Global Mesoscale Ocean Variability from Multiyear Altimetry: An Analysis of the Influencing Factors

YAO YU,[a] SARAH T. GILLE,[a] DAVID T. SANDWELL,[a] AND JULIAN MCAULEY[b]

[a] *Scripps Institution of Oceanography, University of California San Diego, La Jolla, California*
[b] *Department of Computer Science, University of California San Diego, La Jolla, California*

ABSTRACT: Sea surface slope (SSS) responds to oceanic processes and other environmental parameters. This study aims to identify the parameters that influence SSS variability. We use SSS calculated from multiyear satellite altimeter observations and focus on small resolvable scales in the 30–100-km wavelength band. First, we revisit the correlation of mesoscale ocean variability with seafloor roughness as a function of depth, as proposed by Gille et al. Our results confirm that in shallow water there is statistically significant positive correlation between rough bathymetry and surface variability, whereas the opposite is true in the deep ocean. In the next step, we assemble 27 features as input variables to fit the SSS with a linear regression model and a boosted trees regression model, and then we make predictions. Model performance metrics for the linear regression model are $R^2 = 0.381$ and mean square error = 0.010 μrad^2. For the boosted trees model, $R^2 = 0.563$ and mean square error = 0.007 μrad^2. Using the hold-out data, we identify the most important influencing factors to be the distance to the nearest thermocline boundary, significant wave height, mean dynamic topography gradient, and M_2 tidal speed. However, there are individual regions, that is, the Amazon outflow, that cannot be predicted by our model, suggesting that these regions are governed by processes that are not represented in our input features. The results highlight both the value of machine learning and its shortcomings in identifying mechanisms governing oceanic phenomena.

KEYWORDS: Mesoscale processes; Satellite observations; Machine learning; Regression

1. Introduction

Sea surface slope (SSS) varies in response to a range of oceanic processes. On scales large enough to represent geostrophic flows, it is a measure of geostrophic velocity; on smaller scales, it varies with tides, surface waves, internal waves, eddies, etc. Large-scale ocean processes are generally well observed and have been studied since the era of satellite altimetry. Mesoscale oceanic variability (30–100 km), however, is less understood, as its signatures on the sea surface generally occur on scales smaller than the 100–200-km resolution of the widely used multimission sea surface height (SSH) product distributed by the Copernicus Marine Service (CMEMS; Ballarotta et al. 2019). Despite their small scales, mesoscale oceanic processes provide an essential link in the ocean's large-scale circulation and are associated with eddy kinetic energy generation and dissipation (Ferrari and Wunsch 2009).

In this paper, our goal is to investigate the sea surface variability of small resolvable scales in the 30–100-km wavelength band and to contrast this with variability at scales greater than 100 km. We use SSS calculated from multiyear satellite altimeter observations as a metric for ocean variability. The major questions that we address are: how well can we characterize SSS variability, and what variables are needed to explain SSS variability? We know from previous studies (e.g., Gille et al. 2000; Nikurashin and Legg 2011) that surface variability is linked to bathymetry, seafloor roughness, and baroclinic instability, among other variables. Using satellite altimeter data, Gille et al. (2000)

found evidence that mesoscale oceanic variability (with spatial scales from 80 to 160 km) is indirectly controlled by bathymetry: in ocean regions that are deeper than 4800 m, seafloor roughness is anticorrelated with SSS variability, implying that rough topography helps to dissipate mesoscale kinetic energy. In contrast, in shallow waters, seafloor roughness and sea surface slope variability are correlated, implying that mesoscale variability is generated over rough topography. Nikurashin and Legg (2011) used numerical simulations to show that the energy from large-scale internal tides to smaller-scale internal waves depends on seafloor roughness, tidal amplitudes, and the Coriolis frequency. While the comparisons of Gille et al. (2000) provided a statistical assessment, depth-dependent impacts of roughness could explain less than 10% of the overall variability in SSS. Here we aim to learn what other parameters might influence small-scale SSS variability.

We address the major question in two ways: (i) How well can we predict oceanic variability on global scales, and what fraction of the global variance can we predict? (ii) Are there particular regions that are unusual and cannot be predicted? For places where the surface variability responds to a particular local effect (e.g., the Amazon outflow), can the variability be represented using a statistical model derived with machine learning?

We aim to answer the above questions by using three statistical methods: 1) Correlation analysis, in which we revisit the conditional correlation between seafloor roughness versus SSS variability as a function of depth, as proposed in Gille et al. (2000), with updated datasets, and we explore linear correlations between sea surface variability and other variables. 2) Machine learning using a linear regression approach. 3) Machine learning with a boosted trees algorithm. Both machine learning methods

Corresponding author: Yao Yu, yayu@ucsd.edu

take in multiple features to predict the sea surface variability, analyze the relevance of each feature, and discuss prediction failures. While conditional correlation analysis quantifies the linear dependence between variables, it lacks the flexibility to handle nonlinear dependencies on multiple variables. The linear regression model is straightforward, and it assumes that labels (what we are attempting to predict or forecast, i.e., SSS variability in this study), are a linear combination of different features. While this assumption will prove to be inadequate to explain all variability, the linear regression is a base model that is able to identify linear relations between the SSS variability and features. In contrast to conventional linear regression, the boosted tree model can capture nonlinear relationships between the features and the outcome. It uses the boosting method that sequentially combines decision trees, in a way that each new tree fits the residuals from the previous step so that the model improves (Friedman 2002). Decision trees use a greedy algorithm that finds the optimal data split solution for each node, which is a split on a feature at a specific value, resulting in the largest information gain (Quinlan 1986). Both models assist in understanding the governing factors of SSS variability in our case. In section 2b we introduce strategies for ranking features, that is, selecting governing factors.

We estimate the global ocean variability in the form of SSS variability using satellite altimetry profiles from *Geosat*, *Environmental Satellite* (*Envisat*), *CryoSat-2*, *Jason-1/Jason-2*, *Satellite with Argos and Altika* (*SARAL*), and *Sentinel-3A/ Sentinel-3B* collected from 1993 to 2021 (an updated version of the dataset as in Sandwell et al. 2019). SSS is the along-track derivative of SSH. Thus, it achieves finer spatial resolution than SSH by "whitening" the red spectrum slope of SSH and is more sensitive to high-wavenumber signals. Note that this is a 1D along-track slope, so it only approximates a 2D slope estimate. The slope estimates come from a variety of directions depending on the inclination of the satellite orbit. Sandwell et al. (2019) showed that SSS variability from multiyear repeat and nonrepeat altimetry missions is able to reveal oceanic processes with scales as small as 25 km. The combination of multiyear satellite altimetry profiles provides a dense ground track coverage at the cost of losing temporal resolution. In contrast, the gridded SSH product using multisatellite data distributed by CMEMS, has a temporal resolution of 10–33 days and a spatial resolution of about 100–200 km (Ballarotta et al. 2019; Taburet et al. 2019). The coarse spatial resolution of the multisatellite data is restricted by the wide cross-track distance and instrument noise (Fu and Ubelmann 2014). This product is not able to fully capture mesoscale oceanic activities and is thus not adopted in our study.

We study the SSS variability in two wavelength bands: 1) a band encompassing mesoscale variability as well as larger submesoscale features (30–100 km). As a shorthand, we refer to this band as the mesoscale; 2) the large-scale band (>100 km). The mesoscale band is generally hard to observe on a global scale. It contains the variation of unbalanced wave motions and the mesoscale eddies that include coherent vortices, filaments, squirts, and spirals. Unbalanced wave motions are mostly attributed to internal tides or waves, and they are generally greater in amplitude over rough topography (i.e.,

the Hawaiian Ridge) or in highly stratified zones (i.e., the Amazon shelf). The mesoscale eddies emerge from the instabilities of strong geostrophic flows, and they contain the majority of oceanic kinetic energy (Zhang and Qiu 2018). Large-scale SSS variability is well characterized, and it is related to balanced geostrophic flows that have large values of mean SSS, that is, western boundary currents and the Antarctic Circumpolar Currents (ACC). Slight perturbations to the large-scale mean SSS lead to variability at large spatial scales. The transition scale that delineates balanced geostrophic flows and unbalanced wave motions depends sensitively on local mesoscale eddy variability and varies with time (Qiu et al. 2018). The 100-km delineation here is an empirical choice. While there is no consensus on the transition scale, it is generally short (below 40 km) in eddy-intensified western boundary currents and in the ACC, and it increases equatorward in relative stable regions (40–100 km in subtropical and subpolar gyres; >200 km in the tropical oceans) (Qiu et al. 2017, 2018). The 30–100-km mesoscale band contains both unbalanced wave motions (internal tides, near-initial flows) and balanced geostrophic flows. The dominant component of the flow is geographically and temporally dependent.

This paper is organized as follows: In section 2, we introduce data and methods used in this study. In section 3, we revisit the approach used by Gille et al. (2000) to assess the correlation between SSS and seafloor roughness as a function of ocean depth. This section also builds a linear regression model and a boosted trees model to infer SSS variability from other influencing factors including stratification, ocean basins, and distance to the nearest thermocline boundary. We make predictions of SSS variability and identify the dominant factors that contribute to the SSS variability. We summarize and discuss perspectives in section 4.

2. Data and methods

a. Data

1) SEA SURFACE SLOPE VARIABILITY

In this paper, we calculate SSS variability in the mesoscale band (30–100 km) and large-scale band (>100 km) using multiyear along-track altimetry. Calculation steps are described below.

We take the along-track profiles from *Geosat*, *Envisat*, *CryoSat-2*, *Jason-1/Jason-2*, *SARAL/Altika*, and *Sentinel-3A/ Sentinel-3B* collected from 1992 to 2021. Standard 1-Hz geophysical data records are inadequate for this analysis because the retracking of the waveform does not account for the high correlation between significant wave height (SWH) and arrival time (Sandwell and Smith 2005; Zaron and DeCarvalho 2016). Moreover, the 1-Hz boxcar averaging aliases noise at less than 1 Hz (~14-km wavelength) into the 30–50-km wavelength band. Last, to achieve a uniform quality among the various altimeters, one must retrack, filter, and edit the raw waveform data in a consistent way. To retrack raw waveform data [except for *CryoSat* Synthetic Aperture Mode (SAR) and SAR + interferometer mode (SIN) and *Sentinel-3A/Sentinel-3B*], we adopt a two-pass method that effectively reduces the wave height noise and improves the

range precision of altimeter echoes by a factor of 1.5–1.7 (García-García and Ummenhofer 2015; Zhang and Sandwell 2017). We edit 20 Hz waveform data using flags provided in the level-1 product. We then apply a Parks–McClellan low-pass filter with half gain at 6.7 km and downsample data to 5 Hz. We apply geophysical corrections, including wet and dry troposphere delay, inverse barometer effect, and solid Earth and ocean tides (FES2014; Carrère et al. 2016). We further edit data with residuals from the EGM 2008 model greater than 3 standard deviations [typically > 30 microradians (μrad)]. We apply a second Parks–McClellan low-pass, derivative filter with half gain at 8.3 km to all profiles and form along-track SSS. We apply local geoid corrections (Sandwell and Smith 2014) and remove the mean SSS to obtain slope anomalies that reflect oceanic variability, wave height noise, and tide model error.

