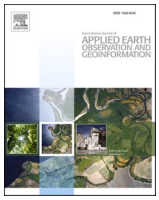
International Journal of Applied Earth Observation and Geoinformation 132 (2024) 104029

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Deep learning-based gap filling for near real-time seamless daily global sea surface salinity using satellite observations

Eunna Jang [a,b,1], Daehyeon Han [b,1], Jungho Im [b,c,d,*], Taejun Sung [b], Young Jun Kim [b]

[a] Korea Ocean Satellite Center, Korea Institute of Ocean Science and Technology, Busan, South Korea
[b] Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea
[c] Artificial Intelligence Graduate School, UNIST, Ulsan, South Korea
[d] Graduate School of Carbon Neutrality, UNIST, Ulsan, South Korea

ARTICLE INFO

Keywords:
SMAP
Sea surface salinity
Gap filling
U-Net

ABSTRACT

Sea surface salinity (SSS) provides crucial information about ocean environments, influencing global hydrological cycles, thermohaline circulation, and climate change. Although L-band passive microwave radiometers have provided satellite-based SSS data, there are gaps due to the limited daily coverage of the sensors. This study proposes a U-Net-based spatial gap-filling model for global SSS using Soil Moisture Active Passive (SMAP) satellite data. The proposed model utilizes SSS swath data from the target and past days to generate daily global SSS maps with full coverage, incorporating only past temporal information. Additionally, bias-corrected data using gradient-boosted regression trees (GBRT) are employed to reduce inherent errors in the SMAP SSS data. We designed 24 schemes using data from the past 3, 5, and 7 days for both GBRT-corrected and original SMAP SSS data, with one to four times oversampling for low-salinity water, where the number of samples is significantly small. Validation results using masked-out pixels indicate that the gap-filling models with GBRT-corrected SSS data from the past 3 and 5 days and four times oversampling yielded the best performance, with root mean square errors (RMSEs) of 0.388 and 0.413 psu, respectively. Compared with the *in situ* Argo data for 2020, the RMSEs were 0.237 and 0.241 psu for the two models, respectively, significantly outperforming the SMAP Level 3 8-day SSS, which requires future data (RMSE of 0.456 psu). Notably, these models successfully filled gaps over coastal areas where low SSS (<30 psu) fluctuated due to freshwater discharge from inland areas. This study proposes an effective, novel gap-filling method for generating bias-corrected global daily SSS without delay, meeting operational needs. Moreover, the proposed gap-filling framework can be applied to other environmental swaths or even grid datasets.

1. Introduction

Sea salinity plays an essential role in the global water cycle as an indicator of evaporation, precipitation, river runoff, ice formation, and melting (Lagerloef and Font, 2010). Monitoring sea surface salinity (SSS) provides critical information about oceanic thermohaline circulation, global hydrological cycles, and climate change (Kim et al., 2023). Since 2009, L-band microwave radiometers [Soil Moisture and Ocean Salinity (SMOS), Aquarius, and Soil Moisture Active Passive (SMAP)] have been used to monitor global SSS from space. The L-band microwave sensor-derived SSS products have a resolution of 25 to 100 km (Dinnat et al., 2019); however, due to their relatively long revisit time (≥3 days) with daily along-track data, they do not cover the global ocean daily. Aquarius ended its mission in June 2015, and since then, SMOS and SMAP have provided SSS observations of the global ocean at various temporal scales (Reul et al., 2020). The Center Aval de Traitement des Données SMOS Expertise Center SSS Level-3 (L3) 8-day running mean SSS products were provided with an average of 9 and 18 days using a slipping Gaussian kernel (Boutin et al., 2018). Furthermore, the Jet Propulsion Laboratory's SMAP SSS L3 daily product was provided with an 8-day running average using Gaussian weight for 3.5 days before and after the target date (Fore et al., 2020; Qin et al., 2020). By November 2021, SMOS L3 products were only available every 4 days, and the SMAP L3 daily product was not provided in real-time

* Corresponding author.
E-mail address: ersgis@unist.ac.kr (J. Im).
[1] The first two authors equally contributed to the paper.

https://doi.org/10.1016/j.jag.2024.104029
Received 5 February 2024; Received in revised form 20 June 2024; Accepted 8 July 2024
Available online 13 July 2024
1569-8432/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

because it requires future data.

Several studies have been conducted to reconstruct SSS products by primarily assimilating *in situ* data into ocean models such as the Nucleus of European Modeling of the Ocean and Regional Ocean Modeling Systems and satellite-retrieved data (Droghei et al., 2018; Martin, 2014; Mu et al., 2019; Nardelli et al., 2016; Zujev et al., 2021). Furthermore, optimal interpolation and three- or four-dimensional variational data assimilation methods have been used to reconstruct SSS. Additionally, Data Empirical Orthogonal Functions and distance weighting are frequently used to reconstruct satellite-retrieved along-track SSS products (Alvera-Azcárate et al., 2016; Li et al., 2019). The Hybrid Coordinate Ocean Model (HYCOM; reanalysis data) provides high spatial resolution (0.08°) SSS data every 3 h; however, it overestimates the SSS of coastal low-salinity water regions (Jang et al., 2022). By taking advantages of data assimilation from multiple sources from both the past and future of the target date, these reconstruction approaches have achieved high performance. However, this assimilation suffers from its innate limitation of lag-time, which fails to produce daily products in near real-time.

With recent advances in deep learning, there have been studies on the reconstruction of satellite data. Convolutional neural networks (CNNs) have played a key role in spatial modeling and have been adopted in various studies for reconstruction. Using deep learning, Wu et al. (2022) produced an all-weather gapless MODIS land surface temperature dataset using CNN and generative adversarial network (GAN). Moreover, Xing et al. (2022) reconstructed the MODIS normalized difference snow index using U-Net and partial convolutions. For oceanic research, Barth et al. (2020) suggested a reconstruction model of sea surface temperature (SST) from an advanced very high-resolution radiometer using a data interpolating convolutional autoencoder (DINCAE), which was further improved by Barth et al. (2022). DINCAE is based on CNN with an encoder-decoder structure similar to U-Net. As DINCAE utilizes and predicts SST anomalies to reconstruct time-series data, the temporal variability of SST was calculated to reduce a widely known smoothing problem (Barth et al., 2020). DINCAE has been widely adopted to reconstruct other oceanic variables, such as chlorophyll-a (Han et al., 2020; Luo et al., 2022), suspended particulate matter (Barth et al., 2021), and SST, with multiple satellite sensors (Jung et al., 2022). The primary advantage of deep learning is its flexibility and fast inference time, which can address the limitations of traditional approaches.

However, when it comes to SMAP data, most gap-filling studies rely on traditional pixel-level machine learning approaches that do not fully learn the spatial features of the input data, failing to capture the spatial dependencies and relationships inherent in the dataset. Machine learning-based gap-filling of SMAP products has primarily been conducted for soil moisture using random forest (Mao et al., 2019; Tong et al., 2021). Although deep learning approaches have demonstrated success in gap-filling missing values in remotely sensed data, SMAP products have been insufficiently explored due to irregular data coverage and wide missing swath areas (Singh and Gaurav, 2023). Moreover, gap-filling of SSS has rarely been conducted, even with traditional machine learning techniques. Recent studies have repeatedly emphasized the critical importance of gap-filled SSS (Kim et al., 2023; Lee et al., 2022; Nardelli et al., 2016; Shin et al., 2023). As most deep learning-based studies have focused on local regions (Shin et al., 2023), research encompassing the global ocean is also required.

In this context, this study aimed to fill the gaps in global SSS using deep learning. By incorporating U-Net, the proposed model can effectively fill gaps in global SSS data without relying on future data. To the best of our knowledge, this is the first attempt to adopt a deep learning approach for filling global gaps in SSS. The proposed model provides gap-free global SSS data in a timely manner, making it suitable for operational purposes unlike the SMAP L3 global SSS product that is not provided in real time. Moreover, rather than using the original SMAP SSS, we further employed gradient-boosted regression trees (GBRT)-corrected SSS by Jang et al. (2022) (detailed in Section 2.1) as the input SSS for gap filling. The uncertainty of SMAP SSS was significantly reduced when using the GBRT-corrected SSS (Jang et al., 2022). The research objectives can be summarized as: (1) producing real-time daily gap-filled global SSS data based on a deep learning approach, (2) analyzing gap-filling results using intentionally masked original SSS data, *in situ* Argo data (an international program that collects information from inside the ocean), and SMAP SSS products, and (3) demonstrating the feasibility of deep learning in SSS applications. This paper is organized as follows: The data used in this study are introduced in Section 2. Section 3 describes the specific gap-filling and validation methods. Next, the deep learning-based gap-filled SSS findings are analyzed and discussed in Section 4, followed by the conclusion in Section 5.

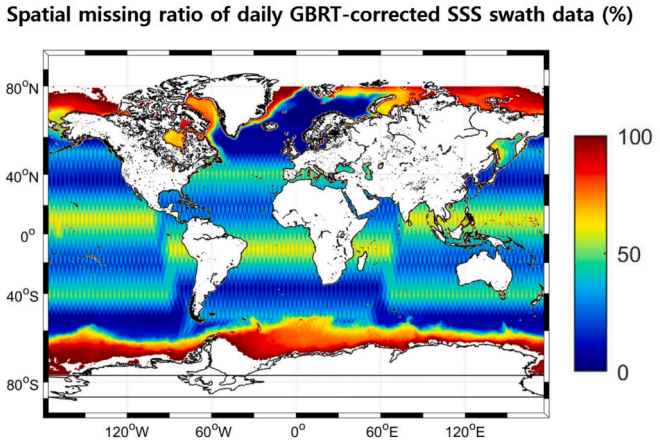


Fig. 1. The percentage of days with no data per pixel for daily GBRT-corrected SSS between April 2015 and December 2020. The pixel spatial resolution is 0.25°, as the GBRT-corrected SSS is based on SMAP SSS data.

2. Data

2.1. GBRT-corrected SSS

Before conducting gap-filling, GBRT-corrected SSS was developed by combining SMAP and HYCOM SSS data using GBRT to enhance the quality of global SSS (Jang et al., 2022). To generate GBRT-corrected SSS, the inputs included data used in the SMAP L2 SSS retrieval algorithm (SMAP brightness temperature of H- and V-polarization and land fraction, HYCOM SSS, sea surface temperature, and wind speed) along with other ancillary data (distance from land and precipitation). The GBRT-corrected SSS outperformed both SMAP and HYCOM SSS across all oceans, providing more accurate data for SSS gap filling. Detailed methods and results of the GBRT-corrected SSS can be found in Jang et al. (2022).

Fig. 1 shows the percentage of days with no data per pixel for daily GBRT-corrected SSS between April 2015 and December 2020 (2,056 days). The GBRT-corrected SSS data above 80°N was excluded due to errors in the input data (Jang et al., 2022). The high missing rate in polar regions is caused by uncertainties related to low sea temperatures and sea ice. Data coverage increases in high-latitude regions due to overlapping footprints by along-tracks. The average daily missing rate was 40.1 % (ranging from 31.4 % to 95.5 % over 2,056 days). We also compared the Jet Propulsion Laboratory SMAP combined active–passive SSS Level 2B (L2B) and L3 8-day running mean SSS (Version 5) products to gap-filled SSS results.

