Reconstructing ocean subsurface salinity at high resolution using a machine learning approach

Tian Tian[1,2], Lijing Cheng[2,3], Gongjie Wang[4], John Abraham[5], Wangxu Wei[2], Shihe Ren[6], Jiang Zhu[2], Junqiang Song[1], Hongze Leng[1]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha, 410073, China
[2]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[3]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, 266071, China
[4]National Climate Center, Chinese Meteorological Administration, Beijing, 100081, China
[5]School of Engineering, University of St. Thomas, Minneapolis, 55105, MN, U.S.A
[6]Key Laboratory of Research on Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Ministry of Natural Resources, Beijing, 100081, China

Correspondence to: Lijing Cheng (chenglij@mail.iap.ac.cn)

Abstract. A gridded ocean subsurface salinity dataset with global coverage is useful for research on climate change and its variability. Here, we explore the feed-forward neural network (FFNN) approach to reconstruct a high-resolution (0.25° × 0.25°) ocean subsurface (1–2000 m) salinity dataset for the period 1993–2018 by merging *in situ* salinity profile observations with high-resolution (0.25° × 0.25°) satellite remote sensing altimetry absolute dynamic topography (ADT), sea surface temperature (SST), sea surface wind (SSW) field data, and a coarse resolution (1° × 1°) gridded salinity product. We show that the FFNN can effectively transfer small-scale spatial variations in ADT, SST and SSW fields into the 0.25° × 0.25° salinity field. The root-mean-square error (RMSE) can be reduced by ~11% on a global-average basis compared with the 1° × 1° salinity gridded field. The reduction in RMSE is much larger in the upper ocean than the deep ocean, because of stronger mesoscale variations in the upper layers. Besides, the new 0.25° × 0.25° reconstruction shows more realistic spatial signals in the regions with strong mesoscale variations, e.g., the Gulf Stream, Kuroshio, and Antarctic Circumpolar Current regions, than the 1° × 1° resolution product, indicating the efficiency of the machine learning approach in bringing satellite observations together with *in situ* observations. The large-scale salinity patterns from 0.25° × 0.25° data are consistent with the 1° × 1° gridded salinity field, suggesting the persistence of the large-scale signals in the high-resolution reconstruction. The successful application of machine learning in this study provides an alternative approach for ocean and climate data reconstruction that can complement the existing data assimilation and objective analysis methods. The reconstructed IAP0.25° dataset is freely available at http://doi.org/10.57760/sciencedb.o00122.00001 (Tian et al., 2022).

1 Introduction

Gridded ocean datasets with complete global ocean coverage are of great importance to marine and climate research(Lyman and Johnson, 2014; Durack et al., 2014; Ciais et al., 2013; Domingues et al., 2008; Bagnell and DeVries, 2021). For example, most climate monitoring applications depend on gridded products (Abram et al., 2019; Ishii et al., 2017; Liang et al., 2021). However, *in situ* salinity observation data are sparse owing to the limitations of observation techniques, which brings difficulties to the generation of ocean salinity data products (Roemmich et al., 2019, 2009).

Currently, the estimation of salinity fields mainly relies on two approaches: (1) objective analysis methods based on *in situ* observations, resulting in so-called objective analysis data products (Gaillard et al., 2016; Hosoda et al., 2008; Roemmich and Gilson, 2009; Cheng and Zhu, 2016; Lu et al., 2020); and (2) data assimilation methods, which combine numerical simulations and observations to result in what is referred to as a reanalysis product (Carton et al., 2018; Forget et al., 2015; Jean-Michel et al., 2021; Balmaseda et al., 2013). The accuracy of the more traditional objective analysis approach is critically dependent on the data coverage and the reliability of spatial covariance, which defines how the information is propagated from data-rich to data-sparse regions (Von Schuckmann et al., 2014; Zhou et al., 2004). Previously available objective analysis products mostly have a horizontal resolution of 1° × 1°. The reanalysis approach relies on model simulations, which uses data assimilation schemes to constrain models with various types of observations, such as *in situ* and satellite remote sensing data (Storto et al., 2019; Jean-Michel et al., 2021). Reanalysis product can be strongly impacted by model biases, especially below the ocean surface (Storto et al., 2019; Balmaseda et al., 2015). . However, Palmer et al. (2017) and Cheng et al. (2020) indicated that such reanalysis products have much larger spread than observational products for ocean heat content (temperature) and ocean salinity, suggesting caution when adopting the data assimilation approach in some applications such as long-term climate change.

As higher resolution observational data are crucial for evaluating models/reanalysis and understanding ocean processes at multiple scales, such as the meso- and sub-mesoscale (McWilliams, 2016), high-quality observational datasets with resolutions higher than 1° × 1° could be useful for ocean and climate research, e.g., resolving the vertical structure of warm and cold eddies. To reconstruct an observational dataset with resolution higher than 1° × 1°, remote sensing observations are essential, as they can be used to incorporate smaller-scale signals that are insufficiently sampled by *in situ* salinity profile observations. Two types of approaches have been used previously to propagate surface information to the subsurface: dynamical and statistical. The dynamical approach adopts physical relationships between surface and subsurface variables. For example, Wang et al. (2013) proposed an isQG (interior + surface quasi geostrophy) method, using information such as sea surface height (SSH), sea surface temperature (SST), and sea surface salinity (SSS) to invert the density and velocity fields of the ocean interior. The statistical approach utilizes historical data to establish a statistical relationship between remote sensing observations and subsurface observations (e.g., linear regression between surface height and subsurface temperature/salinity). Some of the available statistical relationships have been derived by multiple linear regression models (Warin, 2019), empirical orthogonal functions (Hannachi et al., 2007), and the Gravest Empirical Modes method (Liu et al., 2021). However, both

dynamic and statistical approaches have obvious limitations. The dynamic approach is overly dependent on physical
assumptions, which are always simplified such as relying on Surface Quasi-Geostrophic dynamics to derive subsurface signals
from surface changes. Caveats of the statistical approach is the lack of physical constraints and the simplified assumptions.
For example, some approaches have simplified the nonlinear relationship between surface dynamic height and subsurface
temperature/salinity to a linear relationship.

Compared with traditional reconstruction methods, machine learning approaches do not rely on any simplified assumptions in
the process of data reconstruction, and are able to learn and fit parameters automatically. There have been some attempts
recently, for example, Wang et al. (2021) used remote sensing data and neural network methods to estimate the subsurface
temperature in the western Pacific. Combining satellite data and Argo gridded products, Su et al. (2020) used neural networks
to estimate ocean heat content anomalies over four different depths down to 2000 m. Lu et al. (2019) estimated the subsurface
temperature through a clustering neural network method based on SST, SSH and wind field measurement data. Although these
studies provide some hints that machine learning approaches can be useful in data reconstruction applications, there are still
some limitations. First, some studies (Lu et al., 2019; Su et al., 2020; Wang et al., 2021) used Argo gridded data as the "truth"
to train the machine learning model, thus the reconstruction error in the Argo gridded data is embedded in the final
reconstruction. Second, the spatial resolution of the most reconstructed data is still 1° × 1°. The added value of high-resolution
remote sensing data is not maximized; for example, the resultant 1° × 1° field is still insufficient to resolve mesoscale signals.
Finally, previous reconstructions have tended to focus mainly on temperature rather than salinity, and a comprehensive
assessment of the uncertainty has always been absent. Besides, due to the lack of interpretability of machine learning approach,
a thorough comparative evaluation with traditional methods is clearly a crucial line of study.

This paper explores the feed-forward neural network (FFNN) approach to reconstruct a high-resolution (0.25° × 0.25°) ocean
subsurface (1–2000 m) salinity dataset for the period 1993–2018 by merging *in situ* profile observations (processed to a gridded
0.25° × 0.25° arithmetic mean field in this study, detailed in the following text) with high-resolution satellite remote sensing
altimetry absolute dynamic topography (ADT), SST, sea surface wind (SSW) data, which included zonal (USSW) and
meridional (VSSW) components, and a coarse resolution IAP1° gridded salinity product from Institute of Atmospheric Physics
(IAP). The first objective is to understand how well the FFNN approach works for data reconstruction using the ocean salinity
field as an example. Second, the new machine-learning-based high-resolution (0.25° × 0.25°) salinity dataset will be
comprehensively evaluated in this study, which facilitate its further applications. The added value of remote sensing data in a
high-resolution salinity reconstruction is demonstrated and discussed.

The rest of the paper is organized as follows: The data and methods employed in our study are presented in section 2. The
performance of the dataset in terms of geographical pattern, regional analysis, and overall reconstruction performance is
assessed in section 3. The uncertainty of the FFNN approach is examined in section 4. An analysis of the major climatic
patterns is conducted in section 5. Importance of each feature for the reconstruction is described in section 6. Data availability
is described in section 7. The results of the study are summarized and discussed in section 8.

2 Data and Methods

2.1 Input Data

2.1.1 Remote sensing and IAP1° observational data

100   The ADT data was extracted from the European Copernicus Marine Environment Monitoring Service (CMEMS). with altimeter data from several satellites (including Jason-3, Sentinel-3A, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, and ERS1/2).(Mertz et al., 2016). The dataset has a spatial resolution of 0.25° × 0.25° and a daily temporal resolution covering the period from January 1993 to December 2020.

The SST data was extracted from the daily Optimum Interpolation Sea Surface Temperature (OISST) v2.1 data, provided by
105   National Oceanic and Atmospheric Administration (NOAA). OISST is produced by interpolating SST observations from different sources, resulting in a smoothed field with complete global ocean coverage. The sources of data are satellite (Advanced Very High Resolution Radiometer) and *in situ* platforms (i.e., ships and buoys) (Banzon et al., 2016; Reynolds et al., 2007; Huang et al., 2021). Data are currently available from 1 September 1981 to the present day, with a spatial resolution of 0.25° × 0.25° and a daily temporal resolution.

110   The SSW data was extracted from the Cross-Calibrated Multi-Platform (CCMP) V2 monthly dataset (Wentz et al., 2016), provided by National Center for Atmospheric Research (NCAR). The processing of CCMP V2 now combines Version-7 Remote Sensing Systems radiometer wind speeds, QuikSCAT and ASCAT scatterometer wind vectors, moored buoy wind data, and ERA-Interim model wind fields using a variational analysis method to produce 0.25° × 0.25° gridded vector winds (Atlas et al., 2011). The dataset covers the period from July 1987 to April 2019.

115   The IAP1° salinity data give a global coverage of the oceans at a horizontal resolution of 1° × 1° in 41 vertical levels from 1 m to 2000 m, and a monthly temporal resolution from 1940 to the present day (Cheng and Zhu, 2016; Cheng et al., 2017). This product combines *in situ* salinity profiles with coupled model simulations (from phase 5 of the Coupled Model Intercomparison Project) to derive an objective analysis with the Ensemble Optimal Interpolation approach. (Cheng et al., 2020). The product is designed to minimize the sampling error in representing large-scale and long-term climate changes and variabilities.

120   All of the above-mentioned products were processed into monthly averages, and IAP1° data were linearly interpolated to unified 0.25° × 0.25° resolution fields, which were used as inputs for the FFNN approach.

2.1.2 *In situ* salinity observations

In situ ocean salinity observations were sourced from the World Ocean Database (WOD) (Boyer et al., 2018). Data from all available instruments [i.e., Argo (Li et al., 2017), Bottle, conductivity–temperature–depth (CTD)] were used in this study.
125   Quality flags from WOD were applied (only "good data" with a flag equal to zero were used). All of the profiles were first interpolated to a standard 41 vertical levels between 1 and 2000 meters, as in the IAP1° data. Then, 12 monthly climatologies were constructed using all data from 1990 to 2010 (centered around 2000), and anomaly profiles are derived by subtracting

4

monthly climatologies from salinity profiles. Finally, the salinity anomalies were averaged into 0.25° × 0.25°, 1-month, and 41-level grid boxes via simple arithmetic averaging, without spatial interpolation and smoothing. The reconstructions are
130 applied to the anomaly fields, similar to previous objective analyses, because of the larger spatial decorrelation length scale of the anomaly fields than the absolute fields (i.e., Levitus et al. 2009; Cheng et al. 2017). The gridded averaged salinity anomalies were regarded as the "truth" to train the FFNN approach during the reconstruction. The gridded averages were used instead of the raw profiles because they were able to minimize the impact of spatial heterogeneity of the raw profiles, reduce the subgrid (<0.25°) variability and observational noises, and improve the computational efficiency (Cheng et al., 2017, 2020; Gaillard et
135 al., 2016; Good et al., 2013; Ishii et al., 2003, 2017; Levitus et al., 2012; Lyman and Johnson, 2014; Roemmich et al., 2009).

2.2 Independent Ocean products for evaluation and comparison

Three independent ocean products were used to assess the quality of the newly reconstructed high-resolution salinity data in this study (hereafter denoted as IAP0.25°), ARMOR3D data, SMAP satellite data, and EN4 gridded data.

The ARMOR3D V4 dataset was extracted from CMEMS, which is a global three-dimensional temperature and salinity dataset
140 with high horizonal (0.25° × 0.25°) spatial resolution from 0–5500 m (Guinehut et al., 2012; Mulet et al., 2012). The spatial coverage ranges from 82.125°S to 89.875°N and 0.125°E to 359.875°E. The ARMOR3D combines satellite observation data (ADT, geostrophic surface currents, SST) and *in situ* temperature and salinity profile data through a multivariate/simple linear regression and an optimal interpolation method. This product offers four versions with different temporal resolutions: near-real-time weekly data, near-real-time monthly data, multi-year reprocessed weekly data, and multi-year reprocessed monthly
145 data. The multi-year reprocessed monthly data were used in this study.

The SMAP V4 dataset is provided by NOAA, and comprises monthly 0.25° × 0.25° resolution SSS data. The SMAP data are based on satellite observations, and have the capability to monitor the global SSS field in near real time, with mesoscale information resolved for salinity (Vinogradova et al., 2019).

The EN4 gridded data were obtained from the UK Met Office. The EN4 dataset uses an optimal interpolation method to
150 construct the ocean subsurface gridded salinity field based on *in situ* salinity profile data. We chose the EN4-GR10 version (Gouretski and Reseghetti, 2010), which has a spatial resolution of 1° × 1° and 42 vertical levels from 5 to 5500 m, and temporal coverage from 1940 to 2018 (Li et al. 2019).

Table 1. A list of the input datasets and validation products used in this study.

Data type	Variable	Dataset	Data source	Horizontal resolution	Vertical coverage and resolution	Time period	Reference	DOI/URL
Input	ADT	CMEMS	CMEMS	0.25° × 0.25°	Sea surface	1993–2020	(Mertz et al., 2016)	https://doi.org/10.48670/moi-00148
Input	SST	OISST	NOAA	0.25° × 0.25°	Sea surface	1981–2022	(Huang et al., 2021)	https://www.ncei.noaa.gov/products/optimum-interpolation-sst
Input	SSW	CCMP	NCAR	0.25° × 0.25°	Sea surface	1987–2019	(Wentz et al., 2016)	https://doi.org/10.5065/4TSY-K140 /
Input	Salinity	IAP1°	IAP	1° × 1°	41 levels (1–2000 m)	1960–2021	(Cheng and Zhu, 2016)	http://www.ocean.iap.ac.cn/
Input	Salinity observations	*In situ* observations	WOD	Averaged into 0.25° × 0.25°	Interpolated to 41 levels (1–2000 m)	1960–2021	(Boyer et al., 2018)	https://www.ncei.noaa.gov/products/world-ocean-database
Validation	Salinity	ARMOR3D	CMEMS	0.25° × 0.25°	50 levels (1–5000 m)	1993–2020	(Mertz et al., 2016)	https ://doi.org/10.48670/moi-00052
Validation	SSS	SMAP	NOAA	0.25° × 0.25°	Sea surface	2015–2019	(Vinogradova et al., 2019)	https://data.remss.com/smap/SSS/V04.0/
Validation	Salinity	EN4	UK Met Office	1° × 1°	42 levels (1–5500 m)	1940–2018	(Gouretski and Reseghetti, 2010)	https://www.metoffice.gov.uk/hadobs/en4/

2.3 Methods

2.3.1 Feed-forward neural network

The FFNN was used to reconstruct the subsurface salinity anomalies in this study. We chose FFNN because it has been shown to be superior to the other three widely used machine learning approaches in reconstructing ocean parameters (Lu et al., 2019; Stamell et al., 2020; Wang et al., 2021). Our own evaluation based on synthetic data and salinity observations (see supplementary material) also reveals that FFNN is a robust approach and leads to the smallest error compared with other approaches [e.g., light gradient boosting machine (LightGBM)] (Gan et al., 2021). An FFNN is a one-way, multi-layer structure network that includes an input layer, hidden layers, and an output layer (Abdar et al., 2021; Contractor and Roughan, 2021; Gabella, 2021). The zero layer is called the input layer, the last layer is called the output layer, and the other intermediate layers are called the hidden layers. The neurons in the FFNN are arranged in layers, with each neuron belonging to a different layer. The neurons of each layer are fully connected; that is, each neuron is connected to the neurons of the previous layer and the next layer. The information in the network will only flow from the input layer, to the hidden layers, and then to the output layer; that is, the output of the previous layer is used as the input of the next layer, and the information of the next layer has no effect on the previous layer. Each layer of the FFNN is equivalent to a function, and the connection of the multi-layer network is equivalent to a composite function to form a linear or nonlinear mapping from input variables to output variables. The complexity of the FFNN depends not only on how many neurons or layers are chosen, but also on type of layers and activation functions.