For isotropic geostrophic flows, SSS variability is linearly related to eddy kinetic energy: $E_k = \langle v'^2 \rangle = \langle (\partial \eta'/\partial l)^2 \rangle g^2/f^2$, where eddy kinetic energy E_k is defined as the time-averaging of squared surface geostrophic velocity perturbations $\langle v'^2 \rangle$, g is gravity, f is the Coriolis parameter, and $\partial \eta'/\partial l$ is the along-track SSS anomaly. Thus, SSS variability $\sqrt{\langle (\partial \eta'/\partial l)^2 \rangle}$ represents eddy kinetic energy for flows on scales large enough to be geostrophically balanced. The relation does not apply for flows that are not geostrophically balanced, either because they represent small-scale ageostrophic motions or because they occur in equatorial regions, where the Coriolis parameter approaches zero. SSS variability with wavelengths shorter than 30 km is mostly buried in wave height noise. Thus, we apply a 30-km Gaussian low-pass filter to reduce noise (Sandwell et al. 2019).

We sort the SSS anomalies into 7 arc min × 5 arc min blocks and use the absolute median value in each block to represent oceanic variability. The mesoscale and large-scale SSS variability are shown in Fig. 1. In the mesoscale band, we can see patterns of variability that are potentially consistent with signals due to internal tides as well as horizontally sheared boundary current motions. For example, there is strong variability associated with internal tides over rough topography (the Mid-Atlantic Ridge, the Southwest Indian Ridge, and the Hawaiian Ridge) and continental shelves (the Amazon shelf and the Mascarene Basin). There is also strong SSS variability in the vicinity of western boundary currents and the ACC. On large scales, the strong SSS variabilities are always associated with western boundary currents and the ACC, where the mean SSSs are large.

2) ENVIRONMENTAL PARAMETERS

For the correlation analysis, we use bathymetry, seafloor roughness, and SSS variability to revisit the relation between roughness and eddy kinetic variability as a function of seafloor depth as in Gille et al. (2000). For machine learning approaches, we use 27 features (listed below) as input variables to build a linear regression model and a boosted tree model to predict the SSS variability. Our analysis screens out regions poleward of 60°, where strong seasonal sea ice contaminates oceanic signals. The *Jason-1*/*Jason-2* orbit was designed with a 66° latitude inclination. We exclude all land, lakes, ponds, and semienclosed seas including the Mediterranean Sea, Gulf of Mexico, and Caribbean Sea. We also exclude coastal regions where ocean depth is less than 100 m. All features are processed to have consistent spatial coverage (60°S–60°N) and resolution (7 arc min by 5 arc min). We normalize features such that the normalized features have similar ranges and comparable variances. We retain the sign of latitude, we linearly scale the absolute latitude to range from 0 to 2, we apply a scaler to all other features that subtracts the median, and we scale each feature to the interquartile range. The centering and scaling statistics of the scaler are based on percentiles and are therefore robust to large marginal outliers. The 27 features are either associated with the solid Earth or the dynamic ocean. Features are described below and shown in Fig. 2.

(i) Seafloor roughness

Seafloor roughness is the root-mean-square (RMS) height of short wavelength bathymetry (we use a wavelength range of 50–160 km in this study). The roughness directly derived from SRTM15+V2.3 predicted bathymetry (Tozer et al. 2019) is underestimated, because the gravity anomalies from small structures including abyssal hills, small seamounts, and so on, are attenuated in altimetry measurements. Goff (2010) put forward that the statistical properties of abyssal morphology can be related to the gravity field that is derived from satellite altimetry using the upward continuation formulation. We produce a roughness map by adding back the latest abyssal hill RMS height from Goff (2020) to predicted bathymetry in the following steps:

1) square the Goff (2020) RMS height and replace with 0 over regions measured by ship soundings;
2) apply a high-pass Gaussian filter to the SRTM15+V2.3 predicted bathymetry at 160 km then square the result;
3) combine the above two datasets and apply a low-pass Gaussian filter with 0.5 gain at 50 km to eliminate contamination from wave height noise in satellite measurements;
4) take the square root of the dataset in step 3 to get RMS roughness, which is the square root of the average squared bathymetry deviation about a linear trend. We recovered the abyssal hill roughness while keeping all ship soundings untouched. The uncharted small seamounts are not taken into consideration in this study.

(ii) Smooth seafloor roughness

Seafloor roughness is low-pass filtered using a Gaussian filter with half gain at 500 km to obtain the smooth seafloor roughness. This captures the locations of large-scale rough seafloor and is less sensitive to estimation errors than roughness itself.

(iii) Bathymetry

We use the STRM15+V2.3 15 arc-s-resolution bathymetry map that includes > 33.6 million multibeam and single beam measurements (covering 15% of the ocean; Wölfl et al. 2019) and retracked range measurements from *Geosat*, *Envisat*, *CryoSat-2*, *Jason-1*/*Jason-2*, and *SARAL*/*Altika* (Tozer et al. 2019).
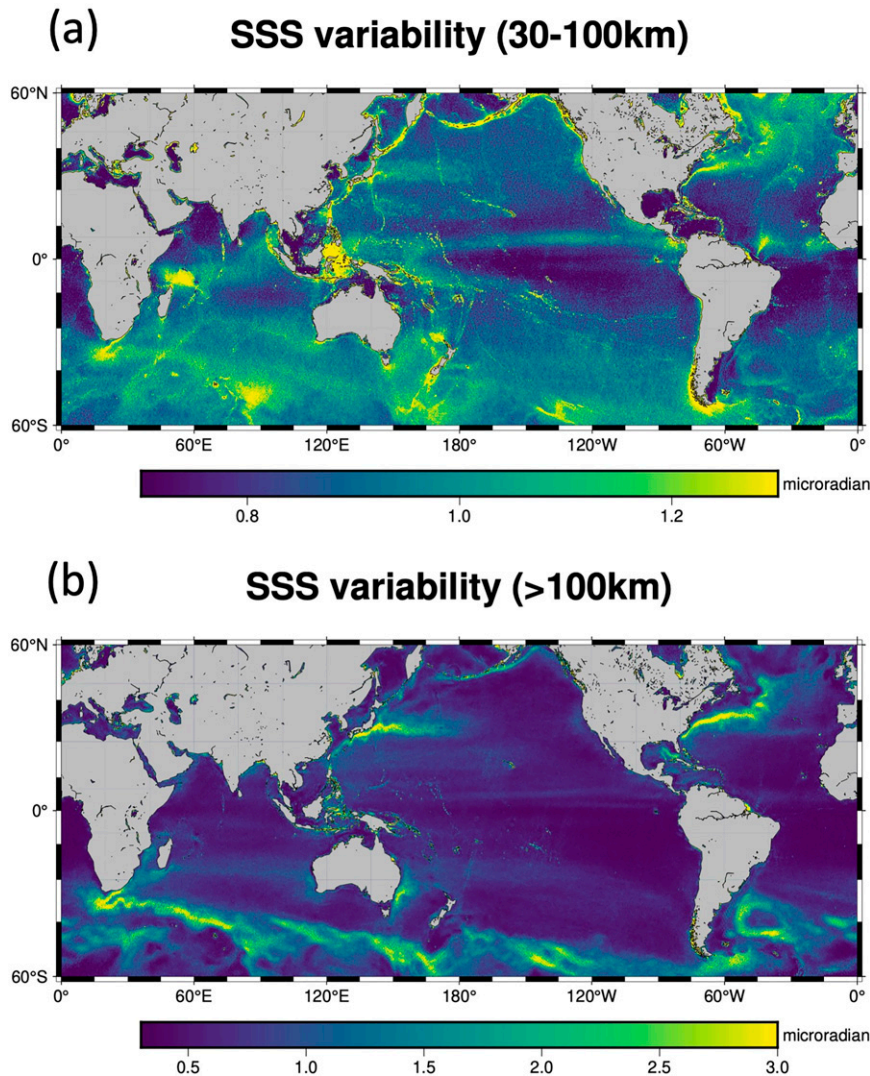
FIG. 1. (a) Mesoscale (30–100 km) and (b) large-scale (>100 km) SSS variability. 1 microradian is equal to a 1-cm change in height over a horizontal distance of 10 km. All plots are at 7 arc min × 5 arc min resolution using a Mercator projection. All land is gray. Note that the color scales for (a) and (b) are different.

(iv) Ocean depth slope

Internal tides are generally generated over variable bottom topography such as continental slopes (Baines 1982). Small topography structures like abyssal hills and seamounts associated with slopes up to 0.2, are not fully captured in the SRTM15+V2.3 bathymetry map. These features are potential sites for internal tide generation. The synthetic bathymetry map (SYNBATH) includes the statistics of abyssal hills and Gaussian-shaped small uncharted seamounts (Sandwell et al. 2022). We use the magnitude of the vector gradient of the SYNBATH to represent the ocean depth slope.

(v) Vertical gravity gradient

Vertical gravity gradient (VGG) is the vertical derivative of gravity anomaly and is linearly related to the derivative of the mean SSS through Laplace's equation (Sandwell 2022). It describes the bumps and dips from the topography of the seafloor. We use the 1 arc min marine VGG of the SRTM15+V2.3 product (Tozer et al. 2019).

(vi) Free-air gravity

Free-air gravity is the negative radial derivative of the disturbing potential evaluated on the geoid (Sandwell 2022). We use the 1 arc min marine free-air gravity anomalies of the SRTM15+V2.3 product (Tozer et al. 2019). (In Fig. 6 "FA_gravity" represents free-air gravity.)

