



Article

STA-GAN: A Spatio-Temporal Attention Generative Adversarial Network for Missing Value Imputation in Satellite Data

Shuyu Wang ^{1,2} , Wengen Li ^{1,*} , Siyun Hou ^{1,2} , Jihong Guan ¹ and Jiamin Yao ^{1,2}¹ Department of Computer Science and Technology, Tongji University, Shanghai 200082, China² Project Management Office of China National Scientific Seafloor Observatory, Tongji University, Shanghai 200082, China

* Correspondence: lwengen@tongji.edu.cn

Abstract: Satellite data is of high importance for ocean environment monitoring and protection. However, due to the missing values in satellite data, caused by various force majeure factors such as cloud cover, bad weather and sensor failure, the quality of satellite data is reduced greatly, which hinders the applications of satellite data in practice. Therefore, a variety of methods have been proposed to conduct missing data imputation for satellite data to improve its quality. However, these methods cannot well learn the short-term temporal dependence and dynamic spatial dependence in satellite data, resulting in bad imputation performance when the data missing rate is large. To address this issue, we propose the Spatio-Temporal Attention Generative Adversarial Network (STA-GAN) for missing value imputation in satellite data. First, we develop the Spatio-Temporal Attention (STA) mechanism based on Graph Attention Network (GAT) to learn features for capturing both short-term temporal dependence and dynamic spatial dependence in satellite data. Then, the learned features from STA are fused to enrich the spatio-temporal information for training the generator and discriminator of STA-GAN. Finally, we use the generated imputation data by the trained generator of STA-GAN to fill the missing values in satellite data. Experimental results on real datasets show that STA-GAN largely outperforms the baseline data imputation methods, especially for filling satellite data with large missing rates.



Citation: Wang, S.; Li, W.; Hou, S.; Guan, J.; Yao, J. STA-GAN: A Spatio-Temporal Attention Generative Adversarial Network for Missing Value Imputation in Satellite Data. *Remote Sens.* **2023**, *15*, 88. <https://doi.org/10.3390/rs15010088>

Academic Editor: Massimo Panella

Received: 18 November 2022

Revised: 14 December 2022

Accepted: 20 December 2022

Published: 23 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

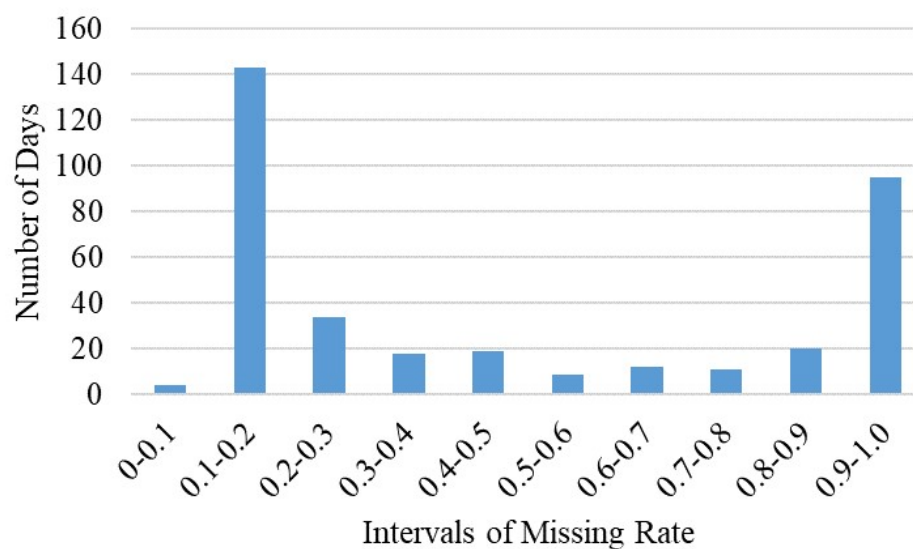
Keywords: satellite data; data imputation; spatio-temporal analytics; generative adversarial network

1. Introduction

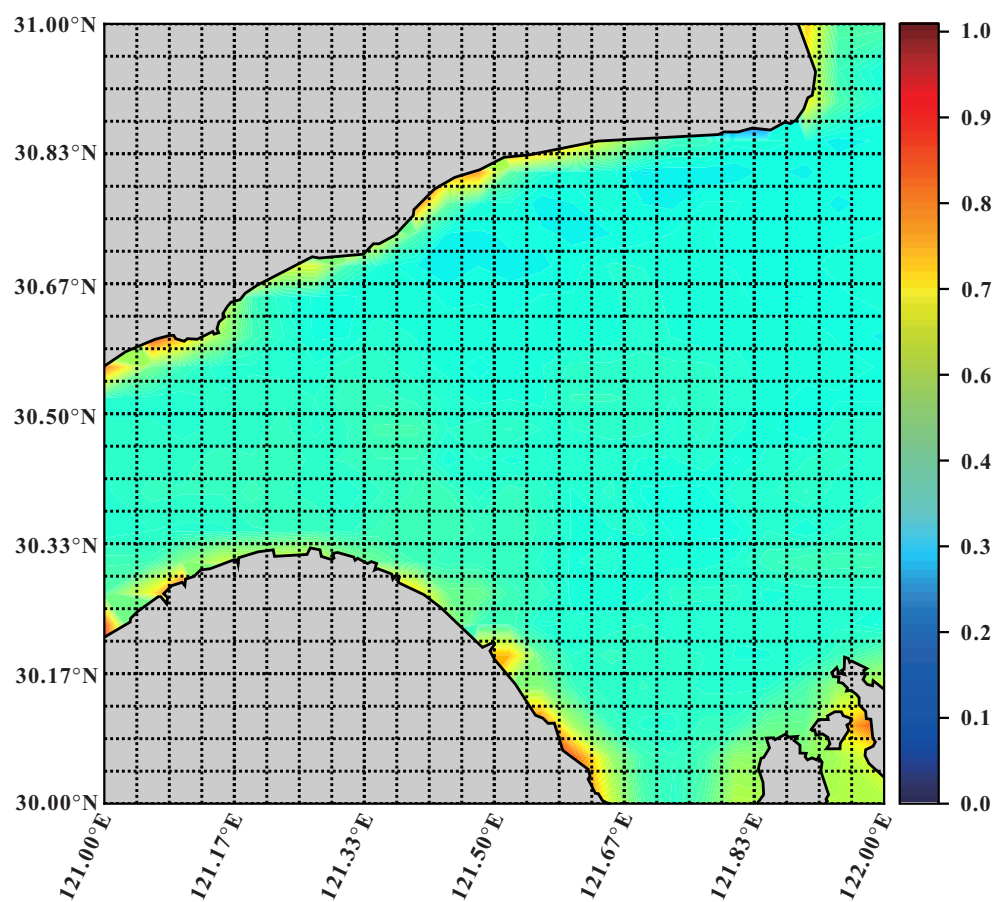
The data obtained by satellites has the advantages of real-time, wide coverage and low cost, and is thus widely used for monitoring the ocean environment and climate [1]. With the rapid development of satellite remote sensing technology, satellite data has grown exponentially, which greatly facilitates the development and application of red tide prediction [2–4], ocean environmental disaster prevention [5–7], typhoon path prediction [8,9], etc.

However, the issue of data missing is very common in satellite data due to the influence of various force majeure factors such as cloud occlusion, bad weather and sensor failure [10,11]. For example, Figure 1 shows the temporal and spatial missing rates of Sea Surface Temperature (SST) data in the region of the East China Sea in 2021 from the Advanced Very High-Resolution Radiometer (AVHRR) sensor. According to Figure 1a, the data missing rate of AVHRR SST data is mainly located in the two intervals of 10–20% and 90–100%, and the average missing rate is actually as high as 48.51%. According to Figure 1b, the missing rates in most grid regions are between 30% and 60%, and are even higher near the coastline. Obviously, compared with other spatio-temporal data, e.g., traffic data and crowd volume data, the missing rate in satellite data is much larger. The high missing values of satellite data make it difficult to conduct data analysis, and greatly impede the real-time monitoring of the ocean environment and climate. Therefore, accurate

and efficient missing value imputation is an important task for improving the quality of satellite remote sensing data.



(a) The statistics of daily missing rates of AVHRR SST data in 2021.



(b) The spatial distribution of the missing rates of AVHRR SST data.

Figure 1. The temporal and spatial distributions of missing rates of AVHRR SST data in the region of East China Sea in 2021.

A variety of methods have been proposed for satellite data imputation, and these methods can be roughly classified into three categories, i.e., statistical methods, traditional machine learning-based methods and deep learning methods. Statistical methods learn the linear or non-linear correlations in satellite data based on prior knowledge to achieve data imputation. Typical statistical methods include Optimal Interpolation (OI) method, Data Interpolation Empirical Orthogonal Function (DINEOF) method and distance-based method (e.g., Kriging and IDW). The OI method derives the optimal unbiased estimate of missing satellite data based on the min-variance and is often used to produce cloud-free products of SST data by fusing data from multiple platforms, e.g., satellite data, buoy data, and in situ data. For example, OI method is used in the OISST product from National Oceanic and Atmospheric Administration (NOAA) [12], Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system [13], and blended foundation SST product [14]. OI method is also used for analyzing daily SST by fusing the AVHRR satellite data and Tropical Rainfall Measuring Mission Microwave Imager (TMI) product [15]. The DINEOF method achieves missing value imputation in oceanographic data [16] based on the Empirical Orthogonal Function (EOF). It is widely used for the reconstruction of Chl-a data [17–19], SST data [19,20], ocean wind data [21] and multivariate reconstruction [22]. The distance-based data imputation method achieves data imputation based on the spatial relevance of satellite data. Representative approaches include Kriging [23,24] and IDW [25]. Statistical methods have been widely used for filling missing values in various satellite data products. However, with the increasing of satellite data, it is time-consuming for statistical methods to construct correlation equations in satellite data. In addition, they cannot well learn the complex nonlinear relationships in satellite data to further improve the accuracy of data imputation.

Representative traditional machine learning methods for satellite data imputation include Random Forest (RF), eXtreme Gradient Boosting (XGBoost), k-Nearest Neighbors (KNN), Support Vector Regression (SVR) and Matrix Factorization (MF). The RF method and XGBoost method are typical tree-based ensemble methods [26], and fill missing values in satellite data by constructing multiple decision trees. Mohebzadeh et al. [27] compared RF method with DINEOR method for the imputation of Chl-a data from the MODIS satellite. Chen et al. [28] used RF to improve the coverage of Chl-a data with two external factors, i.e., the Ocean Color Index and the Rayleigh-corrected Reflectivity. In addition, RF and XGBoost are also used for the reconstruction of Chl-a satellite data [29–32]. The KNN method and SVR method fill missing data based on distance measurement, and have been widely applied for satellite data imputation [29,30,33,34]. Hankel Matrix Factorization (HMF) is used for data imputation in the Global Navigation Satellite System (GNSS) [35]. In practice, the traditional machine learning methods described above has some limitations for missing value imputation in satellite data. First, it is hard for them to fully mine the hidden information in the long time series satellite data. Second, the short-term temporal dependence and dynamic spatial dependence are also not well considered by these methods.