2.2. In situ data

Argo, a global network of drifting profile floats that measures sea

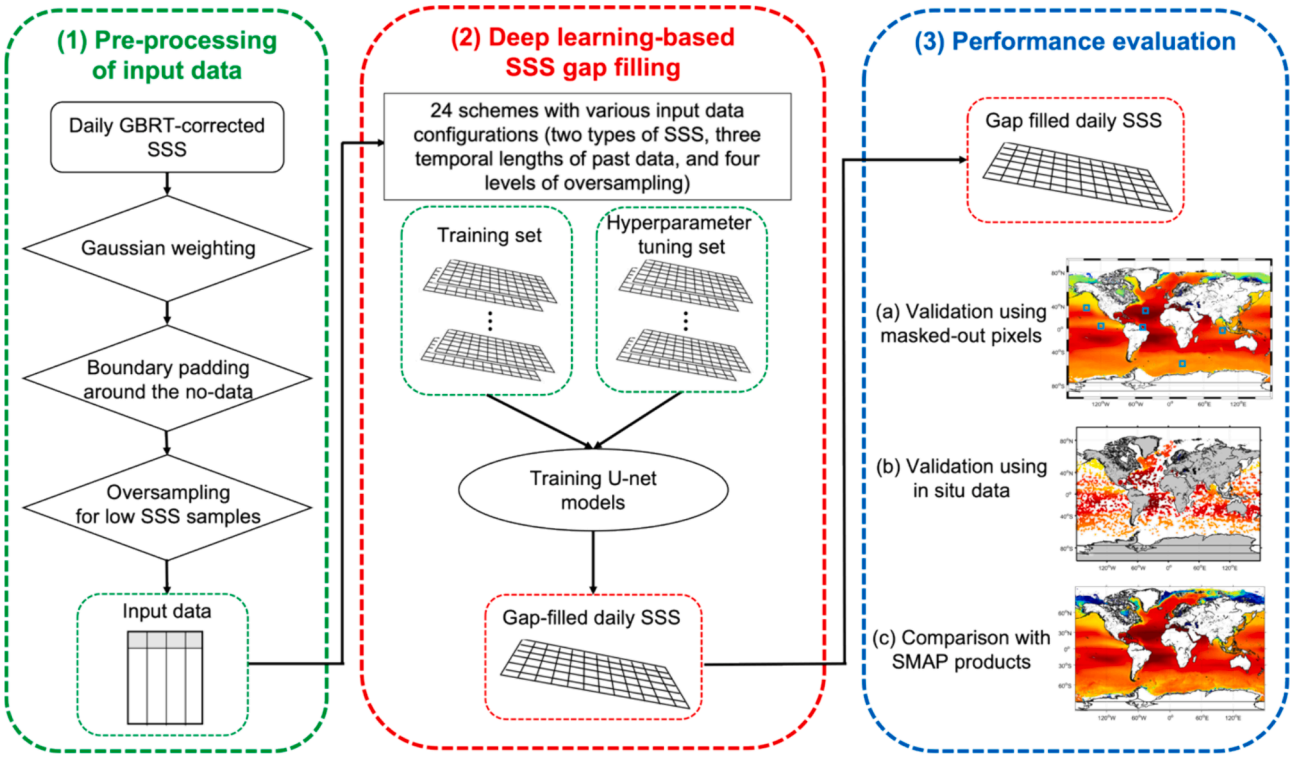


Fig. 2. The flow chart for gap filling of global sea surface salinity (SSS) data.

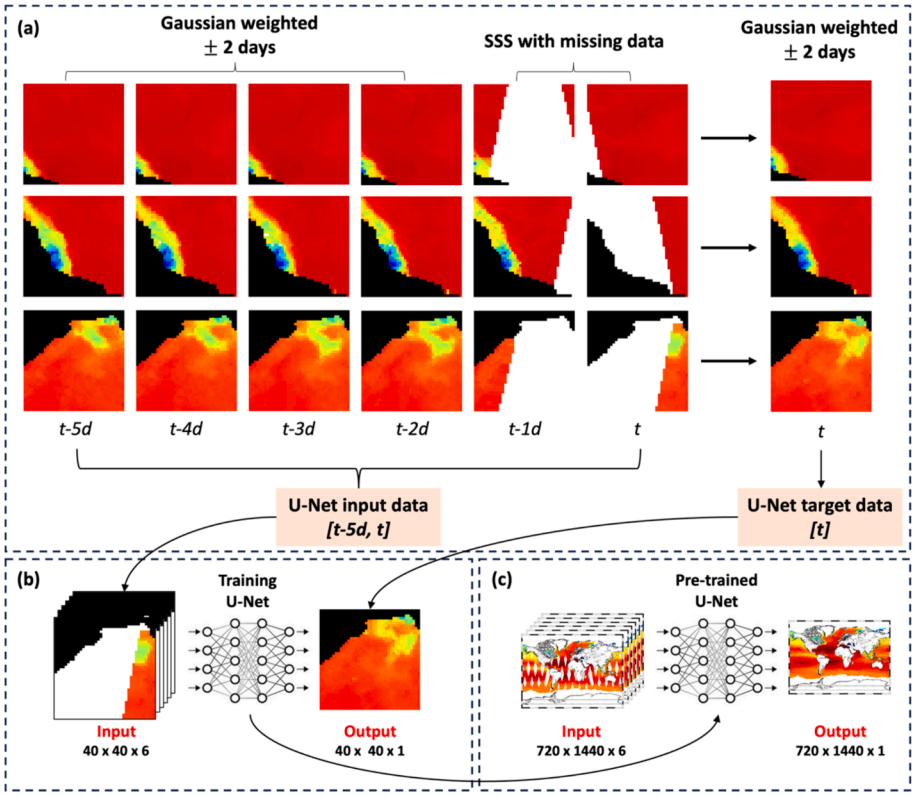


Fig. 3. Comprehensive processes for (a) building patch samples with historical data, (b) training the model, and (c) inferring global SSS using Gaussian weighted ± 2 days and past 5 days for input data as examples. The land mask is represented in black for patch data in (a) and (b).

temperature and salinity in the upper 2000 m of the ocean, provided data obtained from the Coriolis Data Center (http://www.coriolis.eu.org/). Data in delayed mode with quality control flags of 1 or 2 (indicating good or probably good quality) were used (Kolodziejczyk et al., 2021). For colocation with SMAP grid data, Argo SSS data were selected with a time delay of around 6 h and distances between satellite

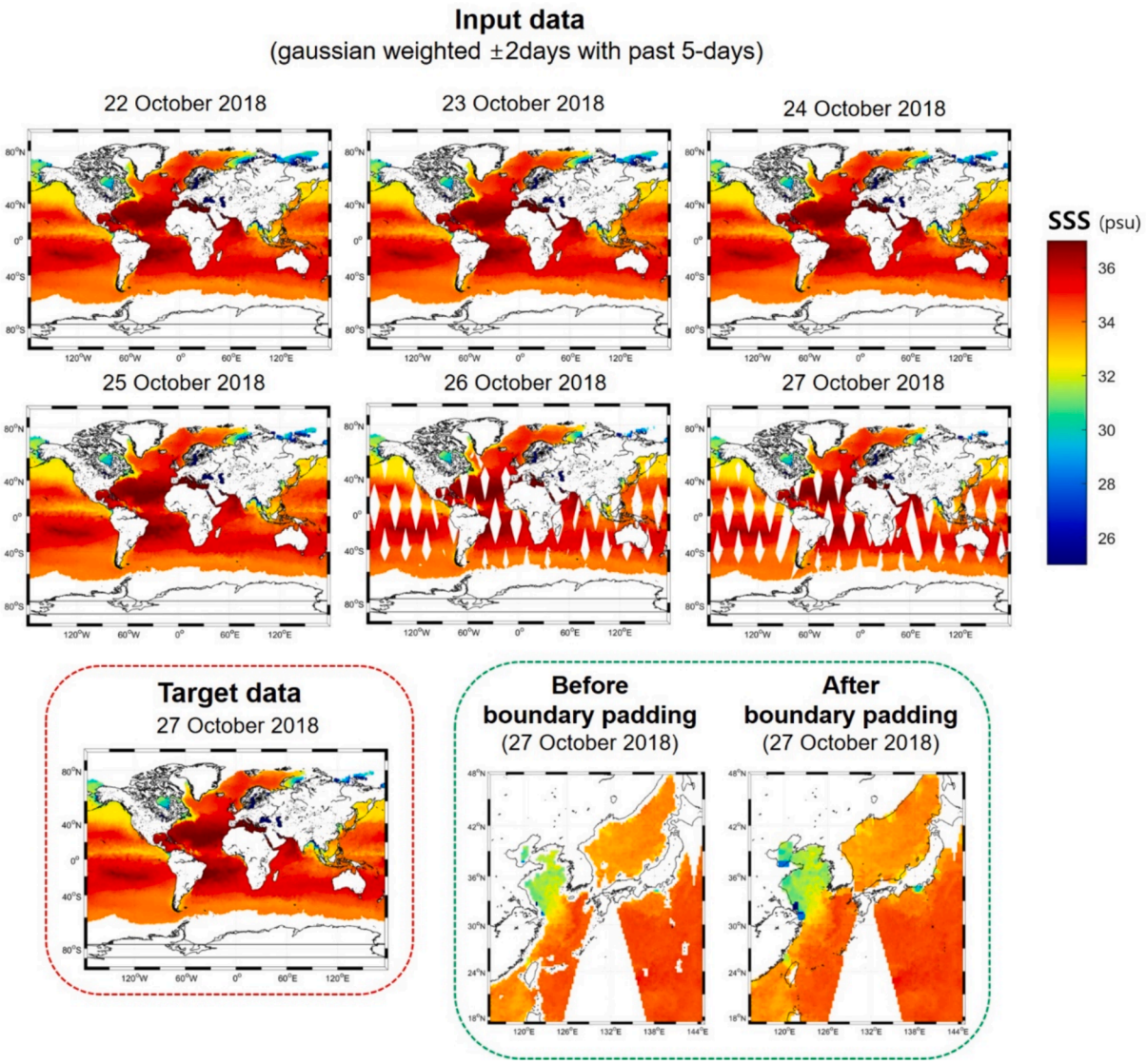


Fig. 4. An example of input and target data and boundary padding when Gaussian-weighted ± 2 days for gap-filled SSS and the past 5 days was used as input data. An example of boundary padding was expanded only in East Asia, as shown in the figure.

and Argo data of less than 12.5 km (based on the SMAP pixel size of 25 km). We used 7,758 global Argo SSS data from 2020 as the validation data.

3. Methodology

The U-Net deep learning model was used to gap-fill the GBRT-corrected SSS product (Jang et al., 2022) to generate daily global SSS data (Fig. 2). The aim of this study is to generate gap-filled SSS data despite some areas being missing at the current timestamp (Fig. 3). To train the gap-filling model, we need both gap-containing and gap-free data pairs, which are difficult to obtain in real-world scenarios. Therefore, gap filling was performed using Gaussian weighting before generating training samples. To minimize the impact of land or coastal pixels where invalid data exist, boundary padding was applied around no-data pixels. Additionally, low SSS samples were oversampled to mitigate the significantly skewed sample distribution. Gaussian weighting and preprocessing steps are described in detail in Sections 3.1 and 3.2.

In addition to the swath data for the target day, past swath scenes were used to create a partially filled input sequence. We tested the preceding 3, 5, and 7 days to determine the effect of the input sequence length. Masked-out pixels and *in situ* data were used to validate the gap-filling model and compare the distribution with the SMAP products. Furthermore, the U-Net model was trained using the original SMAP L2B daily along-track data and compared with the model based on GBRT-corrected SSS.

3.1. Gaussian weighting for sample generation

The first step is to apply Gaussian weighting to the daily along-track data. The proposed U-Net model aims to generate a gap-filled global SSS from the sequence of the along-track SSS data. Therefore, the target data in the U-Net model must be gap-filled. Furthermore, due to the long revisit time and quality flags of SMAP, some pixels from the previous collection times have no value when used as input data; thus, gap-filled past data are required. To generate gap-filled SSS data for past data, we employed the Gaussian weighting method to generate the SMAP 8-day running mean data (Fore et al., 2020; Qin et al., 2020). By utilizing past and future information, the Gaussian weight function (Equation (1) was used to perform the gap filling.

$$G(t) = e^{-t^2/2\sigma^2} \quad (1)$$

where t is the daily time step bounded by $-n \leq t \leq n$ with n being the maximum number of days for past and future data, and σ stands for a standard deviation of Gaussian distribution. The mean value and the current time step are set to zero. We verified the data existence $E(i, j, t)$ for each time step within the range $-n \leq t \leq n$ to exclude missing values in interpolation (Equation (2). To ensure the sum of all Gaussian function values equals 1, normalized Gaussian weights $w(i, j, t)$ were calculated for only the valid time steps for each pixel at location (i, j) as shown in Equation (3). Finally, we used Gaussian weights to sum and generate gap-filled SSS data for model training (Equation (4). Since the Gaussian window places more weight on data closest to the center, which in this study is the target date set as zero, it enables the generation of accurate and consistent results when past and future data are available.

$$E(i,j,t) = \begin{cases} 0, if SSS(i,j,t) isNull \\ 1, if SSS(i,j,t) exists \end{cases} \quad (2)$$

$$w(i,j,t) = \frac{G(t)}{\sum_{t=-n}^{n} G(t) \bullet E(i,j,t)} \quad (3)$$

$$Gaussian\_SSS(i,j,0) = \sum_{t=-n}^{n} (w(i,j,t) \bullet SSS(i,j,t)) \quad (4)$$

Although the SMAP algorithm uses data from ± 3.5 days from the target date, we used Gaussian weighted data from ± 2 days. Fig. 3 shows an example of input and target data using Gaussian weighted ± 2 days for the gap-filled SSS and the past 5 days of the SSS sequence with the target date. For example, when future data were available, gap-filled data with Gaussian weighting were used as input data between October 22 and October 25, with October 27 as a target date. Next, when future data were unavailable between October 26 and October 27, along-track data were used. Notably, gap filling with Gaussian weighting on the target date was performed only during the training phase. Therefore, the gap-filling model could generate real-time daily global SSS data.