(vii)–(viii) Seafloor spreading rate and oceanic crustal age

Seafloor spreading rate and oceanic crustal age are two fundamental geophysical variables. Oceanic crusts are young at
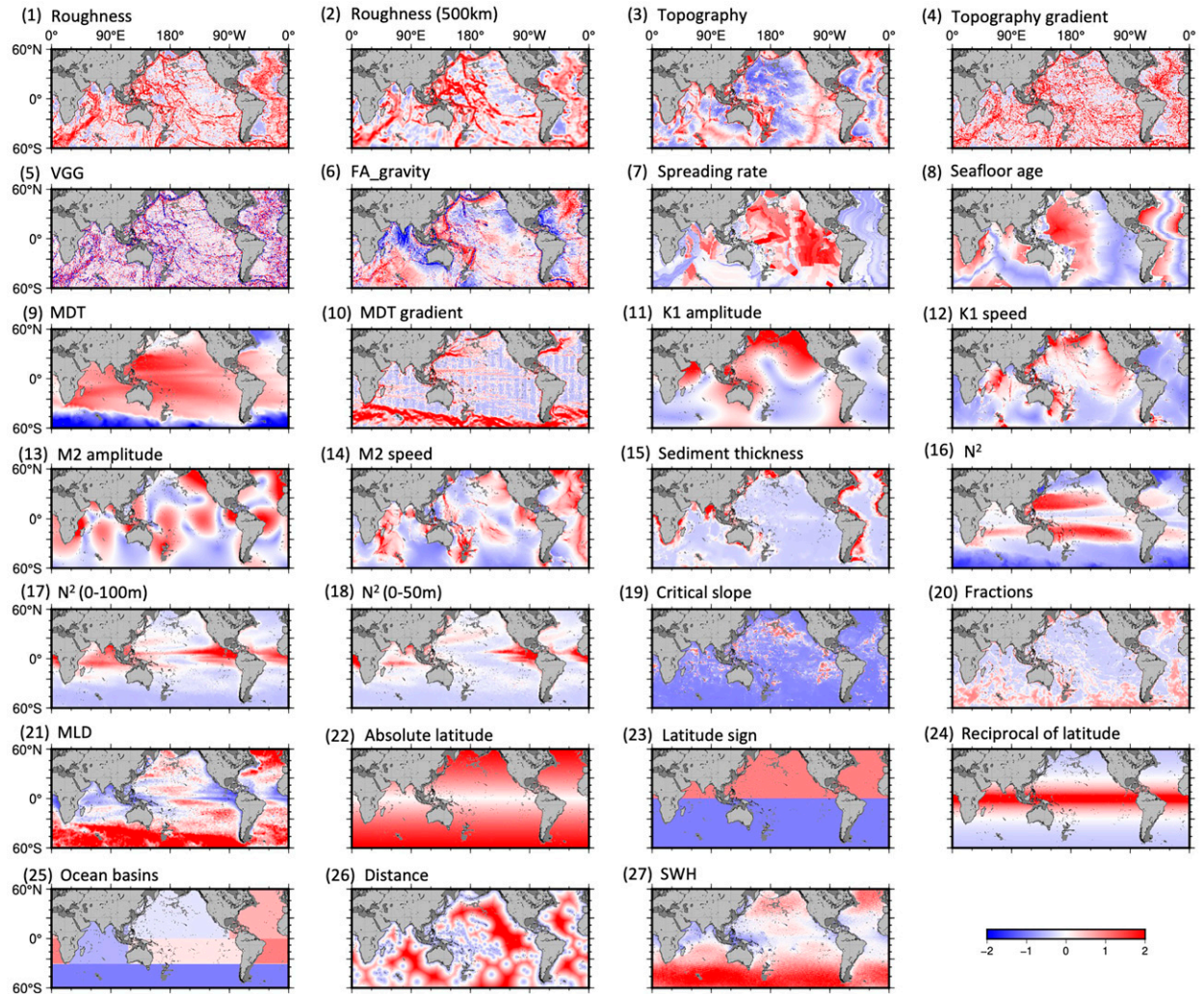
FIG. 2. The 27 normalized environmental parameters used in this study. Land is masked as light gray. All areas that were excluded (lakes, ponds, inland ocean, and coastal areas) are masked as dark gray. All plots are at 7 arc min × 5 arc min resolution using a Mercator projection.

newly generated midocean ridges. Spreading rate is the rate at which an ocean basin widens due to seafloor spreading. It ranges from less than 40 mm yr^{-1} at the Mid-Atlantic Ridge to more than 100 mm yr^{-1} at the east Pacific Rise. We use the seafloor spreading rate and oceanic crustal age from Seton et al. (2020). The dataset is based on magnetic anomaly identifications and the plate tectonic model of Müller et al. (2019). Regions of present-day deformation are not available and are replaced with zeros in the normalized datasets.

(ix) Mean dynamic topography

The mean dynamic topography (MDT) is the current relief that shows steady-state general circulation with gyres and associated western boundary currents. We use the DTU10 MDT, which is the difference between the 12-yr averaged sea surface and the EGM2008 geoid (Andersen and Knudsen 2009). It measures the expected sea surface

height due to currents like the Gulf Stream and the Kuroshio.

(x) Mean dynamic topography gradient

The gradient of dynamic topography is proportional to the geostrophic component of ocean surface current speed. We take the vector gradient of the MDT topography and use the magnitude as MDT gradient.

(xi)–(xiv) K_1 and M_2 tidal amplitude and current speed

Barotropic tides are the major origin of internal tide generation, which leads to SSS variability. We use the two largest components, the K_1 and M_2 tides, with their tidal amplitude and surface current speed from the FES2014 tide model as features. FES2014 tide model is the latest finite element solution tide model assimilating long-term altimetry data and tidal gauges (Carrère et al. 2016). Note that

barotropic tides are removed from along-track altimetry observations.

(xv) Sediment thickness

The seafloor is covered in varying amounts of sediment, and the thickness ranges from a few tens of meters in the open ocean, to several kilometers near the coasts. We use the global ocean sediment thickness map (GlobSed) derived from seismic reflection data (Straume et al. 2019).

(xvi)–(xviii) Stratification N^2

Internal tides are generated in stratified water by the interaction of barotropic tides over rough bottom topography (Garrett and Kunze 2007), and stratification is a key factor in learning the SSS variability. Stratification can be represented by the buoyancy frequency N, or the Brunt–Väisälä frequency. Using the annual statistical mean salinity and temperature data from the *World Ocean Atlas 2018* (*WOA18*), we evaluate the mean buoyancy frequency N for the mixed layer (0–100 m), upper ocean (100–300 m) and deep ocean (300–2000 m) with the Gibbs Seawater oceanographic toolbox (McDougall and Barker 2011).

(xix) M_2 tide critical slope

Critical slope is the bottom slope that equals the angle at which rays of internal waves of tidal frequency propagate. It is a key parameter governing the internal tide generation. We estimate the critical slope of the M_2 tide following Eq. (1) of Becker and Sandwell (2008). The calculation uses *WOA18* salinity and temperature data to calculate the buoyancy frequency N at different depths and then extrapolates N to the seafloor, assuming an exponential function of depth (St. Laurent and Garrett 2002).

(xx) Fractions of slope above critical

The smallest seamounts that are detectable in the satellite altimetry could be 800 m in height and 4 km in radius (Gevorgian et al. 2021). Our spatial resolution of 7 arc min × 5 arc min is coarser than the scales of small tectonic structures including seamounts. We use the 15-arc-s SYNBATH bathymetry map (Sandwell et al. 2022) to calculate seafloor slope, then calculate the fractions of super critical slope of M_2 tide in each 7 arc min × 5 arc min grid. (In Fig. 6, we use "fractions" to represent fractions of slope above critical.)

(xxi) Mixed layer depth

The ocean mixed layer is a surface layer of nearly uniform density resulting from stirring of surface waters by the wind or heat fluxes. As a feature, we use the 12-month average of the monthly mean mixed layer depth (MLD) product derived from almost 2 450 000 Argo profiles collected through March 2021 (Holte et al. 2017).

(xxii)–(xxiii) Absolute latitude and the sign of latitude

Some ocean activities are tied to the latitude; for example, ocean eddies scale with the Rossby radius, which varies with latitude; zonal jets are also shown to populate every part of the ocean (Maximenko et al. 2005). We use the absolute latitude and the sign of latitude (1 for the Northern Hemisphere and −1 for the Southern Hemisphere) as features in this study. We avoid using longitude as a feature. Using both the longitude and latitude as features would allow the model to take a shortcut using geographic coordinates in training, instead of learning the relations between input physical features and output labels as we expect.

(xxiv) Reciprocal of latitude

The reciprocal of Coriolis frequency connects the SSS to the EKE, or the average geostrophic flow speed. Coriolis frequency is defined as $f = 2\omega \sin(\phi)$, where ϕ is the latitude and ω is Earth's angular speed. Coriolis frequency f sets the lower bound for the frequency of internal wave motions. For this study, we neglect the constants and slightly modify the term to be $1/[\sin(|\phi|) + 0.2]$, where 0.2 is added to the denominator to avoid a singularity at the equator.

(xxv) Ocean basins

Different basins of the ocean exhibit large-scale differences in stratification and circulation. To allow for the possibility of basin-scale variability that is not readily represented by the other variables, we identify each ocean basin with an integer from −1 to 4 to distinguish the Southern Ocean, the Indian Ocean, the North Pacific Ocean, the South Pacific Ocean, the North Atlantic Ocean, and the South Atlantic Ocean.

(xxvi) Distance to the nearest thermocline boundary

A thermocline is the transition layer where temperature decreases rapidly from the mixed upper layer of the ocean to much colder deep water. It is associated with high stratification and creates conditions for internal tide generation. We use the 12-month average of the monthly maximum mean mixed layer depth product collected from Argo profiles to represent the thermocline depth (Holte et al. 2017). We pick out the boundary where the thermocline intersects the ocean floor, and we calculate the nearest distance to the boundary as a feature. (In Fig. 6, "distance" denotes distance to the nearest thermocline boundary.)

(xxvii) Significant wave height

Winds and wave heights are highly correlated. Most ocean surface currents are caused by wind, and surface gravity waves are also generated by the friction between wind and water. We use the multiyear mean SWH as a feature. SWH is provided in the waveform data in the along-track satellite altimetry product.

b. Methods

1) CORRELATION ANALYSIS: REVISITING GILLE ET AL. (2000)

Correlation is a statistical method that measures the strength of association between two linearly related variables and the direction of the relationship. Gille et al. (2000) calculated the Pearson correlation coefficients between seafloor roughness and eddy kinetic energy as a function of depth. Their analysis built on a hypothesis that seafloor roughness could serve either
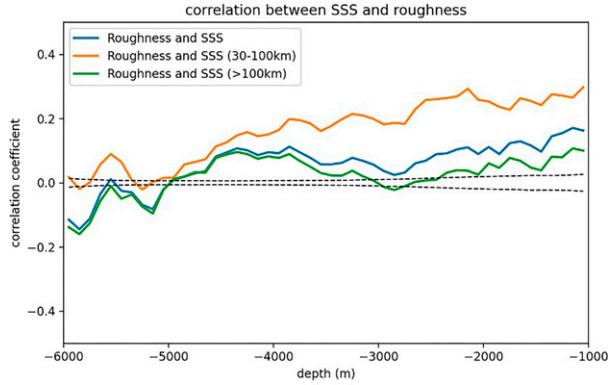
FIG. 3. Correlation coefficients of seafloor roughness vs SSS variability in three wavelength bands (>30 km or full band, 30–100 km, and >100 km) as a function of ocean depth. The black dashed lines are the 95% confidence intervals beyond which correlation coefficients are statistically reliable.

to dissipate eddy kinetic energy by exerting friction at observed scales, or it could be a source of energy by generating lee waves or instabilities or steering eddies. Roughness was computed by bandpass filtering Smith and Sandwell (1997) bathymetry to retain wavelengths between 80 and 160 km and then computing RMS height. Eddy kinetic energy E_k was derived from along-track slopes of TOPEX, *European Remote-Sensing Satellite-1* (*ERS-1*), and *ERS-2* using the geostrophic relationship and assuming eddy variability to be isotropic. Along-track slopes were low-pass filtered to retain signals with wavelengths longer than 80 km, and data equatorward of 20° were omitted because of errors associated with small Coriolis parameter f. They found a positive correlation between roughness and $E_k^{1/2}$ in regions shallower than 3000 m and a negative correlation at depths greater than 4800 m.

In this study, we repeat the correlation analysis of Gille et al. (2000) using updated SSS variability, bathymetry and roughness datasets as described in section 2a. We bin the roughness and SSS variability (30–100 km; >100 km) by local depth in each 100-m range, and as a function of depth calculate the Pearson correlation coefficient between roughness and SSS variability. The result is shown in Fig. 3.

For each 100-m depth range, we fit the corresponding 30–100-km SSS variability as a linear function of roughness and use this information to predict SSS variability. We then combine the predictions made over each depth range to map predicted SSS variability (Fig. 4a) and the differences with the observed SSS variability (Fig. 4b). We also predicted the large-scale SSS variability (>100 km) (Figs. 5a and 5b).