All of the input parameters of the model training were converted to anomalies by subtracting their respective climatologies for 1993–2015. For example, the input data were the IAP1° salinity anomalies (IAP1SA), the absolute dynamic topography anomalies (ADTA), sea surface temperature anomalies (SSTA), and sea surface wind anomalies (SSWA), which included zonal and meridional anomalies (USSWA/VSSWA). The standardization of data can improve the convergence speed of the neural network method (LeCun et al., 2012). The method proposed by Denvil-Sommer et al. (2019) was used to perform sine/cosine transformations of latitude, longitude, and time. The Z-score standardization method was used to process the ADTA, SSTA, USSWA, VSSWA, IAP1SA, and standard layer depth parameters.

The structure of the FFNN used in this study is illustrated in Fig. 1, in which y is the gridded salinity anomalies are regarded as "truth values". The input x includes longitude, latitude, time, depth, IAP1SA, ADTA, SSTA, USSWA, VSSWA.

A training/testing approach was adopted to determine the key parameters and settings for the FFNN. The full salinity dataset was randomly divided into a training set (80% of the whole dataset) and a test set (20% of the whole dataset). The training set was used to fit the FFNN algorithm, and the test set was used for performance evaluation. The root-mean-square error (RMSE) was calculated between the reconstruction and the test dataset as a model evaluation metric. Once the RMSE begins to increase, the training iteration stops and the best parameters and settings are stored. A grid search strategy (Liashchynskyi et al. 2019) was used to optimize the structure of the neural network. The optimized neural network we used consists of one input layer,

one output layer, and four hidden layers; the number of neurons in each hidden layer was set to 256, 128, 64, and 32; the activation function was the Rectified Linear Unit; the optimizer was the root mean square propagation (RMSProp) (Vitaly Bushaev, 2018); the learning rate was 0.001;the cost function minimized for training as follows:

190
$$J(\theta_0, \theta_1, \ldots \theta_n) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(X^{(i)}) - y^{(i)})^2, \tag{1}$$

where X = (longitude, latitude, depth, time, IAP1SA, ADTA, SSTA, USSWA, VSSWA), y is the "truth values", m is the number of samples, h_θ is the FFNN model for training, θ is the parameter of the model, and the training objective is the minimum J.

Bayesian regularization (Foresee and Hagan. 1997) and dropout was used during model training to maintain generalization
195 and prevent overfitting. Since the Bayesian regularization algorithm does not require cross-validation to ensure generalization, no validation set was defined in this process (Lu et al., 2019; Wang et al., 2021). Adopting these settings and parameters, the final FFNN model used for reconstruction was trained by the full salinity dataset to ensure the best performance.
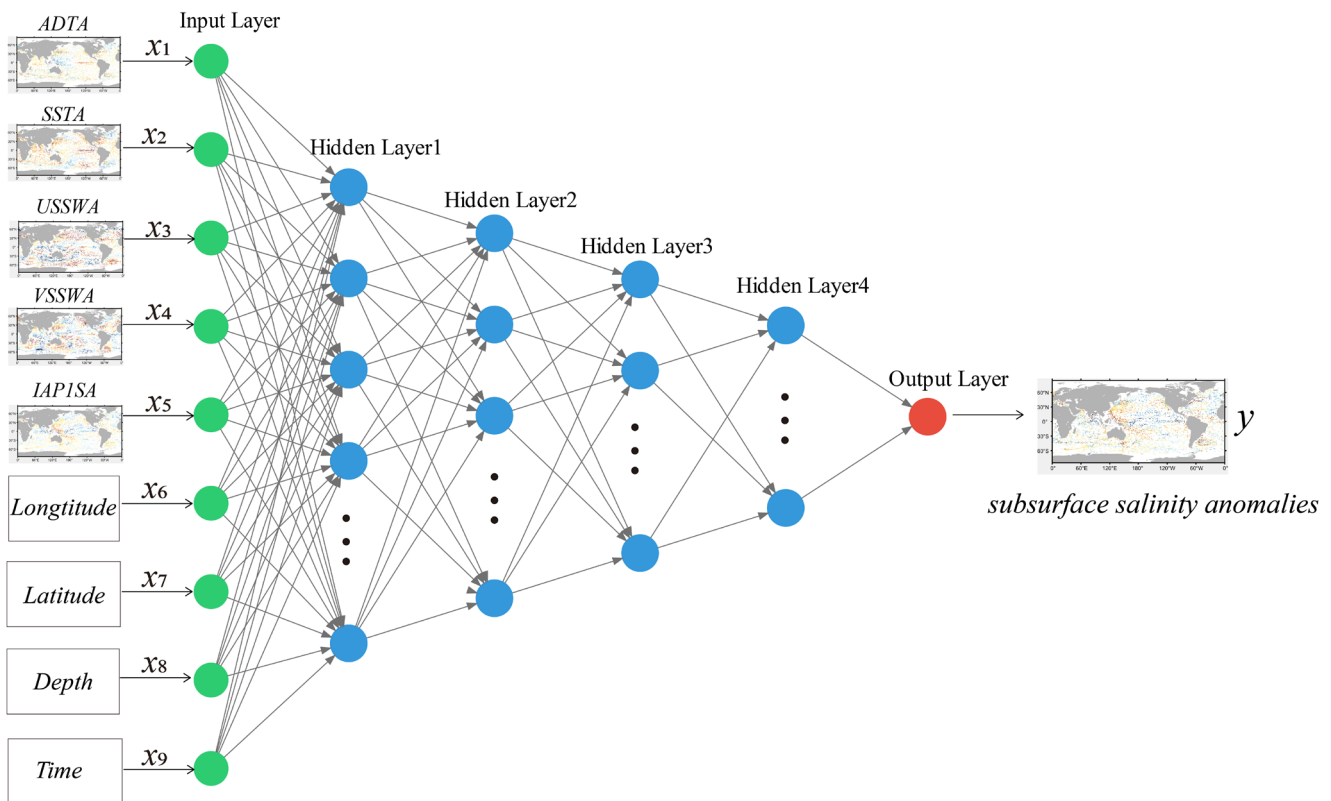

Figure 1: Schematic diagram of subsurface salinity reconstruction by the FFNN approach.

200 **2.3.2 Five-fold cross validation**

To evaluate the reconstruction using independent test data, a 5-fold cross validation approach was used (Gui et al., 2020). This type of validation is one of the most popular models in statistics and machine learning (Berrar, 2018; Lei, 2020). In the five-

8

fold cross validation, the full observational datasets were split into training (80%) and testing (20%) datasets. The training set was used to train the machine learning model and then to derive the reconstruction. The testing dataset was used as an independent dataset to verify the performance of the reconstruction. To effectively construct the training and testing datasets, the full observational data 0.25° × 0.25° gridded average fields, as introduced before, were randomly split into five subsets. In each run, four subsets were used for training, and the remaining subset was used for testing. This process was repeated four times so that each of the five subsets could be used as the testing dataset. In this way, the five-fold cross validation needed to be trained five times to ensure that all the data participated in both the training and testing. By calculating the difference between the reconstructed data and the truth value of the independent testing data, the machine learning method could be evaluated.

2.3.3 Uncertainty quantification

The uncertainty of the final reconstruction will stem from three sources: instrumental error, representativeness error, and reconstruction error. The instrumental error is related to the instrument accuracy of a salinity measurement. For instance, the accuracy of most Argo salinity data is about ±0.01 psu. However, many Argo floats suffer from sensor drift, so the error is much larger for some data, which has yet to be quantified (Wong et al., 2020). For the validated marine mammal data, the accuracy can be ±0.03 psu (Siegelman et al., 2019). The CTD data accuracy is about ±0.01 psu. The representativeness error defines the accuracy of the gridded average in representing the true average of salinity in this grid. This error is associated with the sampling of ocean changes inside each 0.25° and 1-month grid: insufficient sampling will lead to errors in the grid average. The magnitude of error depends on the spatiotemporal variability in each grid box: larger subgrid variability requires more data to accurately estimate the grid average. The reconstruction error defines the accuracy of the gap-filling, i.e., the machine learning approach in this study. This error is related to the errors in the FFNN model's design and parameter choice, etc. To quantify the uncertainty in the reconstruction, these three major error sources had to be properly accounted for.

This study adopted a probabilistic approach to the propagation of errors and to quantify the uncertainty. We perturbed the interpolated 0.25° gridded average fields five times with prescribed instrumental and representativeness uncertainty, and then obtained five perturbed gridded fields. The FFNN was then applied to these five perturbed fields, separately, to result in five reconstructions at each time with a Monte Carlo dropout approach (Gal and Ghahramani, 2016; Abdar et al., 2021). This process led to 25 ensemble members in the final reconstruction, the spread of which was used to quantify the uncertainty.

The perturbations were performed as follows. For instrumental error, we perturbed each individual salinity profile by its instrumental accuracy, assuming a Gaussian distribution with zero mean and the instrumental accuracy as the standard deviation. The perturbed profiles were then used to construct the gridded average fields (0.25° and 1-month grid). On this basis, the representativeness error was included by perturbing each gridded average by a Gaussian distribution with zero mean and standard deviation of $Var/\sqrt{n}$, where Var is a measure of subgrid variability (<0.25° and <1-month) and n is the number of profiles to calculate the grid mean. The $Variance$ is static (non-time-varying) and calculated by the standard deviation of the salinity data in each 1° grid box when there were >10 measurements. Grid boxes of 1° were used here because of the data

scarcity, which would overestimate the subgrid variability of <0.25°. However, this overestimation could compensate for the unresolved/unknown errors, such as salinity drift. The results are presented in Figs. 2a and c. The figures reveal incomplete ocean coverage. Thus, we interpolated this field using an objective interpolation method (Figs. 2b and d) (Cheng et al., 2017). A spatial decorrelation length scale of 8° was used, which is consistent with a previous gap-filling approach for ocean

240 temperature and salinity (Levitus et al., 2012). A global estimate of salinity *Var* in Fig. 2b and Fig. 2d shows that the coastal regions, boundary current systems, and Antarctic Circumpolar Current (ACC) regions have larger *Var* than other places, because of either river runoff or mesoscale and sub-mesoscale variabilities. The estimated *Var* and the number of data *n* in each 0.25° and 1-month grid box were combined to perturb the local 0.25° and 1-month grid mean assuming a Gaussian distribution, which resulted in five perturbed 0.25° and 1-month gridded average fields and inputs into the FFNN approach.

245 The Monte Carlo dropout approach was employed to estimate the uncertainty due to the FFNN approach. This is one of the most popular ways to estimate machine learning method uncertainty, and it does not require any change to the existing model architecture (Stoean et al., 2020; Milanés-Hermosilla et al., 2021). Monte Carlo dropout was operated by randomly dropping some units of the neural network during the model training process. In this way, the approach and parameter uncertainty could be accounted for. In this study, we applied the FFNN to the five above-derived perturbed fields separately to create five

250 reconstructions each time with the Monte Carlo dropout approach. Then, the spread (i.e., standard deviation) of the $5 \times 5 = 25$ final reconstructions was used to quantify the total uncertainty. With this approach, all major error sources were propagated to the final fields, where the Monte Carlo dropout approach takes into account the reconstruction uncertainty of the FFNN, and the five perturbed gridded fields represent the instrumental and representativeness uncertainty.

255

10
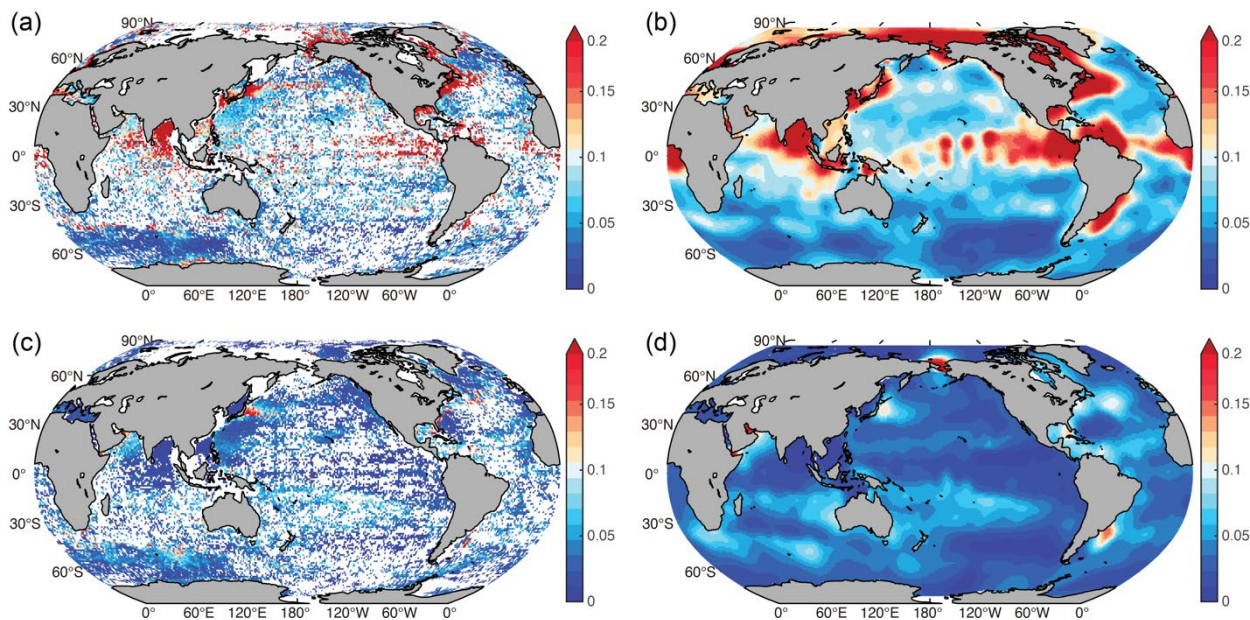
Figure 2: The estimated *Var* (a quantification of salinity subgrid variability) of *in situ* observations in this study. The original (left) and objectively analyzed fields (right) are presented for the depths of 20 m (upper panel) and 300 m (lower panel), respectively, as two examples.

2.3.4 Evaluating the relative importance of different inputs

260 In this paper, we used an approach named Kernel SHAPley Additive exPlanations (SHAP) to evaluate the contribution of different input parameters (Lundberg and Lee, 2017; Lundberg, 2019).

Shap is a method inspired by game theory to explain contribution of each feature on the model output. It works for any model and is especially useful for interpreting black box models (e.g., FFNN). It approximates the original model with a sum of linear terms. Similar to linear regression model, each term is contribution of corresponding feature on model output. To compute the

265 linear terms, combinations of features are examined. Assuming a total of p features. For a given combination of k features out of the original p, feature j is dropped and added back to the combination. The change of model performance is marginal contribution of the feature j. Repeat the same process for all combinations of features from k=1 to p. The aggregated marginal contribution over all combinations is contribution of feature j on model output.

With this approach, SHAP can quantify the average impact of an input on the final output (reconstruction in our case). The

270 change in the output is representative of the importance of the input for predicting the output, which is called SHAP value. By comparing the SHAP value for each input, the relative contribution to the final reconstruction can be assessed.

To implement SHAP, the Kernel SHAP algorithm was employed, which makes no additional assumption about the model type (e.g., linear models, tree models and deep network models). The disadvantage of the SHAP algorithm is that it is slower than other model type specific algorithms. The SHAP algorithm is too computationally expensive to apply for the full dataset (Chau

275 et al., 2021). Pauthenet et al. (2022) indicated that ~0.44% of the total samples is sufficient to obtain stable results for ocean temperature and salinity reconstruction in the Gulf Stream region. Therefore, we follow their choice and randomly selected 0.5% of data to calculate the Shapley value for each input parameter (expanding it to 1% did not make significant difference based on our test). The input importance of each input is estimated by the average of absolute Shapley values for each input, which is then normalized by the sum of the absolute Shapley values to derive the relative importance of each input.

280 ## 3 Reconstruction results

This section presents the reconstruction results and discusses their reliability using multiple examples and thorough statistical analyses.