3.2. Data pre-processing

After the Gaussian weighting, a boundary padding was adopted around the no-data pixels. Due to the nature of the convolutional process, there is a possibility that dummy values over land areas might be used near coastlines, which is undesirable. For this study, a value of 0 was assigned to pixels lacking data over land, coastal, and polar regions, as well as in the along-track gap areas. These zero-filled pixels along the coastlines and along-track gaps can cause underestimation during the convolutional process. To avoid a significant disparity between the dummy zero and neighboring SSS values, we padded two pixels adjacent to pixels with no data to the nearest SSS value (Fig. 4).

The U-Net can accept various input and output sizes under specific conditions detailed in Section 3.3, which is one of its primary advantages. Owing to this flexibility, we used different SSS scales for the training and inference phases. Considering the extensive land areas in global SSS, it is more efficient to focus on training only the relevant and valid ocean areas rather than the entire global dataset (Fig. 3). Moreover, given the constraints on the number of days in the study period for global-level sampling, patch-level sampling offers significant advantages in terms of sample size. Therefore, we cropped the pre-processed global SSS data into small patches to exclude inland areas and increase the number of samples. Following a series of tests, the patch size was set to 40 × 40 pixels, corresponding to a spatial coverage of 1,000 × 1,000 km with a 20-pixel moving window. As the missing ratio of SSS varies between the coastal region and the open ocean, we set different sampling conditions to avoid sparse patches that were almost filled with missing values. Therefore, the patch size was 40 × 40 pixels, each containing 1,600 pixels. Over the open ocean, only patches with no missing values were collected to achieve the highest gap-filling performance. In contrast, sample patches with missing values in coastal or inland areas are unavoidable.

Table 1
The number of sample patches per SSS range. The representative SSS value was calculated by averaging each 40 × 40 sample patch. The plain text indicates the count of original samples, while numbers that have been oversampled four times are presented in italics and within parentheses.

Input past days	Total number of patch samples per mean SSS range					
	< 20	20–25	25–30	30–35	> 35	Total
3	16,580 *(66,320)*	3,537 *(14,148)*	12,809 *(51,236)*	218,109	108,946	359,981
5	15,451 *(61,804)*	3,080 *(12,320)*	11,227 *(44,908)*	195,126	104,948	329,832
7	14,653 *(58,612)*	2,816 *(11,264)*	10,195 *(40,780)*	179,870	101,998	309,532

To minimize the performance degradation caused by sparse patch samples in the open ocean, patches with less than 50 % missing pixels were collected when at least one land pixel was included. Data between April 2015 and December 2019 were used as the training datasets, whereas data from 2020 were used as the test dataset. Table 1 lists the total number of sample patches extracted from the training dataset. In training dataset, 80 % of the total samples were randomly selected to fit the model, while the remaining 20 % were used to determine the optimal training iteration. The skewed distribution of training samples by SSS values could significantly affect model performance (Buda et al., 2018; Mellor et al., 2015). As most SSS samples were collected in the open ocean, the sample distribution was skewed toward relatively high SSS values of over 30 (Table 1). The lack of low-SSS samples can result in a performance drop, especially over enclosed seas (such as the Black Sea, and Baltic Sea) and coastal areas where salinity is generally low. Therefore, sample patches with mean SSS < 30 psu were augmented up to four times with random noise between −0.5 and 0.5 added to each input pixel and target SSS samples as Table 1.

3.3. U-Net

Due to its ability to extract spatial features, the convolutional neural network (CNN) has recently become one of the most effective methods for spatial modeling (Ahishali et al., 2021; Chen and Tsou, 2022; Kang et al., 2022; Lee et al., 2021; Zhou et al., 2024). Features in a CNN consist of a set of $n \times n$ filters in multiple convolutional layers that are gradually fitted to the given dataset from random initial weights. Similar to a moving window, convolutional filters calculate dot products over the corresponding area of an input image. Another set of convolutional filters performs the same calculation with the output from the previous layer serving as an input for the subsequent layer. This process is repeated throughout the entire convolutional layer. An activation function is used after each convolutional layer to add nonlinearity to the deep learning model because the dot product in a convolutional layer is a linear operation. We used the rectified linear unit (ReLU) for an activation function due to its simple structure, rapid computation, and robust performance (Ma et al., 2023; Nair and Hinton, 2010; Zhang et al., 2023).

U-Net (Ronneberger et al., 2015) is one of the most widely adopted image-to-image models in diverse remote sensing fields (Dedring and Rienow, 2024; Gao et al., 2021; Masolele et al., 2022; Mukherjee and Liu, 2021; Zhang et al., 2024). U-Net converts input images into key abstract features in the encoder and then rebuilds them in the decoder. The encoder captures important details at different levels, gradually reducing the image size to focus on key features. The decoder then rebuilds the image, increasing the size back to the original while keeping important details. Skip connections link layers at the same level in the encoder and decoder to prevent the loss of details of original data,

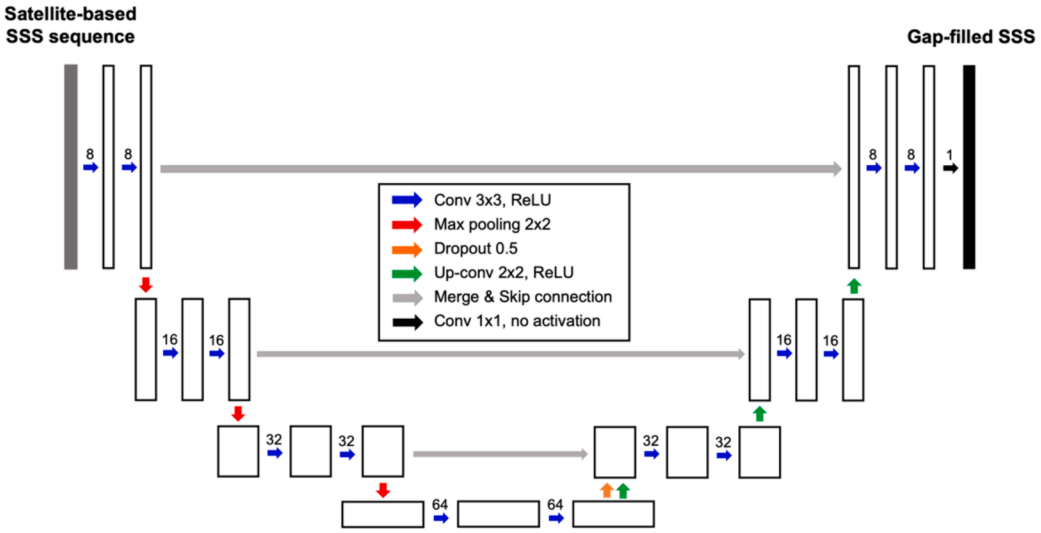


Fig. 5. The study's model structure. Numbers upon the blue arrow represent the number of convolutional filters per level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of the schemes used in this study. Schemes were designed with various input data configurations (i.e., either GBRT-corrected or original SMAP SSS data, three temporal lengths [3, 5, and 7 days] of past data, and four levels [one to four times] of oversampling for low salinity water).

		U-Net (without oversampling)	U-Net (twice oversampling)	U-Net (thrice oversampling)	U-Net (four times oversampling)
GBRT-corrected SSS	Past 3 days	U3G1	U3G2	U3G3	U3G4
	Past 5 days	U5G1	U5G2	U5G3	U5G4
	Past 7 days	U7G1	U7G2	U7G3	U7G4
SMAP L2B SSS	Past 3 days	U3S1	U3S2	U3S3	U3S4
	Past 5 days	U5S1	U5S2	U5S3	U5S4
	Past 7 days	U7S1	U7S2	U7S3	U7S4

allowing U-Net to produce accurate and detailed images. Additionally, a U-Net model can handle arbitrary data sizes that are multiples of k^n, when n is the number of pairs of k-fold upsample and downsample layers. We adopted U-Net for its flexibility and the comparable performance demonstrated in previous studies.

In this study, the proposed model consists of four spatial levels, with the size halved at each successive level. Each level has two convolutional layers with 3×3 filters and 8, 16, 32, and 64 convolutional filters per level. The ReLU was used as an activation function after each convolutional layer. To obtain the different scales in the encoder and decoder,

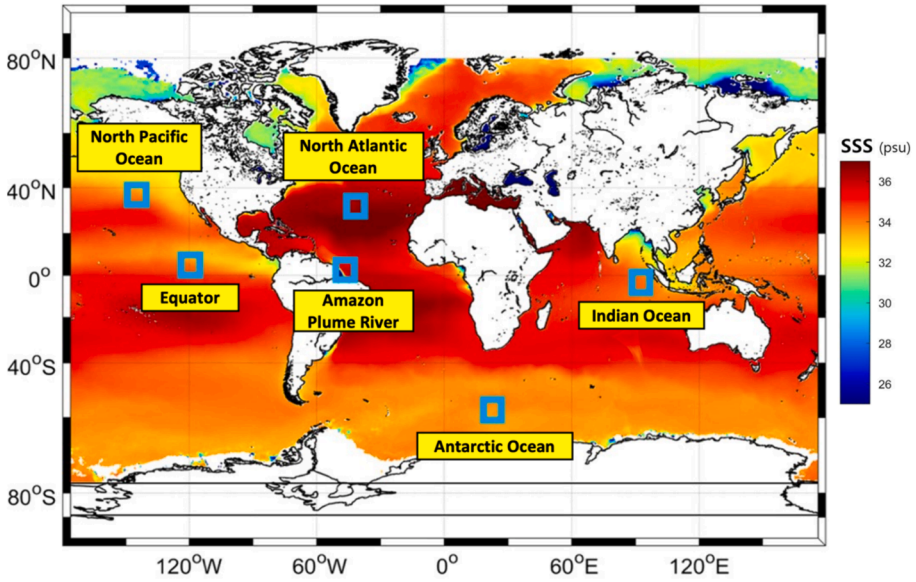


Fig. 6. Six regions (North Pacific Ocean, North Atlantic Ocean, Equator, Amazon River Plume, Indian Ocean, and Antarctic Ocean) with masked-out pixels (blue boxes) for model performance evaluation. The background is the average gradient boosted regression trees (GBRT)-corrected SSS between April 2015 and December 2020. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