Mean-square error (MSE) and R^2 serve as metrics of model performance. MSE is the second moment of the error (L2 norm) and incorporates both the variance of the estimator and its bias. It is computed as (Sammut and Webb 2011)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2. \tag{1}$$

The R^2 score is a measure of the amount of variance in the predictions explained by the dataset; R^2 values near 1

indicate better predictions. The R^2 is defined as (Devore 2011)

$$R^2 = 1 - \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \Big/ \sum_{i=1}^{N} (y_i - \overline{y})^2, \tag{2}$$

where y_i is the ith observation, $\hat{y}_i$ is the ith prediction, and $\overline{y}$ is the mean value of observed SSS variability. The total number of observations N is 2 677 475 grid points in this study. We assume that the prediction at each data point has the same uncertainty.

2) LINEAR REGRESSION MODEL

Linear regression assumes linear relationships between the input features and the output labels. It can be computed via ordinary least squares fitting, which minimizes the sum of the squared residuals. Although SSS variability may have nonlinear or high-order dependences on the environmental parameters, we do not consider higher-order terms and focus only on the linear correlation between SSS variability and environmental variables. We adopt the linear regression model from the scikit-learn library (Pedregosa et al. 2011) to build models between the SSS variability and 27 environmental parameters and to evaluate model performance as well as feature importance.

We compute a global mesoscale SSS variability (30–100 km) prediction map using the linear regression model and evaluate model performance using the R^2 and MSE metrics. Steps to construct the predicted map are as follows: 1) First, we divide the ocean into 64 blocks of equal area (see Fig. A1 in the appendix for the block separation). 2) We make predictions over each block. For each unique block that is used as a test set, we randomly select 44 of the remaining 63 blocks as training datasets (around 70%). We adopt all 27 environmental parameters as features and SSS variability (30–100 km) as labels (i.e., the predicted variable) and train a linear regression model, then use the model to predict SSS variability over the selected block. 3) We then shuffle the training datasets 30 times. Each time we randomly select 44 of 63 blocks as training datasets and repeat the above step to get predictions over the chosen block or test set. 4) We use the average of the 30 predictions, concatenate the averaged predictions over each unique block, and make a global prediction-only map (Fig. 4c). The training strategy of splitting the data into 64 geographical bins and using the input from different separate regions to make predictions over a specific region ensures model generalizability and prevents the model from simply learning local information. The differences between predicted and observed SSS variability (30–100 km) are shown in Fig. 4d. We also run the same processes to obtain the prediction map for the large-scale SSS variability (>100 km; Figs. 5c and 5d). Note that R^2 and MSE are calculated using the prediction map (test sets only) and observations.

We use two methods to evaluate the feature importance of the linear regression model: (i) feature forward selection; (ii) feature ablation. The details are described below.
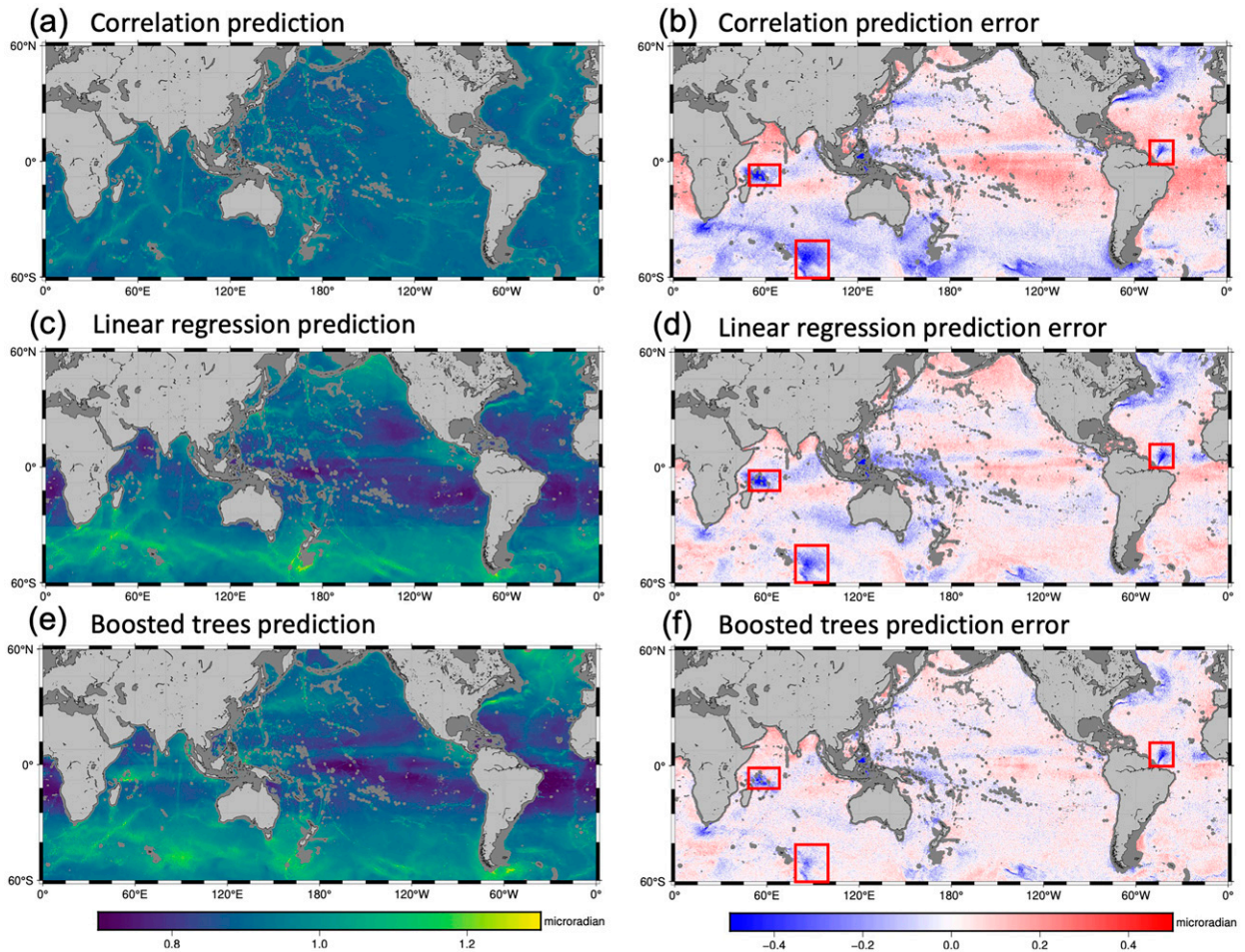
FIG. 4. (left) The predicted mesoscale SSS variability map (30–100 km) and (right) the associated prediction error using (a),(b) the correlation between seafloor roughness and SSS variability (30–100 km) as a function of seafloor depth; (c),(d) the linear regression model; and (e),(f) the boosted trees model. Regions with prediction failures are enclosed in red-outlined boxes in (b), (d), and (e).

(i) Forward selection

The forward selection algorithm starts with a model with zero features and iterates through all single-feature models to find one that is most predictive (e.g., with the smallest L2 norm); picks the feature; then for all the remaining features, considers adding each of them to build a two-feature model, by including whichever second feature is most predictive; always includes the feature giving the biggest marginal gain in predictive power in the presence of all other selected features. The order in which features are selected is a measure of feature importance. We use the sequential forward selection method from the scikit-learn library to select and rank features that are most relevant to the SSS variability. We evaluate model performance using fivefold cross validation in which data are split into five groups; each unique group will be used as a test set and remaining groups will be used as a training set; we train a model, retain the evaluation score (L2 norm), and then summarize the model performance using the five evaluation scores.

(ii) Feature ablation

Feature ablation measures how much the model performance is degraded by deleting one feature. First, we use all M (27 in this study) features and 70% of data as training datasets to train a linear regression model. (The process can also be applied with any other machine learning model.) Models with $M - 1$ features obtain worse performance over the test dataset, which comprises the remaining 30% of the data, for example, with increased MSE relative to the model using all available M features. The reduction in performance, as quantified by the L2 norm, provides a measure of the importance of the deleted feature.

3) BOOSTED TREES ALGORITHM

The gradient boosted trees algorithm is an ensemble of decision trees as weak learners. Each tree tries to fit the residuals from previous models. All those trees are trained by propagating the gradients of errors throughout the system. We implement the boosted trees algorithm using the Light
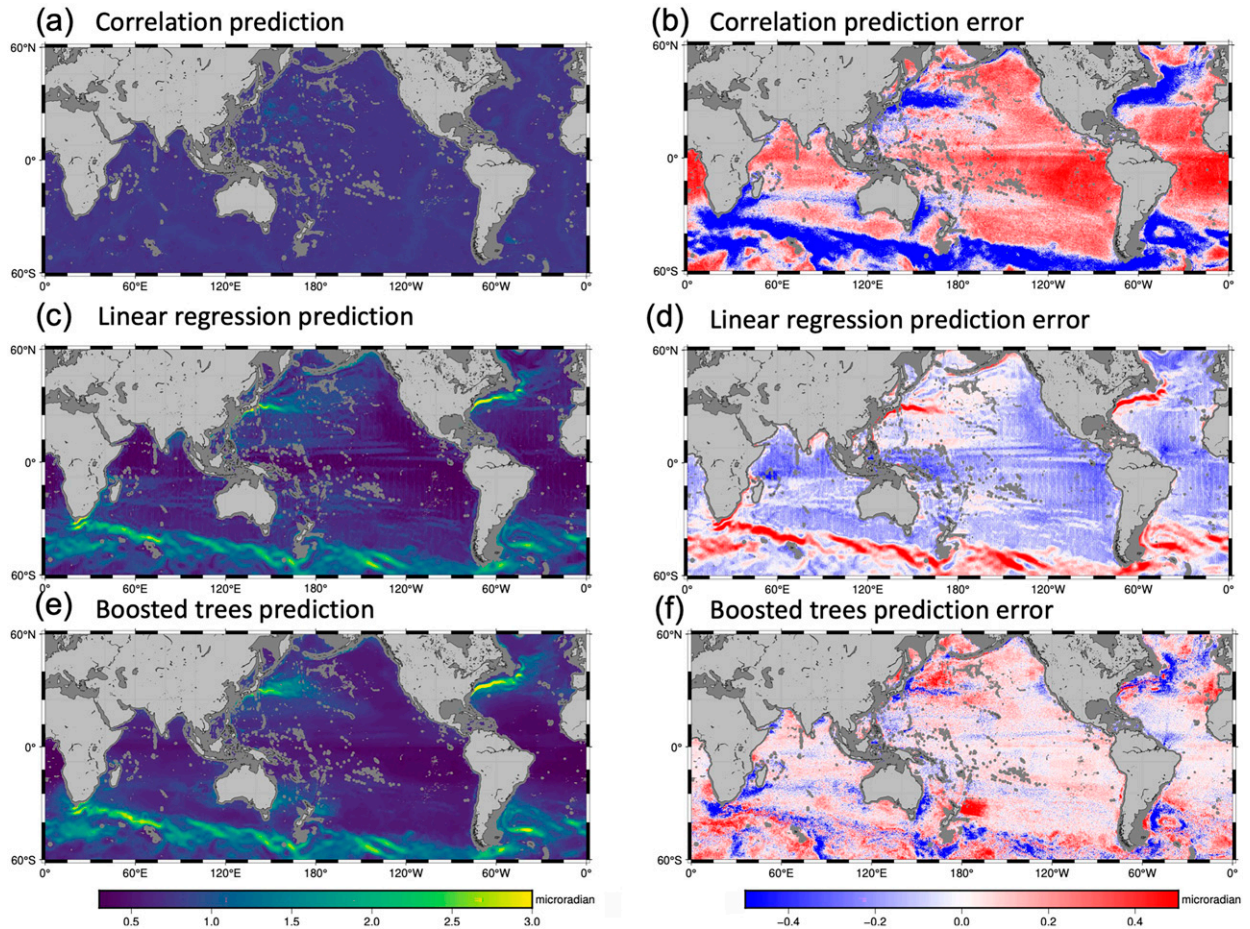
FIG. 5. As in Fig. 4, but for >100 km.

gradient boosted machine (LightGBM; Ke et al. 2017). LightGBM is a fast and efficient framework. It splits the tree leafwise using a histogram-based method for selecting the best split and buckets continuous feature values into discrete bins thus lowering memory usage. It works with large datasets with ease and results in much better accuracy, which can rarely be achieved by any of the existing boosting algorithms. The main drawback of gradient boosted trees is that finding the best split points in each tree is time consuming.

