remote sensing

MDPI

Communication

On the Reconstruction of Missing Sea Surface Temperature Data from Himawari-8 in Adjacent Waters of Taiwan Using DINEOF Conducted with 25-h Data

Yi-Chung Yang [1], Ching-Yuan Lu [1], Shih-Jen Huang [1], Thwong-Zong Yang [2], Yu-Cheng Chang [2] and Chung-Ru Ho [1,*]

[1] Department of Marine Environmental Informatics, National Taiwan Ocean University, 2 Pei-Ning Road, Keelung 202301, Taiwan; 29981001@mail.ntou.edu.tw (Y.-C.Y.); 20681002@mail.ntou.edu.tw (C.-Y.L.); huangsj@mail.ntou.edu.tw (S.-J.H.)

[2] Meteorological Satellite Center, Central Weather Bureau, 64 Gongyuan Road, Taipei 100006, Taiwan; yangth@cwb.gov.tw (T.-Z.Y.); luu@cwb.gov.tw (Y.-C.C.)

* Correspondence: b0211@mail.ntou.edu.tw

Abstract: Satellite remote sensing sea surface temperature (SST) data are lost due to cloud cover. Missing data often cause inconvenience in subsequent applications and thus need to be reconstructed. In this study, the Data Interpolating Empirical Orthogonal Function (DINEOF) method was used to reconstruct the hourly SST data missing from the Himawari-8 satellite in the waters near Taiwan. The SST characteristics in the waters around Taiwan are quite complex, with high SST at Kuroshio in the east of Taiwan and great variation in the SST west of Taiwan due to the influence of tides. Therefore, the analysis with DINEOF was conducted using 25-h data to match the tidal cycle. The influence of SST characteristics on the accuracy of SST reconstruction is also discussed. The results show that in the western sea area where the variation of SST is large, the average root-mean-square error of SST between the original SST and the reconstructed SST is the lowest and the average coefficient of determination is the highest. The accuracy of the reconstructed SST is positively correlated with the SST variation. Furthermore, the statistical results also show that the DINEOF method can effectively reconstruct the SST regardless of the missing data rate.

Keywords: Himawari-8; sea surface temperature; missing data; DINEOF; Taiwan

check for updates

Citation: Yang, Y.-C.; Lu, C.-Y.; Huang, S.-J.; Yang, T.-Z.; Chang, Y.-C.; Ho, C.-R. On the Reconstruction of Missing Sea Surface Temperature Data from Himawari-8 in Adjacent Waters of Taiwan Using DINEOF Conducted with 25-h Data. *Remote Sens.* **2022**, *14*, 2818. https://doi.org/10.3390/rs14122818

Academic Editors: Wonkook Kim and JongKuk Choi

Received: 16 May 2022
Accepted: 10 June 2022
Published: 12 June 2022

1. Introduction

Sea surface temperature (SST) is the basis for understanding, monitoring, and forecasting heat, momentum, and gas fluxes and it also provides basic information about the global climate system. According to the sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC), the ocean temperature has increased by about 0.88 °C from the beginning of the 21st century to the present compared with the 19th century. The rate of increase in sea level from 2006 to 2018 increased to 3.7 mm per year, representing a rise in the global mean sea level with increasing ocean temperatures [1]. SST can clearly show the characteristics of water mass front, upwelling, and ocean eddies in biology-related studies [2] and also provide the initial boundary conditions for the ocean dynamic model. SST data can be obtained from on-site anchor observations, ship surveys, drifting buoys, or satellites. However, the data obtained by various methods have limitations. For example, satellite SST data can provide information for long-term and large-scale observation [3] but are limited by orbital scanning or the acquisition rate of SST data may be poor due to weather effects. Therefore, to explore the SST data situation in the adjacent waters of Taiwan and for subsequent application, in this study, we used the SST data of Himawari-8, a geosynchronous satellite that is not restricted by orbital movement, and the Data Interpolating Empirical Orthogonal Function (DINEOF) method was applied to reconstruct the values missing due to cloud cover. The influence of the characteristics of the original data on the accuracy of the filling values was then compared.

