sanford, T. B., Niiler, P. P., & Terrill, E. J. (2007). Cold wake of Hurricane Frances. *Geophysical Research Letters*, *34*, L15609. <https://doi.org/10.1029/2007GL030160>
- de Boer, G. J., Pietrzak, J. D., & Winterwerp, J. C. (2008). Using the potential energy anomaly equation to investigate tidal straining and advection of stratification in a region of freshwater influence. *Ocean Modelling*, *22*(1–2), 1–11. <https://doi.org/10.1016/j.ocemod.2007.12.003>
- de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*, C12003. <https://doi.org/10.1029/2004JC002378>
- de Boyer Montégut, C., Mignot, J., Lazar, A., & Cravatte, S. (2007). Control of salinity on the mixed layer depth in the world ocean: 1. General description. *Journal of Geophysical Research*, *112*, C06011. <https://doi.org/10.1029/2006JC003953>
- Emanuel, K. (2018). 100 years of progress in tropical cyclone research. *Meteorological Monographs*, *59*, 15.1–15.68. <https://doi.org/10.1175/AMSMONOGRAPHIS-D-18-0016.1>

- Emanuel, K. A. (1999). Thermodynamic control of hurricane intensity. *Nature*, *401*(6754), 665–669. <https://doi.org/10.1038/44326>
- Fischer, M. (2000). Multivariate projection of ocean surface data onto subsurface sections. *Geophysical Research Letters*, *27*(6), 755–757. <https://doi.org/10.1029/1999GL010451>
- Fofonoff, N. P., & Millard, R. C. Jr. (1983). *Algorithms for the computation of fundamental properties of seawater*. Paris, France: UNESCO. 53pp. (UNESCO technical papers in marine sciences; 44), <http://hdl.handle.net/11329/109>
- Garcia, H. E., Boyer, T. P., Baranova, O. K., Locarnini, R. A., Mishonov, A. V., Grodsky, A., et al. (2019). In A. Mishonov (Ed.), *World Ocean Atlas 2018: Product documentation*. Technical Editor Silver Spring, MD USA: U.S. Department of Commerce.
- Garcia, H. E., Boyer, T. P., Locarnini, R. A., Baranova, O. K., & Zweng, M. M. (2018). In A. V. Mishonov (Ed.), *World ocean database 2018: User's manual (prerelease)*. Silver Spring, MD: NOAA. (Available at [https://www.NCEI.noaa.gov/OC5/WOD/pr\\_wod.html](https://www.NCEI.noaa.gov/OC5/WOD/pr_wod.html))
- Goni, G., DeMaria, M., Knaff, J., Sampson, C., Ginis, I., Bringas, F., et al. (2009). Applications of satellite-derived ocean measurements to tropical cyclone intensity forecasting. *Oceanography*, *22*(3), 190–197. <https://doi.org/10.5670/oceanog.2009.78>
- Guinehut, S., Dhomp, A. L., Larnicol, G., & Traon, P. Y. L. (2012). High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Science*, *14*. <https://doi.org/10.5194/os-8-845-2012>
- Hosoda, S., Ohira, T., Sato, K., & Suga, T. (2010). Improved description of global mixed-layer depth using Argo profiling floats. *Journal of Oceanography*, *66*(6), 773–787. <https://doi.org/10.1007/s10872-010-0063-3>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds) (2013). *An introduction to statistical learning: With applications in R*. New York: Springer.
- Kara, A. B. (2003). Mixed layer depth variability over the global ocean. *Journal of Geophysical Research*, *108*(C3), 3079. <https://doi.org/10.1029/2000JC000736>
- Khedouri, E., Szczechowski, C., & Cheney, R. (1983). Potential oceanographic applications of satellite altimetry for inferring subsurface thermal structure. In *proceedings OCEANS'83* (pp. 274–280). New York, NY USA: IEEE.
- Klemas, V., & Yan, X.-H. (2014). Subsurface and deeper ocean remote sensing from satellites: An overview and new results. *Progress in Oceanography*, *122*, 1–9. <https://doi.org/10.1016/j.pocean.2013.11.010>
- Landsea, C. W., & Frank, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, *141*(10), 3576–3592. <https://doi.org/10.1175/MWR-D-12-00254.1>
- Li, Y., Han, W., Shinoda, T., Wang, C., Ravichandran, M., & Wang, J.-W. (2014). Revisiting the wintertime intraseasonal SST variability in the tropical South Indian Ocean: Impact of the ocean interannual variation. *Journal of Physical Oceanography*, *44*(7), 1886–1907. <https://doi.org/10.1175/JPO-D-13-0238.1>
- Li, Y., Han, W., Wang, W., Ravichandran, M., Lee, R., & Shinoda, T. (2017). Bay of Bengal salinity stratification and Indian summer monsoon intraseasonal oscillation: 2 Impact on SST and convection. *Journal of Geophysical Research*, *122*, 4312–4328. <https://doi.org/10.1002/2017JC012692>
- Lin, I.-I., Black, P., Price, J. F., Yang, C.-Y., Chen, S. S., Lien, C.-C., et al. (2013). An ocean coupling potential intensity index for tropical cyclones. *Geophysical Research Letters*, *40*, 1878–1882. <https://doi.org/10.1002/grl.50091>
- Liu, L., Peng, S., & Huang, R. X. (2017). Reconstruction of ocean's interior from observed sea surface information. *Journal of Geophysical Research: Oceans*, *122*, 1042–1056. <https://doi.org/10.1002/2016JC011927>