We compute a global prediction-only SSS variability (30–100 km) map with the boosted trees model, again using 27 environmental parameters as features as described in section 2a(2). The steps to calculate predictions are basically the same as for the linear regression model [as discussed in section 2b(2)] with the exception that for each model we randomly select 44 of the remaining 63 blocks as training datasets, and the other 19 blocks as validation datasets. Figure 4e shows the prediction-only map, and Fig. 4f shows differences with observations. We evaluate the model performance through the R^2 score and the MSE.

We identify the dominant features in the boosted trees model using two methods: (i) feature ablation and (ii) feature importance inbuilt by the boosted trees model. Feature ablation

is introduced in section 2b(2). For method ii, we train a boosted trees linear regression model with feature subsampling (0.8) and bagging (bagging fraction 0.7 and bagging frequency 5) to avoid overfitting and use the feature importance returned by the model's split gain.

3. Results

a. Correlation with roughness

As a baseline for the machine learning analyses, we first assess the skill of the roughness correlation of Gille et al. (2000) in predicting SSS. Figure 3 shows the correlation coefficients between SSS variability and seafloor roughness as a function of depth. The blue curve, indicating the correlation between roughness and full-scale SSS variability (>30 km), resembles Fig. 2a in Gille et al. (2000), which focused on the 80–160-km wavelength band. Both show that there is positive correlation between roughness and SSS variability in regions shallower than 3000 m and negative correlation in regions deeper than 5000 m. The correlation with large-scale SSS variability (>100 km; green curve) is similar to the correlation with full-scale SSS variability (blue curve) in the deep ocean. It has reduced

values at shallow water, and the correlation is sometimes insignificant. The negative correlation between roughness SSS variability (>100 km) and roughness at depths greater than 5000 m are mostly attributed to the Argentine Basin, where the energetic Zapiola Anticyclone circulates counterclockwise over smooth abyssal plains (Saraceno et al. 2009). The negative correlation also indicates that the variation of geostrophic flows is not related to seafloor roughness. At mesoscales (30–100 km), the correlation between seafloor roughness and SSS variability is basically positive at all depth levels, and the correlation is much higher in shallow water when compared with large-scale flows. This pattern of positive correlation suggests that the mesoscale oceanic variability is generated as a response to rough topography.

b. Predicted SSS variability

1) MESOSCALE SSS VARIABILITY PREDICTION

We have introduced three statistical methods for producing predictions of SSS variability in section 2b. We show the predicted SSS variability and prediction errors for the mesoscale band in Fig. 4.

Using the correlation between seafloor roughness and SSS variability as a function of ocean depth, we map global predicted mesoscale SSS variability (Fig. 4a). The prediction is far from being realistic. It is only capable of capturing SSS variability over regions with rough topography, for example, over the slow-spreading ridges (the southwest Indian Ridge and the Mid-Atlantic Ridge), fracture zones (the Challenger Fracture Zone), and hotspot chains (the Emperor Seamount Chain and the Louisville Seamount Chain). The difference with the observed SSS variability (30–100 km) is shown in Fig. 4b. The prediction error is large over western boundary currents and the ACC. This model has an R^2 score of 0.064 and an MSE of 0.015 μrad^2 on a global scale.

The linear regression model, which uses 27 features as variables to fit the SSS variability (30–100 km), makes predictions over each individual block. The prediction map that concatenates the predictions is shown in Fig. 4c and the prediction errors in Fig. 4d. This time the predicted variability map is able to capture signatures of most geostrophic flow instabilities, and unbalanced flows over rough topography. It has an R^2 score of 0.362 and an MSE of 0.010 μrad^2. It also shows significantly reduced prediction errors in the Southern Ocean when compared with the correlation analysis (cf. Figs. 4d and 4b).

The nonlinear boosted trees model provides the best prediction (Fig. 4e) among the three models. On global scales, the R^2 score is 0.563, the MSE is 0.007 μrad^2, and the model provides realistic predictions. For example, it predicts high SSS variability in the Mascarene Basin; the prediction in the tropical Pacific agrees well with observations. However, this model, like the other two models, fails to predict the internal tide variability in the Amazon outflow. We hypothesize that this local variability is related to physical processes that are not accounted for in our features. There are some discontinuities at the block edges (e.g., a zonal stripe above the Mascarene Ridge) arising from the fact that predictions for

each block are independent. Using the average from 30 rounds of training/predicting greatly reduces the discontinuities.

2) LARGE-SCALE SSS VARIABILITY PREDICTION

At large scales, SSS variability (>100 km) is well categorized, and it is expected to be dominated by western boundary currents and the ACC, regions where the mean SSS is large. In general, large-scale SSS variability (>100 km) is linearly related to the MDT gradient, which represents the strength of geostrophic flows. We expect that both the linear regression model and the boosted trees algorithm, which use MDT gradient as a feature, should capture the relations with the large-scale SSS variability and make realistic predictions. We use the three models introduced in section 2b to predict large-scale SSS variability (>100 km) and to check if the model performance is consistent with our current knowledge. The prediction and associated errors are shown in Fig. 5.

We expect that by employing the correlation between seafloor roughness and SSS variability, we will not be able to make good predictions since the MDT gradient is not correlated with seafloor roughness. This expectation is borne out in Fig. 5a, which captures almost no large-scale SSS variability. The prediction has an R^2 score of 0.010 and an MSE of 0.208 μrad^2.

The linear model uses MDT gradient as one of the 27 input variables. Overall, it captures the variations of geostrophic flows relatively well and shows strong variability in the vicinity of strong geostrophic flow. The predicted SSS variability (Fig. 5c) is somewhat biased: it is larger than observations over regions where the SSS is larger and smaller in background regions where SSS is small. This model has an R^2 score of 0.667 and an MSE of 0.070 μrad^2.

The boosted trees model makes the best predictions for the large-scale SSS variability (>100 km). It makes realistic predictions (Fig. 5e), and the prediction error (Fig. 5f) is reduced and is less biased relative to the linear regression model (Fig. 5d). This model has a R^2 score of 0.776 and an MSE of 0.047 μrad^2.

c. Feature importance

One key question that arises from using the 27 environmental parameters described in section 2a(2) is to determine which features are most critical in training the linear regression model and the boosted trees model. As described in sections 2b(2) and 2b(3), we evaluate the feature importance of the linear regression model using feature forward selection and feature ablation, and we evaluate the boosted tree model using feature ablation and the boosted tree embedded feature importance. We list feature importance in the form of ranks (from 1 to 27) for the mesoscale SSS variability (30–100 km) in Fig. 6. Features with smaller ranks have higher importance.

These four methods provide different ranks of feature importance. At least two out of four show that distance to the nearest thermocline boundary, SWH, MDT, MDT gradient, absolute latitude, seafloor roughness, ocean basins and M_2 tidal speed, K_1 tidal amplitude, and N^2 (0–100 m) are the governing features in predicting the mesoscale SSS variability. The boosted trees ablation methods show that the model performance would decrease the most when removing SWH as a
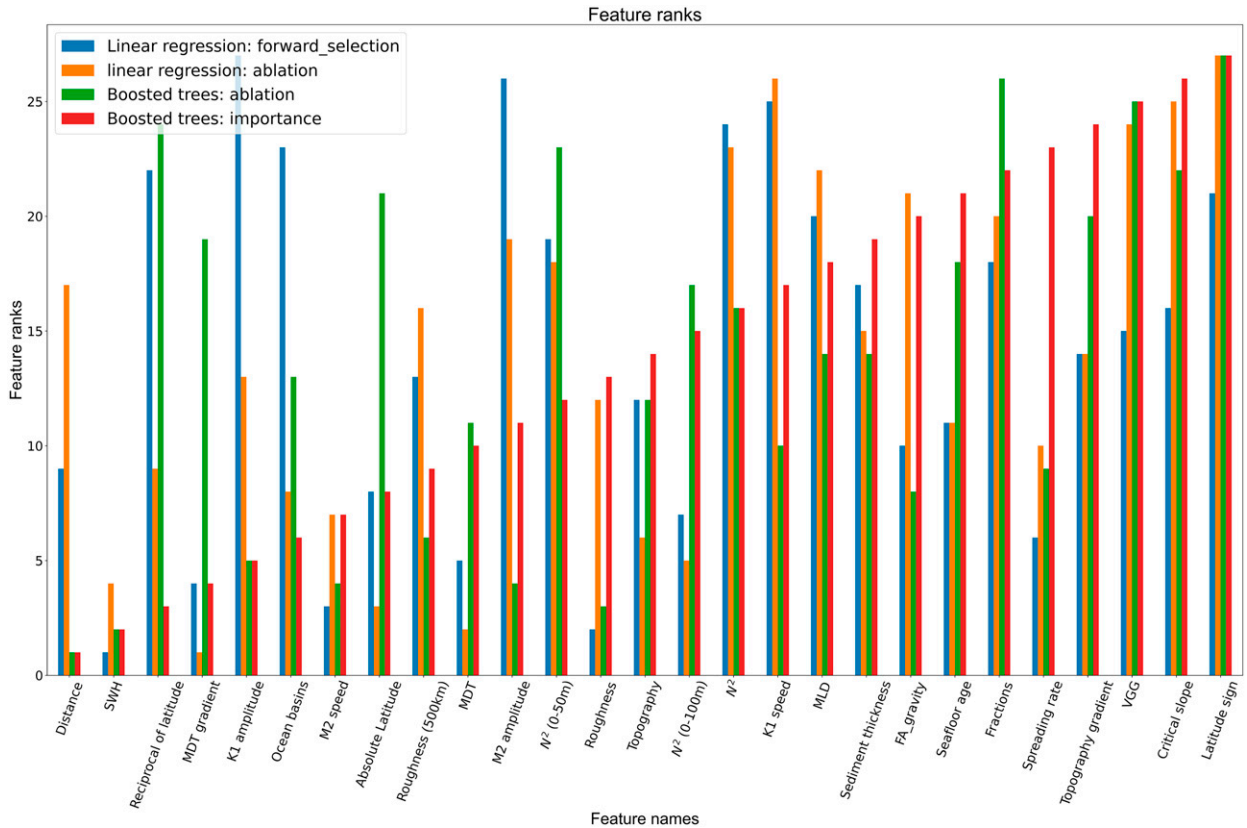
FIG. 6. Ranks of features of the mesoscale SSS variability (30–100 km) from the linear regression model using the forward selection and the feature ablation and from the boosted trees model using the feature ablation and the feature importance returned by the model. Features with smaller ranks have higher importance. Features and their full name are introduced in section 2a(2).

feature (L2 norm increases by 35.6%). The sign of latitude, the critical slope of the M_2 tide, VGG, topography gradient, and fractions of slope above critical, are not important in any of the four methods. While SSS variability is relevant to ocean basins, it is not hemispherically related. Note that the lack of correlation with VGG is an indication that the mean SSS model removed from the profiles accurately captures the small-scale gravity features that are represented by VGG. The remaining SSS variability, which is used as a label in this study, represents oceanic signals, a main assumption of the entire analysis.