3.1 Reconstruction of the geographical pattern

We first examine the geographical distribution of the reconstructions for January 2016 as an example. January 2016 was chosen

285 arbitrarily for illustration purposes. Using other times yielded similar results. Figs. 3 and 4 show the salinity anomalies of the new reconstruction, IAP0.25°, in comparison with the IAP1°, ARMOR3D, ADTA, SSTA and *in situ* observations of January

2016 at 100 m and 300 m, respectively. It appears that the large-scale pattern of the IAP0.25° salinity field is closely consistent with the distribution of IAP1° and ARMOR3D, as revealed by *in situ* data. There are significant negative salinity anomalies in the tropical Pacific, Intertropical Convergence Zone (ITCZ), and the Pacific north of the equator, as well as positive salinity anomalies in the Southeast Pacific, Northwest Pacific, and the east coast of Australia. Compared with the very smooth field of IAP1°, the IAP0.25° salinity field includes more small-scale signals. Such information has to come from *in situ* data, ADT/SST and USSW/VSSW fields, as they have finer resolution. The previous IAP1° salinity data mainly reveal large-scale patterns and there are almost no eddies in the Gulf Stream and the Kuroshio extension regions. The comparison between IAP0.25° and IAP1° reveals a bigger difference in boundary currents and ACC regions with rich eddies.
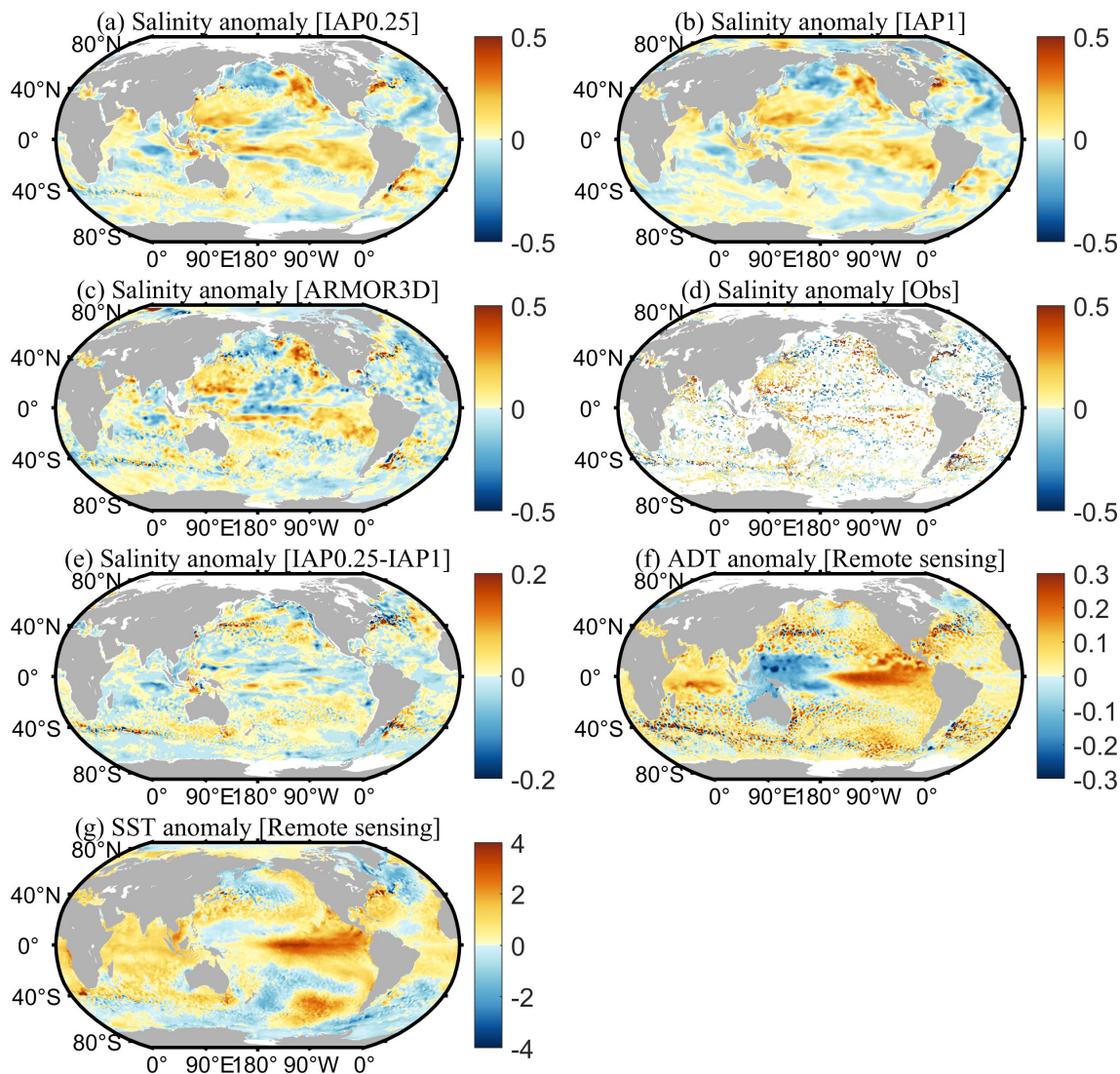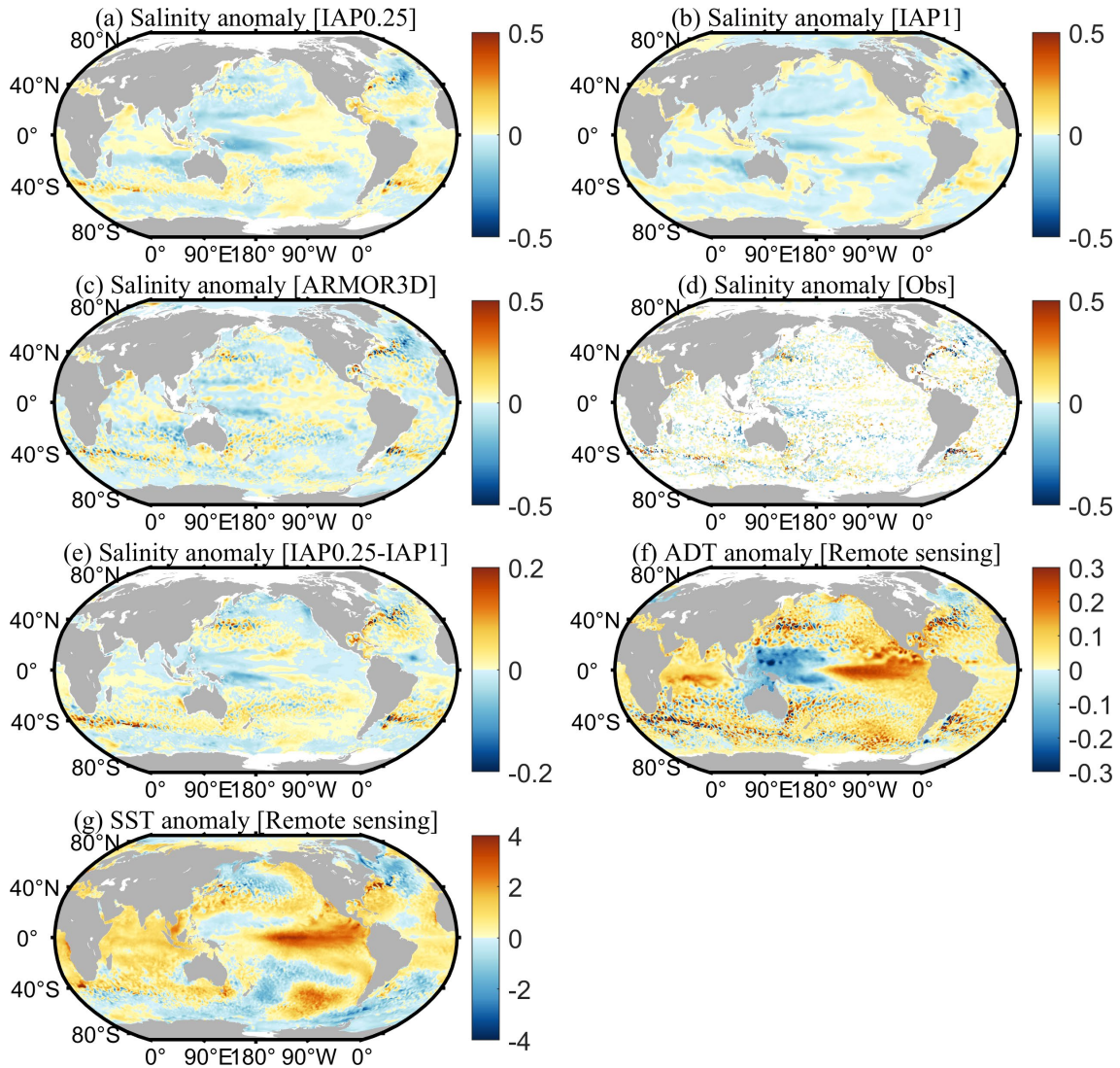

Figure 3: Spatial distribution of salinity anomalies of (a) IAP0.25°, (b) IAP1°, (c) ARMOR3D, (d) *in situ* observations, and (e) IAP1° minus IAP0.25° at 100 m, as well as the spatial distribution of (f) ADTA and (g) SSTA in January 2016.
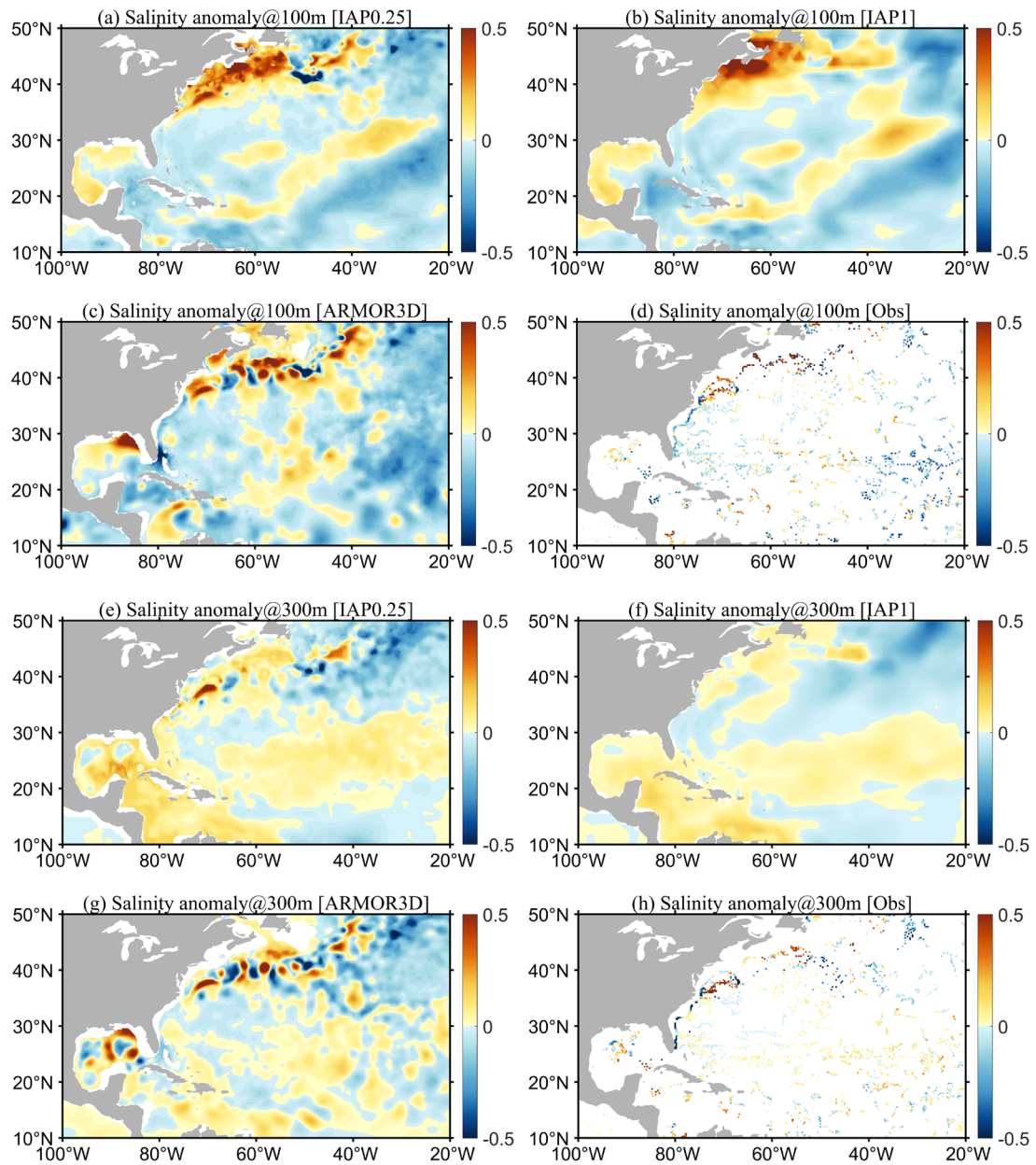
Figure 4: Spatial distribution of salinity anomalies of (a) IAP0.25°, (b) IAP1°, (c) ARMOR3D, (d) *in situ* observations, and (e) IAP1° minus IAP0.25° at 300 m, as well as the spatial distribution of (f) ADTA and (g) SSTA in January 2016.

To get a closer look at these critical regions, we concentrate on the salinity reconstruction in the boundary currents regions, e.g., the Gulf Stream (Fig. 5) and the Kuroshio Extension (Fig. 6). Both of these areas are active with mesoscale eddies (Chassignet et al., 2020; Cheng et al., 2014; Frenger et al., 2015; Rhines, 2019), and with very strong net energy growth and net dissipation of mesoscale eddies (Xu et al., 2013).

In the Gulf Stream region (Fig. 5), at a depth of 100 m, IAP1°, IAP0.25°, and ARMOR3D all reveal large-scale positive salinity anomalies along the North Atlantic coastal regions, negative anomalies in the middle and east, and positive anomalies extending from southwest (80°W, 15°N) to the northeast (50°W, 30°N). The 300 m salinity shows similar patterns as that at

13

100 m, except for larger areas of positive anomalies from southwest to northeast. The similarity among the large-scale patterns

310 yielded by the datasets suggests the large-scale reconstruction is reliable. However, there are obvious differences between the four datasets in their ability to describe the fine-scale structures. For example, at 100 m, the reconstructed (IAP0.25°) field and ARMOR3D show more detail in the salinity structure, especially near the boundary currents and coastal regions, compared to IAP1°. These anomalies are more consistent with the *in situ* salinity observations (Figs. 5a–c vs. 5d; Figs. 5e–g vs. 5h).

In the Northwest Pacific Ocean (Fig. 6), IAP1°, IAP0.25°, and ARMOR3D indicate a large area of positive salinity anomalies

315 in the Northwest Pacific south of 40°N at a depth of 100 m, and negative anomalies north of 40°N. A very smooth transition zone along ~40°N is apparent for IAP1° (Fig. 6b), but IAP0.25° and ARMOR3D suggest a zone with rich mesoscale variabilities (Figs. 6a and c). These findings are more physically plausible because it is the Kuroshio extension region (Chassignet et al., 2020). Even down to 300 m, there are still rich mesoscale structures seen in the observations within 30°–40°N (Fig. 6h), as represented by both IAP0.25° and ARMOR3D.

320 A closer look into the two regions suggests that the new IAP0.25° data are capable of resolving the mesoscale signals in the boundary flow area. However, it also appears that the small-scale anomalies are stronger in the ARMOR3D data than the IAP0.25° data, which is likely associated with the methodological differences. A comparison between *in situ* observations and IAP0.25° (Fig. 5a vs. 5c; 5e vs. 5h; 6a vs. 6c; 6e vs. 6h) suggests a closer correspondence of IAP0.25° data with observations than ARMOR3D, and a more thorough statistical analysis of the errors will be presented in subsequent sections.