Since the thermal infrared band from satellites, used for detecting SST, is often affected by clouds or water vapor, data gaps occur frequently. To fill in the gaps in the images, the empirical orthogonal function (EOF) method has been used as a basis [4] for analyzing the characteristics of incomplete data to reconstruct the missing data, and this method is called DINEOF [5]. DINEOF was systematically developed to understand the accuracy of the reconstructed data, and it was applied to the SST data of the Advanced Very High Resolution Radiometer (AVHRR) onboard National Oceanographic and Atmospheric Administration (NOAA) satellites, with artificial addition of 40%, 60%, and 80% missing data to simulate the phenomenon of cloud cover [6]. The root-mean-square errors (RMSEs) compared with the original data were 0.89 K, 0.78 K, and 1.25 K, respectively. The results of the DINEOF method were also compared with optimal interpolation [7] and were found to be quite close, showing that DINEOF can be effectively applied to fill data gaps [8,9]. In addition to simply using AVHRR SST data, there are also multivariate DINEOF calculations that combine different satellite SST data, chlorophyll data, and wind field data [10]. The DINEOF method to fill in missing SST data has progressively been improved after several attempts. Therefore, to explore the variations in SST of adjacent waters in Taiwan, this study applied the DINEOF method to the high temporal resolution Hemawari-8 SST data to reconstruct the missing values in the data. The accuracy of the reconstructed data was then compared in different sea areas to understand how different characteristics of the SST influence the accuracy of the reconstructed data.

2. Materials and Methods

2.1. SST Data

For this study, we used the SST data of Himawari-8 in 2018 for analysis. The data range is from 117°E to 125°E longitude and 20°N to 27°N latitude, as shown in the red box in Figure 1. The spatial resolution of the SST dataset is 2 km, and the temporal resolution is one hour.

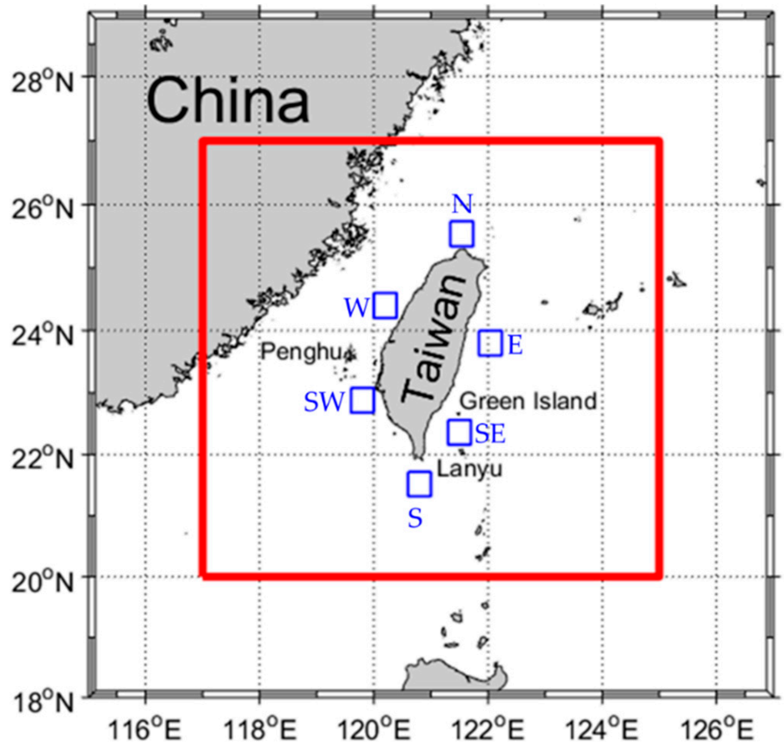

Figure 1. The inside of the red square is the study area. The blue squares are the areas selected for the discussion of DINEOF performance.