Deep learning methods can effectively learn the hidden regularity in satellite data [36,37], thus being introduced for satellite data imputation. Jean-Marie et al. [38] achieved the interpolation of SST data using a neural network, and proved that neural network is superior to the OI and EOF methods. Artificial Neural Network (ANN) has been also applied for the reconstruction of satellite data [30,33,39]. Jouini et al. [40] applied Self-Organizing Map (SOM) network to reconstruct Chl-a data by integrating SST and sea surface height (SSH). Data-Interpolating Convolutional Auto-Encoder (DINCAE) method is proposed for filling the missing values in SST data [41]. The Auto-Encoder in DINCAE is similar to the EOF for data reduction, and can effectively capture non-linear information of satellite data. DINCAE has been widely used for the reconstruction of Chl-a data [42] and SST data [43].

However, existing satellite data imputation methods still face some big challenges. First, the spatial dependence in satellite data is dynamic and affected by various uncontrollable factors around the target regions. For example, SST is usually affected by wind

and solar radiation, and Chl-a concentration is vulnerable to the reproduction direction of phytoplankton. Second, satellite data usually show bidirectional short-term temporal dependence, and has large fluctuations in short term. For satellite data imputation, in addition to historical data, future data should be also taken into account because it is already known. For example, the occurrence of red tides causes the Chl-a concentration to be higher than usual. In this case, the imputation of missing values in Chl-a concentration is highly dependent on future data. Therefore, the bidirectional short-term temporal dependence should be fully utilized to improve the performance of satellite data imputation. Unfortunately, existing satellite data imputation methods cannot well learn the dynamic spatio-temporal dependence in satellite data to achieve accurate data imputation.

Therefore, in this work, we propose the Spatio-Temporal Attention Generative Adversarial Network (STA-GAN), which integrates Graph Attention Network (GAT) and Generative Adversarial Network (GAN) for missing value imputation in satellite data. GAN can efficiently learn the complex distribution of the data so that the generated data well conform to the original data distribution, and is suitable for time series data imputation [44]. Therefore, multiple GAN models, e.g., Generative Adversarial Imputation Network (GAIN) [45], GAN-2-stage [46] and SolarGAN [47]), have been introduced for missing data imputation. Meanwhile, the attention mechanism [48,49] can achieve dynamic aggregation of effective information and is widely used for the prediction of SST and Chl-a concentration [50–52]. Inspired by the attention mechanism, the GAT network processes graph nodes of different degrees and gives higher weights to the influential neighbor nodes [53]. GAT is suitable for mining the spatio-temporal dependence of ocean remote sensing data.

Concretely, STA-GAN first learns the short-term temporal dependence and dynamic spatial dependence with a Spatio-Temporal Attention mechanism based on GAT, and introduces GAN to learn the underlying distribution of satellite data. Then, we train the generator and discriminator of STA-GAN by fusing the learned spatio-temporal dependence features. Finally, the missing data is filled with the generated data by STA-GAN.

In sum, our contributions are summarized as follows:

- We identified the challenges in satellite data imputation and proposed the STA-GAN model that integrates GAT and GAN to achieve accurate data imputation.
- We developed a new spatio-temporal attention mechanism based on GAT to capture the short-term temporal dependence and dynamic spatial dependence of satellite data in parallel.
- We re-designed the structure of GAN to achieve data imputation by learning the distribution of satellite data with the learned spatio-temporal dependence information.

In addition, we evaluated STA-GAN model on both SST satellite data and Chl-a satellite data. The results demonstrate that STA-GAN outperforms a variety of existing data imputation methods, especially for filling satellite data with high missing rates.

The rest of this article is organized as follows. Section 2 describes the data materials, the pre-processing and methods details of STA-GAN model for satellite data imputation. Then, we present the results in Section 3 and give the discussion in Section 4. Finally, we draw the conclusions in Section 5.

2. Materials and Methods

2.1. Study Area and Data

We use the satellite data for both SST and Chl-a as examples to study the problem of missing value imputation in satellite data. Satellite data are usually grid data partitioned according to the latitude and longitude, and the value of each grid is obtained by retrieving the satellite observation images from the sensors deployed on satellites.

The SST data are from the Advanced Very High-Resolution Radiometer (AVHRR) and collected twice, i.e., daytime and nighttime, one day. The spatial resolution is 4 km ($1/24^\circ \times 1/24^\circ$) and the time range is from 2000 to 2021. We select two study areas, i.e., $[121.00^\circ\text{E}–122.00^\circ\text{E}, 30.00^\circ\text{N}–31.00^\circ\text{N}]$ in the East China Sea (referred to as SST-EAST)

and $[109.00^{\circ}\text{E}\text{--}110.00^{\circ}\text{E}, 13.00^{\circ}\text{N}\text{--}14.00^{\circ}\text{N}]$ in the South China Sea (referred to as SST-SOUTH). For SST-EAST dataset, we remove the grid regions with missing rates larger than 95% since they lack enough ground-truth samples to validate the filled values. In addition, the land region is not taken into account. Finally, there are 368 grid regions in SST-EAST dataset and the overall missing rate is 8.43%. Similarly, there are 428 grid regions in the SST-SOUTH dataset after the filtering operation, and the overall missing rate is 11.72%. Figures 2 and 3 illustrate the spatial distributions of SST-EAST and SST-SOUTH, respectively, on a selected day.

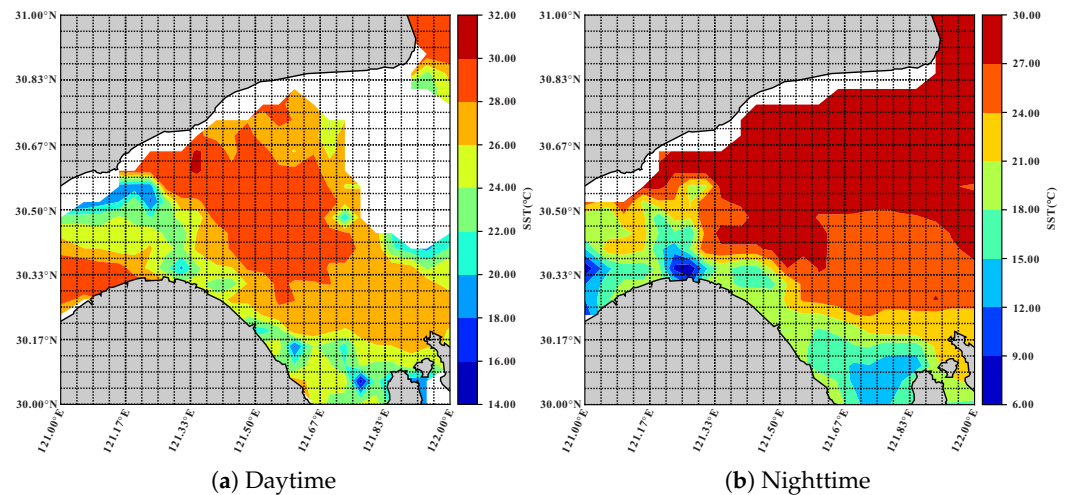


Figure 2. The spatial distribution of SST-EAST dataset on Jul 18, 2021, where the white areas have no data, and the gray areas are land.

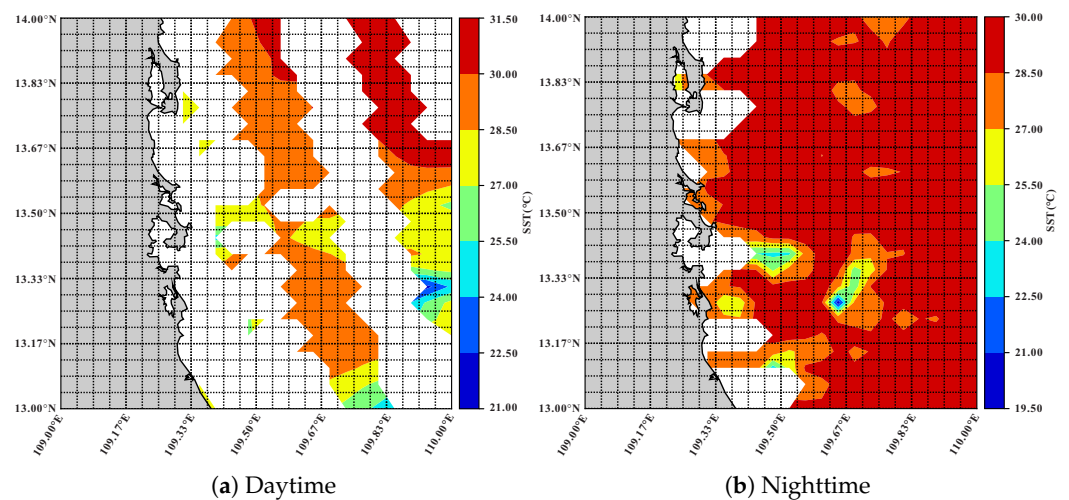


Figure 3. The spatial distribution of SST-SOUTH dataset on Jul 17, 2021, where the white areas have no data, and the gray areas are land.

The Chl-a data is from the Ocean Color Climate Change Initiative 5.0 version (OC-CCI) product and collected once everyday. This product fuses the satellite data from SeaWiFS, MERIS, MODIS and VIIRS sensors to produce daily Chl-a data. The spatial resolution is also 4 km ($1/24^{\circ} \times 1/24^{\circ}$) and the time range is from 2000 to 2021. The study area is $[128.00^{\circ}\text{E}\text{--}130.00^{\circ}\text{E}, 24.00^{\circ}\text{N}\text{--}26.00^{\circ}\text{N}]$ in the East China Sea (referred to as CHA-EAST). The CHA-EAST dataset has 309 grid regions after removing the grid regions with missing rates larger than 95%, and its overall missing rate is 54.97%. Figure 4 shows the spatial distribution of CHA-EAST on 25 July 2021.