325

Figure 5: The geographical distribution of salinity anomalies in the Gulf Stream region for IAP1°, IAP0.25°, ARMOR3D, and *in situ* observations in January 2016: (a–d) 100 m; (e–f) 300 m.
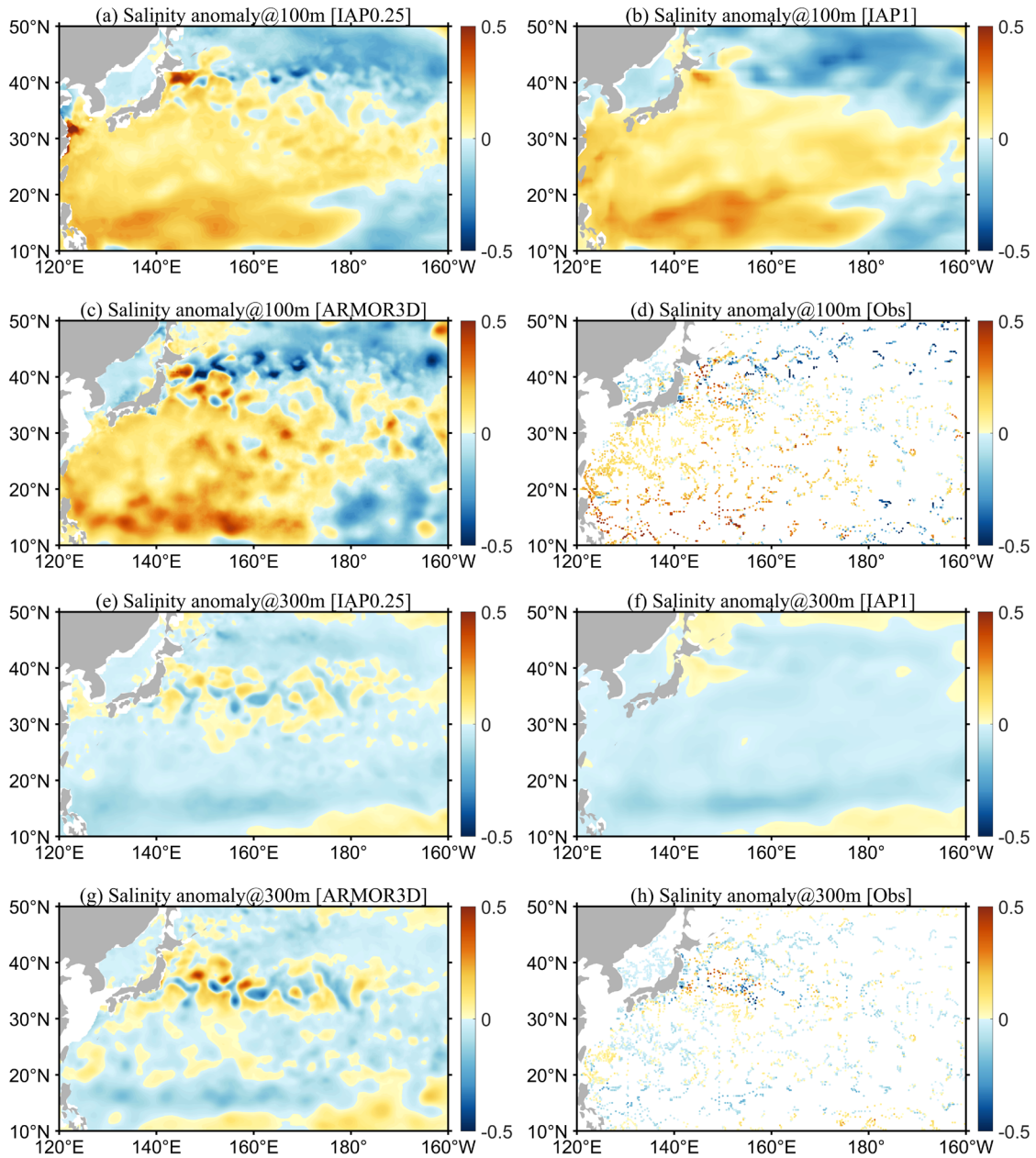
Figure 6: The geographical distribution of salinity anomalies in the Kuroshio and its extension region for IAP1°, IAP0.25°, ARMOR3D, and *in situ* observations in January 2016: (a–d) 100 m; (e–f) 300 m.

3.2 Reconstruction of the vertical structure

In addition to the spatial distribution at a given layer, it is desirable to investigate the vertical structure of the salinity field (i.e., to what extent the vertical structure can be restored by the new approach). This section seeks to address this question with two examples: in the Gulf Stream region (along 35.375°N), and in the Kuroshio extension region (along 35.375°N) (Figs. 7 and 8).

335 Along the Gulf Stream section, satellite remote sensing SSS data (Fig. 7a), independent from our analysis, suggest a general negative salinity anomaly associated with substantial variability. This reveals the existence of both large-scale change and mesoscale variability. It can be seen from Figs. 7b and d that both the IAP0.25° and IAP1° anomalies are generally negative near the sea surface, indicating the near-surface signals can be restored with IAP0.25°. The smoothness of the IAP1° data is associated with its mapping method (Cheng and Zhu, 2016), which focuses more on the large-scale patterns and effectively

340 filters out small-scale variations. Moreover, the small-scale, along-section fluctuations revealed by the *in situ* profile data (Fig. 7c) are well reflected in the IAP0.25° (Fig. 7d), but not in the IAP1° data (Fig. 7b), highlighting that IAP0.25° is capable of correctly reflecting the small-scale features. ADT data reflect the integrated effect or thermosteric (linked to temperature) and halosteric (linked to salinity) effects over the full volume (Llovel and Lee, 2015; Wang et al., 2017). To the first order, in the thermocline regions, the ADT change should correspond better with thermocline change (first baroclinic mode), thus the

345 salinity change near the thermocline should resemble the ADT in many places. This is supported by Fig. 7 (red line in Fig. 7a vs. 7d); for example, the large positive anomaly near 48°W, 60°W, and 68°W revealed by both the *in situ* salinity profile and ADT data. This indicates a positive contribution of ADT to the reconstruction.

Along the Kuroshio extension section (Fig. 8), the SST and ADT are more consistent with each other than in the Gulf Stream (Fig. 7). There is a deep mixed layer (deeper than 500 m) in this region, so ADT is more influenced by mixed-layer dynamics.

350 In this case, both ADT and SST could support the high-resolution salinity reconstruction because mixed-layer dynamics are associated with strong air–sea heat and freshwater exchanges. For example, the high ADT, SST, and salinity anomalies at 155°–160°E, around 180°E, and negative anomalies at 178°E and other places in the upper 500 m are all consistent (Figs. 8a, c, d).

In summary, the investigation of these two specific cases increases the degree of confidence in the reconstruction. It seems

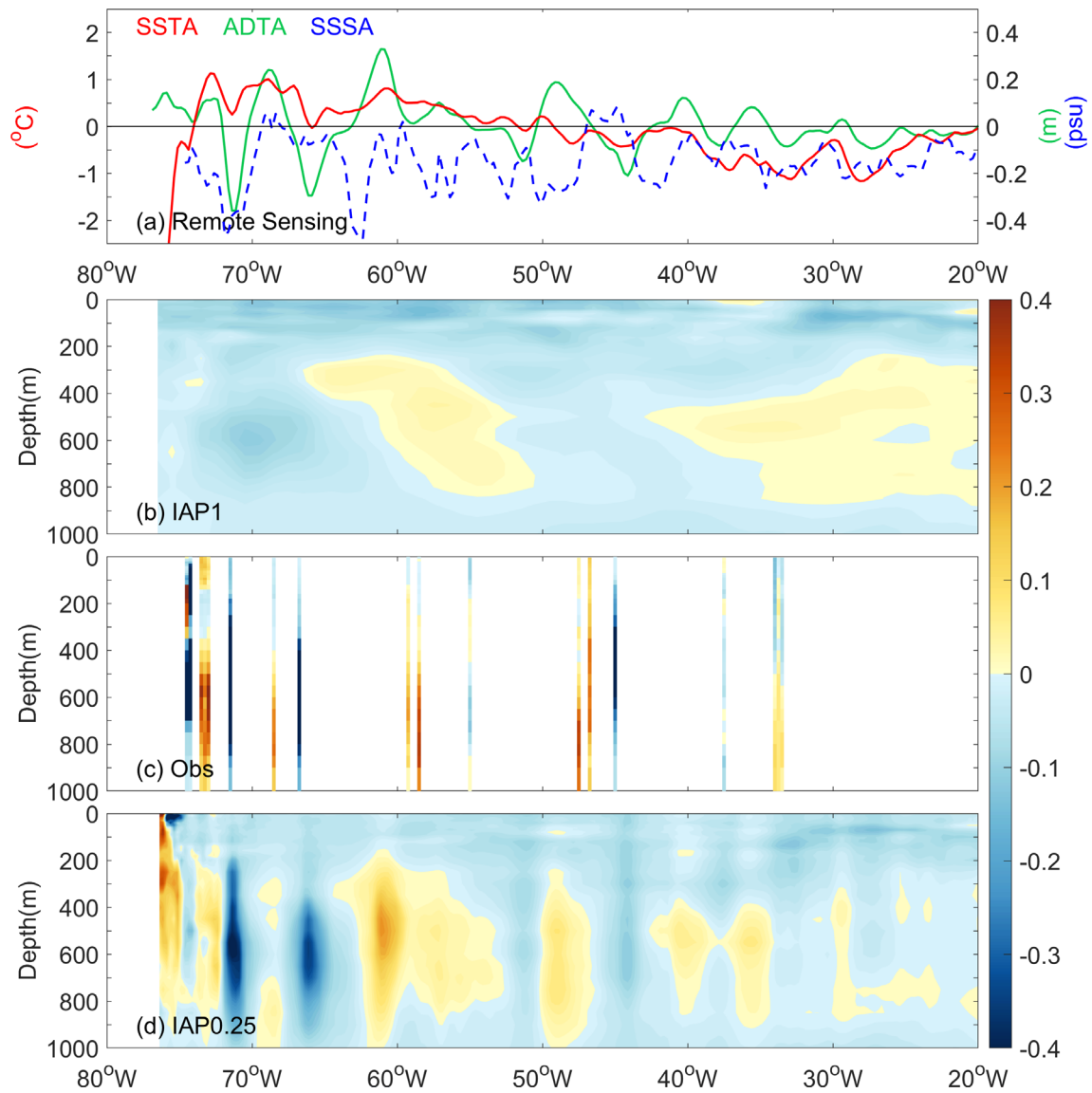355 that mesoscale structures can be realistically restored.

Figure 7: The distribution of SSTA (left *y*-axis), ADTA, and SSSA observed by SMAP on the sea surface (ADTA and SSTA share the right *y*-axis), as well as the vertical section along 35.375°N of the observational *in situ* salinity anomalies, subsurface salinity anomalies from IAP1°, and IAP0.25° gridded data in the Gulf Stream region.
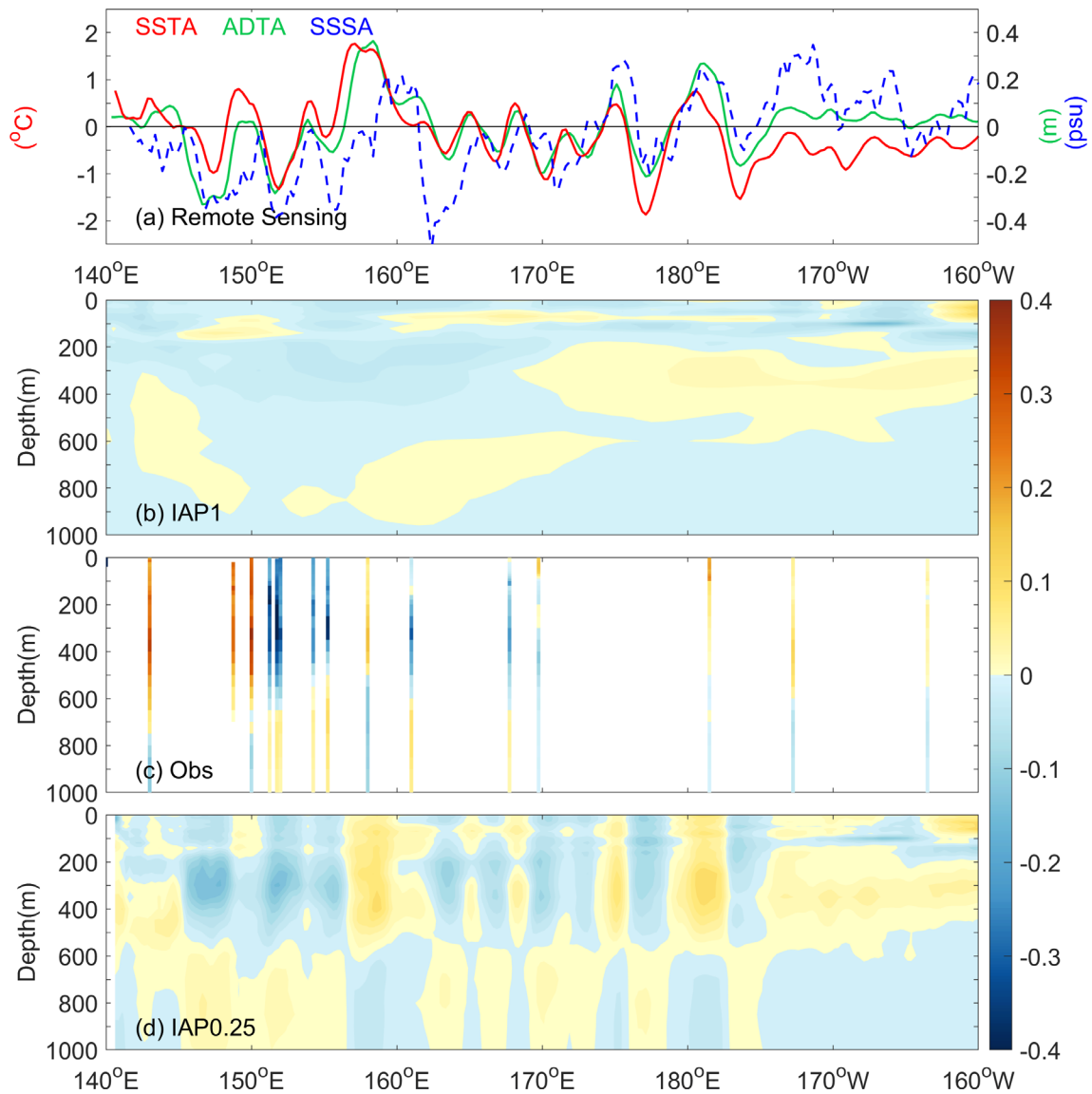
Figure 8: As in Fig. 7, but for the Kuroshio and its extension region along 35.375°N in January 2016.

3.3 Overall reconstruction performance

365 In the previous two sections, examples were presented to illustrate the reconstruction performance. However, a more thorough statistical examination is still required, which is provided in this section. To quantify the overall error compared with observations, we calculated the RMSE between IAP0.25°, IAP1°, ARMOR3D, EN4, and the observation data, respectively (denoted as IAP0.25_RMSE, IAP1_RMSE, ARM_RMSE, and EN4_RMSE), and the results are presented in Figs. 9a–d at different layers from the surface down to 2000 m over the globe and in three major basin regions during 1993–2018. Figs 9e–

370 h shows the differences between IAP1_RMSE, ARM_RMSE, EN4_RMSE, and IAP0.25_RMSE. First, IAP0.25_RMSE is

smaller than all other data products for both global- and basinal -average. At a global scale, the maximum values are 0.37 psu for IAP0.25_RMSE, 0.50 psu for IAP1_RMSE, 0.55 psu for ARM_RMSE, and 0.56 psu for EN4_RMSE near the sea surface. Globally, although a reduction in RMSE can be seen from the surface down to at least 1000 m, the upper 1–200 m shows a more significant reduction for IAP0.25_RMSE compared with other data, consistent with more mesoscale variability in the

375 upper ocean. Globally, the 1–2000 m mean IAP0.25_RMSE is 0.016 psu, 0.019 psu and 0.021 psu lower than IAP1_RMSE, ARM_RMSE and EN4_RMSE, respectively (Table 2). The smallest RMSE is apparent for IAP0.25° at all 41 vertical levels over the globe and in the three major basins, suggesting that IAP0.25° best represents the *in situ* observations.