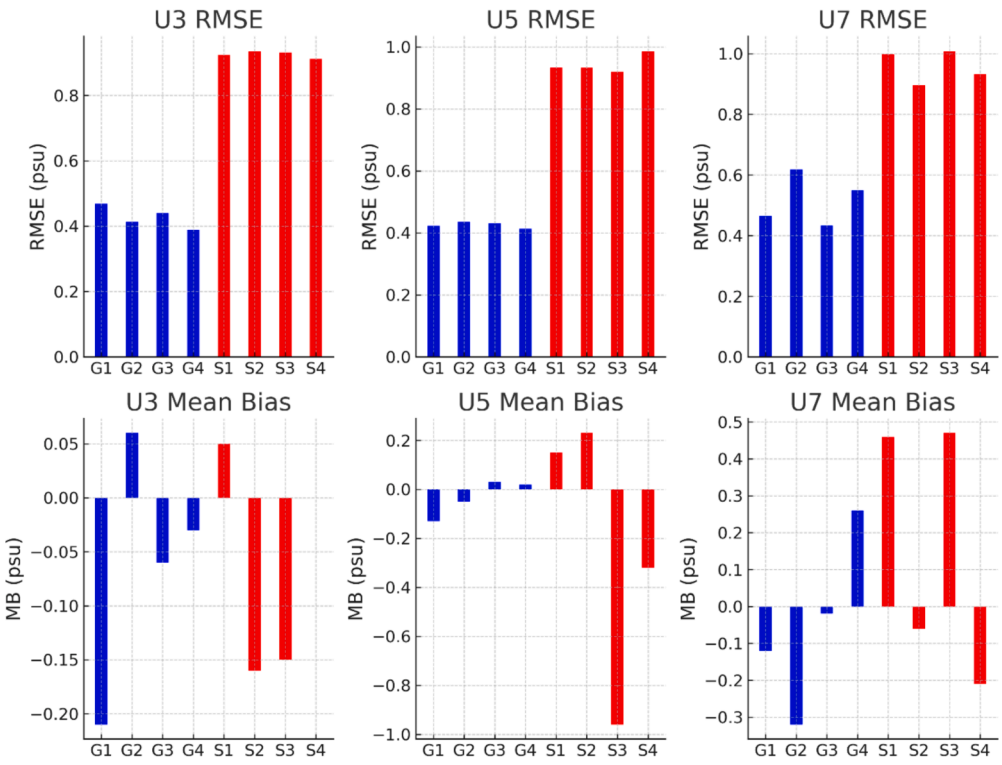


Fig. 7. The root mean square error (RMSE) and mean bias (MB) for masked-out pixels of the results of U-Net models *versus* original pixel values with test datasets for 2020. The results were derived from the U-Net model using GBRT-corrected SSS and SMAP L2B daily along-track data. The blue bar represents the results from GBRT-corrected SSS (G1-G4) and the red bar corresponds to SMAP L2B SSS (S1-S4). Refer to Table 2 for schemes. In each subplot, U stands for U-Net, and the following number is for number of past days used for gap filling. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

three max pooling and up-convolutional layers with 2 × 2 filter size were used in the encoder and decoder, respectively (Fig. 5). As all trainable parameters in U-Net are simply a set of convolutional filters and their respective biases, U-Net can handle various data sizes because the convolutional operations are applied locally and do not depend on the overall size of the input data. Specifically, with three max pooling and up-convolution layers using a 2 × 2 filter, the minimum data size must be a multiple of 2^3, or 8. Thus, any data size that is a multiple of 8 is feasible. By maximizing this benefit, the proposed model can be trained with patch-level samples (for example, 40 × 40) that satisfy the condition for gap filling (the composite area of the input data should be equal to that of the target data). The error between the prediction and reference values is calculated using the loss function, which should be minimized. Randomly initialized convolutional filters were gradually fitted as the number of iterations increased during the backpropagation procedure based on stochastic optimization. In this study, we used the mean squared error as the loss function. Furthermore, we used an adaptive momentum optimizer (Kingma and Ba, 2014), which is one of the most widely used methods due to its robust performance (Han et al., 2023; Wu et al., 2022). The proposed model was trained with an Nvidia RTX 3090 graphics processing unit (GPU) to speed up the training time, with a batch size of 128. We terminated the model training when the validation loss did not decrease over three iterations, and the maximum number of epochs was set to 100 to prevent overfitting. At the time of inference, the fitted model took an input of 720 × 1440 to generate a global gap-filled satellite-retrieved SSS directly. Fig. 5 shows the U-Net structure used in this study.

We designed 24 schemes with various input data configurations (i.e., either GBRT-corrected or original SMAP SSS data, three temporal lengths [3, 5, and 7 days] of past data, and four levels [one to four times] of oversampling for low salinity water) to examine the effects of various factors for the gap-filling models. Schemes, expressed using acronyms, are summarized in Table 2: For example, the U-Net model using GBRT-corrected SSS data from the past 5 days and four times oversampling is referred to as U5G4 and the U-Net model using SMAP SSS data from the past 3 days and without oversampling is referred to as U3S1.

3.4. Performance evaluation

The training and test datasets were separated by year (training data from April 2015 to December 2019 and test data from 2020) to evaluate model performance. Two methods were used to assess the gap-filled SSS: using masks of the original SSS data and *in situ* Argo data. First, the original SSS values for the selected regions were masked out before model training, and the gap-filled values of the regions were compared with the original values. Six regions (North Pacific Ocean, North Atlantic Ocean, Equator, Amazon River Plume, Indian Ocean, and Antarctic Ocean) were chosen based on SSS distribution and the missing data rates (Fig. 6). Values were eliminated with a 20 × 20 window size every 30 days in the 2020 test dataset. We eliminated 28,057 pixels for the validation. Second, the gap-filled SSS data from the U-Net model were further validated using *in situ* Argo data. We used 7,758 global Argo SSS data points from 2020 as the validation data. The gap-filled SSS was assessed using three statistical metrics: root mean square error [RMSE; Equation (5)], coefficient of determination [R^2; Equation (6)], and mean bias [MB; Equation (7)].

$$\mathrm{RMSE} = \sqrt{\frac{\sum (M-O)^2}{n}} \tag{5}$$

$$R^2 = \left(\frac{\sum (M-\overline{M})(O-\overline{O})}{\sqrt{\sum (M-\overline{M})^2 \sum (O-\overline{O})^2}} \right)^2 \tag{6}$$

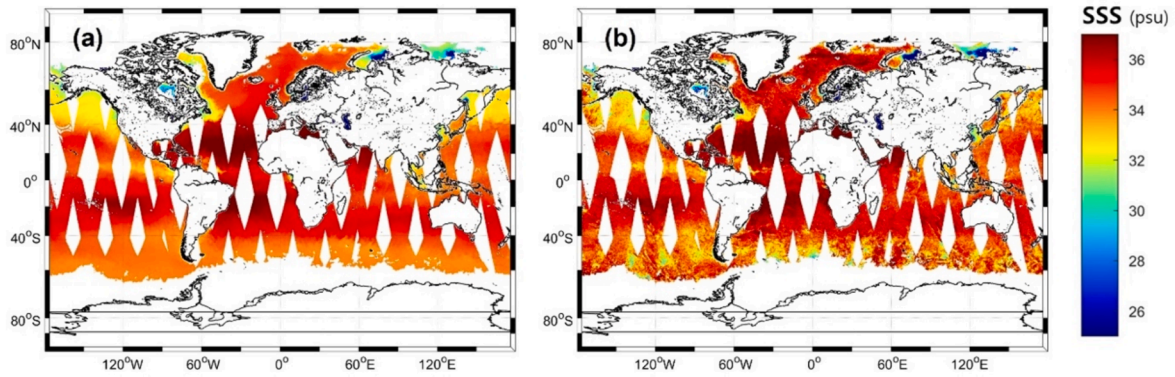


Fig. 8. (a) GBRT-corrected SSS and (b) SMAP SSS for July 19, 2020.

$$\mathrm{MB} = \frac{1}{n}\sum (M - O) \tag{7}$$

where M is the gap-filled SSS data, O is the masked area of the original SSS data or the *in situ* Argo data, n is the total number of data points, and $\overline{M}$ and $\overline{O}$ are the mean values of each SSS data set.

Additional comparisons with the original SMAP SSS product were conducted. To justify the use of the GBRT-corrected SSS, the U-Net model trained with the SMAP L2B daily along-track data was compared with the model based on the GBRT-corrected SSS (the proposed model). Furthermore, the gap-filled SSS was compared with the SMAP L3 8-day running mean SSS data incorporating future observations to demonstrate that gap-filling with only past data can yield robust performance.

4. Results and discussions

4.1. Comparison with masked-out pixels

The validation results using the masked-out pixels for 2020 (Fig. 6) are summarized in Fig. 7. U-Net models with GBRT-corrected SSS input data for the past 3, 5, and 7 days were tested. Oversampling improved the gap-filling results using GBRT-corrected past SSS data for 3 and 5 days as input data. The R^2, RMSE, and MB of the U5G4 model were 0.94, 0.413 psu, and 0.02 psu, respectively, while those of the U5G1 model were 0.94, 0.423 psu, and −0.13 psu, respectively. As the amount of oversampled data increased, smoothing in the resultant SSS was mitigated, increasing accuracy. Increasing the number of low SSS (<30 psu) samples up to four times, as detailed in Section 3.2, effectively improved performance in both quantitative and qualitative evaluations. This is further discussed in Section 4.3 with distribution maps. However, oversampling did not always improve performance when using GBRT-corrected past data for 7 days. This suggests that too long past sequences can fail to take advantage of oversampling because of accumulated random noise and high dependency on past SSS data. We observed no significant difference in accuracy when the SSS data from the past 3 and 5 days were used as input data. However, compared to the model using the past 3 days, the model using the past 5 days produced gap-filled results closer to the original values; the results are discussed in detail in Section 4.3.

The results of the U-Net model trained using SMAP L2B daily along-track data were compared with those obtained using the GBRT-corrected SSS (Fig. 7). The RMSE of the U-Net model with the SMAP SSS ranged from 0.896 to 1.007 psu, whereas the model with GBRT-corrected SSS ranged from 0.388 to 0.618 psu. SMAP L2B data have significant spatial noise and inconsistency, particularly in areas of high uncertainty, such as polar regions with low sea surface temperatures (Lagerloef and Font, 2010). When comparing GBRT-corrected SSS with SMAP SSS (Fig. 8), the higher spatial inconsistency in the SMAP SSS influenced the image learning of the U-Net model, resulting in relatively low accuracy of the gap filling. We used the GBRT-corrected SSS to improve this limitation of SMAP SSS through the synergistic use of HYCOM reanalysis model data and other ancillary data based on GBRT (Jang et al., 2022), leading to successful gap-filling. As the evaluation over masked pixels measures the performance of gap-filling models (Zhu et al., 2017), the model design plays a crucial role. However, data noise also significantly affects the performance (Fig. 7). Therefore, studies on gap filling should accompany the enhancement of sea surface salinity (SSS).

4.2. Comparison with in situ data

Table 3 compares the values of the pixel gaps filled by the U-Net models with the Argo *in situ* data. The results of the models were further compared with the SMAP L3 8-day running mean SSS. Similar to the validation based on the masked-out pixels, oversampling improved the

Table 3
The root mean square error (RMSE) and mean bias (MB) for the results of U-Net models *versus in situ* data with validation datasets for 2020. The lowest RMSE per input data and past day are in bold.

SMAP L3 8-day running mean SSS				RMSE (psu) 0.456				MB (psu) 0.09
	GBRT-corrected SSS				SMAP L2B along-track SSS			
	U3G1	U3G2	U3G3	U3G4	U3S1	U3S2	U3S3	U3S4
RMSE (psu)	0.316	0.246	0.244	**0.237**	0.648	0.654	0.674	**0.646**
MB (psu)	−0.21	0.07	−0.06	0.01	0.09	−0.17	−0.16	0.07
	U5G1	U5G2	U5G3	U5G4	U5S1	U5S2	U5S3	U5S4
RMSE (psu)	0.256	0.245	0.247	**0.241**	0.667	0.645	**0.631**	0.680
MB (psu)	−0.09	−0.07	0.07	0.03	0.14	0.19	−0.04	−0.23
	U7G1	U7G2	U7G3	U7G4	U7S1	U7S2	U7S3	U7S4
RMSE (psu)	0.263	0.505	**0.238**	0.384	0.809	**0.627**	0.813	0.650
MB (psu)	−0.11	−0.33	0.00	0.29	0.50	−0.04	0.50	−0.14