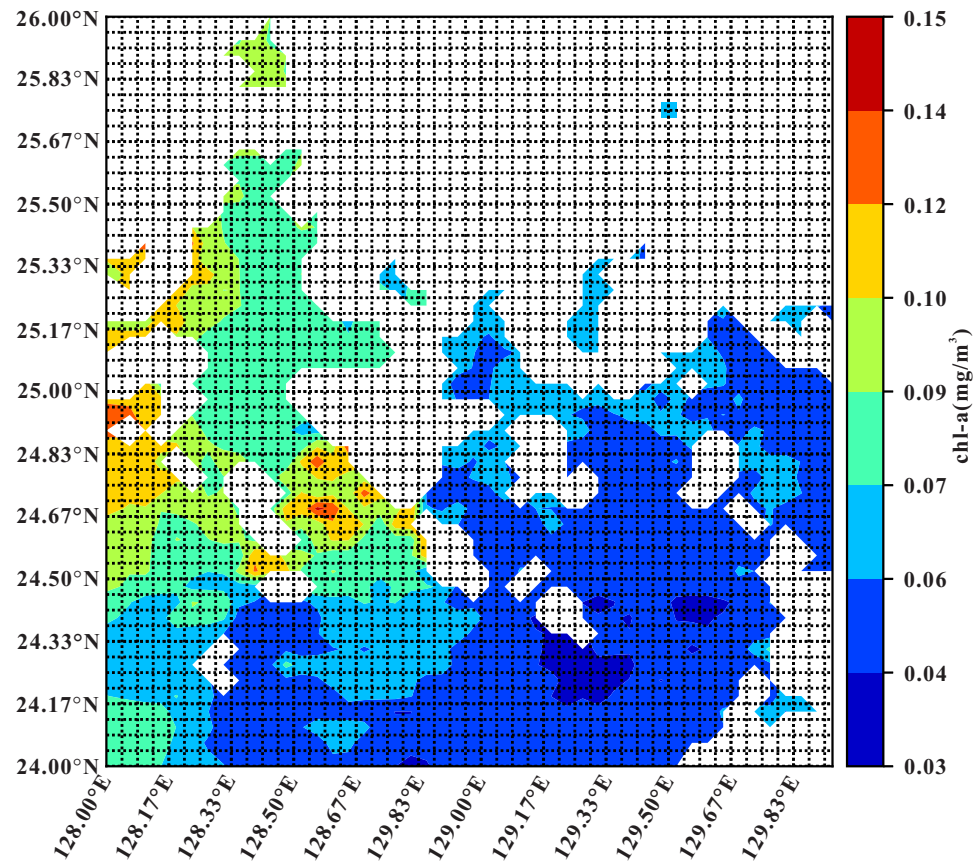


Figure 4. The spatial distribution of CHA-EAST dataset on 25 July 2021, where the white areas have no data.

2.2. Data Processing

As illustrated in Figure 5, we use a dynamic graph $G = \langle V, E, A, X, M \rangle$ to represent satellite data, where V is a set of N nodes and each node corresponds to a grid region; E is the set of edges and two nodes have an edge if their spatial distance is less than a given threshold. Given the locations (l_i^{lat}, l_i^{lon}) and (l_j^{lat}, l_j^{lon}) of nodes $v_i \in V$ and $v_j \in V$, respectively, the distance $d_{i,j}$ between v_i and v_j is computed as below.

$$d_{i,j} = \sqrt{(l_i^{lat} - l_j^{lat})^2 + (l_i^{lon} - l_j^{lon})^2}. \quad (1)$$

We use an adjacency matrix $A \in \mathbb{R}^{N \times N}$ to represent the topology structure of G . In matrix A , each entry $a_{i,j} = 1$ indicates that there is an edge between nodes v_i and v_j , otherwise $a_{i,j} = 0$. The value of $a_{i,j}$ is determined as below.

$$a_{i,j} = \begin{cases} 1 & \text{if } \exp\left\{\frac{-d_{i,j}^2}{\delta}\right\} \geq \varepsilon, \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where δ is a scaling factor, and ε is the threshold that controls the sparsity of matrix A . Two nodes are connected if their scaled distance is larger than or equal to the threshold ε . We also assume that the topology structure of graph G does not change over time, i.e., fixing matrix A .

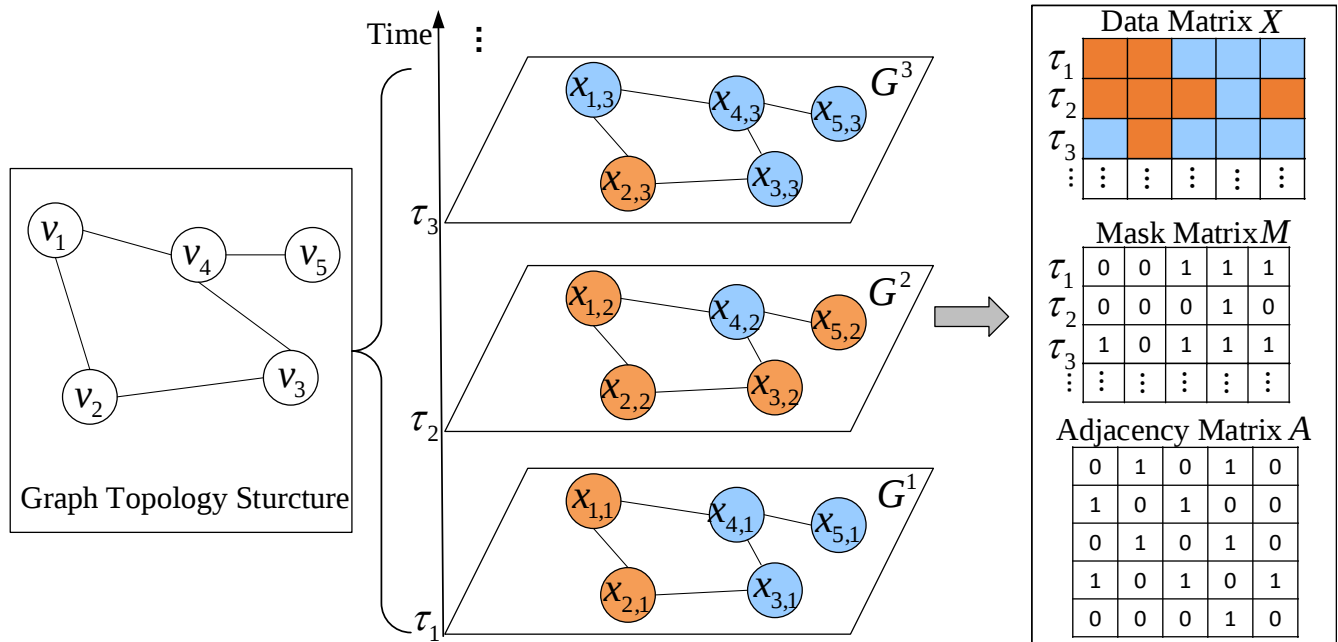


Figure 5. The graph structure of satellite data, where orange color represents missing data, blue color represents observed data, and solid line indicates that two nodes are spatially connected.

As illustrated in Figure 5, for each timestamp τ_j , G^j is used to represent the corresponding instant graph, which may contain missing values at some nodes. For example, in G^1 , nodes v_1 and v_2 have no values. For the whole graph G , we use matrix $X = (X_1, \dots, X_j, \dots, X_T) \in \mathbb{R}^{N \times T}$ to record the time-series satellite data for N nodes at T timestamps, where X_j is the observation values of satellite data for all nodes (i.e., grid regions) at timestamp τ_j , and $x_{i,j}$ is the observation value of node v_i at timestamp τ_j . In addition, we introduce a masked matrix $M = (M_1, \dots, M_j, \dots, M_T) \in \mathbb{R}^{N \times T}$ for X , where each entry $m_{i,j} = 0$ if the entry $x_{i,j} \in X$ is missing, otherwise $m_{i,j} = 1$.

Given the incomplete satellite data matrix X and its masked matrix M , we aim to fill the missing values in X and ensure the filled values are close to the real values. To this end, we design the STA-GAN model to obtain an imputed matrix $\tilde{X}$ that is close to X , i.e.,

$$\tilde{X} = X + \hat{X} \odot (O - M), \quad (3)$$

where $\hat{X}$ is the generated matrix by STA-GAN model, O is a matrix whose entries are all ones, and $\odot$ denotes Hadamard product.

2.3. Methods

2.3.1. Overview

Figure 6 illustrates the structure of the STA-GAN model, which consists of spatio-temporal attention (STA) module and Generative Adversarial Network (GAN) module. First, STA learns the short-term temporal dependence and dynamic spatial dependence in satellite data based on GAT and produces the short-term temporal dependence representation matrix F and the dynamic spatial dependence representation matrix S . Then, the generator and discriminator of GAN are trained by fusing the learned spatio-temporal dependence features. Finally, the missing data is filled with the generated data by GAN module.

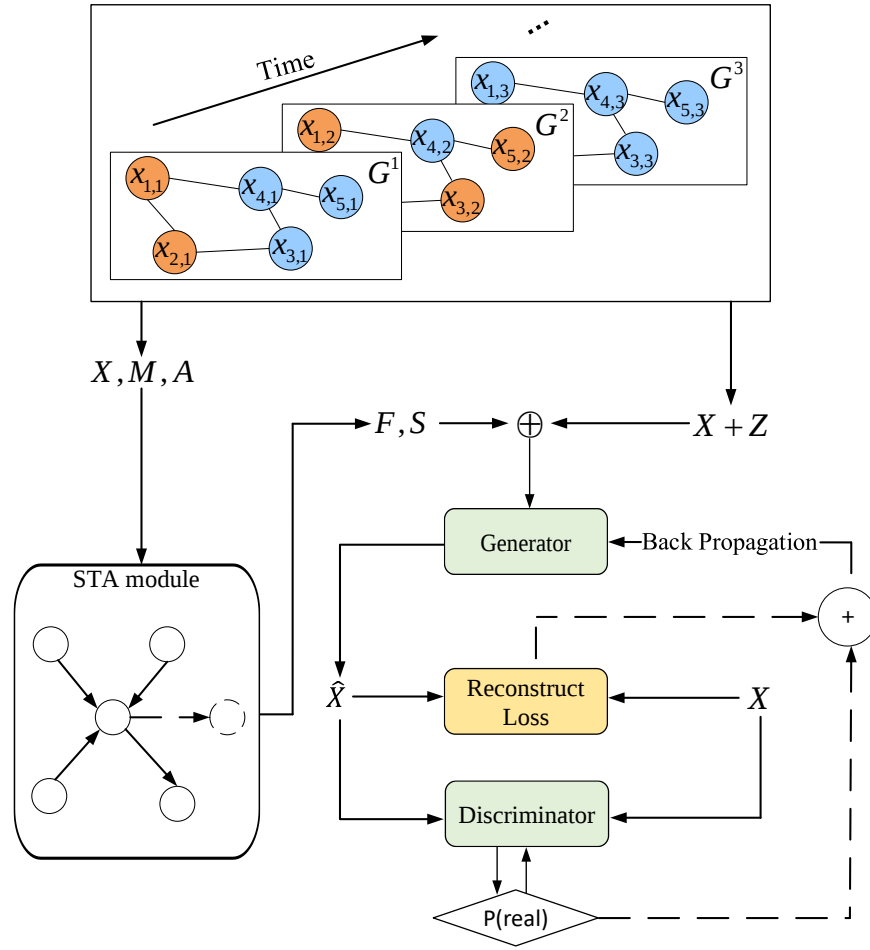


Figure 6. The architecture of the STA-GAN model, where $\oplus$ denotes the concatenation operator, and $+$ indicates the addition of two values.