- Lumpkin, R., & Garzoli, S. L. (2005). Near-surface circulation in the tropical Atlantic Ocean. *Deep-Sea Research I*, *52*, 495–518. <https://doi.org/10.1016/j.dsr.2004.09.001>
- MacKenzie, L., & Adamson, J. (2004). Water column stratification and the spatial and temporal distribution of phytoplankton biomass in Tasman Bay, New Zealand: Implications for aquaculture. *New Zealand Journal of Marine and Freshwater Research*, *38*(4), 705–728. <https://doi.org/10.1080/00288330.2004.9517271>
- Mayer, D. A., Molinari, R. L., Baringer, M. O., & Goni, G. J. (2001). Transition regions and their role in the relationship between sea surface height and subsurface temperature structure in the Atlantic Ocean. *Geophysical Research Letters*, *28*(20), 3943–3946. <https://doi.org/10.1029/2001GL013331>
- Price, J. F. (1981). Upper Ocean response to a hurricane. *Journal of Physical Oceanography*, *11*(2), 153–175. [https://doi.org/10.1175/1520-0485\(1981\)011<0153:UORTAH>2.0.CO;2](https://doi.org/10.1175/1520-0485(1981)011<0153:UORTAH>2.0.CO;2)
- Price, J. F. (2009). Metrics of hurricane-ocean interaction: Vertically-integrated or vertically-averaged ocean temperature? *Ocean Science*, *5*(3), 351–368. <https://doi.org/10.5194/os-5-351-2009>
- Pun, I.-F., Lin, I.-I., & Ko, D. S. (2014). New generation of satellite-derived ocean thermal structure for the western North Pacific typhoon intensity forecasting. *Progress in Oceanography*, *121*, 109–124. <https://doi.org/10.1016/j.pocean.2013.10.004>
- Pun, I.-F., Price, J. F., & Jayne, S. R. (2016). Satellite-derived ocean thermal structure for the North Atlantic hurricane season. *Monthly Weather Review*, *144*(3), 877–896. <https://doi.org/10.1175/MWR-D-15-0275.1>
- Rappaport, E. N., Jiing, J.-G., Landsea, C. W., Murillo, S. T., & Franklin, J. L. (2012). The joint hurricane test bed: Its first decade of tropical cyclone research-to-operations activities reviewed. *Bulletin of the American Meteorological Society*, *93*(3), 371–380. <https://doi.org/10.1175/BAMS-D-11-00037.1>
- Simpson, J. H. (1981). The shelf-sea fronts: Implications of their existence and behaviour. *Philosophical Transactions of the Royal Society of London Series A, Mathematical and Physical Sciences*, *302*(1472), 531–546. <https://doi.org/10.1098/rsta.1981.0181>
- Su, H., Huang, L., Li, W., Yang, X., & Yan, X. (2018). Retrieving ocean subsurface temperature using a satellite-based geographically weighted regression model. *Journal of Geophysical Research: Oceans*, *123*, 5180–5193. <https://doi.org/10.1029/2018JC014246>
- Su, H., Li, W., & Yan, X.-H. (2018). Retrieving temperature anomaly in the global subsurface and deeper ocean from satellite observations. *Journal of Geophysical Research: Oceans*, *123*, 399–410. <https://doi.org/10.1002/2017JC013631>
- Su, H., Wu, X., Yan, X.-H., & Kidwell, A. (2015). Estimation of subsurface temperature anomaly in the Indian Ocean during recent global surface warming hiatus from satellite measurements: A support vector machine approach. *Remote Sensing of Environment*, *160*, 63–71. <https://doi.org/10.1016/j.rse.2015.01.001>
- Vincent, E. M., Lengaigne, M., Madec, G., Vialard, J., Samson, G., Jourdain, N. C., et al. (2012). Processes setting the characteristics of sea surface cooling induced by tropical cyclone. *Journal of Geophysical Research*, *117*, C02020. <https://doi.org/10.1029/2011JC007396>
- Vincent, E. M., Lengaigne, M., Vialard, J., Madec, G., Jourdain, N. C., & Masson, S. (2012). Assessing the oceanic control on the amplitude of sea surface cooling induced by tropical cyclones. *Journal of Geophysical Research*, *117*, C05023. <https://doi.org/10.1029/2011JC007705>
- Wu, X., Yan, X.-H., Jo, Y.-H., & Liu, W. T. (2012). Estimation of subsurface temperature anomaly in the North Atlantic using a self-organizing map neural network. *Journal of Atmospheric and Oceanic Technology*, *29*(11), 1675–1688. <https://doi.org/10.1175/JTECH-D-12-00013.1>

- Yamaguchi, R., & Suga, T. (2019). Trend and variability in global Upper-Ocean stratification since the 1960s. *Journal of Geophysical Research: Oceans*, *124*, 8933–8948. <https://doi.org/10.1029/2019JC015439>
- Yamaguchi, R., Suga, T., Richards, K. J., & Qiu, B. (2019). Diagnosing the development of seasonal stratification using the potential energy anomaly in the North Pacific. *Climate Dynamics*, *53*(7–8), 4667–4681. <https://doi.org/10.1007/s00382-019-04816-y>
- Zeng, L., Wang, D., Chen, J., Wang, W., & Chen, R. (2016). SCSPD14, a South China Sea physical oceanographic dataset derived from in situ measurements during 1919–2014. *Scientific Data*, *3*(1), 160029. <https://doi.org/10.1038/sdata.2016.29>