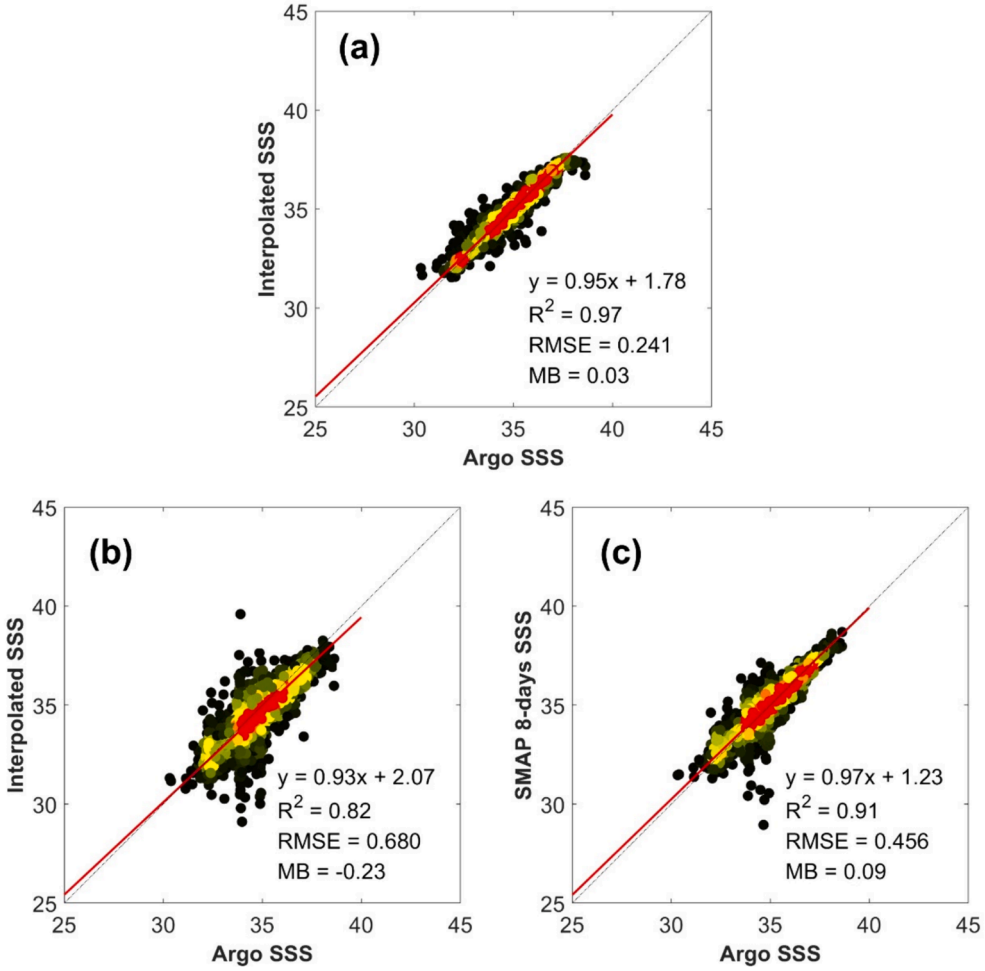


Fig. 9. Scatter plots of Argo *in situ* data *versus* results of U-Net models using (a) U5G4 (with GBRT-corrected SSS) and (b) U5S4 (with SMAP L2B SSS), and (c) SMAP L3 8-day running mean SSS. Refer to Table 2 for schemes.

gap-filling results. The R^2, RMSE, and MB of the U5G4 model were 0.97, 0.241 psu, and 0.03 psu, respectively, whereas those of the U5G1 model were 0.97, 0.256 psu, and −0.09 psu, respectively. For *in situ* data-based validation, the U-Net models using GBRT-corrected SSS were more accurate than those using SMAP L2 SSS. Improving SSS by GBRT (Jang et al., 2022) likely enhanced gap-filling through the bias correction process based on *in situ* measurements, as described in Section 2.1. Li et al. (2019) gap-filled weekly near-global SSS using the Aquarius L2 along-track SSS data based on a weighted average fitting approach, resulting in accuracy comparable to our U5G4 model. It should be noted that the U5G4 model proposed in this study has a higher temporal resolution (i.e., daily) than the Aquarius gap-filled SSS (i.e., weekly), capturing detailed temporal variations in SSS.

Fig. 9 depicts scatterplots of Argo *in situ* data *versus* SMAP L3 8-day running mean SSS and the results of the U-Net models using GBRT-corrected and SMAP L2B SSS (i.e., U5G4 and U5S4). The U-Net models based solely on past GBRT-corrected SSS outperformed the gap-filled SMAP 8-day mean SSS based on past and future data. The results based on SMAP SSS showed that U5S4 (Fig. 9b) performed worse than SMAP L3 SSS (Fig. 9c) in terms of R^2 and RMSE. This suggests that Gaussian weighting could be effective if future data were available. Since Gaussian weighting was employed to construct training samples from historical data, this validates the rationale of Gaussian weighting in sample building for this study. By integrating GBRT-based correction (Jang et al., 2022), the proposed model outperformed SMAP L3 SSS (Fig. 9a), eliminating the need for future data in near real-time applications.

Time-series validation of the proposed U-Net models for gap filling was conducted using Argo data periodically measured in four selected areas: the Amazon River plume and the Bay of Bengal, where low salinity waters are primarily observed; the Yellow Sea, where high and low salinity waters are dynamically mixed; and the equatorial region, where high salinity waters predominate (Fig. 10). The results of U5G4 and U5S4, and SMAP L3 8-day running mean SSS were compared. In April, the U5G4- and U5S4-based SSS data in the Amazon River plume had a similar temporal pattern, but were 1 psu lower than the Argo data (Fig. 10b). In the Argo location, low salinity water and high salinity water were mixed on April 14, so the difference between Argo and gap-filled SSS data was large due to the smoothing effect caused by the surrounding pixels in the gap-filled SSS. Between the two models, U5G4 performed slightly better than U5S4. On the other hand, the SMAP L3 8-day running mean SSS did not exhibit the temporal pattern of low salinity water (Fig. 10b). Similar results were found for the Bay of Bengal (Fig. 10c). The SMAP L3 8-day running mean SSS estimated low salinity water (28–29 psu) in this region but did not exhibit SSS variations over the dates similar to Argo or gap-filled SSS data. Note that the 8-day running mean often results in a smoothed SSS by considering data from multiple days, which may contain variations. The SSS gap-filled by Alvera-Azcárate et al. (2016) using the Data Empirical Orthogonal Function in the North Atlantic Ocean and the Mediterranean Sea exhibited a similar difference when compared to the SMOS SSS in the Douro and Gironde River plumes. The U5G4 model proposed in this study showed similar accuracy with previous studies focusing on river plumes with dynamic changes (Alvera-Azcárate et al., 2016). In the

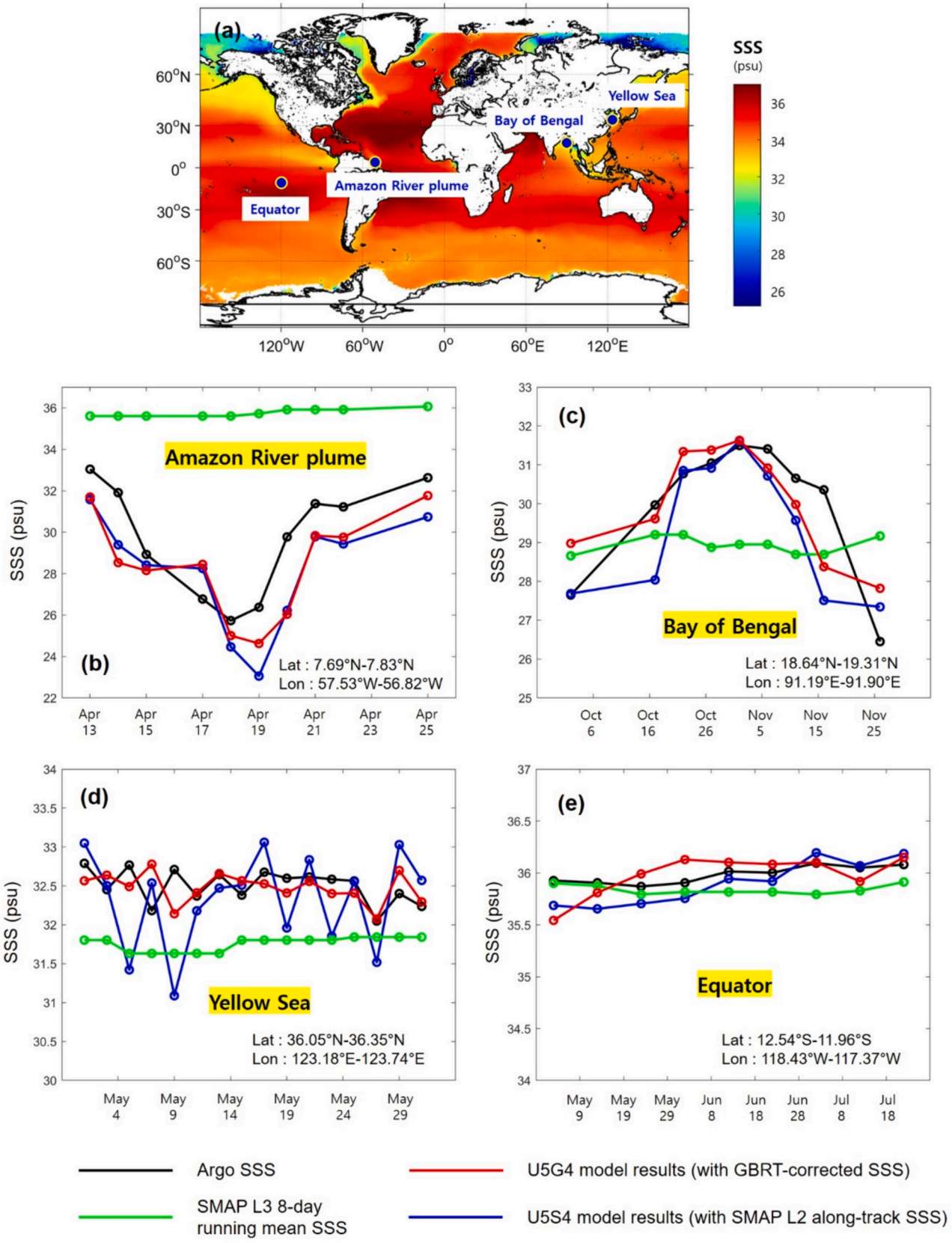


Fig. 10. Day-SSS plots and (a) the location of *in situ* Argo data used for temporal comparison marked as blue circles. The x-axis of the plots is day of 2020. The Argo data were compared to the gap-filled SSS using the proposed U5G4 model, the U5S4 model, and the SMAP L3 8-day running mean SSS ((b), (c), (d), and (e)). The background of (a) is the averaged GBRT-corrected SSS from April 2015 to December 2020. The range of locations for the Argo floats used in each region is indicated in the graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

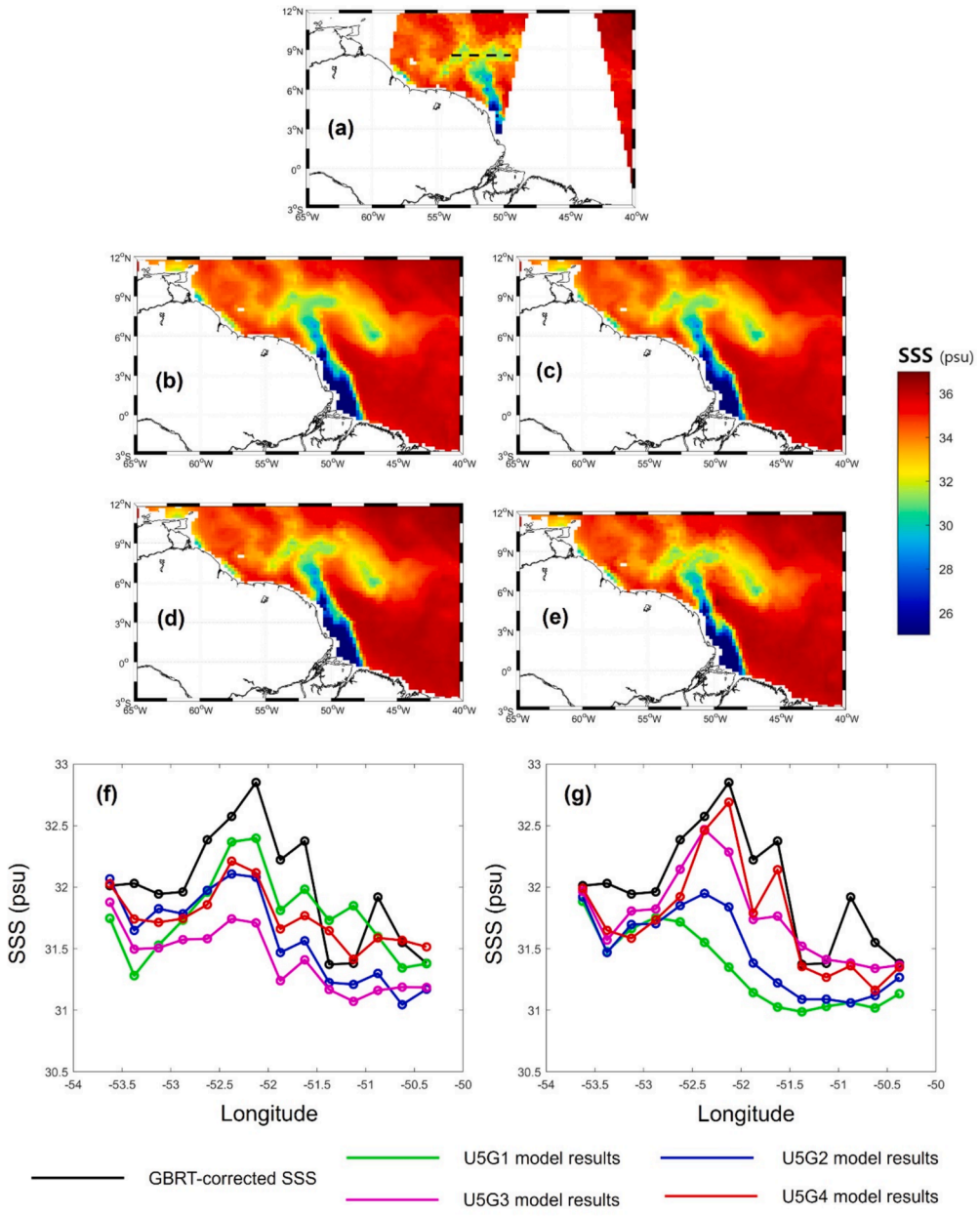


Fig. 11. Spatial distribution and comparison plots of the SSS data gap-filled with different oversampling levels in the Amazon River plume region for July 29, 2020: (a) GBRT-corrected SSS and (b) U5G1 without oversampling, (c) U5G2 with twice oversampling, (d) U5G3 with thrice oversampling, and (e) U5G4 with four times oversampling using data from the past 5 days as input. The plots (f) and (g) compare the results of each model for pixels along the black dashed line in (a) on the same latitude using data from the past 3 and 5 days, respectively. The x-axis of the plot represents the longitude of each pixel.