Besides, the reduction in IAP0.25_RMSE is more remarkable in the Atlantic Ocean than in the Pacific and Indian oceans. A previous dataset intercomparison study suggested that the Atlantic Ocean shows larger spread among different data products,

380 even with more observations (Frederikse et al., 2018). This larger error reduction in the Atlantic basin for IAP0.25° suggests that signals smaller than 1° might play an important role for a reliable reconstruction.
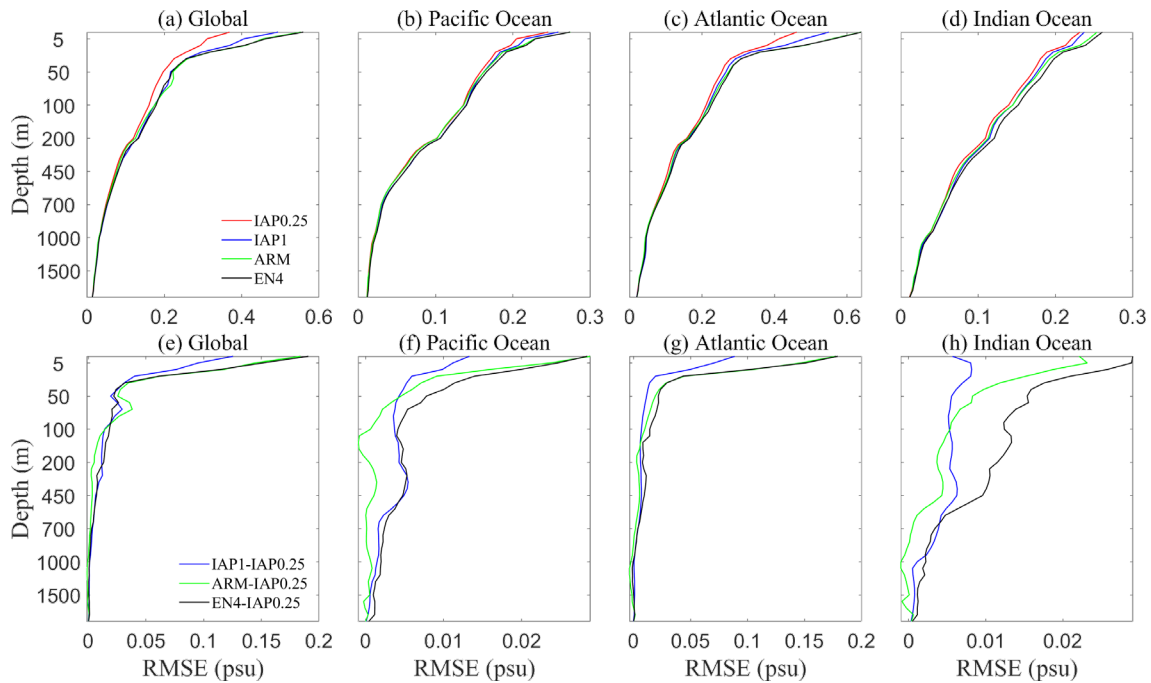

Figure 9: (a–d) Vertical distribution of RMSE for IAP1°, ARMOR3D, EN4, and IAP0.25° for the globe and three major basin regions during 1993–2018. (e–h) Vertical distribution of RMSE for IAP1°, ARMOR3D, and EN4 minus IAP0.25°, respectively.

385

Table 2. RMSE of 1–2000 m mean salinity for the IAP1°, ARMOR3D, EN4, and IAP0.25° datasets (unit: psu).

Region	IAP1_RMSE-IAP0.25_RMSE	ARM_RMSE-IAP0.25_RMSE	EN4_RMSE-IAP0.25_RMSE
Global	0.016	0.019	0.021
Pacific	0.004	0.003	0.006
Atlantic	0.010	0.016	0.019

Indian	0.004	0.005	0.009

Furthermore, to identify the causes of the most significant errors, the spatial distribution of RMSE for IAP0.25° and IAP1° minus IAP0.25° for three representative layers and 1–2000 m averages are presented in Fig. 10. Compared with IAP1°, the

390 reduction in RMSE for the IAP0.25° data is more dramatic in the western boundary currents, ACC region, and coastal regions (Figs. 10e–h). However, the RMSE for IAP0.25° remains larger in the western boundary currents, such as the Kuroshio, Gulf Stream, and Brazil Current regions, as well as the ACC regions, compared to the open ocean. This finding is consistent with stronger mesoscale variations in these regions in the upper ocean (Frenger et al., 2015; Chassignet et al., 2020; Rhines, 2019) (Figs. 10a–d). In particular, at 5 m, IAP0.25_RMSE is larger in the Bay of Bengal and Arabian Sea, mainly because, in these

395 areas, salinity changes are strongly affected by river runoff, surface fluxes, and ocean currents (Skliris et al., 2014; Adler et al., 2018; Liu et al., 2022). Larger errors in the ITCZ regions are also apparent at 5 m and 100 m, corresponding to strong surface freshwater fluxes (Liu et al., 2022).
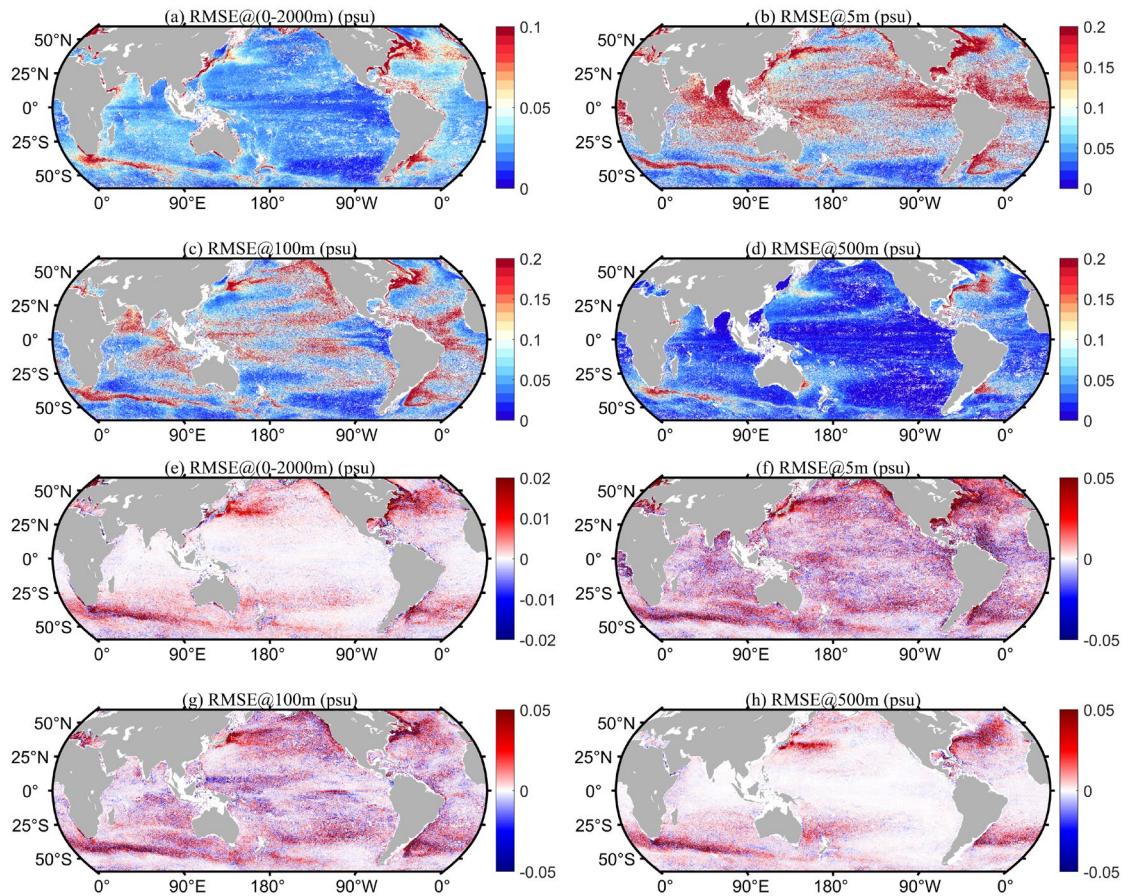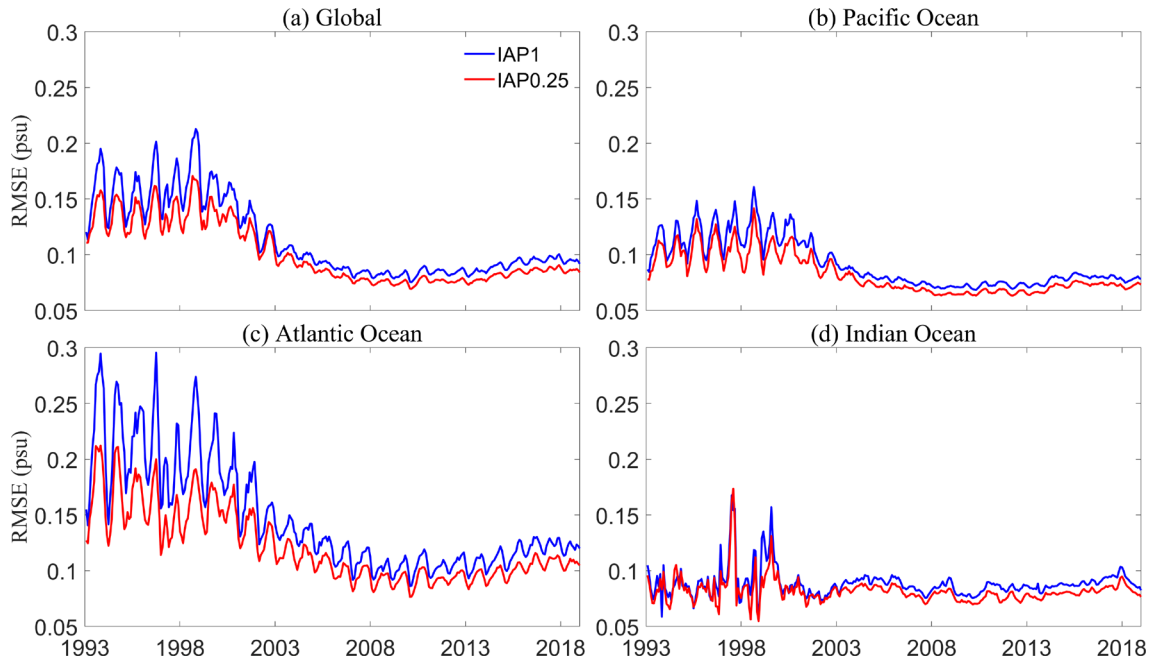

Figure 10: Geographical distribution of RMSE: (a–d) IAP0.25_RMSE; (e–h) IAP1_RMSE minus IAP0.25_RMSE.

21

The temporal variation of the global 1–2000 m mean RMSE from 1993 to 2018 is presented in Fig. 11 for both IAP1° and IAP0.25°, to show the change in reconstruction performance with time. It appears that the error reduces with time, especially since the early 2000s, probably in association with there being more salinity observations available from the Argo network (Yan et al., 2021; Chen et al., 2018; Wong et al., 2020; Roemmich et al., 2019), which helps to increase the reconstruction accuracy. The RMSE has been increasing slightly in recent years, probably because of the inclusion of more real-time Argo data in our analyses. Nevertheless, IAP0.25_RMSE is smaller than IAP1_RMSE for the global ocean, by ~11% on average. The reduction is about 0.016 psu. The lowest RMSE reduction is seen in the Indian Ocean, likely associated with smaller area of the western boundary current systems and relatively less meso-scale activities compared with the other two basins. The biggest reduction in RMSE is seen in the Atlantic Ocean (~17%), confirming our results in previous sections. It is interesting that there is a strong seasonal fluctuation of RMSE: it is larger in the Northern (Southern) Hemisphere summer (winter), likely because of there being less data in the Southern Ocean owing to the sea-ice coverage and severe weather (Zweng et al., 2019; Gould et al., 2013; Auger et al., 2021; Durack, 2015).

Figure 11: The 1–2000 m average RMSE time series for the IAP0.25° and IAP1° datasets for the globe and three ocean basins from 1993 to 2018.

In summary, statistical analysis reveals a small RMSE between the reconstruction field of IAP0.25° and the *in situ* salinity observations, indicating an improved performance of the high-resolution reconstruction (IAP0.25°) in this study compared with the other products considered (IAP1°, ARMOR3D, and EN4). However, it should be noted here that the observations were processed by our research group and used to train the FFNN model, meaning the observations are not independent of the

420   final reconstruction. Thus, besides the analyses reported so far, an independent validation was warranted, the results of which we report in the next section.


4 Five-fold cross validation and uncertainty estimate

In this section, we use the results from a five-fold cross validation to further evaluate the new reconstruction with the withheld independent observations. For this test, we trained the FFNN model with the training set (80% of the full set) and use the

425   withheld 20% of data to evaluate the performance. This process was repeated five times, so the full set could be used for evaluation (i.e., the full set was divided into five sets and each time one set was used as the testing data and the other four sets were training sets).

Fig. 12 shows the density distribution between the reconstructed salinity value and the *in situ* observations for the training set (a, c) and testing set (b, d) at the depths of 10 m and 100 m, separately. The diagonal line denotes a perfect reconstruction, i.e.,

430   the reconstruction perfectly matches the observations. Any errors in observations and/or reconstructions can lead to scattering of the distribution. It appears that the highest density locations are along the diagonal line and most of the reconstruction–observation are within ±0.2 psu for both the training and testing set, indicating that the reconstructions are obviously not biased. The density distribution for the testing set is not visibly different from the training set, suggesting that our FFNN model does not overfit the data in the training set, meaning the reconstruction method is robust in the independent data.

435   Besides, we calculated the RMSE and its degradation between the reconstructed salinity fields and *in situ* observations of the training and testing sets at each depth layer. The degradation rate is defined as: (RMSE of the testing set - RMSE of the training set) / (RMSE of the training set), to quantify the generalization of the model. Fig. 12(a) shows that RMSE of the testing set is consistent (only marginally higher than) with that of the training set. The degradation rate decreases rapidly with depth, about 5.49% at the surface and 0.10% at 100 m. Specifically, at 10 m, the RMSE is 0.261 psu for the

440   training set and 0.269 psu for the testing set, the degradation rate is 3% (Fig. 12b, c); and at 100 m, the RMSE is decreases from 0.1403 psu (training set) to 0.1401 psu (testing set) (Fig. 12d, e). Besides, the correlation coefficient is slightly lower for the testing set: for example, 0.686 at 10 m but 0.707 for training set at the same depth; at 100 m, it decreases from 0.625 (training) to 0.623 (testing). As the testing set is independent from training set, this test indicates that the FFNN model does not experience serious overfitting, and the method is valid.

445
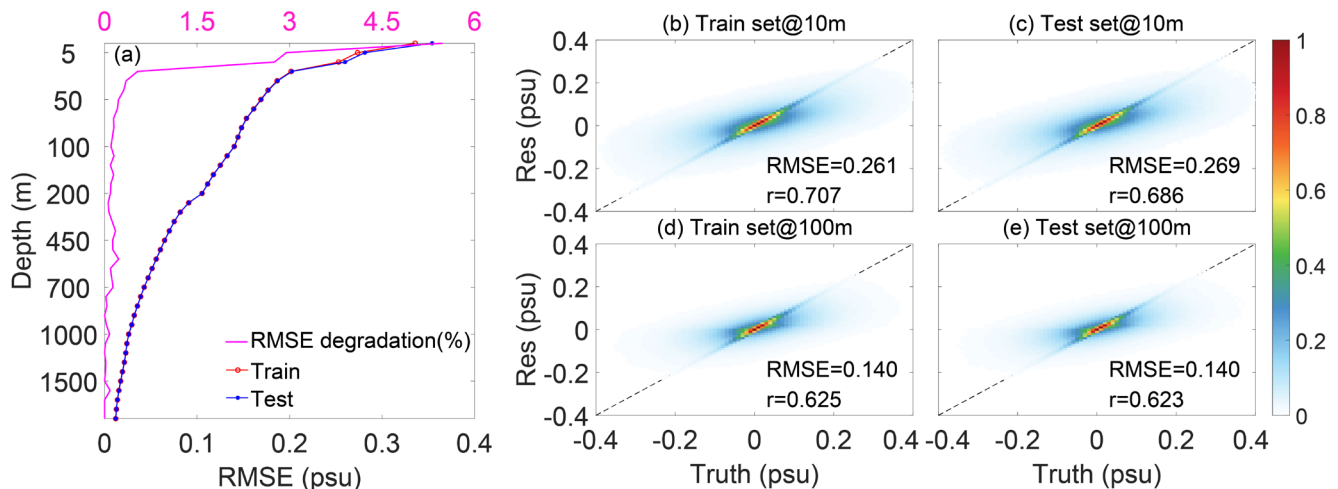
Figure 12**: (a) Vertical distribution of RMSE for the training set and testing set for global region (bottom *x*-axis) and the RMSE degradation from the training set to the testing set (top *x*-axis). (b–e) Density distribution diagrams at depths of 10 m and 100 m for the training set and testing set, as well as RMSE and correlation coefficients (r) for the corresponding layers. The color-coded blocks represent the density of samples.5 Evaluation of the major climatic patterns**

It was shown in the previous section that IAP0.25° is capable of reconstructing salinity with more mesoscale signals. However, it remains to be quantified whether the large-scale pattern of salinity change associated with climate change can be well represented in the new reconstruction. Here, we examine the long-term trends in ocean salinity using the Salinity Contrast (SC) index proposed by Cheng et al. (2020). The SC is defined as the difference between the salinity averaged over high-salinity regions (V_{High}, where salinity is higher than a climatological global median, S_{clim}) and low-salinity regions (V_{Low}, where salinity is below S_{clim}). It was calculated each month over the three-dimensional (x, y, z) ocean salinity field:

$$SC(t) = \frac{\iiint_{V_{High}} S(x,y,z,t)dV}{\iiint_{V_{High}} dV} - \frac{\iiint_{V_{Low}} S(x,y,z,t)dV}{\iiint_{V_{Low}} dV}, \tag{2}$$

where x, y, and z are the three dimensions of latitude, longitude, and depth, respectively. The terms V_{High} and V_{Low} were both determined on the basis of the climatological salinity field.