2.3.2. Spatio-Temporal Attention for Dependence Learning

The STA consists of two operations, i.e., Temporal Attention (TA) operation and Spatial Attention (SA) operation, which are used to learn the short-term temporal dependence and dynamic spatial dependence, respectively.

(a) Temporal Attention

The TA operation focuses on learning the short-term temporal dependence for the time series T_i of each node v_i in graph G . Concretely, we first divide T_i into consecutive sub-sequences of the same length l , thus producing T_i/l sub-sequences. For each node v_i , we learn its short-term temporal dependence representation $f_{i,j}$ at timestamp τ_j as below.

$$f_{i,j} = \sigma \left(\sum_{\tau_k \in N_{i,j}} \alpha_i^{(j,k)} \cdot x_{i,k} \right), \quad (4)$$

where $\sigma(\cdot)$ is the sigmoid activation function, $x_{i,k}$ represents the value of node v_i at timestamp τ_k , $N_{i,j}$ is the set of timestamps that temporally affects node v_i at timestamp τ_j , and $\alpha_i^{(j,k)}$ is the normalized temporal dependence coefficient between timestamps τ_j and τ_k for node v_i , i.e.,

$$\alpha_i^{(j,k)} = \frac{\exp(p_i^{(j,k)})}{\sum_{\tau_k \in N_{i,j}} \exp(p_i^{(j,k)})}, \quad (5)$$

where $p_i^{(j,k)}$ represents the temporal dependence coefficient between timestamps τ_j and τ_k for node v_i , and is calculated by

$$p_i^{(j,k)} = \text{LeakyReLU}(w^t \cdot ((x_{i,j} \oplus m_{i,j}) \oplus (x_{i,k} \oplus m_{i,k}))). \quad (6)$$

where $\text{LeakyReLU}(\cdot)$ is a nonlinear activation function, $\oplus$ denotes the concatenation operator, w^t is the learnable parameters of TA operation, $m_{i,j}$ and $m_{i,k}$ are the masked values of timestamps τ_j and τ_k , respectively, for node v_i .

Figure 7 illustrates an example of TA operation. We set the length l to 3, and the first sub-sequence of node v_5 contains the values at timestamps τ_1 , τ_2 and τ_3 . The corresponding values of v_5 at these three timestamps are $x_{5,1}$, $x_{5,2}$ and $x_{5,3}$, respectively. The value $x_{5,1}$ of node v_5 at timestamp τ_1 is affected by values $x_{5,2}$ and $x_{5,3}$ with temporal dependence coefficient $\alpha_5^{(1,2)}$ and $\alpha_5^{(1,3)}$, respectively. The short-term temporal dependence representation $f_{5,1}$ for node v_5 at time stamp τ_1 is obtained by calculating $\sigma(\alpha_5^{(1,2)} \cdot x_{5,2} + \alpha_5^{(1,3)} \cdot x_{5,3})$.

With temporal attention, we first calculate the short-term temporal dependence representation matrix F_j for all nodes at each timestamp τ_j , i.e., $F_j = (f_{1,j}, \dots, f_{i,j}, \dots, f_{N,j})^T \in \mathbb{R}^{N \times 1}$, where $(\cdot)^T$ represents matrix transpose. Then, we concatenate matrices $F_1, \dots, F_j, \dots, F_T$ to obtain the short-term temporal dependence representation matrix $F \in \mathbb{R}^{N \times T}$.

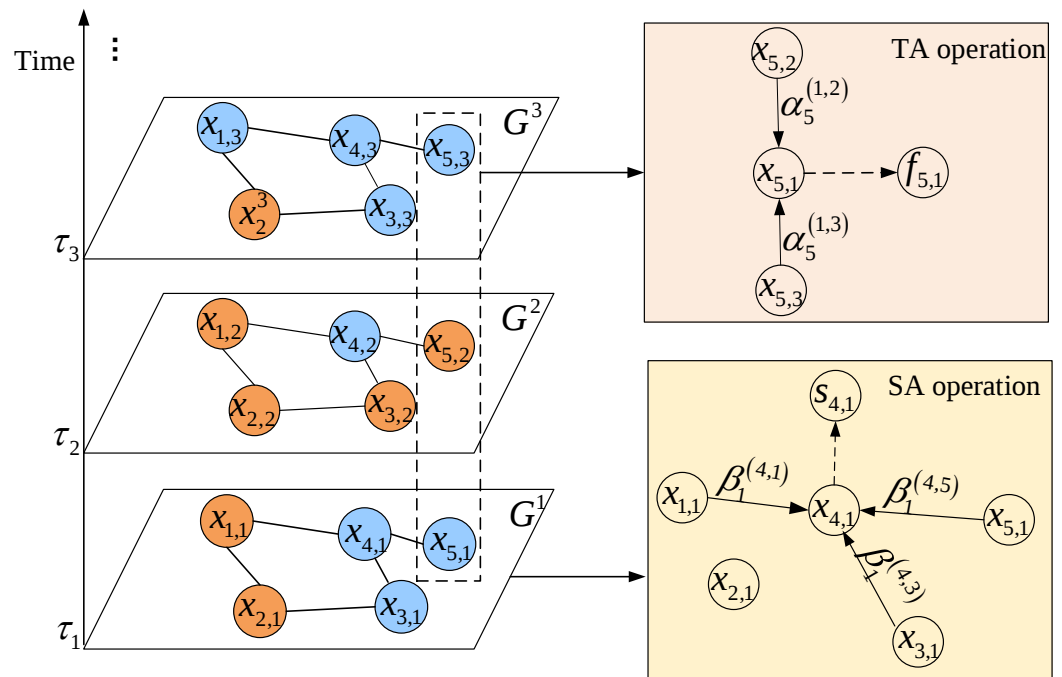


Figure 7. Diagram of STA for temporal attention operation and spatial attention operation.

(b) Spatial Attention

The SA operation learns the dynamic spatial dependence for each node v_i in graph G . Concretely, for each node v_i , we learn its dynamic spatial dependence representation $s_{i,j}$ at timestamp τ_j as below.

$$s_{i,j} = \sigma \left(\sum_{v_k \in N_{i,j}} \beta_j^{(i,k)} \cdot x_{k,j} \right), \quad (7)$$

where $\sigma(\cdot)$ is the sigmoid activation function, $x_{k,j}$ represents the value of node v_k at timestamp τ_j , $N_{i,j}$ is the set of nodes that spatially affects node v_i at timestamp τ_j , and $\beta_j^{(i,k)}$ is the normalized spatial dependence coefficient between nodes v_i and v_k at timestamp τ_j , i.e.,

$$\beta_j^{(i,k)} = \frac{\exp(q_j^{(i,k)})}{\sum_{v_k \in N_{i,j}} \exp(q_j^{(i,k)})}, \quad (8)$$

where $q_j^{(i,k)}$ represents the spatial dependence coefficient between nodes v_i and v_k at timestamp τ_j and is calculated by

$$q_j^{(i,k)} = \text{LeakyReLU}\left(w^s \cdot \left((x_{i,j} \oplus m_{i,j}) \oplus (x_{k,j} \oplus m_{k,j})\right)\right). \quad (9)$$

where $\text{LeakyReLU}(\cdot)$ is a nonlinear activation function, w^s is the learnable parameter of SA operation, and $m_{i,j}$ and $m_{k,j}$ are the masked values of nodes v_i and v_k , respectively, at timestamp τ_j .

Figure 7 illustrates an example of SA operation. For timestamp τ_1 , node v_4 is affected by nodes v_1 , v_3 and v_5 with spatial dependence coefficient $\beta_1^{(4,1)}$, $\beta_1^{(4,3)}$ and $\beta_1^{(4,5)}$, respectively. The dynamic spatial dependence representation $s_{4,1}$ for node v_4 at timestamp τ_1 is obtained by calculating $\sigma\left(\beta_1^{(4,1)} \cdot x_{1,1} + \beta_1^{(4,3)} \cdot x_{3,1} + \beta_1^{(4,5)} \cdot x_{5,1}\right)$.

With spatial attention, we first obtain the dynamic spatial dependence representation matrix S_j for all nodes at each timestamp τ_j , i.e., $S_j = (s_{1,j}, \dots, s_{i,j}, \dots, s_{N,j})^T \in \mathbb{R}^{N \times 1}$. Then, we combine $S_1, \dots, S_j, \dots, S_T$ to obtain the dynamic spatial dependence representation matrix $S \in \mathbb{R}^{N \times T}$.

2.3.3. Generative Adversarial Network for Data Imputation

GAN module achieves missing value imputation of satellite data by learning data distribution with the learned short-term temporal dependence matrix F and dynamic spatial dependence matrix S . We introduce Wasserstein GAN (WGAN) [54] to fill missing values in satellite data. In traditional WGAN, a pure noise matrix Z is usually used as the input of the generator to generate a matrix $\hat{X}$, where $Z \sim N(0, 1)$ is a pure noise matrix. This process does not consider the information about data, thus making it time-consuming to generate a matrix $\hat{X}$ that is close to the real data distribution [47,55]. To address this issue, we integrate $X + Z$, the short-term temporal dependence representation matrix F and the dynamic spatial dependence representation matrix S as the inputs of the generator to provide comprehensive spatio-temporal dependence information for generating matrix $\hat{X}$. In addition, we re-design the structures of the generator and discriminator to enable them to handle the multiple inputs and to better learn the spatio-temporal distribution of satellite data. Figure 8 shows the structure of the generator of STA-GAN model that consists of a sequence of GRU units, and Figure 9 shows the structure of the GRU unit.

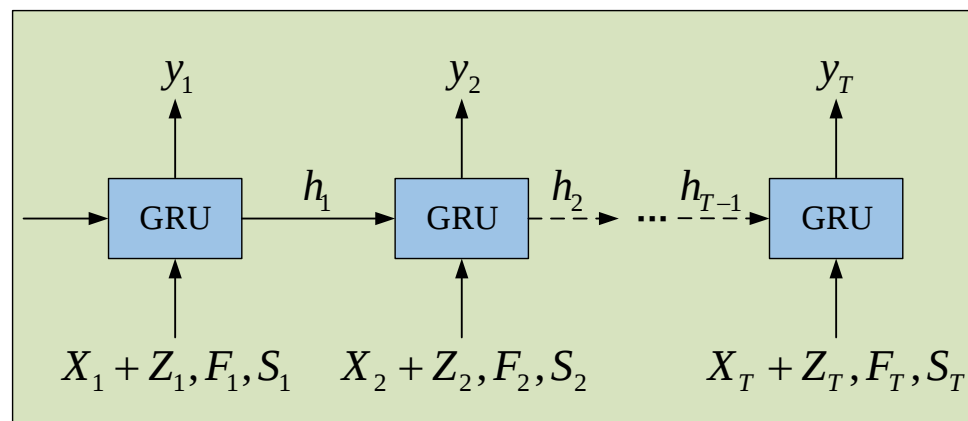


Figure 8. The structure of the generator of STA-GAN model.