Yellow Sea, where low salinity water and high salinity water are dynamically mixed, the U5G4 model performed better than the U5S4 model and the SMAP L3 8-day running mean SSS (Fig. 10d). In the equatorial region, the Argo and gap-filled SSS values differed by less than 0.5 psu (Fig. 10e), and both models successfully gap-filled high salinity waters without significant temporal variations. These temporal validations over multiple regions revealed that the U5G4 model using the GBRT-corrected SSS resulted in better performance than the U5S4 model using the SMAP L2 daily along-track data and the SMAP L3 8-day running mean SSS.

4.3. Spatial distribution of the gap-filled SSS

The spatial distribution of the SSS gap-filled by the U-Net models based on the past 5 days in the Amazon River plume region for July 29, 2020, is depicted in Fig. 11. Compared with the original GBRT-corrected SSS (Fig. 11a), the smoothing pattern of low-salinity water areas in the gap-filled data (Fig. 11b–11e) was mitigated by increasing the oversampled data. Fig. 11f and 11 g show the SSS along a row of pixels with the same longitude (black dashed line in Fig. 11a) when input data from the past 3 and 5 days were used in the models, respectively. When data from the past 3 days were used as inputs to the models, all four oversampling levels resulted in significant deviations from the original GBRT-corrected SSS values. However, when using data from the past 5 days, greater oversampling resulted in SSS values closer to the original GBRT-corrected values. The outcome of four times oversampling (the red line in Fig. 11g) was comparable with that of the original GBRT-corrected SSS (black line in Fig. 11g). Samples of < 30 psu per patch were oversampled; however, the number of low-salinity pixels was substantially smaller than that of the other pixels. In the patches, the

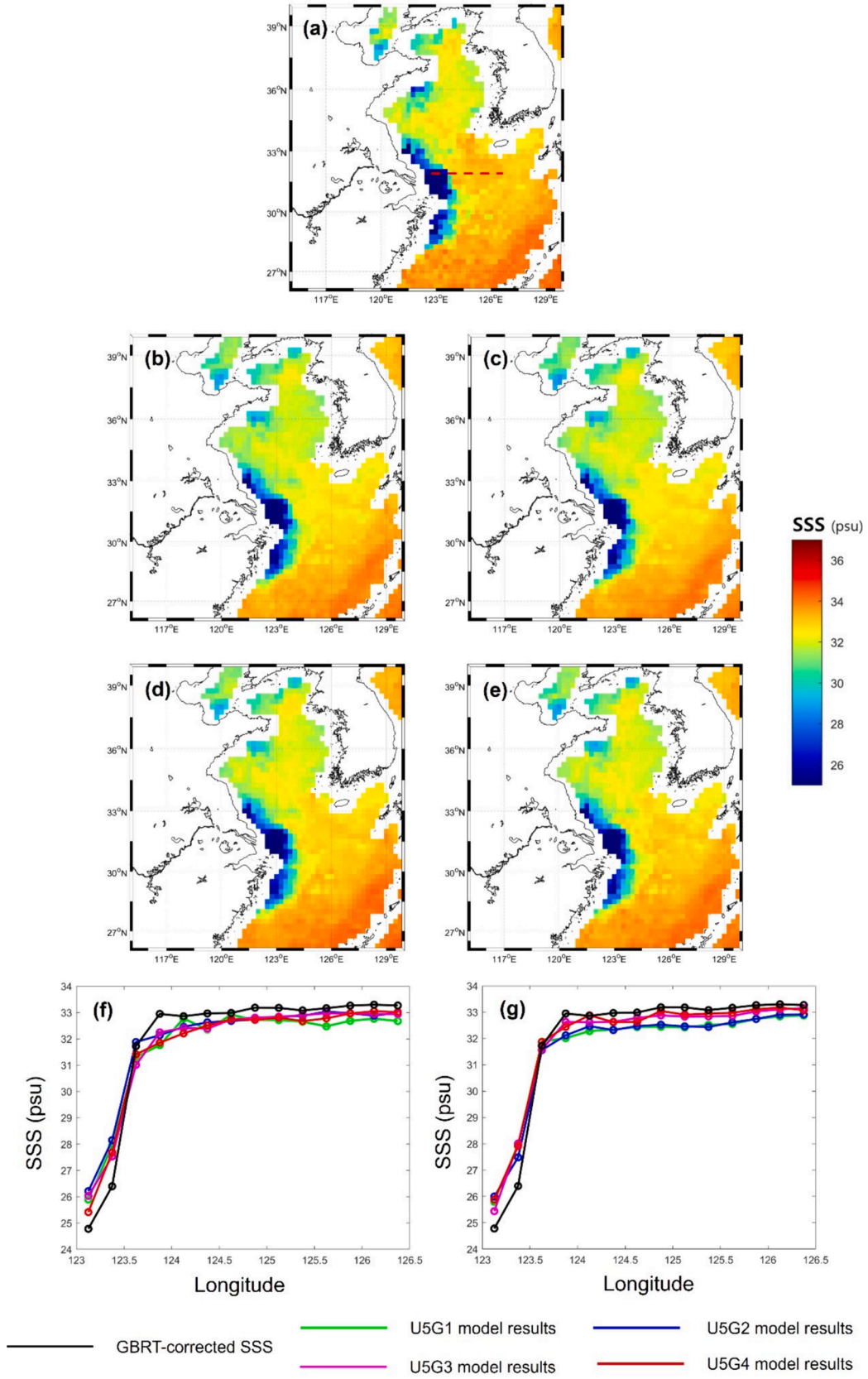


Fig. 12. Spatial distribution and comparison plots of the SSS data gap-filled with different oversampling levels in the East China Sea for July 19, 2020: (a) GBRT-corrected SSS and (b) U5G1 without oversampling, (c) U5G2 with twice oversampling, (d) U5G3 with thrice oversampling, and (e) U5G4 with four times oversampling using data from the past 5 days as input. The plots (f) and (g) compare the results of each model for pixels along the red dashed line in (a) on the same latitude using data from the past 3 and 5 days, respectively. The x-axis of the plot represents the longitude of each pixel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

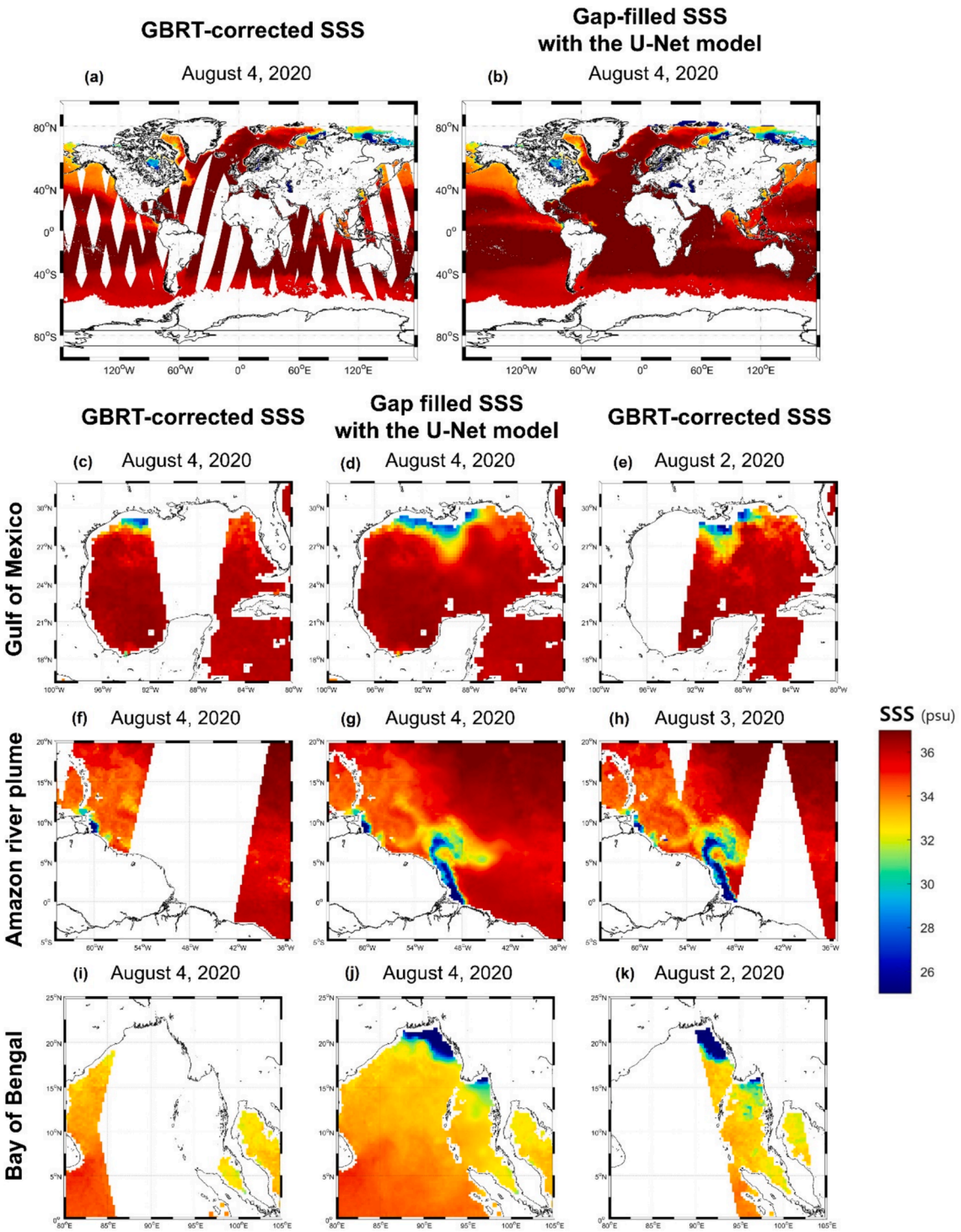


Fig. 13. Spatial distribution of the (a) GBRT-corrected SSS and (b) gap-filled SSS with the U-Net model using data from the past 5 days and four times oversampling (U5G4) for the global ocean on August 4, 2020. Three low salinity regions are shown in detail: the Gulf of Mexico (c, d, and e), the Amazon River plume (f, g, and h), and the Bay of Bengal (i, j, and k). The gap-filled results were compared with images from the recent past (e, h, and k).

average proportion of pixels with a value < 30 psu was 18.8 %. Since low-salinity water and seawater do not mix well due to their different densities (Akhil et al., 2020), their shape was preserved, and gap filling was fairly effective. Smoothing after gap-filling mostly occurred in the 30–33 psu range, where low-salinity water and seawater were mixed. Fig. 11 shows how the gap-filling model effectively used oversampling of low-salinity samples while minimizing the smoothing problem. Fig. 12 depicts low-salinity water from the Yangtze River in the East China Sea using the gap-filling models of this study. As more oversampling and data from the past 5 days were used, the results were fairly comparable to the original values (GBRT-corrected SSS), similar to the Amazon River plume region shown in Fig. 11.