At large scales (>100 km), feature rankings in Fig. 7 indicate that all methods show MDT gradient to be the most dominant feature in predicting SSS variability, in accord with our initial expectation. Large-scale SSS variability can be represented by the strength of the geostrophic flow, which scales with MDT gradient. Where the mean flow is large, the variation is large as well.

4. Discussion and conclusions

We have used three statistical methods to build models and predict the SSS variability in the mesoscale band (30–100 km) and in the large-scale band (>100 km). We have revisited the correlation analysis of Gille et al. (2000) and then extended our analysis to incorporate more environmental parameters and to test additional methods to understand the governing factors of sea surface variability. Both the linear regression model and the boosted trees model incorporate 27 features, and they significantly outperform the correlation analysis that only accounts for the seafloor roughness and ocean depth. The boosted trees model also has advantages over the linear regression model in that it is capable of building more complicated nonlinear relations between environmental parameters and the mesoscale SSS variability. Thus, it has a higher R^2 score (0.563 as compared with 0.381) and results in a smaller MSE in prediction errors (0.007 μrad^2 as compared with 0.010 μrad^2) for the mesoscale SSS variability. The large-scale SSS variability is largely consistent with geostrophy, with both the linear regression and the boosted trees model able to explain more than 60% of the variance, and MDT gradient serving as the leading-order predictor.

We divide data into 64 geographical blocks and use data from separate blocks to train models and make predictions. This approach is essential for the boosted trees model because it prevents the model from simply using local information to make unrealistically good predictions. Although this approach is unnecessary for the linear regression model, for consistency,
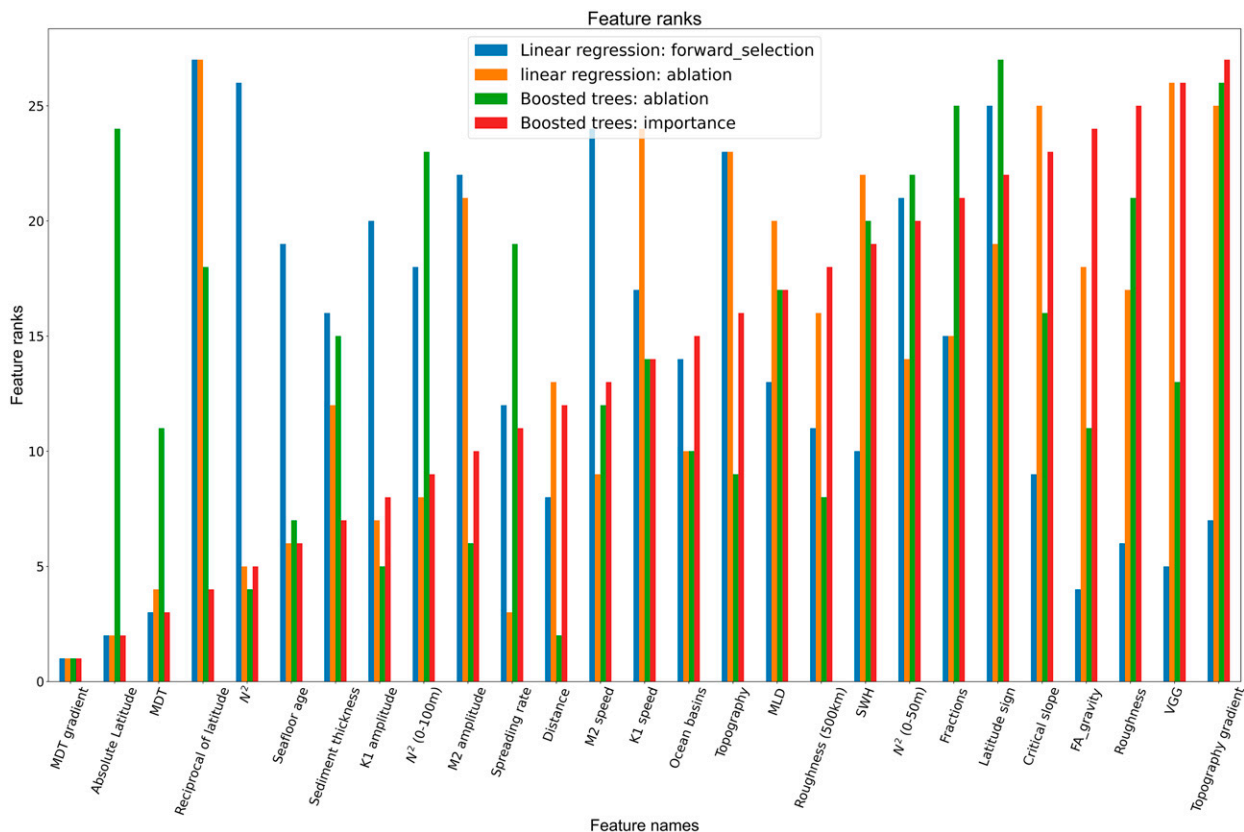
FIG. 7. As in Fig. 6, but for the large-scale SSS variability (>100 km).

we use it for both the boosted trees model and the linear regression model to train the models, make predictions, and evaluate model performance. For each test block, we train the model 30 times as introduced in sections 2b(2) and 2b(3). The prediction map concatenated from 30 realization average greatly reduces the discontinuities at block edges and has little impact on evaluating model performance. The variance of 30 realization of predictions is a measure of model uncertainty. We calculate the standard deviation (std) and use 1/std as weights to recalculate the R^2 score. The updated R^2 scores using 1/std as weights are 0.407 for the linear regression model, and 0.583 for the boosted trees model, which are similar to R^2 scores assuming uniform weights (0.381 for the linear regression model and 0.563 for the boosted trees model).

Although the machine learning models in this study predict SSS variability, our fundamental goal is to use machine learning to identify the physical processes governing SSS variability. There are two categories of conclusions that we can reach from this study: (i) What environmental parameters matter in predicting SSS variability? (ii) In which regions of the ocean do our models fall short, and why? We hypothesize that processes or regions that are not readily represented by a simple model might indicate the presence of unusual or complicated physical mechanisms.

We have used four methods to rank the importance of environmental parameters, that is to calculate the feature importance.

Rankings from the four methods can diverge substantially, although the linear regression and the boosted trees model show some overlap. Overall, the rankings show that distance to the nearest thermocline boundary, SWH, MDT, MDT gradient, seafloor roughness, ocean basins and M_2 tidal speed, K_1 tidal amplitude, and N^2 (0–100 m) are key features in predicting the mesoscale SSS variability (30–100 km). For large scales (>100 km), all methods show that MDT gradient is the dominant feature in predicting SSS variability.

The high feature importance assigned to the distance to the nearest thermocline boundary suggests that internal wave generation plays a role in generating SSS variability. This is because internal waves have the largest amplitudes at the base of the thermocline. They can be generated when tides disturb water to move up and down the steep seafloor boundary, so waves are larger close to the boundary and dissipate as they move away. Tidal wave beams interact with the thermocline and will generate large-amplitude solitary waves (Akylas et al. 2007).

The role of SWH in governing mesoscale SSS variability (30–100 km) could have two interpretations: first it could mean that wave height noise leaks into the 30–100-km wavelength band; alternatively, it could indicate that SWH reflects increased wind forcing, which produces more mesoscale ocean variability. We correlate the mean SWH with multiple bandpass filtered SSS variability and find that they are highly

correlated (coefficients > 0.5) when the SSS variability has a wavelength less than 30 km (Fig. A2 in the appendix). The correlation coefficient drops below 0.5 where the SSS wavelength is longer than 30 km, which indicates SWH noise is not the dominant force anymore, but its influence cannot be ruled out.

There are straightforward relations between mesoscale SSS variability and MDT gradient, seafloor roughness, absolute latitude, ocean basins, stratification N^2 (0–100 m), M_2 tidal speed, and K_1 tidal amplitude. MDT gradient represents the strength of geostrophic flows. The mesoscale eddies emerging from the instabilities of strong geostrophic flows could appear in smaller spatial scales. Thus, there is strong mesoscale SSS variability in the vicinity of western boundary currents and the ACC, and the MDT gradient is highly correlated with mesoscale SSS variability. As discussed by Gille et al. (2000), seafloor roughness could dissipate eddy kinetic energy in the deep ocean, or it could be a source of energy by generating internal waves or instabilities in the shallow ocean. Some ocean activities, for example, ocean eddies, are tied to the latitude. The physical processes and the generation mechanisms are different in different dynamical zones or ocean basins. The M_2 and K_1 barotropic tides can convert to internal tides when they impinge on a steep seafloor in stratified water.

We have identified a number of regions where machine learning consistently fails to yield good predictions for the mesoscale SSS variability: 1) the Amazon outflow (310°–325°E, 0°–15°N) in the tropical Atlantic Ocean; 2) the Mascarene Ridge (45°–63°E, 3°–15°S) in the Indian Ocean; 3) the Kerguelen Plateau (75°–95°E, 45°–60°S) in the Southern Ocean. Each of these regions displays large negative prediction errors (red boxes in Figs. 4b,d and f). The Amazon outflow is associated with strong local and seasonal freshwater input. This freshwater input affects stratification, likely in ways that are not replicated elsewhere in the global stratification data that we use in the study. There is also a strong tidal impact that is not necessarily well captured. The Mascarene Ridge is located in an area of energetic barotropic tidal currents that are normal to the ridge (Lozovatsky et al. 2003; Morozov 2006). The ridge is characterized by guyots with flat summits straddled by channels. The shallow banks, the shallow channel between the banks, and deep waters around the Mascarene Ridge provide unique conditions for the generation of intense internal tides (Morozov 2006). The Kerguelen Plateau is a major topographic feature in the Southern Ocean where the main fronts of the ACC encounter rough topography. The strong geostrophic flow converts to upward-propagating internal waves over rough bottom topography. Strong wind forcing generates near-initial downward-propagating internal waves (Meyer et al. 2015).

The difficulties that the boosted trees approach encounters in modeling these three regions, suggest that they are unusual areas relative to the training datasets, and in the language of machine learning can be considered to be "out of distribution" (Hendrycks and Gimpel 2016) unless they are specifically included as training data. The R^2 scores excluding the above three outlier regions are 0.407 and 0.563 for the linear regression and the boosted trees model, which are slightly better than R^2 scores for the global ocean (0.381 and 0.563, respectively). The outlier regions have little impact on skewing model accuracy. The physical processes within these regions are possibly strongly governed by regionally specific environmental parameters that are not in our features. Local parameters like freshwater input are hard to incorporate because they are concentrated near the coast and sparsely distributed in space. Our model has no temporal resolution, and some important environmental parameters, that is, wind stress, are also not included. This is consistent with findings in other studies showing that machine learning algorithms are less likely to succeed when validation or test data do not lie within the distribution of the training data (Liang et al. 2017; Partee et al. 2021; Sonnewald et al. 2021; Sinha and Abernathey 2021). In these regions, a physical model that incorporates existing physical knowledge may have better predictive skills.