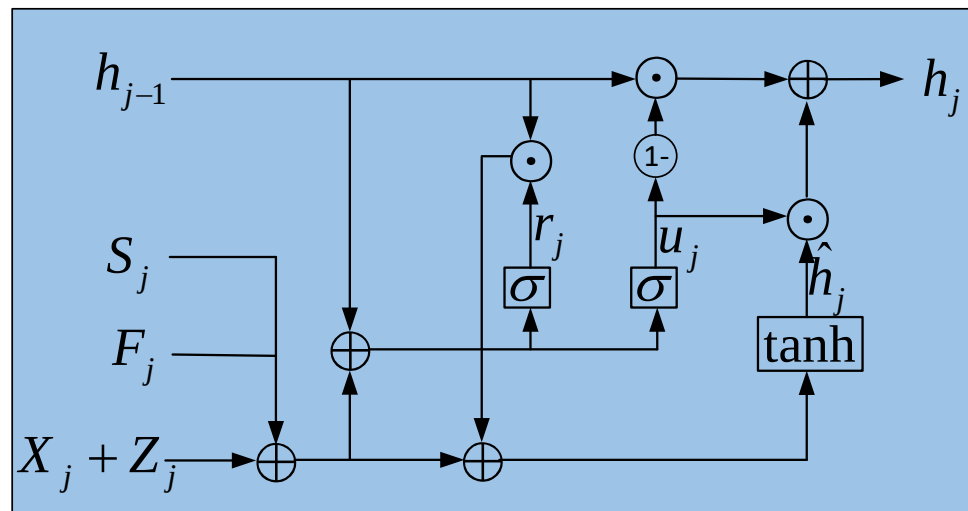


Figure 9. The structure of GRU unit, where $\odot$ is Hadamard product, and $\oplus$ is the concatenation operator.

The updating functions of GRU as listed as below.

$$u_j = \sigma(W_u[X_j + Z_j, F_j, S_j, h_{j-1}] + b_u). \quad (10)$$

$$r_j = \sigma(W_r[X_j + Z_j, F_j, S_j, h_{j-1}] + b_r). \quad (11)$$

$$\hat{h}_j = \tanh(W_h[X_j + X_j, F_j, S_j, h_{j-1} \odot r_j] + b_h). \quad (12)$$

$$h_j = (1 - u_j) \odot h_{j-1} + u_j \odot \hat{h}_j. \quad (13)$$

where u_j and r_j are the update gates and reset gates, $\hat{h}_j$ is the candidate hidden state, $\sigma(\cdot)$ is the sigmoid function, W_u , W_r , W_h , b_u , b_r and b_h are the trainable parameters.

The discriminator has a similar structure to generator. The discriminator takes the original data matrix X and the generated matrix $\hat{X}$ of the generator as inputs to obtain the probability of the authenticity of the generated matrix $\hat{X}$.

The loss function L_G of the generator consists of reconstruction loss L_R and the probability that the discriminator regards the generated matrix $\hat{X}$ as the real data matrix, i.e.,

$$L_G = \lambda L_R - D(\hat{X}), \quad (14)$$

where λ is the hyper-parameter that balances reconstruction loss L_R and the output probability $D(\hat{X})$. The reconstruction loss L_R is the average absolute error between $\hat{X}$ and X , i.e.,

$$L_R = \|\hat{X} - X\|_2. \quad (15)$$

The discriminator is designed to distinguish real data X and the generated matrix $\hat{X}$, and the loss function of the discriminator is

$$L_D = D(\hat{X}) - D(X). \quad (16)$$

where $D(\hat{X})$ and $D(X)$ are the output probabilities of discriminator for matrices $\hat{X}$ and X , respectively.

The back propagation method is used to optimize the loss function of the generator and discriminator by optimizing the corresponding parameters. Finally, we use the generated matrix $\hat{X}$ to fill the missing values in the data matrix X , and obtain the imputation matrix $\tilde{X}$.

3. Results

3.1. Experimental Settings

To evaluate the effectiveness and demonstrate the superiority of our STA-GAN model, we compare it with multiple baseline methods, including:

- (1) MEAN: achieves data imputation by using the mean values of historical records.
- (2) KNN [34]: finds the k nearest neighbors and fills in the missing value with the average of these neighbors.
- (3) GRU-D [56]: achieves the missing value imputation by introducing a decay mechanism in the input variable and hidden state to capture missing information.
- (4) GAIN [45]: a GAN-based data imputation method that introduces a hinting mechanism in discriminator to distinguish the original data and the generated data.
- (5) GAN-2-stage [46]: a data imputation method based on GAN through two-stage training. First, the decay mechanism is introduced into the generator and discriminator to consider time irregularity. Data imputation is then achieved based on the generated matrix from the trained GAN model.
- (6) SolarGAN [47]: an imputation method similar to GAN-2-stage that is used for solar data imputation. The difference with GAN-2-stage is that the noise matrix and the data matrix are fused as the inputs of the generator in SolarGAN.

Among these methods, GAIN, GAN-2-stage and SolarGAN are based on GAN and widely used for data imputation.

We use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as performance evaluation metrics. For STA-GAN model, the length l of the divided sub-sequences is set to 12, and the learning rate is set to 0.001 for all methods. The batch sizes on SST-EAST, SST-SOUTH and CHA-EAST datasets are set to 64, 64 and 32, respectively. For all the experiments, the partition ratio of training, validation and testing data in each data set is set to 8:1:1. We use the Adam optimizer to update the parameters of model. All the data imputation models are implemented on TensorFlow 1.7.1. The dimension of pure noise matrix Z and GRU hidden state are both 64. In addition, we use a 64-core Intel Xeon processor with 256 GB RAM and 3 NVIDIA RTX 2080Ti GPUs.

3.2. Hyper-Parameter Selection

The adjacency matrix A is a representation of the neighbor nodes for each node. For each node, selecting a small number of neighbor nodes cannot fully cover all the spatial dependence information while selecting a large number of neighbor nodes may cause information bias and result in errors in the learned spatial dependence information. According to Equation (2), parameters ε and δ are used to control the number of neighbor nodes for each node. We thus conduct experiments on three datasets, respectively, to identify a good setting for the two parameters. Figure 10 presents the results of STA-GAN model while varying ε and δ , where the missing rate is 0.5. To generate data with different missing rates, we randomly mask some real data values in satellite data. According to the results, we set $\varepsilon = 0.3$, $\delta = 0.05$ for the SST-EAST and CHA-EAST datasets, and set $\varepsilon = 0.9$, $\delta = 0.01$ for the SST-SOUTH dataset.

The parameter λ is used to balance the importance of reconstruction loss and the output probability of discriminator in STA-GAN model. Figure 11 illustrates the results of STA-GAN while varying λ , where the missing rate is 0.5. According to the results, we set $\lambda = 200$ for SST-EAST and SST-SOUTH datasets and $\lambda = 300$ for CHA-EAST dataset.

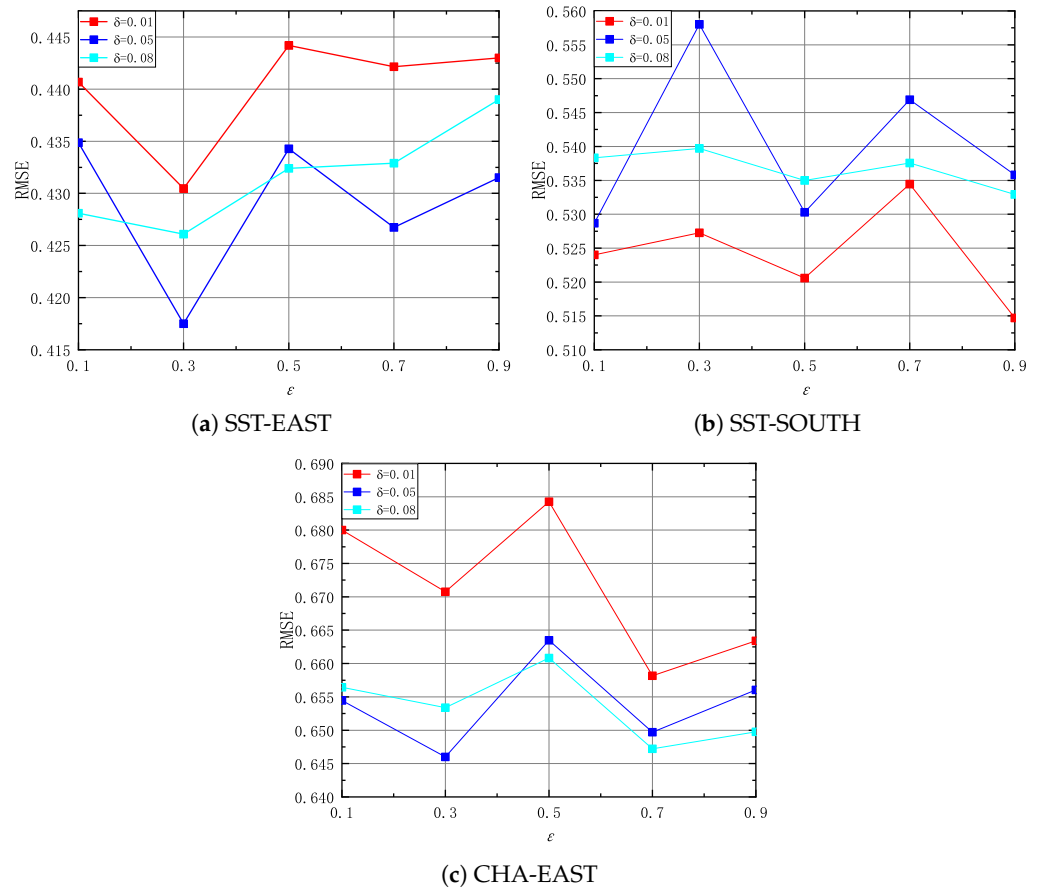


Figure 10. The RMSE of STA-GAN model while varying ε and δ , where the missing rate is 0.5.