The characteristics of SST in the waters around Taiwan are varied. The northern waters of Taiwan are affected by east–west tidal currents, showing a complex background current and temperature frontal changes [11]. Due to topographic factors, the western waters of Taiwan have become the area with the largest tidal range change along the coast of central Taiwan [12]. The waters of Penghu are affected by the high-temperature seawater from south to north in the strait [13] and, thus, have obvious SST characteristics. The eastern sea area of Taiwan is the Kuroshio region with long-term high temperature, so the SST does not change significantly [14]. Due to the island effect, the Kuroshio is divided into a dual-axis phenomenon in the waters of Lanyu and Green Island in the southeastern part of Taiwan [15]. The southern waters of Taiwan are located on the north side of the South China Sea, where cold water occasionally intrudes [16]. Therefore, based on six water areas with different SST around Taiwan with a size of 0.4 × 0.4 degrees (the blue boxes in Figure 1) were selected for discussion of how different SST characteristics affect the performance of the DINEOF method. The location information of the six water areas is listed in Table 1.

Table 1. The selected sea area for the DINEOF performance.

Sea Area	Latitude (°N)	Longitude (°E)	Number of Data
East (E)	121.84–122.24	23.6–24.0	738
South (S)	120.6–121.0	21.32–21.72	1141
West (W)	120.0–120.4	24.2–24.6	1359
North (N)	121.34–121.74	25.34–25.74	1116
Southeast (SE)	121.30–121.70	22.16–22.56	1065
Southwest (SW)	119.6–120.0	22.68–23.08	1480

2.2. DINEOF Method

DINEOF is a statistical method based on empirical orthogonal function (EOF) [17], which complements incomplete data and analyzes its characteristics. It is often used to fill in missing data in the field of geophysics. For example, the satellite infrared SST image is obscured by clouds, resulting in missing values. DINEOF obtains the eigenvalues and eigenvectors through EOF analysis, determines the optimal reconstruction order of EOFs for the complete dataset or incomplete dataset through a cross-validation procedure, and infers the data of the missing value points to be reconstructed [5]. After a series of modifications and improvements [6,8], the GeoHydrodynamics and Environment Research (GHER) team was established to provide the DINEOF algorithm and codes. This study performed DINEOF based on the codes provided by GHER on its website. The principle of DINEOF is described as follows.

Before performing the DINEOF operation, the data are first arranged into a two-dimensional matrix to form a two-dimensional array **X** with the number of columns M (the number of spatial grid points) and the number of rows N (the number of time data) as

$$\mathbf{X} = \begin{bmatrix} Data_{11} & Data_{12} & \cdots & Data_{1n} \\ Data_{21} & Data_{22} & \cdots & Data_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Data_{m1} & Data_{m2} & \cdots & Data_{mn} \end{bmatrix} \quad (1)$$

where $M = 1, 2, \ldots, m$ and $N = 1, 2, \ldots, n$. To complement the value, first, the set of missing points in the data is identified as a marker. The form of the set is represented by **I**. It is assumed that (i, j) corresponds to the missing point of **X**. For elements (i, j) in the original data **X**, if $(i, j) \in \mathbf{I}$, it means that this element is missing. The number of missing data points is represented by n_0 and the values of missing data points of the original data are then set to 0. The points without missing values maintain their original values to form a new matrix $\mathbf{X_0}$, and the singular value decomposition (SVD) method is then used to obtain the eigenvalue of the matrix $\mathbf{X_0}$ [18] and the EOFs of space (**U**) and time (**V**) and the diagonal elements of matrix S are initially estimated as

$$\mathbf{X_0} = \mathbf{USV}^{\mathrm{T}} \tag{2}$$

where **S** is a diagonal matrix composed of singular values, and the matrix size is $r \times r$; **U** is the spatial decomposition of the original data matrix, and the matrix size is $m \times r$; **V** is the time principal component (PC) of the original data matrix, the matrix size is $n \times r$. m represents the size of the data point; n represents the number of data points; r is the rank of the matrix **X**, which represents the number of linearly independent columns of the matrix $\mathbf{X_0}$; and the value of r must not exceed the values of m and n.