Fig. 13 shows the results gap-filled by the U5G4 model on August 4, 2020. The proposed U-Net model produced successful gap-filling results compared with the original global ocean data (Fig. 13a and 13b). Three areas with significant salinity changes—the Gulf of Mexico (Fig. 13c, 13d, and 13e), the Amazon River plume (Fig. 13f, 13 g, and 13 h), and the Bay of Bengal (Fig. 13i, 13j, and 13 k)—were compared in further

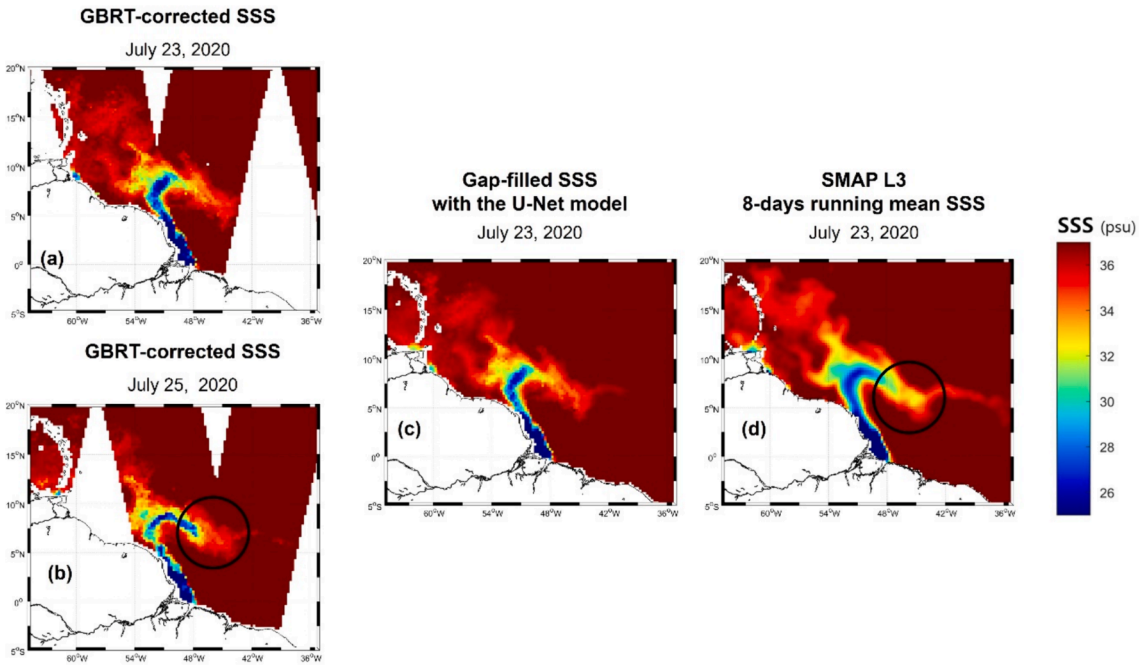


Fig. 14. Spatial distribution of the (a and b) GBRT-corrected SSS, (c) gap-filled SSS using the proposed U-Net model, and (d) SMAP L3 8-day running mean SSS data in the Amazon River plume region. The expanded low salinity water is marked with a black circle (b and d).

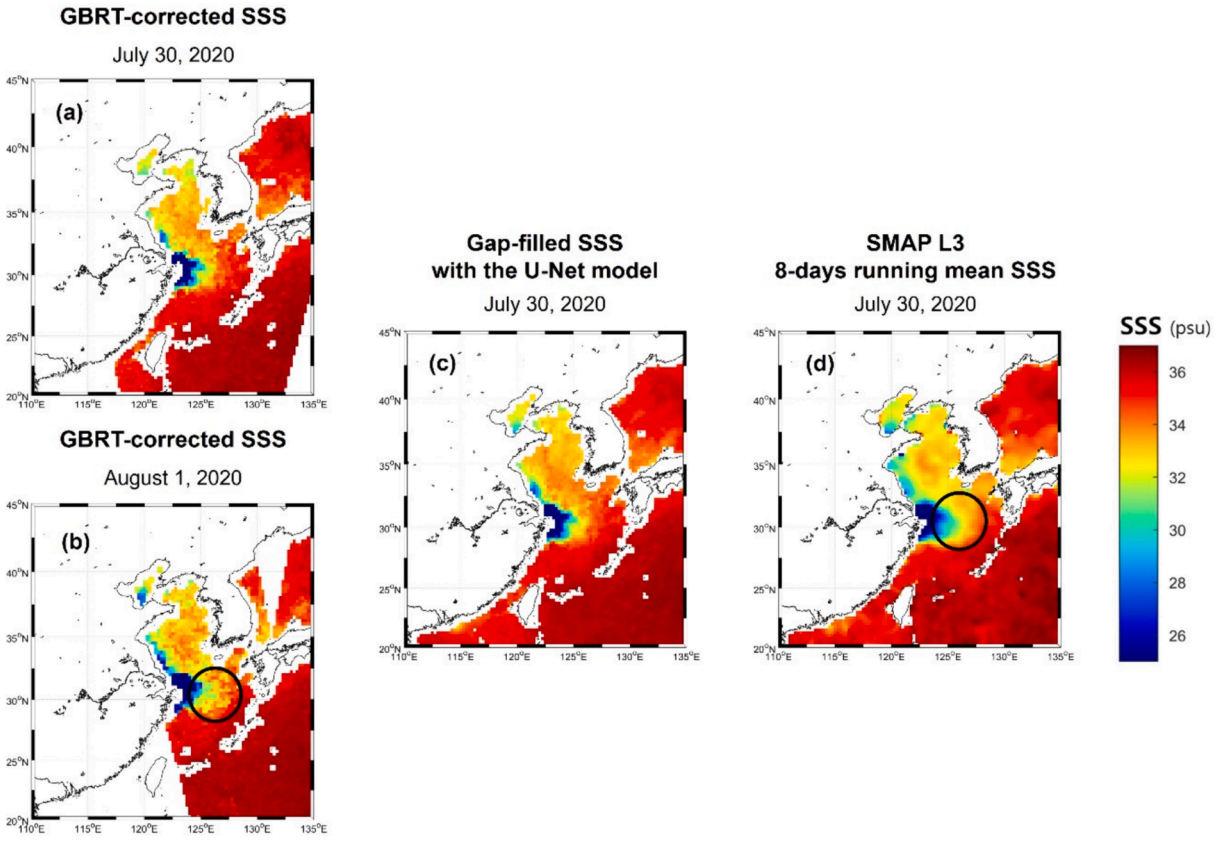


Fig. 15. Spatial distribution of the (a and b) GBRT-corrected SSS, (c) gap-filled SSS using the proposed U-Net model, and (d) SMAP L3 8-day running mean SSS data in the East China Sea. The expanded low salinity water is marked with a black circle (b and d).

detail. Low salinity water is formed by significant freshwater input from the Atchafalaya and Mississippi rivers in the Gulf of Mexico, the Amazon River in the plume region, and the Ganges and Irrawaddy Rivers in the Bay of Bengal. The low-salinity water patches in the original images demonstrate that these patches were filled in all regions. Our proposed gap-filling model performed effectively in open oceans with low saline

changes and low-salinity water regions with significant salinity distribution changes due to freshwater influx.

Figs. 14 and 15 compare the results of the U5G4 model with SMAP L3 8-day running mean SSS in the Amazon River plume region and the East China Sea, respectively. Fig. 14c and 15c show how the proposed gap-filling model produced results similar to the low-salinity water patches shown in the original GBRT-corrected SSS images (Fig. 14a and 15a). In contrast, in the SMAP L3 8-day running mean SSS, the low-salinity water patches appeared to expand more (black circles in Fig. 14d and 15d). Two days later, expanding low-salinity water patches were visible in the real-time GBRT-corrected images (Fig. 14b and 15b). This demonstrates how the future expansion of low-salinity water impacts the SMAP L3 8-day running mean SSS, which used future data as input. Our proposed gap-filling model only employed past data, effectively filling the gap in the global SSS without considering future events.

4.4. Novelty and limitations

The novelty of this study can be summarized as follows: First, it is one of the few studies to adopt a CNN for spatial gap filling of global SSS. Despite recent explorations of deep learning in oceanic studies, such as SST gap filling, there has been little research on its SSS applications. Second, by using the U-Net structure with arbitrary input sizes, the proposed model can accept various input sizes during the training and inference phases. Therefore, this model can be used for either regional or global applications, allowing easy fine-tuning. From the pretrained global model, it is possible to train the model with local samples, enhancing the predicted results when additional data are available. Additionally, the arbitrary input size reduces computational costs. Using smaller patches (40×40) in the training phase reduced the required memory compared with larger inputs (720×1440 for global coverage). Once the model is trained, the SSS gap filling can be completed within a few seconds. Moreover, our model showed a high correlation with *in situ* measurements with the GBRT-corrected SSS from Jang et al. (2022). Lastly, our model produced a daily global gap-filled SSS without delay. From an operational perspective, this can significantly contribute to SSS research and other models that use SSS as their input.

Despite these novelties, some limitations remain. Even with the ability of U-Net to retain details from the input, some blurry predictions were observed. This might be caused by convolutional operations, whose return values represent receptive input filters. Such a smoothing issue can be mitigated using generative models or super-resolution techniques. Another limitation is that an additional gap-filling process is required to construct the target samples. In this study, we utilized Gaussian gap-filling with 2-day buffers for past and future data. The quality of the training samples largely depends on the performance of the interpolation method. Although no significant performance drop was observed in this study due to Gaussian interpolation, the inherent error associated with the additional interpolation process still remains. Lastly, the pre-processing gap filling near the coastline is also a limitation because the suggested model could not automatically deal with dummy zero values.

5. Conclusions

This study proposed a daily gap-filling model for SSS using a sequence of satellite swath data with the U-Net deep learning model. Unlike the SMAP L3 8-day running mean SSS, the proposed model does not rely on future datasets. The model can accept arbitrary input sizes, so we used 40×40 patch samples in the training phase and 720×1440 global images in the inference step. This flexibility reduced the time and computational resources required during training and significantly increased the number of samples. When evaluating the gap-filling models with the masked-out test pixels, the overall RMSE was lower for the model using GBRT-corrected SSS than that for the one using SMAP SSS. This is likely due to the GBRT-corrected SSS having lower spatial and temporal noise than the SMAP SSS data, thanks to bias correction with *in-situ* samples. Different levels of SSS sequences and the oversampling of low-salinity water were investigated. Oversampling significantly improved the two types of validations (using masked-out pixels and Argo *in situ* data). Based on quantitative performance and visual investigation, the U5G4 model yielded the best performance among multiple schemes. Compared with the SMAP L3 8-day running mean SSS, which used past and future data, the proposed model was unaffected by significant SSS changes from future data over coastal areas. With the fast prediction of SSS using the pretrained model, it is highly expected to produce bias-corrected daily gap-filled SSS with the GBRT-corrected SSS dataset and U-Net models. Generating high-quality, daily gap-filled global SSS data for near real-time applications will greatly enhance the monitoring and modeling of oceanic environments. Although this study addresses gaps in SSS swath data, it can be further extended to other sensors, such as the Geostationary Ocean Color Imager (GOCI)-II, to monitor real-time low-salinity water in the East China Sea, where highly fluctuating SSS patterns are found, especially in summer (Jang et al., 2024; Sung et al., 2022).

CRediT authorship contribution statement

Eunna Jang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Daehyeon Han:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Jungho Im:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Taejun Sung:** Formal analysis, Data curation. **Young Jun Kim:** Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the “Development of technology using analysis of ocean satellite images (20210046)” and “Development of risk managing technology tackling ocean and fisheries crisis around Korean Peninsula by Kuroshio Current” (RS-2023-00256330) funded by the Korea Institute of Marine Science & Technology Promotion (KIMST) funded by Ministry of Oceans and Fisheries.

References

Ahishali, M., Kiranyaz, S., Ince, T., Gabbouj, M., 2021. Classification of polarimetric SAR images using compact convolutional neural networks. Gisci. Remote Sens. 58 (1), 28–47. https://doi.org/10.1080/15481603.2020.1853948.

Akhil, V.P., Vialard, J., Lengaigne, M., Keerthi, M.G., Boutin, J., Vergely, J.-L., Papa, F., 2020. Bay of Bengal Sea surface salinity variability using a decade of improved SMOS re-processing. Remote Sens. Environ. 248, 111964 https://doi.org/10.1016/j.rse.2020.111964.

Alvera-Azcárate, A., Barth, A., Parard, G., Beckers, J.-M., 2016. Analysis of SMOS sea surface salinity data using DINEOF. Remote Sens. Environ. 180, 137–145. https://doi.org/10.1016/j.rse.2016.02.044.