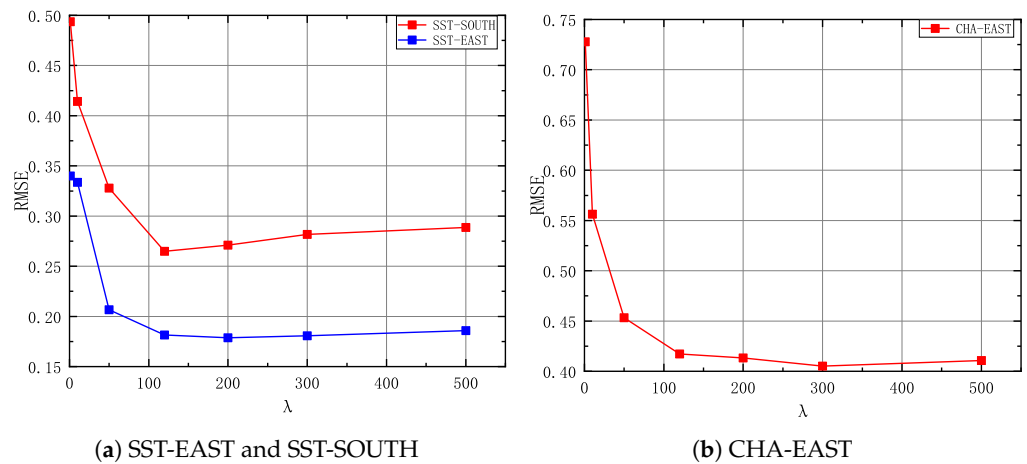


Figure 11. The RMSE of STA-GAN model while varying λ , where the missing rate is 0.5.

3.3. Performance Comparison with Baseline Methods

Figures 12–14 show the results of STA-GAN model and six baseline methods on the SST-EAST, SST-SOUTH, and CHA-EAST datasets, respectively, while varying the missing rate. Meanwhile, Tables 1–3 show the corresponding quantitative experimental results. In general, with the increase of the missing rate, the performance of all the data imputation methods, except for MEAN method, degrades. However, the performance decline trend of the STA-GAN model is significantly lower than that of other baseline methods, which indicates that STA-GAN effectively learns the spatio-temporal information in satellite data. Among the baseline methods, MEAN and KNN have the worst performance since they do not consider

the correlations between time series. Meanwhile, GAN-2-stage, GRU-D and SolarGAN perform better than the other baseline methods. However, STA-GAN model outperforms all these baseline methods at various missing rates and the superiority increases rapidly when the missing rate increases. The reason is that STA-GAN provides enriched spatio-temporal information for data imputation (especially at high missing rates) by simultaneously learning the short-term temporal dependence and dynamic spatial dependence features.

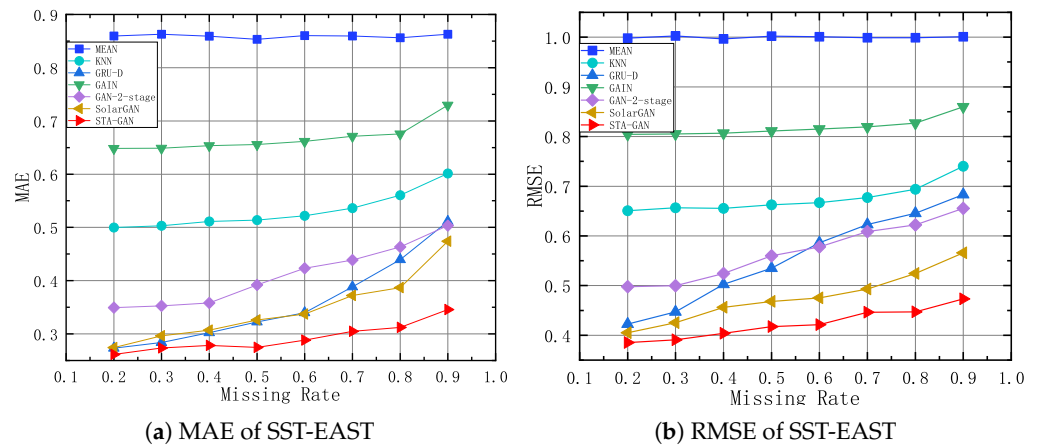


Figure 12. MAE and RMSE of SST-EAST dataset with different missing rates.

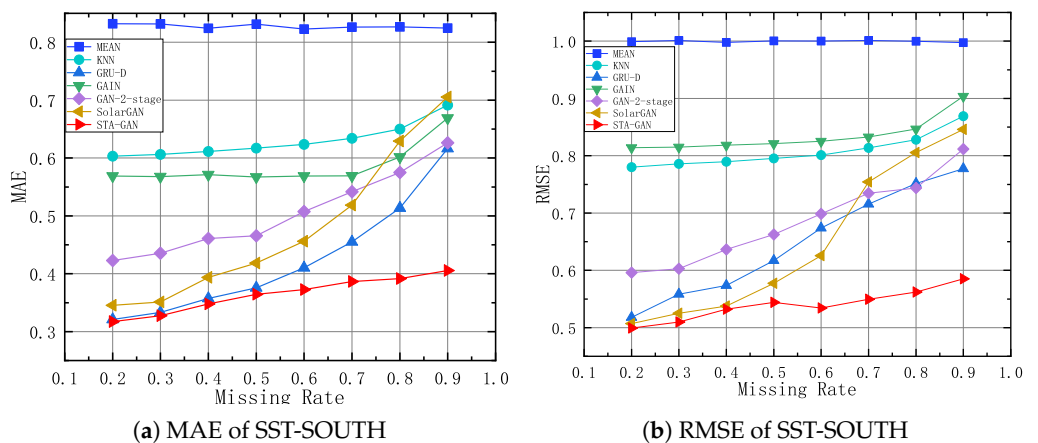


Figure 13. MAE and RMSE of SST-SOUTH dataset with different missing rates.

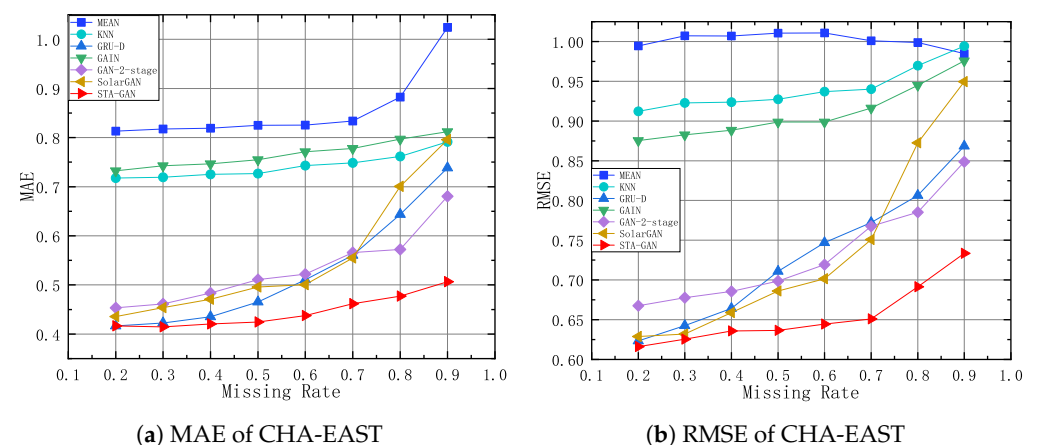


Figure 14. MAE and RMSE of CHA-EAST dataset with different missing rates.

Table 1. The results (RMSE and MAE) of SST-EAST dataset with different missing rates.

Missing Rate	Metris	MEAN	KNN	GRU-D	GAIN	GAN-2-Stage	SolarGAN	STA-GAN
0.2	RMSE	0.9982	0.6509	0.4225	0.8043	0.4977	0.4051	0.3851
	MAE	0.8595	0.4999	0.2731	0.6483	0.3494	0.2743	0.2613
0.3	RMSE	1.0024	0.6567	0.4472	0.8052	0.4998	0.4252	0.3908
	MAE	0.8627	0.5032	0.2837	0.6489	0.3526	0.2962	0.2734
0.4	RMSE	0.9964	0.6554	0.5024	0.8067	0.5245	0.4562	0.4037
	MAE	0.8592	0.5111	0.3024	0.6536	0.3581	0.3071	0.2784
0.5	RMSE	1.0019	0.6627	0.5352	0.8112	0.5599	0.4682	0.4175
	MAE	0.8534	0.5137	0.3225	0.6556	0.3920	0.3261	0.2747
0.6	RMSE	1.0009	0.6669	0.5862	0.8152	0.5779	0.4753	0.4215
	MAE	0.8604	0.5217	0.3399	0.6615	0.4234	0.3373	0.2882
0.7	RMSE	0.9990	0.6772	0.6231	0.8193	0.6087	0.4930	0.4464
	MAE	0.8597	0.5363	0.3883	0.6712	0.4386	0.3722	0.3047
0.8	RMSE	0.9905	0.6938	0.6455	0.8270	0.6220	0.5245	0.4472
	MAE	0.8563	0.5606	0.4396	0.6756	0.4634	0.3868	0.3121
0.9	RMSE	1.0010	0.7402	0.6831	0.8598	0.6554	0.5658	0.4734
	MAE	0.8627	0.6013	0.5111	0.7298	0.5035	0.4739	0.3458

Table 2. The results (RMSE and MAE) of SST-SOUTH dataset with different missing rates.

Missing Rate	Metris	MEAN	KNN	GRU-D	GAIN	GAN-2-Stage	SolarGAN	STA-GAN
0.2	RMSE	0.9989	0.7799	0.5179	0.8139	0.5960	0.5072	0.4991
	MAE	0.8321	0.6032	0.3211	0.5690	0.4227	0.3453	0.3174
0.3	RMSE	1.0012	0.7857	0.5583	0.8150	0.6028	0.5249	0.5096
	MAE	0.8316	0.6062	0.3333	0.5677	0.4353	0.3511	0.3274
0.4	RMSE	0.9976	0.7897	0.5737	0.8184	0.6365	0.5377	0.5322
	MAE	0.8241	0.6113	0.3573	0.5712	0.4610	0.3935	0.3478
0.5	RMSE	1.0005	0.7952	0.6172	0.8213	0.6626	0.5774	0.5439
	MAE	0.8313	0.6170	0.3756	0.5671	0.4655	0.4184	0.3645
0.6	RMSE	1.0002	0.8012	0.6739	0.8252	0.6984	0.6256	0.5342
	MAE	0.8227	0.6235	0.4104	0.5687	0.5075	0.4562	0.3727
0.7	RMSE	1.0010	0.8138	0.7155	0.8327	0.7346	0.7541	0.5495
	MAE	0.8262	0.6339	0.4550	0.5693	0.5416	0.5187	0.3865
0.8	RMSE	0.9998	0.8279	0.7515	0.8465	0.7436	0.8054	0.5621
	MAE	0.8267	0.6500	0.5134	0.6019	0.5748	0.6293	0.3917
0.9	RMSE	0.9972	0.8691	0.7774	0.9036	0.8115	0.8460	0.5852
	MAE	0.8246	0.6915	0.6165	0.6689	0.6262	0.7055	0.4055

Table 3. The results (RMSE and MAE) of CHA-EAST dataset with different missing rates.