The SC at the sea surface is called SC0; and for the 1–2000 m volume, the SC is termed SC2000. It can be seen from Fig. 13 that the IAP0.25° and IAP1° datasets have consistent long-term changes of SC. The global-scale SC2000 increases significantly from 1993 to 2018, with a linear trend of 0.045 ± 0.0058 psu century^{-1} (at 90% confidence level, the reduction of degree of freedom has been accounted for in this calculation). The increase of this index indicates that the phenomenon of "fresh gets fresher, salty gets saltier" is more obvious, which is mainly driven by the changes of "wet get wetter, dry get drier" in the global water cycle (Cheng et al., 2020; Marvel et al., 2017; Skliris et al., 2014).

24
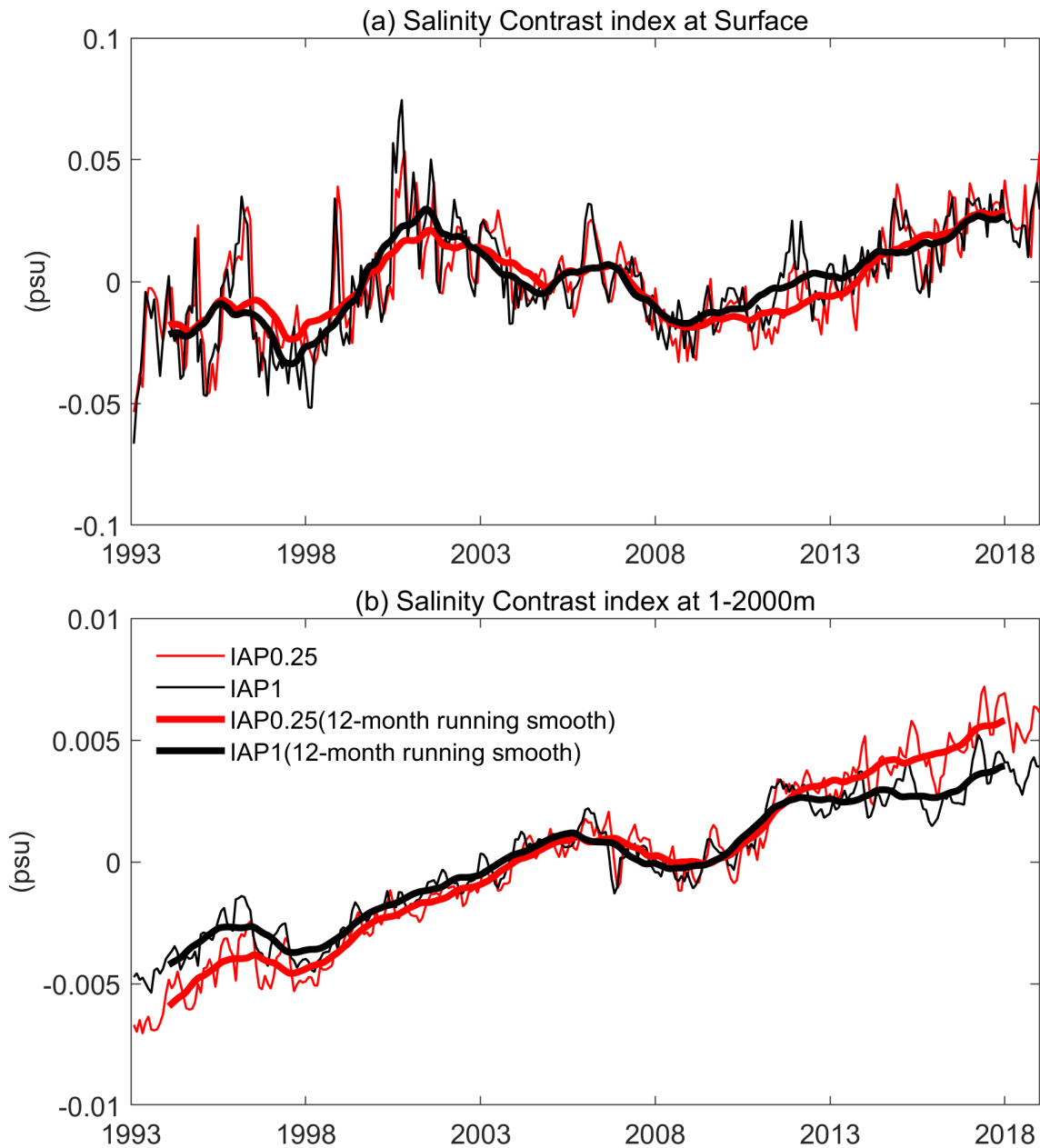
Figure 13: Salinity contrast time series for IAP1° and IAP0.25° data in global regions from 1993 to 2018: (a) at the surface; and (b) in the upper 2000 m.

Besides the global-scale SC metric, we also present the time series over the globe and in different ocean basins for IAP1° and IAP0.25° from 1993 to 2018 for both the salinity anomalies at the surface (S0) and averaged over the 1–2000 m volume (S2000) in Figs. 14 and 15. We divided the oceans into the Pacific Ocean (35°S to 60°N), Atlantic Ocean (35°S to 75°N), Southern Ocean (70°S to 35°S), North Indian Ocean (0° to 30°N), and South Indian Ocean (35°S to 0°). The uncertainty was quantified

by the approach described in section 2.3.3, where the ±2 standard deviation error range is provided, which corresponds to a 95% confidence level. The two datasets continue to show high consistency:

Globally, and in the major ocean basins, IAP0.25° and IAP1° are consistent for S0 (Fig. 14). The S0 of both IAP1° and IAP0.25° shows an upward but statistically insignificant trend from 1993 to 2018. The S0 increases steadily in the Atlantic Ocean (35°S–75°N) and Southern Ocean (70°S–35°S) (Figs. 14c and d). The S0 in Pacific indicates a strong decadal-scale fluctuation. In both the North and South Indian Ocean, the S0 exhibits a weak long-term decreasing trend (Figs. 14e and f).

The global mean S2000 time series (Fig. 15a) shows an increasing trend between 1993 and 2018, consistent with previous studies (Ponte et al., 2021). With more freshwater input into the ocean, the ocean salinity is expected to decrease, and thus this increase of S2000 indicates that the impact of terrestrial ice-melt is difficult to resolve from salinity observations. The data drift in salinity observations is most likely responsible for the observed increase, but this issue has not been resolved yet (https://argo.ucsd.edu/faq). S2000 declines steadily in the Pacific basin, but increases sharply in the Atlantic basin from 1990 (Figs. 15b and c). The contrast salinity change between Pacific Ocean (decreasing, Fig. 15b) and Atlantic Ocean (increasing, Fig. 15c) is associated with increased inter-basin transport of water vapor from the Atlantic to the Pacific (Reagan et al., 2018; Curry et al., 2003). The decadal fluctuations of S2000 in the Atlantic are greater than those in the Pacific, especially in the North Atlantic, and are in-phase with the Atlantic Multidecadal Oscillation (Skliris et al., 2020; Reverdin et al., 2019). S2000 increases in the North Indian Ocean (Fig. 15e) but decreases in the South Indian Ocean (Fig. 15f), showing a "salty gets saltier, fresh gets fresher" change, mainly because of the amplified global hydrological cycle.

In summary, through our investigation of the global and basin salinity time series, we have been able to further confirm that IAP0.25° can capture the integrated property of salinity changes compared with IAP1°. The estimated 95% confidence intervals are generally consistent for different basins, which is about ±0.01 psu for S0 and ±0.002 psu for S2000. Although the error range of this new estimate is larger than IAP1°, because more sources of uncertainty are accounted for (IAP1° mainly accounts for mapping and instrumental error), this new estimate could still be an underestimate because of the neglect of systematic biases (which are currently poorly known) in this study.
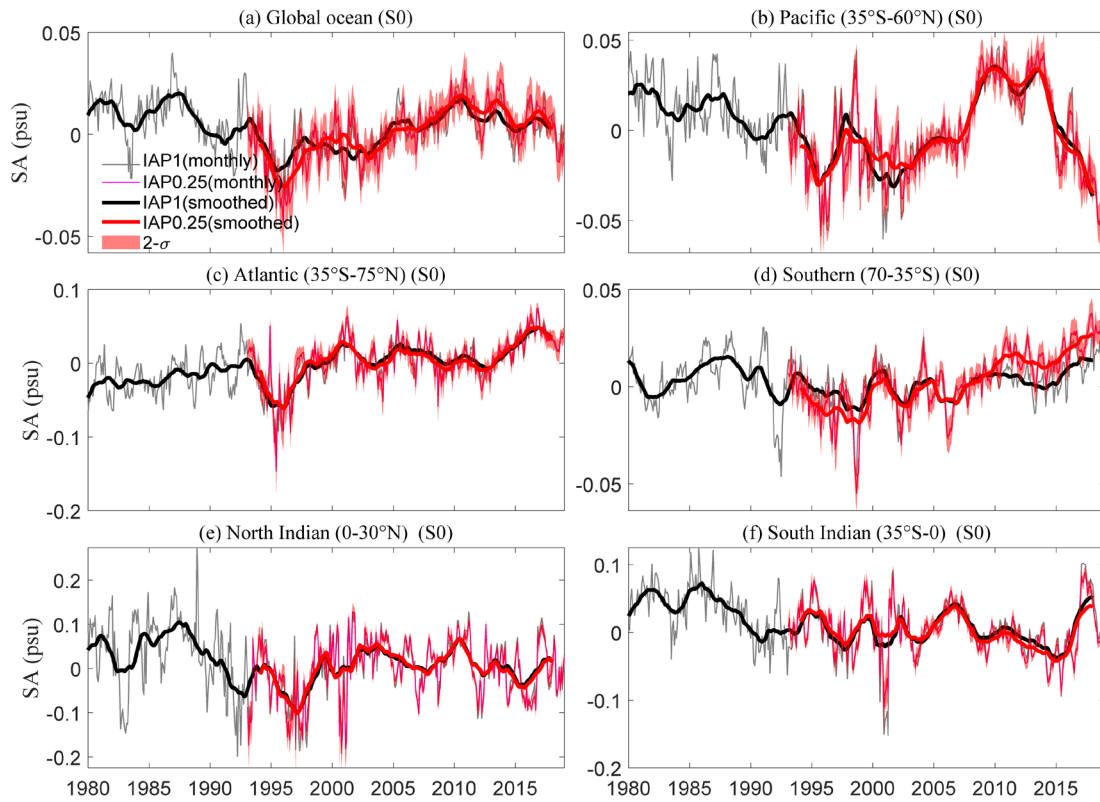
Figure 14: The global (a) and basinal (b-f) salinity anomalies (SA) time series at surface from 1993 to 2018 for IAP1° and IAP0.25° respectively. Both monthly and 12-month running smoothed time series are presented, all data are relative to a 1993–2015 baseline.
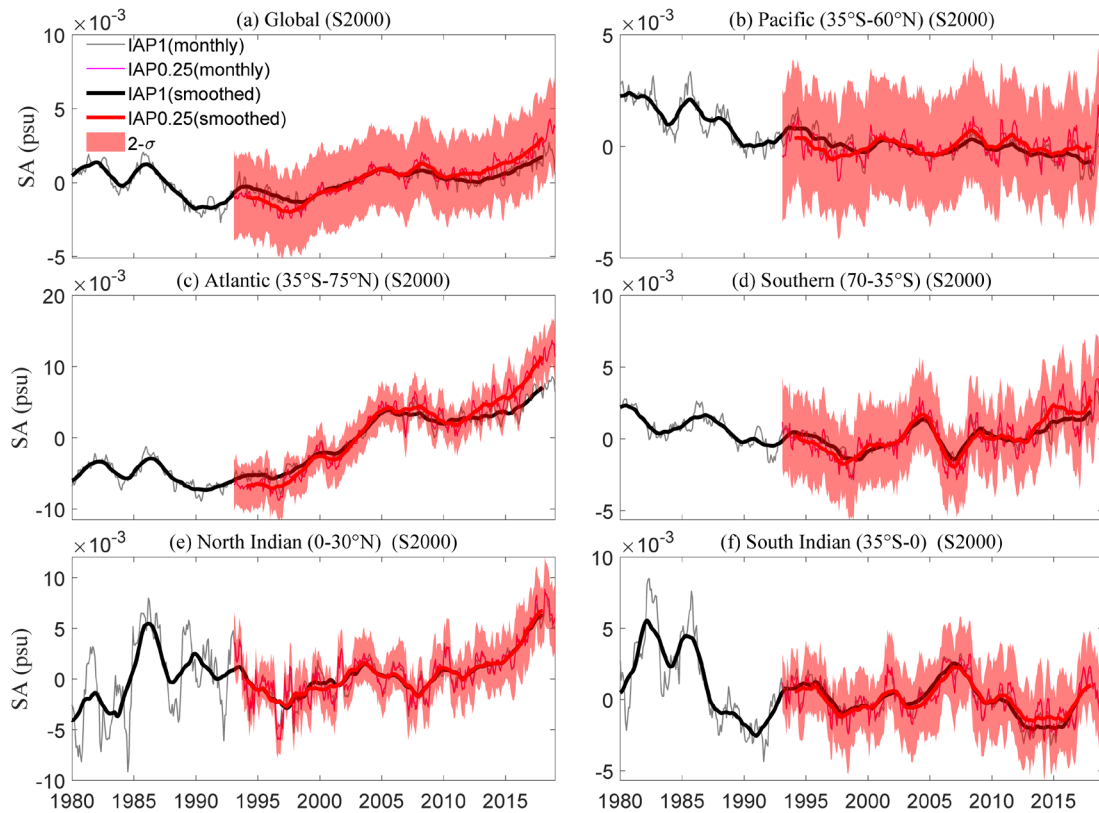
27

Figure 15: The global (a) and basinal (b-f) 1-2000 m averaged salinity anomalies (SA) time series from 1993 to 2018 for IAP1° and IAP0.25° respectively. Both monthly and 12-month running smoothed time series are presented, all data are relative to a 1993–2015 baseline.
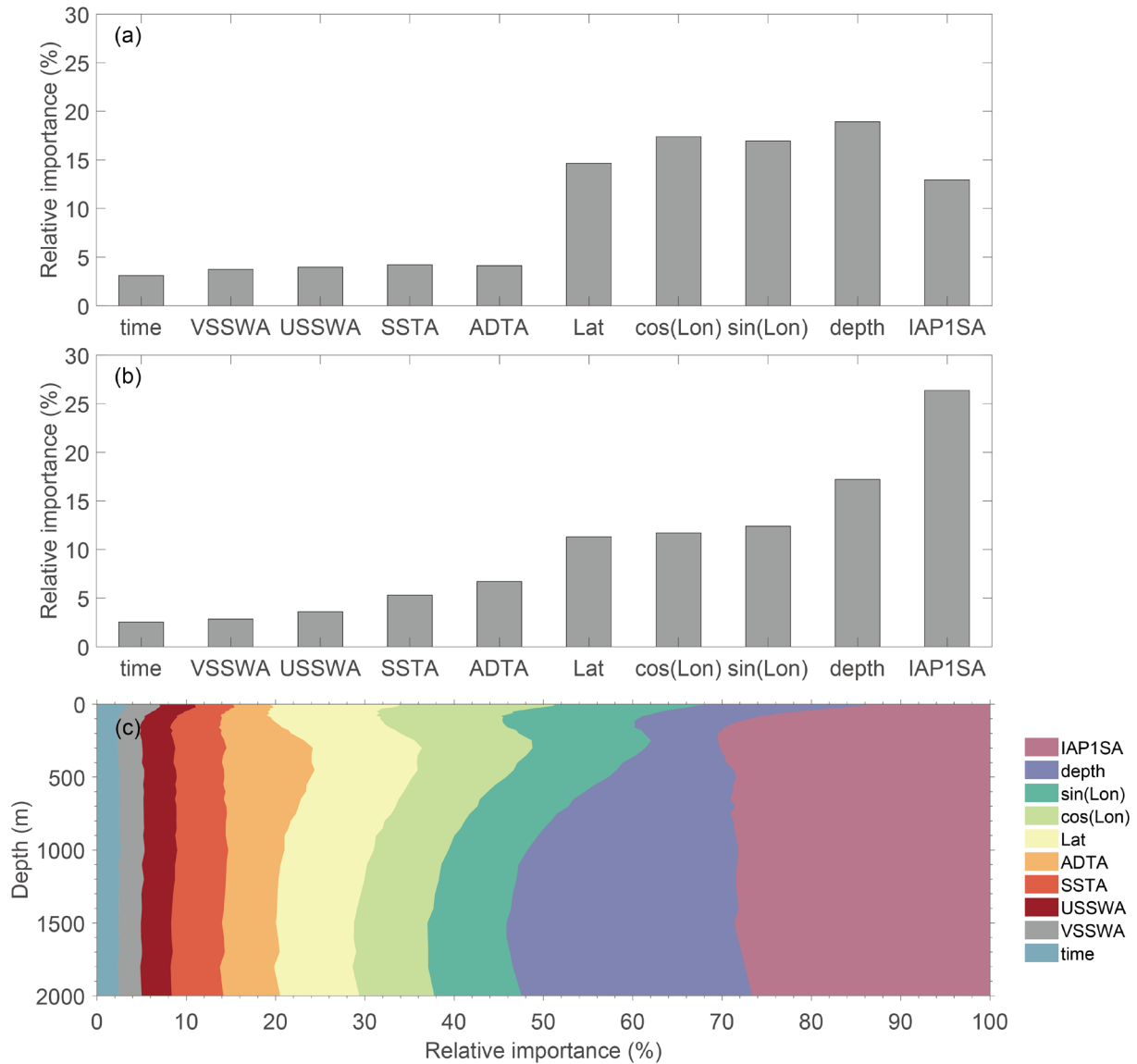
6 Importance of each feature for the reconstruction

The impact of different inputs on the reconstruction of IAP0.25° using the FFNN model is shown in Fig. 16 using the SHAP method. At the surface (Fig. 16a and Fig. 16c), the location parameters (latitude, longitude, depth) are the most important inputs and are probably linked to the strong spatial variability of salinity near sea surface. The IAP1° plays a secondary role near the surface because it provides direct information of salinity and represents the large-scale salinity changes. Accumulatively, the remote sensing data contributes to ~20% of the reconstruction. For the subsurface (Fig. 16c), IAP1° plays a more important role than that near the surface (~26% for 1–2000 m average, Fig. 16b), and this is physically meaningful because there are fewer meso-scale variability in the deeper ocean and large-scale variability becomes more important at the sea subsurface. ADTA becomes more important within 100–700 m than the other layers, because both salinity and ADTA are strongly associated with thermocline variations. VSSWA, USSWA, SSTA plays a similar role from surface to 2000 m (<5% for each), and smaller than most of other inputs, probably because their changes are only weakly coupled with salinity

compared with other parameters. It is interesting that time information (<3%) plays a smallest role in reconstruction, implying that the FFNN can be applied in other time periods without losing too much accuracy.