Other than the linear regression and the boosted trees model, we also tested the possibility of training with the lasso regularization, and the random forest models. The regularization parameter for the lasso model is too small to make it different from the linear regression. The performance of random forest proved to be worse than boosted trees, and thus we discarded its use as a nonlinear model in this study. Deep learning or neural networks can also be a potential class of models to explore. We adopted the boosted trees model as our preferred nonlinear machine learning model because it is efficient, it converges well relative to random forests, and it provides enough flexibility to test multiple inputs and multiple scenarios efficiently.

Data quality goes a long way toward determining the performance of the machine learning model and the predictions. Machine learning algorithms are not able to train a good model by identifying all possible connections between poorly selected input features. To explore the importance of using physically relevant input features, we chose as features 27 pictures of animals that have no relation to the SSS variability, although the animal pictures do have large-scale patterns of spatial variability that are comparable to our environmental parameters. We used these pictures to train a linear regression model and a boosted trees model to predict SSS variability following the same procedures that we used with the environmental parameters (see Fig. A3 in the appendix). Since the animal pictures have no physical relation to environmental parameters, we would expect a large MSE and an R^2 of 0. Indeed, if we use 27 pictures of Gaussian noise as inputs, the MSE and R^2 are 0.025 μrad^2 and -0.007 for both models. However, counter to expectations, when we use animal pictures for the linear regression model the MSE and R^2 are 0.023 μrad^2 and 0.08, and for the boosted trees model they are 0.018 μrad^2 and 0.273. Overall, with unphysical data, both models perform worse than they do when using environmental parameters. For the boosted trees model, the fact that irrelevant data with feature scales that are comparable to observed spatial scales can explain 0.273 of the variance should serve as a cautionary warning that spatially correlated irrelevant fields can yield artificially high skill metrics.
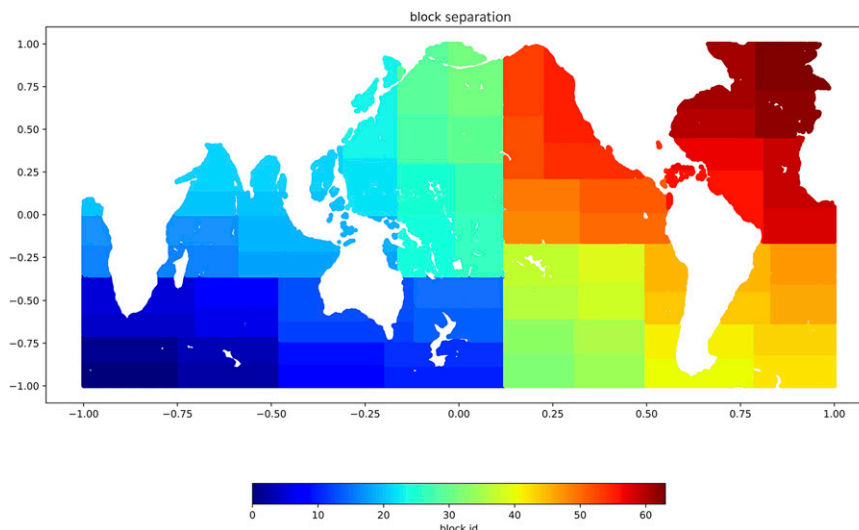
FIG. A1. Equal-area block separation. Each block is assigned with a unique identifier (id) from 0 to 63. The longitude (from 0° to 360°) and latitude (from −60° to 60°) are linearly normalized to −1 and 1.

The swath Surface Water and Ocean Topography (SWOT) mission to be launched in late 2022, will have the ability to resolve scales of a few tens of kilometers where internal tides/waves are mixed with geostrophic currents (Morrow et al. 2019). SWOT's spatial resolution capabilities call for understanding unbalanced waves and mesoscale ocean activities. Our study shows both the potential and limitations of using machine learning to unveil the driving forces and to make global predictions of mesoscale SSS variability. Machine learning is a powerful tool, and this study is a step forward in using machine learning to advance our understanding of Earth system science.

Data availability statement. Links to the openly accessible data are as follows: SYNBATH (https://topex.ucsd.edu/pub/synbath/), STRM15+ (https://topex.ucsd.edu/pub/srtm15_plus/), VGG and free-air gravity (https://figshare.com/articles/online_resource/Tozer_et_al_2019_SRTM15_GMT_Grids/7979780), seafloor spreading rate and oceanic crustal age (https://earthbyte.org/webdav/ftp/earthbyte/agegrid/2020/), mean dynamic topography (https://www.space.dtu.dk/english/Research/Scientific_data_and_models/downloaddata), salinity and temperature data (https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/), mixed layer depth (http://mixedlayer.ucsd.edu/), and sediment thickness (http://earthdynamics.org/data/). All other data, including the sea surface slope variability and

processed features, are available upon request from the corresponding author.

APPENDIX

Ancillary Figures

Figure A1 shows the equal-area block separation. The Pearson correlation coefficients for the mean significant wave height and the SSS variability are shown in Fig. A2. Figure A3 presents the predicted SSS variability maps and associated prediction errors.
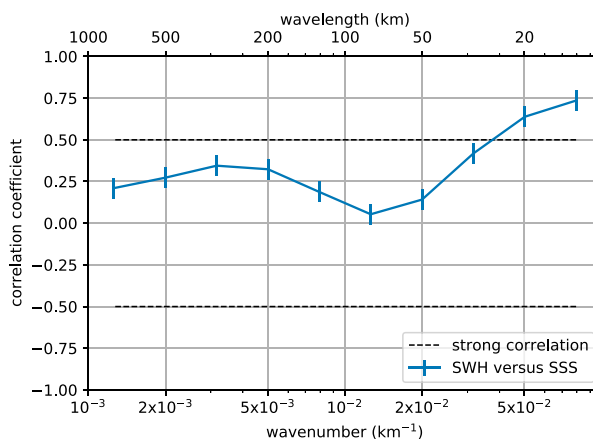

FIG. A2. The Pearson correlation coefficient for the mean significant wave height and the SSS variability in multiple subbands (10–15, 15–25, 25–39, 39–63, 63–100, 100–158, 158–251, 251–398, 398–630, and 630–1000 km). A coefficient of at least ±0.5 means strong correlation.

(a) Animals: linear regression prediction

(b) Animals: linear regression prediction error

(c) Animals: boosted trees prediction

(d) Animals: boosted trees prediction error

(e) Gaussian noise: linear regression prediction

(f) Gaussian noise: linear regression prediction error

(g) Gaussian noise: boosted trees prediction

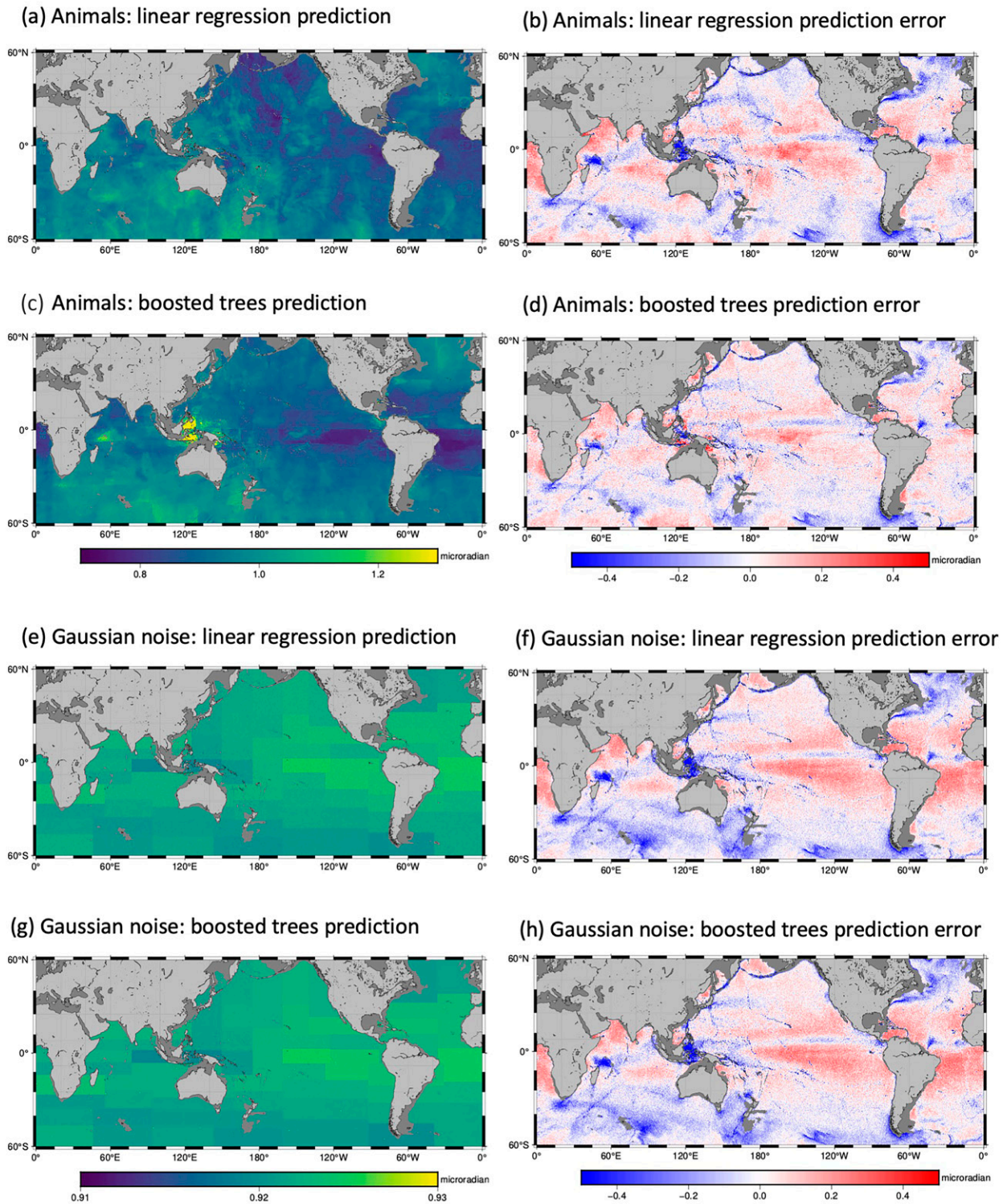## (h) Gaussian noise: boosted trees prediction error

Fig. A3. (a),(c) The predicted SSS variability map (30–100 km) and (b),(d) the associated prediction error using (top) the linear regression model and (top middle) the boosted trees model with 27 features of random animals. The MSE and R^2 for the linear regression model are 0.023 μrad^2 and 0.080 and for the boosted trees model are 0.018 μrad^2 and 0.273. Also shown are (e),(g) the predicted SSS variability map (30–100 km) and (f),(h) the associated prediction error using (bottom middle) the linear regression model and (bottom) the boosted trees model with 27 features of random Gaussian noise. The MSE and R^2 for both models are 0.025 μrad^2 and −0.007.

REFERENCES

Akylas, T. R., R. H. Grimshaw, S. R. Clarke, and A. Tabaei, 2007: Reflecting tidal wave beams and local generation of solitary waves in the ocean thermocline. *J. Fluid Mech.*, **593**, 297–313, https://doi.org/10.1017/S0022112007008786.

Andersen, O. B., and P. Knudsen, 2009: The DNSC08 mean sea surface and mean dynamic topography. *J. Geophys. Res.*, **114**, C11001, https://doi.org/10.1029/2008JC005179.