Missing Rate	Metris	MEAN	KNN	GRU-D	GAIN	GAN-2-Stage	SolarGAN	STA-GAN
0.2	RMSE	0.9944	0.9121	0.6233	0.8755	0.6677	0.6289	0.6162
	MAE	0.8130	0.7178	0.4164	0.7321	0.4532	0.4357	0.4167
0.3	RMSE	1.0073	0.9227	0.6426	0.8827	0.6776	0.6321	0.6254
	MAE	0.8174	0.7194	0.4224	0.7425	0.4614	0.4536	0.4144
0.4	RMSE	1.0070	0.9238	0.6641	0.8885	0.6857	0.6589	0.6358
	MAE	0.8192	0.7253	0.4351	0.7466	0.4835	0.4707	0.4204
0.5	RMSE	1.0106	0.9274	0.7108	0.8987	0.6986	0.6861	0.6366
	MAE	0.8250	0.7268	0.4652	0.7550	0.5108	0.4959	0.4246
0.6	RMSE	1.0109	0.9370	0.7469	0.8988	0.7193	0.7016	0.6446
	MAE	0.8255	0.7430	0.5106	0.7711	0.5219	0.5004	0.4374
0.7	RMSE	1.0010	0.9401	0.7721	0.9163	0.7676	0.7509	0.6510
	MAE	0.8336	0.7487	0.5604	0.7780	0.5658	0.5549	0.4617
0.8	RMSE	0.9987	0.9698	0.8063	0.9448	0.7853	0.8726	0.6913
	MAE	0.8825	0.7615	0.6437	0.7969	0.5723	0.7007	0.4770
0.9	RMSE	0.9849	0.9943	0.8684	0.9756	0.8486	0.9495	0.7336
	MAE	1.0242	0.7913	0.7383	0.8121	0.6804	0.7957	0.5065

3.4. Ablation Experiment

Table 4 shows the results of an ablation experiment for the STA-GAN model on three datasets, where the data missing rate is set to 0.5. According to the results, the performance of the STA-GAN model degenerates after removing the spatial attention and temporal attention, which validates the effectiveness of each module, and indicates that considering the spatio-temporal dependence in satellite data can help reduce the data imputation error.

Table 4. The results of the ablation experiment.

Model	SST-EAST		SST-SOUTH		CHA-EAST	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
w/o SA	0.4326	0.2869	0.5548	0.3800	0.6461	0.4368
w/o TA	0.4245	0.2783	0.5487	0.3719	0.6421	0.4279
STA-GAN	0.4175	0.2747	0.5439	0.3645	0.6366	0.4246

4. Discussion

STA-GAN model is tailored for filling missing values in spatio-temporal data (e.g., SST satellite data and Chl-a satellite data) with high missing rates. To this end, it tries to learn comprehensive spatio-temporal dependence information from original data by combining Spatio-Temporal Attention and Generative Adversarial Network.

Figure 15 plots the original satellite data and the generated data by STA-GAN model while setting the missing rate to 0.2, 0.5 and 0.9, respectively. Three cases are from the CHA-EAST dataset and select one location without missing data in 2021. In all three cases, the imputed data has the similar trend with the original data, and the difference between the imputed data and the original data is small, especially at the missing rates of 0.2 and 0.5. For the data imputation at missing rate of 0.9, the error is large since quite limited knowledge can be learned from the real satellite data to support accurate data imputation.

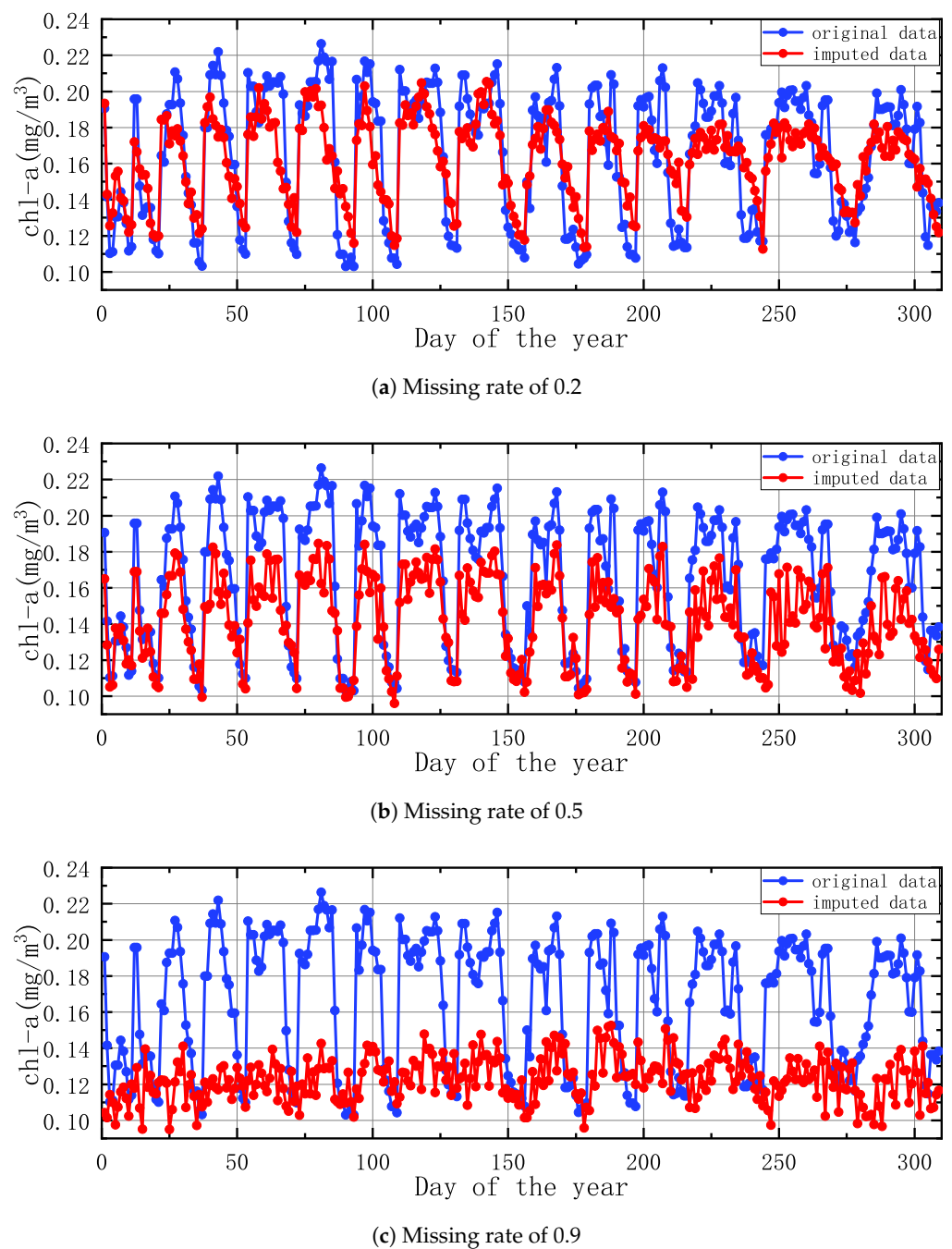


Figure 15. The plotting of the original satellite data and the generated data by STA-GAN model with different missing rates.

According to the results in Section 3.3, compared with existing data imputation methods, STA-GAN model still has good data imputation performance under the high data missing rates, which is mainly due to the abundant spatio-temporal dependence information extracted by the model. On the one hand, the STA-GAN model considers the dynamic spatial dependence. The spatial dependence is dynamic over time. For example, the change of wind direction could result in the change of spatial dependence of SST, and the direction of red tide reproduction also affects the spatial dependence of chlorophyll concentration. Therefore, extracting dynamic spatial dependence information can effectively improve the data imputation performance. On the other hand, the bidirectional short-term temporal dependence is considered in STA-GAN model. In general, the missing data is usually similar to the data from the past few days, which is why existing data imputation models such as

GRU-D [56], GAN-2-stage [46] and SolarGAN [47] consider historical data in the process of data imputation. However, for satellite remote sensing data, only considering the impacts of historical data is not enough since the missing data is also related to the patterns of future data. Therefore, considering the dynamic bidirectional short-term temporal dependence in STA-GAN model can better capture the dynamic changes of satellite data over time, thus bringing better data imputation performance.

5. Conclusions

In this article, we proposed the STA-GAN model for missing value imputation in satellite data. STA-GAN can effectively learn features for capturing both short-term temporal dependence and dynamic spatial dependence in satellite data and exploit the learned features to train a generative model to fill the missing values. According to the experimental results on three real datasets, STA-GAN model achieves much better performance than existing methods, especially for filling satellite data with large missing rates. For future work, we would like to consider some external factors, e.g., humidity and wind, to further enhance the performance of STA-GAN model.

Author Contributions: Conceptualization, S.W. and W.L.; data curation, S.W. and J.Y.; formal analysis, S.W. and W.L.; methodology, S.W., W.L. and S.H.; supervision, J.G.; validation, S.W.; visualization, S.W.; writing—original draft, S.W. and J.Y.; writing—review and editing, S.W. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by National Natural Science Foundation of China (No. 62202336, No. U1936205, No. 62172300), Shanghai Pujiang Program (No. 20PJ1414300), National Key R&D Program of China (No. 2021YFC3300300), Open Research Projects of Zhejiang Lab (No. 2021KH0AB04), and the Fundamental Research Funds for the Central Universities (No. ZD-21-202101).