Figure 16: A quantification of the relative importance of each input in reconstruction of IAP0.25°: (a) at the surface; (b)1–2000 m average; and (c) at each depth from 1 m to 2000 m. The input features are ranked in terms of importance, i.e., the higher is the SHAP value, the more important is the features.

7 Data availability

525 The code used in this paper includes data processing, optimal model building and the result prediction. The reconstructed IAP0.25° dataset is available at http://doi.org/10.57760/sciencedb.o00122.00001 (Tian et al., 2022) or http://www.ocean.iap.ac.cn/ftp/cheng/IAP_v0_Ocean_Salinity_0p25_FFNN_0_2000m/. Here we provide global ocean salinity gridded product at 0.25° × 0.25° horizontal resolution on 41 vertical levels from 1–2000 m, and at a monthly resolution from 1993 to 2018.

530 **8 Summary and Discussion**

This study used an FFNN approach to reconstruct a high-resolution (0.25° × 0.25°) ocean subsurface salinity dataset (1–2000 m) for the period 1993–2018, in which the spatial and temporal information (time, longitude, latitude, depth), previously available the 1°×1° resolution salinity dataset, and satellite remote sensing data (ADT, SST, SSW) were given as inputs to the FFNN algorithm for reconstruction. By training the functional relationship between input variables and truth values (observed 535 gridded averaged salinity), the reconstruction model was established. For the IAP0.25° salinity data, the global and regional reconstruction performance was evaluated using several different tests. In brief, we show that: (1) IAP0.25° salinity data maintain the large-scale information from IAP1° gridded data. Because previous evaluations suggest that IAP1° provides more physically tenable large-scale patterns and long-term climate change and variabilities compared to many available datasets (Cheng et al., 2020; Reed et al., 2022; Sohail et al., 2022), thus the reliability of large-scale signals also becomes an advantage 540 of the new IAP0.25° product. (2) Compared with IAP1°, the RMSE of IAP0.25° can be reduced by ~11% on a global average. Besides, IAP0.25° shows more realistic spatial signals in the Gulf Stream, Kuroshio, and Antarctic Circumpolar Current regions with strong mesoscale variations than the IAP1°product, indicating that FFNN can effectively transfer small-scale spatial variations in ADT, SST and SSW fields into the 0.25° × 0.25° salinity field. It thus serves as an improvement on the currently available IAP data. (3) We show that the FFNN approach is effective in merging different kinds of Earth observations, 545 and the method is robust and can be reliably used for ocean state reconstruction, thus can complement the existing data assimilation and objective analysis methods.

Although the validity and advantages of the FFNN approach have been demonstrated, there are some caveats and limitations to this approach, related to which we provide some future guidance. This study used a probabilistic approach to quantify the uncertainty, which is a practical strategy, and with this approach all major error sources are propagated to the final fields. 550 However, uncertainty assessment with machine learning is still a challenge for many available products, and this issue requires future analysis: i.e., how sensitive the FFNN approach is for parameter choices such as the number of hidden layers and neurons.

Compared with the ARMOR3D data, it seems that the small-scale signals are weaker in IAP0.25° (for example Fig. 5 and Fig. 6), indicating a difference of efficiency with which signals of the remote sensing data are transmitted into the salinity

reconstruction. Therefore, an intriguing scientific question is what is the key factor determining the efficiency? Addressing this question requires a dedicated study that incorporates an intercomparison among other approaches.

As a coarse-resolution product, IAP1° is important in high-resolution reconstructions and provides a critical large-scale precondition. The uncertainty (or biases) in this coarse-resolution field will propagate into the reconstruction, but how the uncertainty propagates and contributes to the final estimate are still open questions. One of the aims of this study was to ensure the continuity from IAP1° to IAP0.25° data, because work has already shown superiority of IAP1° in large-scale reconstructions compared with many available datasets (Cheng et al., 2020). Thus, this study further advances the prior findings of Cheng et al. (2020). Further experiments are needed to quantify the sensitivity of the results to this preconditioning.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges, https://doi.org/10.1016/j.inffus.2021.05.008, 2021.

Abram, N. J., Gattuso, P., Prakash, A., Cheng, L., Chidichimo, M., Crate, S., Enomoto, H., Garschagen, M., Gruber, N., Harper, S., Holland, E., Kudela, R. M., Rice, J., Steffen, K., and von Schuckmann, K.: Framing and Context of the Report. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, IPCC, 2019.

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J. J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D. Bin: The Global Precipitation Climatology Project (GPCP) monthly analysis (New Version 2.3) and a review of 2017 global precipitation, Atmosphere (Basel)., 9, https://doi.org/10.3390/atmos9040138, 2018.

585   Atlas, R., Hoffman, R. N., Ardizzone, J., Leidner, S. M., Jusem, J. C., Smith, D. K., and Gombos, D.: A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications, Bull. Am. Meteorol. Soc., 92, 157–174, https://doi.org/10.1175/2010BAMS2946.1, 2011.

Auger, M., Morrow, R., Kestenare, E., Sallée, J. B., and Cowley, R.: Southern Ocean in-situ temperature trends over 25 years emerge from interannual variability, Nat. Commun., 12, https://doi.org/10.1038/s41467-020-20781-1, 2021.

590   Bagnell, A. and DeVries, T.: 20th century cooling of the deep ocean contributed to delayed acceleration of Earth's energy imbalance, Nat. Commun., 12, 4604, https://doi.org/10.1038/s41467-021-24472-3, 2021.

Balmaseda, M. A., Trenberth, K. E., and Källén, E.: Distinctive climate signals in reanalysis of global ocean heat content, Geophys. Res. Lett., 40, https://doi.org/10.1002/grl.50382, 2013.

Balmaseda, M. A., Hernandez, F., Storto, A., Palmer, M. D., Alves, O., Shi, L., Smith, G. C., Toyoda, T., Valdivieso, M.,
595   Barnier, B., Behringer, D., Boyer, T., Chang, Y. S., Chepurin, G. A., Ferry, N., Forget, G., Fujii, Y., Good, S., Guinehut, S., Haines, K., Ishikawa, Y., Keeley, S., Köhl, A., Lee, T., Martin, M. J., Masina, S., Masuda, S., Meyssignac, B., Mogensen, K., Parent, L., Peterson, K. A., Tang, Y. M., Yin, Y., Vernieres, G., Wang, X., Waters, J., Wedd, R., Wang, O., Xue, Y., Chevallier, M., Lemieux, J. F., Dupont, F., Kuragano, T., Kamachi, M., Awaji, T., Caltabiano, A., Wilmer-Becker, K., and Gaillard, F.: The ocean reanalyses intercomparison project (ORA-IP), J. Oper. Oceanogr., 8,
600   https://doi.org/10.1080/1755876X.2015.1022329, 2015.

Banzon, V., Smith, T. M., Mike Chin, T., Liu, C., and Hankins, W.: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies, https://doi.org/10.5194/essd-8-165-2016, 2016.

Berrar, D.: Cross-validation, in: Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, vol. 1–
605   3, 542–545, https://doi.org/10.1016/B978-0-12-809633-8.20349-X, 2018.

Boyer, T. P., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Locarnini, R. A., Mishonov, A. V., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., Weathers, K. W., and Zweng, M. M.: World Ocean Database 2018, A. V. Mishonov, Tech. Ed. NOAA Atlas NESDIS 87, 1–207, 2018.

Carton, J. A., Chepurin, G. A., and Chen, L.: SODA3: A new ocean climate reanalysis, J. Clim., 31, 6967–6983,
610   https://doi.org/10.1175/jcli-d-18-0149.1, 2018.

Chassignet, E. P., Fox-Kemper, B., Yeager, S. G., and Bozec, A.: Sources and Sinks of Ocean Mesoscale Eddy Energy, CLIVAR Exch. CLIVAR Var., 18, 3–8, 2020.

Chau, S. L., Gonzalez, J., and Sejdinovic, D.: RKHS-SHAP: Shapley Values for Kernel Methods, CoRR, abs/2110.0, 2021.

Chen, G., Peng, L., and Ma, C.: Climatology and seasonality of upper ocean salinity: a three-dimensional view from argo
615   floats, Clim. Dyn., 50, 2169–2182, https://doi.org/10.1007/s00382-017-3742-6, 2018.

Cheng, L. and Zhu, J.: Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations, J. Clim., 29, 5393–5416, https://doi.org/10.1175/JCLI-D-15-0730.1, 2016.

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., and Zhu, J.: Improved estimates of ocean heat content from 1960 to 2015, Sci. Adv., 3, e1601545, https://doi.org/10.1126/sciadv.1601545, 2017.

620 Cheng, L., Trenberth, K. E., Gruber, N., Abraham, J. P., Fasullo, J. T., Li, G., Mann, M. E., Zhao, X., and Zhu, J.: Improved estimates of changes in upper ocean salinity and the hydrological cycle, J. Clim., 33, 10357–10381, https://doi.org/10.1175/JCLI-D-20-0366.1, 2020.

Cheng, Y. H., Ho, C. R., Zheng, Q., and Kuo, N. J.: Statistical characteristics of mesoscale eddies in the north pacific derived from satellite altimetry, https://doi.org/10.3390/rs6065164, 2014.

625 Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C. Le, Myneni, R. B., Piao, S., and Thornton, P.: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, Chang. IPCC Clim., https://doi.org/10.1017/CBO9781107415324.015, 2013.

Contractor, S. and Roughan, M.: Efficacy of Feedforward and LSTM Neural Networks at Predicting and Gap Filling Coastal
630 Ocean Timeseries: Oxygen, Nutrients, and Temperature, Front. Mar. Sci., 8, https://doi.org/10.3389/fmars.2021.637759, 2021.

Curry, R., Dickson, B., and Yashayaev, I.: A change in the freshwater balance of the Atlantic Ocean over the past four decades, Nature, 426, 826–829, https://doi.org/10.1038/nature02206, 2003.

Dan Foresee, F. and Hagan, M. T.: Gauss-Newton approximation to bayesian learning, in: IEEE International Conference on Neural Networks - Conference Proceedings, 1930–1935, https://doi.org/10.1109/ICNN.1997.614194, 1997.

635 Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: A two-step neural network model for the reconstruction of surface ocean pCO2 over the global ocean, Geosci. Model Dev., 12, 2091–2105, https://doi.org/10.5194/gmd-12-2091-2019, 2019.

Domingues, C. M., Church, J. A., White, N. J., Gleckler, P. J., Wijffels, S. E., Barker, P. M., and Dunn, J. R.: Improved estimates of upper-ocean warming and multi-decadal sea-level rise, Nature, 453, https://doi.org/10.1038/nature07080, 2008.

640 Durack, P. J.: Ocean salinity and the global water cycle, Oceanography, 28, 20–31, https://doi.org/10.5670/oceanog.2015.03, 2015.

Durack, P. J., Gleckler, P. J., Landerer, F. W., and Taylor, K. E.: Quantifying underestimates of long-term upper-ocean warming, Nat. Clim. Chang., 4, https://doi.org/10.1038/nclimate2389, 2014.

Forget, G., Campin, J. M., Heimbach, P., Hill, C. N., Ponte, R. M., and Wunsch, C.: ECCO version 4: An integrated framework
645 for non-linear inverse modeling and global ocean state estimation, Geosci. Model Dev., 8, 3071–3104, https://doi.org/10.5194/gmd-8-3071-2015, 2015.

Frederikse, T., Jevrejeva, S., Riva, R. E. M., and Dangendorf, S.: A consistent sea-level reconstruction and its budget on basin and global scales over 1958-2014, J. Clim., 31, 1267–1280, https://doi.org/10.1175/JCLI-D-17-0502.1, 2018.

Frenger, I., Münnich, M., Gruber, N., and Knutti, R.: Southern Ocean eddy phenomenology, J. Geophys. Res. Ocean., 120,
650 7413–7449, https://doi.org/10.1002/2015JC011047, 2015.

Gabella, M.: Topology of Learning in Feedforward Neural Networks, IEEE Trans. Neural Networks Learn. Syst., 32, 3588–3592, https://doi.org/10.1109/TNNLS.2020.3015790, 2021.

Gaillard, F., Reynaud, T., Thierry, V., Kolodziejczyk, N., and Von Schuckmann, K.: In situ-based reanalysis of the global ocean temperature and salinity with ISAS: Variability of the heat content and steric height, J. Clim., 29, https://doi.org/10.1175/JCLI-D-15-0028.1, 2016.

Gal, Y. and Ghahramani, Z.: Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, in: 33rd International Conference on Machine Learning, ICML 2016, 1651–1660, 2016.

Gan, M., Pan, S., Chen, Y. ping, Cheng, C., Pan, H., and Zhu, X.: Application of the Machine Learning LightGBM Model to the Prediction of the Water Levels of the Lower Columbia River, J. Mar. Sci. Eng., 9, 496, https://doi.org/10.3390/jmse9050496, 2021.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Ocean., 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Gould, J., Sloyan, B., and Visbeck, M.: In situ ocean observations. A brief history, present status, and future directions., in: International Geophysics, vol. 103, edited by: Siedler, G., Griffies, S. M., Gould, J., and Church, J. A. B. T.-I. G., Academic Press, 59–81, https://doi.org/10.1016/B978-0-12-391851-2.00003-9, 2013.

Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database, Deep. Res. Part I Oceanogr. Res. Pap., 57, 812–833, https://doi.org/10.1016/j.dsr.2010.03.011, 2010.

Gui, K., Che, H., Zeng, Z., Wang, Y., Zhai, S., Wang, Z., Luo, M., Zhang, L., Liao, T., Zhao, H., Li, L., Zheng, Y., and Zhang, X.: Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model, Environ. Int., 141, https://doi.org/10.1016/j.envint.2020.105801, 2020.

Guinehut, S., Dhomps, A. L., Larnicol, G., and Le Traon, P. Y.: High resolution 3-D temperature and salinity fields derived from in situ and satellite observations, Ocean Sci., 8, 845–857, https://doi.org/10.5194/os-8-845-2012, 2012.

Hannachi, A., Jolliffe, I. T., and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, Int. J. Climatol., 27, 1119–1152, https://doi.org/10.1002/joc.1499, 2007.

Hosoda, S., Ohira, T., and Nakamura, T.: A monthly mean dataset of global oceanic temperature and salinity derived from Argo float observations, JAMSTEC Rep. Res. Dev., 8, https://doi.org/10.5918/jamstecr.8.47, 2008.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., and Zhang, H.-M.: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1, J. Clim., 34, 2923–2939, https://doi.org/10.1175/JCLI-D-20-0166.1, 2021.