Baines, P. G., 1982: On internal tide generation models. *Deep-Sea Res.*, **29A**, 307–338, https://doi.org/10.1016/0198-0149(82)90098-X.

Ballarotta, M., and Coauthors, 2019: On the resolutions of ocean altimetry maps. *Ocean Sci.*, **15**, 1091–1109, https://doi.org/10.5194/os-15-1091-2019.

Becker, J. J., and D. T. Sandwell, 2008: Global estimates of sea-floor slope from single-beam ship soundings. *J. Geophys. Res.*, **113**, C05028, https://doi.org/10.1029/2006JC003879.

Carrère, L., F. Lyard, M. Cancet, A. Guillot, and N. Picot, 2016: FES 2014: A new tidal model—Validation results and perspectives for improvements. *Proc. ESA Living Planet Symp.*, Prague, Czech Republic, European Space Agency, 9–13.

Devore, J. L., 2011: *Probability and Statistics for Engineering and the Sciences.* Cengage Learning, 768 pp.

Ferrari, R., and C. Wunsch, 2009: Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annu. Rev. Fluid Mech.*, **41**, 253–282, https://doi.org/10.1146/annurev.fluid.40.111406.102139.

Friedman, J. H., 2002: Stochastic gradient boosting. *Comput. Stat. Data Anal.*, **38**, 367–378, https://doi.org/10.1016/S0167-9473(01)00065-2.

Fu, L. L., and C. Ubelmann, 2014: On the transition from profile altimeter to swath altimeter for observing global ocean surface topography. *J. Atmos. Oceanic Technol.*, **31**, 560–568, https://doi.org/10.1175/JTECH-D-13-00109.1.

García-García, D., and C. C. Ummenhofer, 2015: Multidecadal variability of the continental precipitation annual amplitude driven by AMO and ENSO. *Geophys. Res. Lett.*, **42**, 526–535, https://doi.org/10.1002/2014GL062451.

Garrett, C., and E. Kunze, 2007: Internal tide generation in the deep ocean. *Annu. Rev. Fluid Mech.*, **39**, 57–87, https://doi.org/10.1146/annurev.fluid.39.050905.110227.

Gevorgian, J., D. T. Sandwell, Y. Yu, S.-S. Kim, and P. Wessel, 2021: Global distribution and morphology of seamounts. *2021 Fall Meeting*, New Orleans, LA, Amer. Geophys. Union, Abstract T45D-0273.

Gille, S. T., M. M. Yale, and D. T. Sandwell, 2000: Global correlation of mesoscale ocean variability with seafloor roughness from satellite altimetry. *Geophys. Res. Lett.*, **27**, 1251–1254, https://doi.org/10.1029/1999GL007003.

Goff, J. A., 2010: Global prediction of abyssal hill root-mean-square heights from small-scale altimetric gravity variability. *J. Geophys. Res.*, **115**, B12104, https://doi.org/10.1029/2010JB007867.

——, 2020: Identifying characteristic and anomalous mantle from the complex relationship between abyssal hill roughness and spreading rates. *Geophys. Res. Lett.*, **47**, e2020GL088162, https://doi.org/10.1029/2020GL088162.

Hendrycks, D., and K. Gimpel, 2016: A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv, 1610.02136, https://doi.org/10.48550/arXiv.1610.02136.

Holte, J., L. D. Talley, J. Gilson, and D. Roemmich, 2017: An Argo mixed layer climatology and database. *Geophys. Res. Lett.*, **44**, 5618–5626, https://doi.org/10.1002/2017GL073426.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017: LightGBM: A highly efficient gradient boosting decision tree. *31st Conf. on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, Association for Computing Machinery, 3146–3154, https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Liang, S., Y. Li, and R. Srikant, 2017: Enhancing the reliability of out-of-distribution image detection in neural networks. arXiv, 1706.02690, https://doi.org/10.48550/arXiv.1706.02690.

Lozovatsky, I. D., E. G. Morozov, and H. J. S. Fernando, 2003: Spatial decay of energy density of tidal internal waves. *J. Geophys. Res.*, **108**, 3201, https://doi.org/10.1029/2001JC001169.

Maximenko, N. A., B. Bang, and H. Sasaki, 2005: Observational evidence of alternating zonal jets in the World Ocean. *Geophys. Res. Lett.*, **32**, L12607, https://doi.org/10.1029/2005GL022728.

McDougall, T. J., and P. M. Barker, 2011: Getting started with TEOS-10 and the Gibbs Seawater (GSW) oceanographic toolbox. SCOR/IAPSO WG 127 Doc., 28 pp., http://www.teos-10.org/pubs/gsw/v3_04/pdf/Getting_Started.pdf.

Meyer, A., B. M. Sloyan, K. L. Polzin, H. E. Phillips, and N. L. Bindoff, 2015: Mixing variability in the Southern Ocean. *J. Phys. Oceanogr.*, **45**, 966–987, https://doi.org/10.1175/JPO-D-14-0110.1.

Morozov, E., 2006: Internal tides: Global field of internal tides and mixing caused by internal tides. *Waves in Geophysical Fluids*, Springer, 271–332.

Morrow, R., and Coauthors, 2019: Global observations of fine-scale ocean surface topography with the Surface Water and Ocean Topography (SWOT) mission. *Front. Mar. Sci.*, **6**, 232, https://doi.org/10.3389/fmars.2019.00232.

Müller, R. D., and Coauthors, 2019: A global plate model including lithospheric deformation along major rifts and orogens since the Triassic. *Tectonics*, **38**, 1884–1907, https://doi.org/10.1029/2018TC005462.

Nikurashin, M., and S. Legg, 2011: A mechanism for local dissipation of internal tides generated at rough topography. *J. Phys. Oceanogr.*, **41**, 378–395, https://doi.org/10.1175/2010JPO4522.1.

Partee, S., M. Ellis, A. Rigazzi, S. Bachman, G. Marques, A. Shao, and B. Robbins, 2021: Using machine learning at scale in HPC simulations with SmartSim: An application to ocean climate modeling. arXiv, 2104.09355, https://doi.org/10.48550/arXiv.2104.09355.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Qiu, B., T. Nakano, S. Chen, and P. Klein, 2017: Submesoscale transition from geostrophic flows to internal waves in the northwestern Pacific upper ocean. *Nat. Commun.*, **8**, 14055, https://doi.org/10.1038/ncomms14055.

——, S. Chen, P. Klein, J. Wang, H. Torres, L. L. Fu, and D. Menemenlis, 2018: Seasonality in transition scale from balanced to unbalanced motions in the World Ocean. *J. Phys. Oceanogr.*, **48**, 591–605, https://doi.org/10.1175/JPO-D-17-0169.1.

Quinlan, J. R., 1986: Induction of decision trees. *Mach. Learn.*, **1**, 81–106, https://doi.org/10.1007/BF00116251.

Sammut, C., and G. I. Webb, Eds., 2011: *Encyclopedia of Machine Learning.* Springer Science and Business Media, 1061 pp.

Sandwell, D. T., 2022: *Advanced Geodynamics*. Cambridge University Press, 282 pp.

——, and W. H. Smith, 2005: Retracking ERS-1 altimeter waveforms for optimal gravity field recovery. *Geophys. J. Int.*, **163**, 79–89, https://doi.org/10.1111/j.1365-246X.2005.02724.x.

——, and ——, 2014: Slope correction for ocean radar altimetry. *J. Geod.*, **88**, 765–771, https://doi.org/10.1007/s00190-014-0720-1.

——, H. Harper, B. Tozer, and W. H. Smith, 2019: Gravity field recovery from geodetic altimeter missions. *Adv. Space Res.*, **68**, 1059–1072, https://doi.org/10.1016/j.asr.2019.09.011.

——, and Coauthors, 2022: Improved bathymetric prediction using geological information: SYNBATH. *Earth Space Sci.*, **9**, e2021EA002069, https://doi.org/10.1029/2021EA002069.

Saraceno, M., C. Provost, and U. Zajaczkovski, 2009: Long-term variation in the anticyclonic ocean circulation over the Zapiola Rise as observed by satellite altimetry: Evidence of possible collapses. *Deep-Sea Res. I*, **56**, 1077–1092, https://doi.org/10.1016/j.dsr.2009.03.005.

Seton, M., and Coauthors, 2020: A global data set of present-day oceanic crustal age and seafloor spreading parameters. *Geochem. Geophys. Geosyst.*, **21**, e2020GC009214, https://doi.org/10.1029/2020GC009214.

Sinha, A., and R. Abernathey, 2021: Estimating ocean surface currents with machine learning. *Front. Mar. Sci.*, **8**, 612, https://doi.org/10.3389/fmars.2021.672477.

Smith, W. H., and D. T. Sandwell, 1997: Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, **277**, 1956–1962, https://doi.org/10.1126/science.277.5334.1956.

Sonnewald, M., R. Lguensat, D. C. Jones, P. Dueben, J. Brajard, and V. Balaji, 2021: Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environ. Res. Lett.*, **16**, 16, https://doi.org/10.1088/1748-9326/ac0eb0.

St. Laurent, L., and C. Garrett, 2002: The role of internal tides in mixing the deep ocean. *J. Phys. Oceanogr.*, **32**, 2882–2899, https://doi.org/10.1175/1520-0485(2002)032<2882:TROITI>2.0.CO;2.

Straume, E. O., and Coauthors, 2019: GlobSed: Updated total sediment thickness in the world's oceans. *Geochem. Geophys. Geosyst.*, **20**, 1756–1772, https://doi.org/10.1029/2018GC008115.

Taburet, G., A. Sanchez-Roman, M. Ballarotta, M. I. Pujol, J. F. Legeais, F. Fournier, Y. Faugere, and G. Dibarboure, 2019: DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci.*, **15**, 1207–1224, https://doi.org/10.5194/os-15-1207-2019.

Tozer, B., D. T. Sandwell, W. H. Smith, C. Olson, J. R. Beale, and P. Wessel, 2019: Global bathymetry and topography at 15 arc sec: SRTM15+. *Earth Space Sci.*, **6**, 1847–1864, https://doi.org/10.1029/2019EA000658.

Wessel, P., W. H. Smith, R. Scharroo, J. Luis, and F. Wobbe, 2013: Generic mapping tools: Improved version released. *Eos, Trans. Amer. Geophys. Union*, **94**, 409–410, https://doi.org/10.1002/2013EO450001.

Wölfl, A. C., and Coauthors, 2019: Seafloor mapping—The challenge of a truly global ocean bathymetry. *Front. Mar. Sci.*, **6**, 283, https://doi.org/10.3389/fmars.2019.00283.

Zaron, E. D., and R. DeCarvalho, 2016: Identification and reduction of retracker-related noise in altimeter-derived sea surface height measurements. *J. Atmos. Oceanic Technol.*, **33**, 201–210, https://doi.org/10.1175/JTECH-D-15-0164.1.

Zhang, S., and D. T. Sandwell, 2017: Retracking of SARAL/AltiKa radar altimetry waveforms for optimal gravity field recovery. *Mar. Geod.*, **40**, 40–56, https://doi.org/10.1080/01490419.2016.1265032.

Zhang, Z., and B. Qiu, 2018: Evolution of submesoscale ageostrophic motions through the life cycle of oceanic mesoscale eddies. *Geophys. Res. Lett.*, **45**, 11–847, https://doi.org/10.1029/2018GL080399.