In the process of reconstructing the data, only the most significant spatial and temporal modes are left and the optimal reconstruction result $\mathbf{X_{rec}}$ can be obtained as

$$(\mathbf{X_{rec}})_{ij} = (\mathbf{U_k S_k V_k^{\mathrm{T}}})_{ij} = (\mathbf{X_0^e})_{ij} = \sum_{p=1}^{k} \rho_p (u_p)_i (v_p^{\mathrm{T}})_j \tag{3}$$

where k is the optimal number of the remaining k-order modes, and $\mathbf{X_0^e}$ represents the reconstruction matrix corresponding to the first k-order EOFs. After obtaining the reconstructed data of the missing data point, $(\mathbf{X_{rec}})_{ij}$, it is added to the original data matrix $\mathbf{X_0}$ as

$$\mathbf{X_{rec}} = \mathbf{X_0} + (\mathbf{X_{rec}})_{ij} \tag{4}$$

$(\mathbf{X_{rec}})_{ij}$ means that all data points are 0 except for the reconstructed data at the missing point (i, j). Since the space–time EOFs and singular values are adjusted after each reconstruction, after the first reconstruction is completed, the same procedure as in the following two equations is repeated until convergence is reached:

$$\mathbf{X_{rec}} = \mathbf{USV}^{\mathrm{T}} \tag{5}$$

$$(\mathbf{X_{rec}})_{ij} = (\mathbf{U_k S_k V_k^{\mathrm{T}}})_{ij} = (\mathbf{X_{rec}^e})_{ij} = \sum_{p=1}^{k} \rho_p (u_p)_i (v_p^{\mathrm{T}})_j \tag{6}$$

where $\mathbf{U_k}$ is an $m \times k$ matrix covering the first k spatial EOFs, $\mathbf{V_k}$ is a $k \times n$ matrix covering the first k temporal EOFs, and $\mathbf{S_k}$ is a $k \times k$ diagonal matrix covering the first k singular values. During each iteration, when a new reconstruction matrix is generated, the optimal number of modes is also replaced [19]; that is, the size of k changes with the singular value diagonal matrix and ρ_p is the p-th singular value of **X** (p = 1, 2, ... , k); u_p represents the spatial EOFs of the p-th mode; v_p represents the temporal EOFs of the p-th mode. The criterion for convergence is that the ratio between the RMSE of missing data and the standard deviation (σ) of existing data without missing data reaches 10^{-3} [20], which is called the accuracy of reconstruction data.

DINEOF is a theory derived based on EOF theory, and the EOF analysis method is based on finding the eigenvectors of continuous data and further represents the principal component content of the data period. Therefore, if the selected data period spans different time scales, the physical phenomenon affects the analyzed principal component results. In terms of SST near Taiwan, the western and northern waters of Taiwan are mainly affected by tides. Therefore, 25-h continuous data was selected to match the tidal cycle as the base data for the DINEOF method to obtain the complete SST.

To verify the performance of the DINEOF method in reconstructing data from different sea areas, this study removed 21 × 21 grid points at specific locations from the 25th original SST (SST_O) data to simulate missing values caused by cloud cover. The selection of images must satisfy the following criteria: (1) the 21 × 21 grid points simulating missing values for cloud cover must have no missing values at hour 25, and (2) hourly data of the previous 24 h cannot be completely missing, that is, they must not have a 100% missing rate. The above criteria were applied to all Himawari images in the study areas in 2018, and images satisfying these two criteria were selected and then reconstructed by DINEOF. All comparisons were for the 25th hour original and reconstructed data of each. The areas selected by the simulated cloud cover were the blue squares shown in Figure 1, on which the DINEOF method was carried out to reconstruct the missing data.

3. Results

3.1. Preliminary Analysis of SST Data

Figure 2 is the time series of the missing data rate of Himawari-8 SST data. The blue line is the original data, and the red line is the 30-day moving average. It can be seen that the missing data rate is higher in winter, which is the northeast monsoon period. The missing data rate is also higher in June and August. According to the 2018 annual report of climate monitoring [21], the rainfall during the 2018 Meiyu (plum rain) season was mainly concentrated in June, while from mid-August 2018, Taiwan was located in a low-pressure belt and rain over an extended period. Both rainy periods were responsible for the high SST missing data rate in June and August.