Barth, A., Alvera-Azcárate, A., Licer, M., Beckers, J.-M., 2020. DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. Geosci. Model Dev. 13 (3), 1609–1622. https://doi.org/10.5194/gmd-13-1609-2020.

Barth, A., Alvera-Azcárate, A., Troupin, C., Beckers, J.-M., 2022. DINCAE 2.0: Multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. Geosci. Model Dev. 15 (5), 2183–2196. https://doi.org/10.5194/gmd-15-2183-2022.

Barth, A., Alvera-Azcárate, A., Troupin, C., Beckers, J.-M., Van der Zande, D., 2021. Reconstruction of missing data in satellite images of the Southern North Sea using a convolutional neural network (DINCAE), 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp. 7493-7496. DOI: 10.1109/igarss47720.2021.9554045.

Boutin, J., Vergely, J.L., Marchand, S., d'Amico, F., Hasson, A., Kolodziejczyk, N., Reul, N., Reverdin, G., Vialard, J., 2018. New SMOS Sea Surface Salinity with reduced systematic errors and improved variability. Remote Sens. Environ. 214, 115–134. https://doi.org/10.1016/j.rse.2018.05.022.

Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 106, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011.

Chen, F., Tsou, J.Y., 2022. Assessing the effects of convolutional neural network architectural factors on model performance for remote sensing image classification: An in-depth investigation. Int. J. Appl. Earth Obs. Geoinf. 112, 102865 https://doi.org/10.1016/j.jag.2022.102865.

Dedring, T., Rienow, A., 2024. Synthesis and evaluation of seamless, large-scale, multispectral satellite images using Generative Adversarial Networks on land use and land cover and Sentinel-2 data. Gisci. Remote Sens. 61 (1), 2364460. https://doi.org/10.1080/15481603.2024.2364460.

Dinnat, E.P., Le Vine, D.M., Boutin, J., Meissner, T., Lagerloef, G., 2019. Remote sensing of sea surface salinity: Comparison of satellite and in situ observations and impact of retrieval parameters. Remote Sens. 11 (7), 750. https://doi.org/10.3390/rs11070750.

Droghei, R., Buongiorno Nardelli, B., Santoleri, R., 2018. A new global sea surface salinity and density dataset from multivariate observations (1993–2016). Front. Mar. Sci. 5, 84. https://doi.org/10.3389/fmars.2018.00084.

Fore, A., Yueh, S., Tang, W., Hayashi, A., 2020. SMAP salinity and wind speed data user's guide. California Institute of Technology: Pasadena, CA, USA, 42.

Gao, J., Yuan, Q., Li, J., Su, X., 2021. Unsupervised missing information reconstruction for single remote sensing image with Deep Code Regression. Int. J. Appl. Earth Obs. Geoinf. 105, 102599 https://doi.org/10.1016/j.jag.2021.102599.

Han, D., Choo, M., Im, J., Shin, Y., Lee, J., Jung, S., 2023. Precipitation nowcasting using ground radar data and simpler yet better video prediction deep learning. GISci. Remote Sens. 60 (1), 2203363. https://doi.org/10.1080/15481603.2023.2203363.

Han, Z., He, Y., Liu, G., Perrie, W., 2020. Application of DINCAE to reconstruct the gaps in chlorophyll-a satellite observations in the South china sea and West philippine sea. Remote Sens. 12 (3), 480. https://doi.org/10.3390/rs12030480.

Jang, E., Kim, Y.J., Im, J., Park, Y.-G., Sung, T., 2022. Global sea surface salinity via the synergistic use of SMAP satellite and HYCOM data based on machine learning. Remote Sens. Environ. 273, 112980 https://doi.org/10.1016/j.rse.2022.112980.

Jang, E., Choi, J.K., Ahn, J.H., 2024. Continuity and enhancements in sea surface salinity estimation in the East China Sea using GOCI and GOCI-II: Challenges and further developments. Remote Sens. 16 (12), 2111. https://doi.org/10.3390/rs16122111.

Jung, S., Yoo, C., Im, J., 2022. High-resolution seamless daily sea surface temperature based on satellite data fusion and machine learning over Kuroshio Extension. Remote Sens. 14 (3), 575. https://doi.org/10.3390/rs14030575.

Kang, Y., Jang, E., Im, J., Kwon, C., 2022. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. Gisci. Remote Sens. 59 (1), 2019–2035. https://doi.org/10.1080/15481603.2022.2143872.

Kim, Y.J., Han, D., Jang, E., Im, J., Sung, T., 2023. Remote sensing of sea surface salinity: challenges and research directions. Gisci. Remote Sens. 60 (1), 2166377. https://doi.org/10.1080/15481603.2023.2166377.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. DOI: 10.48550/arXiv.1412.6980.

Kolodziejczyk, N., Hamon, M., Boutin, J., Vergely, J.-L., Reverdin, G., Supply, A., Reul, N., 2021. Objective analysis of SMOS and SMAP sea surface salinity to reduce large-scale and time-dependent biases from low to high latitudes. J. Atmos. Oceanic Technol. 38 (3), 405–421. https://doi.org/10.1175/jtech-d-20-0093.1.

Lagerloef, G., and Font, J., 2010. SMOS and Aquarius/SAC-D missions: The era of spaceborne salinity measurements is about to begin. Oceanography from space, 35-58. DOI: 10.1007/978-90-481-8681-5_3.

Lee, J., Kim, M., Im, J., Han, H., Han, D., 2021. Pre-trained feature aggregated deep learning-based monitoring of overshooting tops using multi-spectral channels of GeoKompsat-2A advanced meteorological imagery. Gisci. Remote Sens. 58 (7), 1052–1071. https://doi.org/10.1080/15481603.2021.1960075.

Lee, M.J., Lee, W.J., Lee, S.K., Jung, H.S., 2022. Deep Learning for Remote Sensing Applications. Korean Journal of Remote Sensing 38 (6), 1581–1587. https://doi.org/10.7780/kjrs.2022.38.6.2.1.

Li, Y., Dong, Q., Ren, Y., 2019. Aquarius sea surface salinity gridding method based on dual quality-distance weighting. Remote Sens. 11 (9), 1131. https://doi.org/10.3390/rs11091131.

Luo, X., Song, J., Guo, J., Fu, Y., Wang, L., Cai, Y., 2022. Reconstruction of chlorophyll-a satellite data in Bohai and Yellow sea based on DINCAE method. Int. J. Remote Sens. 43 (9), 3336–3358. https://doi.org/10.1080/01431161.2022.2090872.

Ma, Y., Zhen, Z., Li, F., Feng, F., Zhao, Y., 2023. An innovative lightweight 1D-CNN model for efficient monitoring of large-scale forest composition: a case study of Heilongjiang Province, China. GISci. Remote Sens. 60 (1), 2271246. https://doi.org/10.1080/15481603.2023.2271246.

Mao, H., Kathuria, D., Duffield, N., Mohanty, B.P., 2019. Gap filling of high-resolution soil moisture for SMAP/sentinel-1: A two-layer machine learning-based framework. Water Resour. Res. 55 (8), 6986–7009. https://doi.org/10.1029/2019wr024902.

Martin, S., 2014. An introduction to ocean remote sensing. Cambridge University Press. DOI: 10.1002/lob.10041.

Masolele, R.N., De Sy, V., Marcos, D., Verbesselt, J., Gieseke, F., Mulatu, K.A., Moges, Y., Sebrala, H., Martius, C., Herold, M., 2022. Using high-resolution imagery and deep learning to classify land-use following deforestation: A case study in Ethiopia. Gisci. Remote Sens. 59 (1), 1446–1472. https://doi.org/10.1080/15481603.2022.2115619.

Mellor, A., Boukir, S., Haywood, A., Jones, S., 2015. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. ISPRS J. Photogramm. Remote Sens. 105, 155–168. https://doi.org/10.1016/j.isprsjprs.2015.03.014.

Mu, Z., Zhang, W., Wang, P., Wang, H., Yang, X., 2019. Assimilation of SMOS sea surface salinity in the regional ocean model for South China Sea. Remote Sens. 11 (8), 919. https://doi.org/10.3390/rs11080919.

Mukherjee, R., Liu, D., 2021. Downscaling MODIS spectral bands using deep learning. Gisci. Remote Sens. 58 (8), 1300–1315. https://doi.org/10.1080/15481603.2021.1984129.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. Omnipress, Haifa, Israel, pp. 807–814.

Nardelli, B.B., Droghei, R., Santoleri, R., 2016. Multi-dimensional interpolation of SMOS sea surface salinity with surface temperature and in situ salinity data. Remote Sens. Environ. 180, 392–402. https://doi.org/10.1016/j.rse.2015.12.052.

Qin, S., Wang, H., Zhu, J., Wan, L., Zhang, Y., Wang, H., 2020. Validation and correction of sea surface salinity retrieval from SMAP. Acta Oceanolog. Sin. 39, 148–158. https://doi.org/10.1007/s13131-020-1533-0.

Reul, N., Grodsky, S., Arias, M., Boutin, J., Catany, R., Chapron, B., d'Amico, F., Dinnat, E., Donlon, C., Fore, A., 2020. Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010–2019). Remote Sens. Environ. 242, 111769 https://doi.org/10.1016/j.rse.2020.111769.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234-241. DOI: 10.1007/978-3-319-24574-4_28.

Shin, J., Kim, D.W., Kim, S.H., Lee, G.S., Kim, B.K., Jo, Y.H., 2023. Gap-filling processes on GOCI-derived daily sea surface salinity product for Changjiang diluted water front in the East China Sea. Earth Syst. Sci. Data Discuss. 2023, 1–29. https://doi.org/10.5194/essd-2023-421.

Singh, A., Gaurav, K., 2023. Deep learning and data fusion to estimate surface soil moisture from multi-sensor satellite images. Sci. Rep. 13 (1), 2251. https://doi.org/10.1038/s41598-023-28939-9.

Sung, T., Sim, S., Jang, E., Im, J., 2022. Estimation of high resolution sea surface salinity using multi satellite data and machine learning. Korean J. Remote Sens. 38 (5), 747–763.

Tong, C., Wang, H., Magagi, R., Goïta, K., Wang, K., 2021. Spatial gap-filling of SMAP soil moisture pixels over Tibetan plateau via machine learning versus geostatistics. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 9899–9912. https://doi.org/10.1109/jstars.2021.3112623.

Wu, S., Heitzler, M., Hurni, L., 2022. Leveraging uncertainty estimation and spatial pyramid pooling for extracting hydrological features from scanned historical topographic maps. GISci. Remote Sens. 59 (1), 200–214. https://doi.org/10.1080/15481603.2021.2023840.

Wu, P., Su, Y., Duan, S.-B., Li, X., Yang, H., Zeng, C., Ma, X., Wu, Y., Shen, H., 2022. A two-step deep learning framework for mapping gapless all-weather land surface temperature using thermal infrared and passive microwave data. Remote Sens. Environ. 277, 113070 https://doi.org/10.1016/j.rse.2022.113070.

Xing, D., Hou, J., Huang, C., Zhang, W., 2022. Spatiotemporal reconstruction of MODIS normalized difference snow index products using U-Net with partial convolutions. Remote Sens. 14 (8), 1795. https://doi.org/10.3390/rs14081795.

Zhang, Q., Liu, X., Peng, T., Yang, X., Tang, M., Zou, X., Liu, M., Wu, L., Zhang, T., 2024. U-SeqNet: Learning spatiotemporal mapping relationships for multimodal multitemporal cloud removal. Gisci. Remote Sens. 61 (1), 2330185. https://doi.org/10.1080/15481603.2024.2330185.

Zhang, R., Zhu, W., Li, Y., Song, T., Li, Z., Yang, W., Yang, L., Zhou, T., Xu, X., 2023. D-FusionNet: road extraction from remote sensing images using dilated convolutional block. GISci. Remote Sens. 60 (1), 2270806. https://doi.org/10.1080/15481603.2023.2270806.

Zhou, X., Yang, X., Ye, X., Li, B., 2024. Dual generative adversarial networks for merging ocean transparency from satellite observations. Gisci. Remote Sens. 61 (1), 2356357. https://doi.org/10.1080/15481603.2024.2356357.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307.

Zujev, M., Elken, J., Lagemaa, P., 2021. Data assimilation of sea surface temperature and salinity using basin-scale reconstruction from empirical orthogonal functions: a feasibility study in the northeastern Baltic Sea. Ocean Sci. 17 (1), 91–109. https://doi.org/10.5194/os-17-91-2021.