Ishii, M., Kimoto, M., and Kachi, M.: Historical ocean subsurfaces temperature analysis with error estimates, Mon. Weather Rev., 131, https://doi.org/10.1175/1520-0493(2003)131<0051:HOSTAW>2.0.CO;2, 2003.

Ishii, M., Fukuda, Y., Hirahara, S., Yasui, S., Suzuki, T., and Sato, K.: Accuracy of Global Upper Ocean Heat Content Estimation Expected from Present Observational Data Sets, SOLA, 13, 163–167, https://doi.org/10.2151/sola.2017-030, 2017.

Jean-Michel, L., Eric, G., Romain, B. B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L. G., Charly, R., Tony, C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D., and Pierre-Yves, L. T.: The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis, Front. Earth Sci., 9, https://doi.org/10.3389/feart.2021.698876, 2021.

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K. R.: Efficient backprop, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7700 LECTU, edited by: Montavon, G., Orr, G. B., and Müller, K.-R., Springer Berlin Heidelberg, Berlin, Heidelberg, 9–48, https://doi.org/10.1007/978-3-642-35289-8_3, 2012.

Lei, J.: Cross-Validation With Confidence, J. Am. Stat. Assoc., 115, 1978–1997, https://doi.org/10.1080/01621459.2019.1672556, 2020.

Levitus, S., Antonov, J. I., Boyer, T. P., Locarnini, R. A., Garcia, H. E., and Mishonov, A. V.: Global ocean heat content 1955-2008 in light of recently revealed instrumentation problems, Geophys. Res. Lett., 36, https://doi.org/10.1029/2008GL037155, 2009.

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J. R., Seidov, D., Yarosh, E. S., and Zweng, M. M.: World ocean heat content and thermosteric sea level change (0-2000m), 1955-2010, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL051106, 2012.

Li, G., Zhang, Y., Xiao, J., Song, X., Abraham, J., Cheng, L., and Zhu, J.: Examining the salinity change in the upper Pacific Ocean during the Argo period, Clim. Dyn., 53, 6055–6074, https://doi.org/10.1007/s00382-019-04912-z, 2019.

Li, H., Xu, F., Zhou, W., Wang, D., Wright, J. S., Liu, Z., and Lin, Y.: Development of a global gridded Argo data set with Barnes successive corrections, J. Geophys. Res. Ocean., 122, https://doi.org/10.1002/2016JC012285, 2017.

Liang, X., Liu, C., Ponte, R. M., and Chambers, D. P.: A comparison of the variability and changes in global ocean heat content from multiple objective analysis products during the Argo period, J. Clim., 34, https://doi.org/10.1175/JCLI-D-20-0794.1, 2021.

Liashchynskyi, P. and Liashchynskyi, P.: Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS, CoRR, abs/1912.0, 2019.

Liu, H., Zhou, H., Yang, W., Liu, X., Li, Y., Yang, Y., Chen, X., and Li, X.: A three-dimensional gravest empirical mode determined from hydrographic observations in the western equatorial Pacific Ocean, J. Mar. Syst., 214, 103487, https://doi.org/https://doi.org/10.1016/j.jmarsys.2020.103487, 2021.

Liu, Y., Cheng, L., Pan, Y., Abraham, J., Zhang, B., Zhu, J., and Song, J.: Climatological seasonal variation of the upper ocean salinity, Int. J. Climatol., 42, 3477–3498, https://doi.org/10.1002/joc.7428, 2022.

Llovel, W. and Lee, T.: Importance and origin of halosteric contribution to sea level change in the southeast Indian Ocean during 2005-2013, Geophys. Res. Lett., 42, 1148–1157, https://doi.org/10.1002/2014GL062611, 2015.

Lu, S., Liu, Z., Li, H., Li, Z., and Xu, J.: Manual of Global Ocean Argo gridded data set (BOA _Argo) (Version 2020), 14, 2020.

720 Lu, W., Su, H., Yang, X., and Yan, X. H.: Subsurface temperature estimation from remote sensing data using a clustering-neural network method, Remote Sens. Environ., 229, 213–222, https://doi.org/10.1016/j.rse.2019.04.009, 2019.

Lundberg, S. M.: A Unified Approach to Interpreting Model Predictions, 31st Conf. Neural Inf. Process. Syst. (NIPS 2017), 32, 2019.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information 725 Processing Systems, 2017.

Lyman, J. M. and Johnson, G. C.: Estimating global ocean heat content changes in the upper 1800m since 1950 and the influence of climatology choice, J. Clim., 27, https://doi.org/10.1175/JCLI-D-12-00752.1, 2014.

Marvel, K., Biasutti, M., Bonfils, C., Taylor, K. E., Kushnir, Y., and Cook, B. I.: Observed and projected changes to the precipitation annual cycle, J. Clim., 30, 4983–4995, https://doi.org/10.1175/JCLI-D-16-0572.1, 2017.

730 McWilliams, J. C.: Submesoscale currents in the ocean, Proc. R. Soc. A Math. Phys. Eng. Sci., 472, 20160117, https://doi.org/10.1098/rspa.2016.0117, 2016.

Mertz, F., Rosmorduc, V., Maheu, C., and Faugère, Y.: CMEMS Product User manual For Sea Level SLA products, Copernicus Mar. Environ. Monit. Serv., 0–41, 2016.

Milanés-Hermosilla, D., Codorniú, R. T., López-Baracaldo, R., Sagaró-Zamora, R., Delisle-Rodriguez, D., Villarejo-Mayor, 735 J. J., and Núñez-álvarez, J. R.: Monte carlo dropout for uncertainty estimation and motor imagery classification, Sensors, 21, https://doi.org/10.3390/s21217241, 2021.

Mulet, S., Rio, M. H., Mignot, A., Guinehut, S., and Morrow, R.: A new estimate of the global 3D geostrophic ocean circulation based on satellite data and in-situ measurements, Deep. Res. Part II Top. Stud. Oceanogr., 77–80, 70–81, https://doi.org/10.1016/j.dsr2.2012.04.012, 2012.

740 Palmer, M. D., Roberts, C. D., Balmaseda, M., Chang, Y. S., Chepurin, G., Ferry, N., Fujii, Y., Good, S. A., Guinehut, S., Haines, K., Hernandez, F., Köhl, A., Lee, T., Martin, M. J., Masina, S., Masuda, S., Peterson, K. A., Storto, A., Toyoda, T., Valdivieso, M., Vernieres, G., Wang, O., and Xue, Y.: Ocean heat content variability and change in an ensemble of ocean reanalyses, Clim. Dyn., 49, 909–930, https://doi.org/10.1007/s00382-015-2801-0, 2017.

Pauthenet, E., Bachelot, L., Balem, K., Maze, G., Tréguier, A.-M., Roquet, F., Fablet, R., and Tandeo, P.: Four-dimensional 745 temperature, salinity and mixed-layer depth in the Gulf Stream, reconstructed from remote-sensing and in situ observations with neural networks, Ocean Sci., 18, 1221–1244, https://doi.org/10.5194/os-18-1221-2022, 2022.

Ponte, R. M., Sun, Q., Liu, C., and Liang, X.: How Salty Is the Global Ocean: Weighing It All or Tasting It a Sip at a Time?, Geophys. Res. Lett., 48, https://doi.org/10.1029/2021GL092935, 2021.

Reagan, J., Seidov, D., and Boyer, T.: Water Vapor Transfer and Near-Surface Salinity Contrasts in the North Atlantic Ocean, 750 Sci. Rep., 8, https://doi.org/10.1038/s41598-018-27052-6, 2018.

Reverdin, G., Friedman, A. R., Chafik, L., Holliday, N. P., Szekely, T., Valdimarsson, H., and Yashayaev, I.: North Atlantic extratropical and subpolar gyre variability during the last 120 years: a gridded dataset of surface temperature, salinity, and density. Part 1: dataset validation and RMS variability, Ocean Dyn., 69, 385–403, https://doi.org/10.1007/s10236-018-1240-y, 2019.

755   Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily high-resolution-blended analyses for sea surface temperature, J. Clim., 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

Rhines, P. B.: Mesoscale eddies, in: Encyclopedia of Ocean Sciences, edited by: Cochran, J. K., Bokuniewicz, H. J., and Yager, P. L. B. T.-E. of O. S. (Third E., Academic Press, Oxford, 115–127, https://doi.org/10.1016/B978-0-12-409548-9.11642-2, 2019.

760   Roemmich, D. and Gilson, J.: The 2004-2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program, Prog. Oceanogr., 82, https://doi.org/10.1016/j.pocean.2009.03.004, 2009.

Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., Garzoli, S. L., Schmid, C., and Ignaszewski, M.: The Argo Program: Observing the global ocean with profiling floats, Oceanography, 22, 34–43, https://doi.org/10.5670/oceanog.2009.36, 2009.

765   Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, M., Belbeoch, M., Bittig, H. C., Boss, E., Calil, P., Carse, F., Carval, T., Chai, F., Conchubhair, D. Ó., d'Ortenzio, F., Dall'Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G. C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Traon, P.-Y. Le, Lenn, Y.-D., Maze, G., Mork, K.

770   A., Morris, T., Nagai, T., Nash, J., Garabato, A. N., Olsen, A., Pattabhi, R. R., Prakash, S., Riser, S., Schmechtig, C., Schmid, C., Shroyer, E., Sterl, A., Sutton, P., Talley, L., Tanhua, T., Thierry, V., Thomalla, S., Toole, J., Troisi, A., Trull, T. W., Turton, J., Velez-Belchi, P. J., Walczowski, W., Wang, H., Wanninkhof, R., Waterhouse, A. F., Waterman, S., Watson, A., Wilson, C., Wong, A. P. S., Xu, J., and Yasuda, I.: On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array   , https://www.frontiersin.org/article/10.3389/fmars.2019.00439, 2019.

775   Von Schuckmann, K., Sallée, J. B., Chambers, D., Le Traon, P. Y., Cabanes, C., Gaillard, F., Speich, S., and Hamon, M.: Consistency of the current global ocean observing systems from an Argo perspective, https://doi.org/10.5194/os-10-547-2014, 2014.

Siegelman, L., Roquet, F., Mensah, V., Rivière, P., Pauthenet, E., Picard, B., and Guinet, C.: Correction and accuracy of high- and low-resolution CTD data from animal-borne instruments, J. Atmos. Ocean. Technol., 36, 745–760, 780   https://doi.org/10.1175/JTECH-D-18-0170.1, 2019.

Skliris, N., Marsh, R., Josey, S. A., Good, S. A., Liu, C., and Allan, R. P.: Salinity changes in the World Ocean since 1950 in relation to changing surface freshwater fluxes, Clim. Dyn., 43, 709–736, https://doi.org/10.1007/s00382-014-2131-7, 2014.

Skliris, N., Marsh, R., Mecking, J. V., and Zika, J. D.: Changing water cycle and freshwater transports in the Atlantic Ocean in observations and CMIP5 models, Clim. Dyn., 54, 4971–4989, https://doi.org/10.1007/s00382-020-05261-y, 2020.

785 Stamell, J., Rustagi, R., Gloege, L., and McKinley, G.: Strengths and weaknesses of three Machine Learning methods for pCO$_2$ interpolation, Geosci. Model Dev. Discuss., 2020, 1–25, https://doi.org/10.5194/gmd-2020-311, 2020.

Stoean, C., Stoean, R., Atencia, M., Abdar, M., Velázquez-Pérez, L., Khosravi, A., Nahavandi, S., Rajendra Acharya, U., and Joya, G.: Automated detection of presymptomatic conditions in spinocerebellar ataxia type 2 using monte carlo dropout and

790 deep neural network techniques with electrooculogram signals, Sensors (Switzerland), 20, https://doi.org/10.3390/s20113032, 2020.

Storto, A., Alvera-Azcárate, A., Balmaseda, M. A., Barth, A., Chevallier, M., Counillon, F., Domingues, C. M., Drévillon, M., Drillet, Y., Forget, G., Garric, G., Haines, K., Hernandez, F., Iovino, D., Jackson, L. C., Lellouche, J. M., Masina, S., Mayer, M., Oke, P. R., Penny, S. G., Peterson, A. K., Yang, C., and Zuo, H.: Ocean reanalyses: Recent advances and unsolved

795 challenges, https://doi.org/10.3389/fmars.2019.00418, 2019.

Su, H., Zhang, H., Geng, X., Qin, T., Lu, W., and Yan, X. H.: OPEN: A new estimation of global ocean heat content for upper 2000 meters from remote sensing data, https://doi.org/10.3390/rs12142294, 2020.

Tian, T., Cheng, L., Wang, G., Abraham, J., Ren, S., Zhu, J., and Song, J.: Reconstructing ocean subsurface salinity at high resolution using a machine learning approach [data set], http://doi.org/10.57760/sciencedb.o00122.00001, 2022.

800 Vinogradova, N., Lee, T., Boutin, J., Drushka, K., Fournier, S., Sabia, R., Stammer, D., Bayler, E., Reul, N., Gordon, A., Melnichenko, O., Li, L., Hackert, E., Martin, M., Kolodziejczyk, N., Hasson, A., Brown, S., Misra, S., and Lindstrom, E.: Satellite salinity observing system: Recent discoveries and the way forward, https://doi.org/10.3389/fmars.2019.243, 2019.

Vitaly Bushaev: Understanding RMSprop — faster neural network learning, Towar. Data Sci., 36, 1–7, 2018.

Wang, G., Cheng, L., Boyer, T., and Li, C.: Halosteric sea level changes during the Argo era, Water (Switzerland), 9,

805 https://doi.org/10.3390/w9070484, 2017.

Wang, H., Song, T., Zhu, S., Yang, S., and Feng, L.: Subsurface temperature estimation from sea surface data using neural network models in the western pacific ocean, https://doi.org/10.3390/math9080852, 2021.

Wang, J., Flierl, G. R., Lacasce, J. H., Mcclean, J. L., and Mahadevan, A.: Reconstructing the ocean's interior from surface data, J. Phys. Oceanogr., 43, 1611–1626, https://doi.org/10.1175/JPO-D-12-0204.1, 2013.

810 Warin, T.: Multiple Linear Regression, https://doi.org/10.6084/M9.FIGSHARE.8178812.V1, 2019.

Wentz, F. J., Scott, J., Hoffman, R., Leidner, M., Atlas, R., and Ardizzone, J.: Cross-Calibrated Multi-Platform Ocean Surface Wind Vector Analysis Product V2, 1987 - ongoing, https://doi.org/10.5065/4TSY-K140, 2016.

Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André,

815 X., Poffa, N., Yashayaev, I., Barker, P. M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, M. O. N., Schmid, C., Lyman, J. M., McTaggart, K. E., Purkey, S. G., Zilberman, N., Alkire, M. B., Swift, D., Owens, W. B., Jayne, S. R., Hersh, C., Robbins, P., West-Mack, D., Bahr, F., Yoshida, S., Sutton, P. J. H., Cancouët, R., Coatanoan, C., Dobbler, D., Juan, A. G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., Claustre, H., D'Ortenzio, F., Le Reste, S., Le Traon, P. Y., Rannou,

J. P., Saout-Grit, C., Speich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I. M., Klein, B., Notarstefano, G., Poulain, P.
820 M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaska, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida, T., Cowley, R., Lovell, J. L., Oke, P. R., van Wijk, E. M., Carse, F., Donnelly, M., Gould, W. J., Gowers, K., King, B. A., Loch, S. G., Mowat, M., Turton, J., Rama Rao, E. P., Ravichandran, M., Freeland, H. J., Gaboury, I., Gilbert, D., Greenan, B. J. W., Ouellet, M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K. R., Jo, H. J., et al.: Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of
825 Profiling Floats, https://doi.org/10.3389/fmars.2020.00700, 2020.

Xu, C., Chen, G., Shang, X.-D., and Huang, R. X.: The spatial distribution of sources and sinks of ocean mesoscale eddies, J. Trop. Oceanogr., 32, 37–46, 2013.

Yan, H., Wang, H., Zhang, R., Bao, S., Chen, J., and Wang, G.: The Inconsistent Pairs Between In Situ Observations of Near Surface Salinity and Multiple Remotely Sensed Salinity Data, Earth Sp. Sci., 8, https://doi.org/10.1029/2020EA001355, 2021.
830 Zhou, G., Fu, W., Zhu, J., and Wang, H.: The impact of location-dependent correlation scales in ocean data assimilation, Geophys. Res. Lett., 31, https://doi.org/10.1029/2004GL020579, 2004.

Zweng, M. ., Reagan, J. R., Seidov, D., Boyer, T. P., Locarnini, R. A., and Garcia, H. E.: World Ocean Atlas 2018, Volume 2: Salinity. A. Mishonov, 50pp pp., 2019.