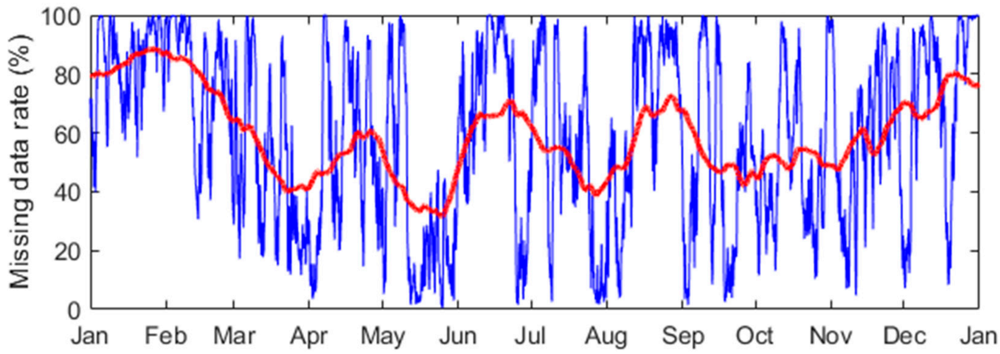

Figure 2. The daily missing data rate in 2018. The red line is the 30-day moving average.

Figure 3a shows the statistics of the spatial distribution of the missing data rate. It can be seen that the northeast and eastern sea areas of Taiwan have a higher missing data rate. This is probably due to the influence of the northeast monsoon and the Kuroshio. In winter, the northeast monsoon brings cold air masses to northern Taiwan, passing the region of the warm Kuroshio. The increased air–sea temperature, moisture differences, and intensified wind speed result in increased evaporation, which eventually condensed into clouds and resulted in precipitation in the area near the Kuroshio [22,23]. The missing data rate is distributed in the interval 45–75% (Figure 3b).

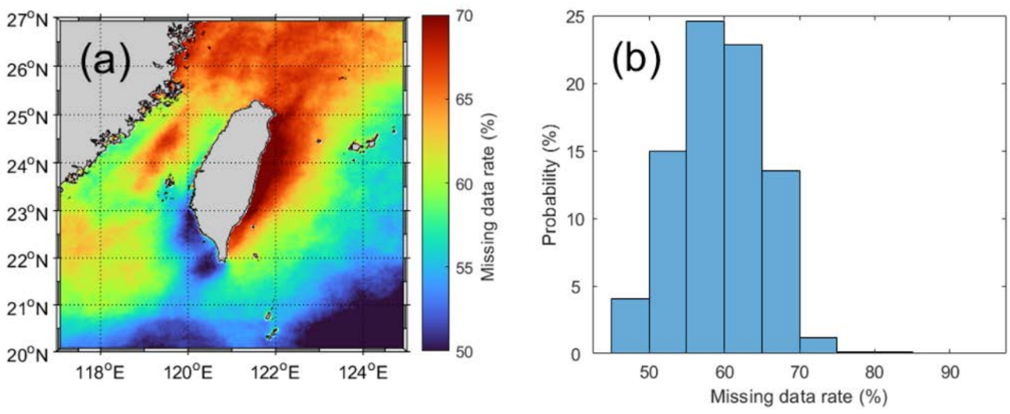

Figure 3. (**a**) Spatial distribution of the missing data rate. (**b**) Histogram of overall missing data rate.

According to the SST characteristics of different sea areas around Taiwan, six areas were selected as artificially excluded values, and the DINEOF method was then used to fill in the missing values. Figure 4 shows the mean and standard deviation of SST throughout the year in 2018. The variability of SST in the eastern and southern waters of Taiwan where the Kuroshio passes is relatively small, while it is more obvious in the northern and western waters of Taiwan. Figure 5 shows the boxplot of the standard deviation of SST_O data in the six areas. It shows relative stability in the eastern, southern, and southeastern sea areas, with larger SST variations in the northern, western, and southwestern sea areas.