Data Availability Statement: The SST satellite data provided by AVHRR is available from <https://www.ncei.noaa.gov/>, accessed on 11 October 2022. The Chl-a satellite data from OC-CCI products is available at <https://www.oceancolour.org/>, accessed on 11 October 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martin, S. *An Introduction to Ocean Remote Sensing*; Cambridge University Press: Cambridge, UK, 2014.
2. He, X.; Shi, S.; Geng, X.; Xu, L.; Zhang, X. Spatial-temporal attention network for multistep-ahead forecasting of chlorophyll. *Appl. Intell.* **2021**, *51*, 4381–4393. [\[CrossRef\]](#)
3. Lee, M.S.; Park, K.A.; Chae, J.; Park, J.E.; Lee, J.S.; Lee, J.H. Red tide detection using deep learning and high-spatial resolution optical satellite imagery. *Int. J. Remote Sens.* **2020**, *41*, 5838–5860. [\[CrossRef\]](#)
4. Qin, M.; Li, Z.; Du, Z. Red tide time series forecasting by combining ARIMA and deep belief network. *Knowl.-Based Syst.* **2017**, *125*, 39–52. [\[CrossRef\]](#)
5. Zheng, G.; Li, X.; Zhang, R.; Liu, B. Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* **2020**, *6*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Woodring, J.; Petersen, M.; Schmeisser, A.; Patchett, J.; Ahrens, J.; Hagen, H. In Situ Eddy Analysis in a High-Resolution Ocean Climate Model. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 857–866. [\[CrossRef\]](#)
7. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
8. Gao, S.; Zhao, P.; Pan, B.; Li, Y.; Zhou, M. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. *Acta Oceanolog. Sin.* **2018**, *37*, 8–12. [\[CrossRef\]](#)
9. Rüttgers, M.; Lee, S.; Jeon, S.; Donghyun, Y. Prediction of a typhoon track using a generative adversarial network and satellite images. *Sci. Rep.* **2019**, *9*, 6057. [\[CrossRef\]](#)
10. Guan, L.; Kawamura, H. SST availabilities of satellite infrared and microwave measurements. *J. Oceanogr.* **2003**, *59*, 201–209. [\[CrossRef\]](#)
11. King, M.D.; Platnick, S.; Menzel, W.P.; Ackerman, S.A.; Hubanks, P.A. Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3826–3852. [\[CrossRef\]](#)
12. Huang, B.Y.; Liu, C.Y.; Banzon, V.; Freeman, E.; Graham, G.; Hankins, B.; Smith, T.; Zhang, H.M. Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *J. Clim.* **2021**, *34*, 2923–2939. [\[CrossRef\]](#)

13. Donlon, C.J.; Martin, M.; Stark, J.; Roberts-Jones, J.; Fiedler, E.; Wimmer, W. The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sens. Environ.* **2012**, *116*, 140–158. [\[CrossRef\]](#)
14. Kohtaro, H.; Futoki, S. Global daily high-resolution satellite-based foundation sea surface temperature dataset: Development and validation against two definitions of foundation SST. *Remote Sens.* **2016**, *8*, 962.
15. He, R.; Weisberg, R.H.; Zhang, H.; Muller-Karger, F.E.; Helber, R.W. A cloud-free, satellite-derived, sea surface temperature analysis for the West Florida Shelf. *Geophys. Res. Lett.* **2003**, *30*, 1–5. [\[CrossRef\]](#)
16. Beckers, J.; Rixen, M. EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Ocean. Technol.* **2003**, *20*, 1839–1856. [\[CrossRef\]](#)
17. Liu, X.; Wang, M. Gap filling of missing data for VIIRS global ocean color products using the DINEOF method. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4464–4476. [\[CrossRef\]](#)
18. Guo, J.; Lu, J.; Zhang, Y.; Zhou, C.; Zhang, S.; Wang, D.; Lv, X. Variability of chlorophyll-a and secchi disk depth (1997–2019) in the Bohai Sea based on monthly cloud-free satellite data reconstructions. *Remote Sens.* **2022**, *14*, 639. [\[CrossRef\]](#)
19. Ma, C.; Zhao, J.; Ai, B.; Sun, S. Two-Decade variability of sea surface temperature and chlorophyll-a in the Northern South China Sea as revealed by reconstructed cloud-free satellite data. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9033–9046. [\[CrossRef\]](#)
20. Li, Y.; Sun, W.; Zhang, J.; Meng, J.; Zhao, Y. Reconstruction of arctic SST data and generation of multi-source satellite fusion products with high temporal and spatial resolutions. *Remote Sens. Lett.* **2021**, *12*, 695–703. [\[CrossRef\]](#)
21. Zhao, X.; Hou, Y.; Qi, P. Interpretation of sea surface wind interannual vector EOFs over the China seas. *Chin. J. Oceanol. Limn.* **2010**, *28*, 340–343. [\[CrossRef\]](#)
22. Alvera-Azcárate, A.; Barth, A.; Beckers, J.M.; Weisberg, R.H. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. *J. Geophys. Res.-Oceans* **2007**, *112*, 1–11. [\[CrossRef\]](#)
23. Bhattacharjee, S.; Mitra, P.; Ghosh, S.K. Spatial interpolation to predict missing attributes in GIS using semantic kriging. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4771–4780. [\[CrossRef\]](#)
24. Rossi, R.E.; Dungan, J.L.; Beck, L.R. Kriging in the shadows: Geostatistical interpolation for remote sensing. *Remote Sens. Environ.* **1994**, *49*, 32–40. [\[CrossRef\]](#)
25. Lu, G.; Wong, D. An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geotech.* **2008**, *34*, 1044–1055. [\[CrossRef\]](#)
26. Shen, J.; Cao, J.; Liu, X.; Zhang, C. DMAD: Data-driven measuring of wi-fi access point deployment in urban spaces. *ACM Trans. Intell. Syst. Technol. (TIST)* **2017**, *9*, 1–29. [\[CrossRef\]](#)
27. Park, J.; Kim, H.C.; Bae, D.; Jo, Y.H. Data reconstruction for remotely sensed chlorophyll-a concentration in the Ross Sea using ensemble-based machine learning. *Remote Sens.* **2020**, *12*, 1898. [\[CrossRef\]](#)
28. Chen, S.; Hu, C.; Barnes, B.; Xie, Y.; Lin, G.; Qiu, Z. Improving ocean color data coverage through machine learning. *IEEE T Geosci Remote.* **2019**, *222*, 286–302. [\[CrossRef\]](#)
29. Hu, C.; Feng, L.; Guan, Q. A machine learning approach to estimate surface chlorophyll a concentrations in global oceans from satellite measurements. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4590–4607. [\[CrossRef\]](#)
30. Mohebzadeh, H.; Mokari, E.; Daggupati, P.; Biswas, A. A machine learning approach for spatiotemporal imputation of MODIS chlorophyll-a. *Int. J. Remote Sens.* **2021**, *42*, 7381–7404. [\[CrossRef\]](#)
31. Park, J.; Kim, J.H.; Kim, H.C.; Kim, B.K.; Bae, D.; Jo, Y.H.; Jo, N.; Lee, S.H. Reconstruction of ocean color data using machine learning techniques in Polar regions: Focusing on Off Cape Hallett, Ross Sea. *Remote Sens.* **2019**, *11*, 1366. [\[CrossRef\]](#)
32. Xing, M.; Yao, F.; Zhang, J.; Meng, X.; Jiang, L.; Bao, Y. Data reconstruction of daily MODIS chlorophyll-a concentration and spatio-temporal variations in the Northwestern Pacific. *Sci. Total Environ.* **2022**, *843*, 156981. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Sunder, S.; Ramsankaran, R.; Ramakrishnan, B. Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from MODIS data. *Isprs J. Photogramm.* **2020**, *166*, 228–240. [\[CrossRef\]](#)
34. Poloczek, J.; Treiber, N.; Kramer, O. KNN Regression as Geo-Imputation Method for Spatio-Temporal Wind Data. In Proceedings of the International Joint Conference SOCO'14-CISIS'14-ICEUTE'14, Bilbao, Spain, 25–27 June 2014; Volume 299, pp. 185–193.
35. Liu, H.; Li, L. Missing Data Imputation in GNSS Monitoring Time Series Using Temporal and Spatial Hankel Matrix Factorization. *Remote Sens.* **2022**, *14*, 1500. [\[CrossRef\]](#)
36. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [\[CrossRef\]](#)
37. Wang, S.; Cao, J.; Yu, P. Deep learning for spatio-temporal data mining: A survey. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 3681–3700. [\[CrossRef\]](#)
38. Jean-Marie, V.; Frederic, J.; Ronan, F.; Baptiste, M.; Ludivine, L.; Christophe, D. Data-Driven interpolation of sea surface suspended concentrations derived from ocean colour remote sensing data. *Remote Sens.* **2021**, *13*, 3537.
39. Pisoni, E.; Pastor, F.; Volta, M. Artificial neural networks to reconstruct incomplete satellite data: Application to the mediterranean sea surface temperature. *Nonlinear Process Geophys.* **2008**, *15*, 61–70. [\[CrossRef\]](#)
40. Jouini, M.; Levy, M.; Crepon, M.; Thiria, S. Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method. *Remote Sens. Environ.* **2013**, *131*, 232–246. [\[CrossRef\]](#)
41. Barth, A.; Alvera-Azcárate, A.; Licer, M.; Beckers, J.M. DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geosci. Model Dev.* **2020**, *13*, 1609–1622. [\[CrossRef\]](#)

42. Han, Z.; He, Y.; Liu, G.; Perrie, W. Application of DINCAE to reconstruct the gaps in chlorophyll-a satellite observations in the South china sea and West philippine sea. *Remote Sens.* **2020**, *12*, 4805. [[CrossRef](#)]
43. Jung, S.; Yoo, C.; Im, J. High-Resolution seamless daily sea surface temperature based on satellite data fusion and machine learning over kuroshio extension. *Remote Sens.* **2022**, *14*, 575. [[CrossRef](#)]
44. Kim, J.; Tae, D.; Seok, J. A survey of missing data imputation using generative adversarial networks. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Fukuoka, Japan, 19–21 February 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 454–456.
45. Yoon, J.; Jordon, J.; Schaar, M. GAIN: Missing Data Imputation using Generative Adversarial Nets. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1–10.
46. Luo, Y.; Cai, X.; Zhang, Y.; Xu, J.; Yuan, X. Multivariate Time Series Imputation with Generative Adversarial Networks. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 1–12.
47. Zhang, W.; Luo, Y.; Zhang, Y.; Srinivasan, D. SolarGAN: Multivariate solar data imputation using generative adversarial network. *IEEE Trans. Sustain. Energy* **2020**, *12*, 743–746. [[CrossRef](#)]
48. Chorowski, J.K.; Bahdanau, D.; Serdyuk, D.; Cho, K.; Bengio, Y. Attention-based models for speech recognition. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9.
49. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
50. Xie, J.; Zhang, J.; Yu, J.; Xu, L. An adaptive scale sea surface temperature predicting method based on deep learning with attention mechanism. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *17*, 740–744. [[CrossRef](#)]
51. Guo, X.; He, J.; Wang, B.; Wu, J. Prediction of Sea Surface Temperature by Combining Interdimensional and Self-Attention with Neural Networks. *Remote Sens.* **2022**, *14*, 4737. [[CrossRef](#)]
52. Wang, X.; Xu, L. Unsteady multi-element time series analysis and prediction based on spatial-temporal attention and error forecast fusion. *Future Internet* **2020**, *12*, 34. [[CrossRef](#)]
53. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
54. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 214–223.
55. Wang, S.; Cao, J.; Chen, H.; Peng, H.; Huang, Z. SeqST-GAN: Seq2Seq generative adversarial nets for multi-step urban crowd flow prediction. *ACM Trans. Spat. Algorithms Syst. (TSAS)* **2020**, *6*, 1–24. [[CrossRef](#)]
56. Che, Z.; Sanjay, P.; Kyunghyun, C.; David, S.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *arXiv* **2016**, arXiv:1606.01865.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.