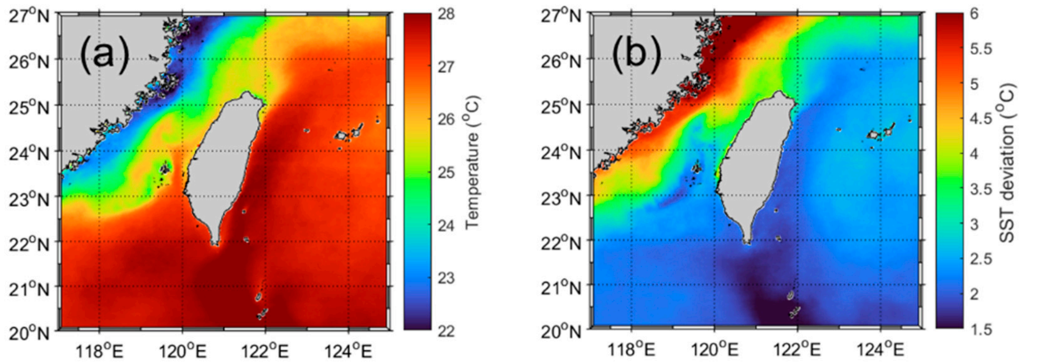

Figure 4. SST data in 2018 (unit: °C): (**a**) mean; (**b**) deviation.

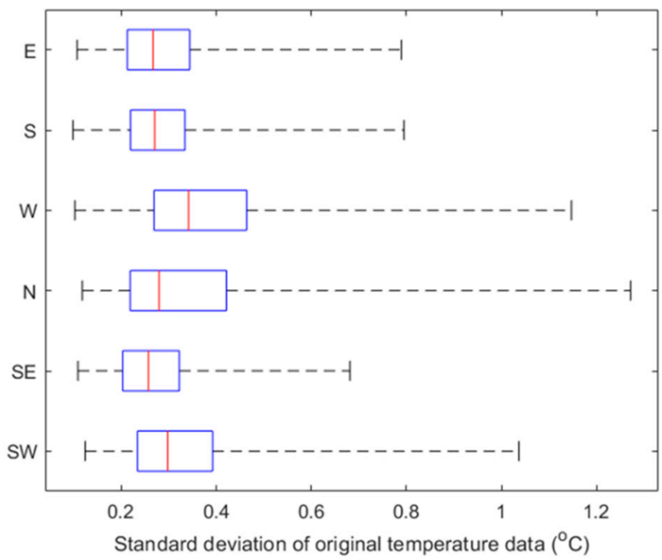

Figure 5. Boxplot of the daily SST standard deviation in 2018 in the study areas. The dashed line represents the standard deviation range. The blue box represents the range between the first and third quartiles and the red line in the box is the median.

3.2. DINEOF Results

Figure 6 shows the most accurate result for DINEOF in reconstructing SST in the six areas. It can be seen that the SST characteristics were well reproduced. The overall SST differences between the SST_O and the reconstructed SST (SST_R) of the six areas were less than ±1.0 °C—less than ±0.5 °C in most cases—and the RMSEs were between 0.13 and 0.29 °C, and the coefficients of determination (r^2) were all above 0.84. The western sea area has the lowest RMSE and the highest r^2, the southern sea area has the highest RMSE, and the southeastern sea area has the lowest r^2. It can also be seen that the areas with less accurate SST_R are mainly in the eastern and southern sea areas, while those with SST_R closer to the actual values are in the western and northern sea areas. The SST_R results were affected by the intensity of the SST gradient. A more accurate SST_R can be obtained with the SST_O if the SST characteristic variations are significant. This may be due to the large variation of SST characteristics in the northern and western waters around Taiwan, which is caused by the influence of tides and results in larger eigenvalues.

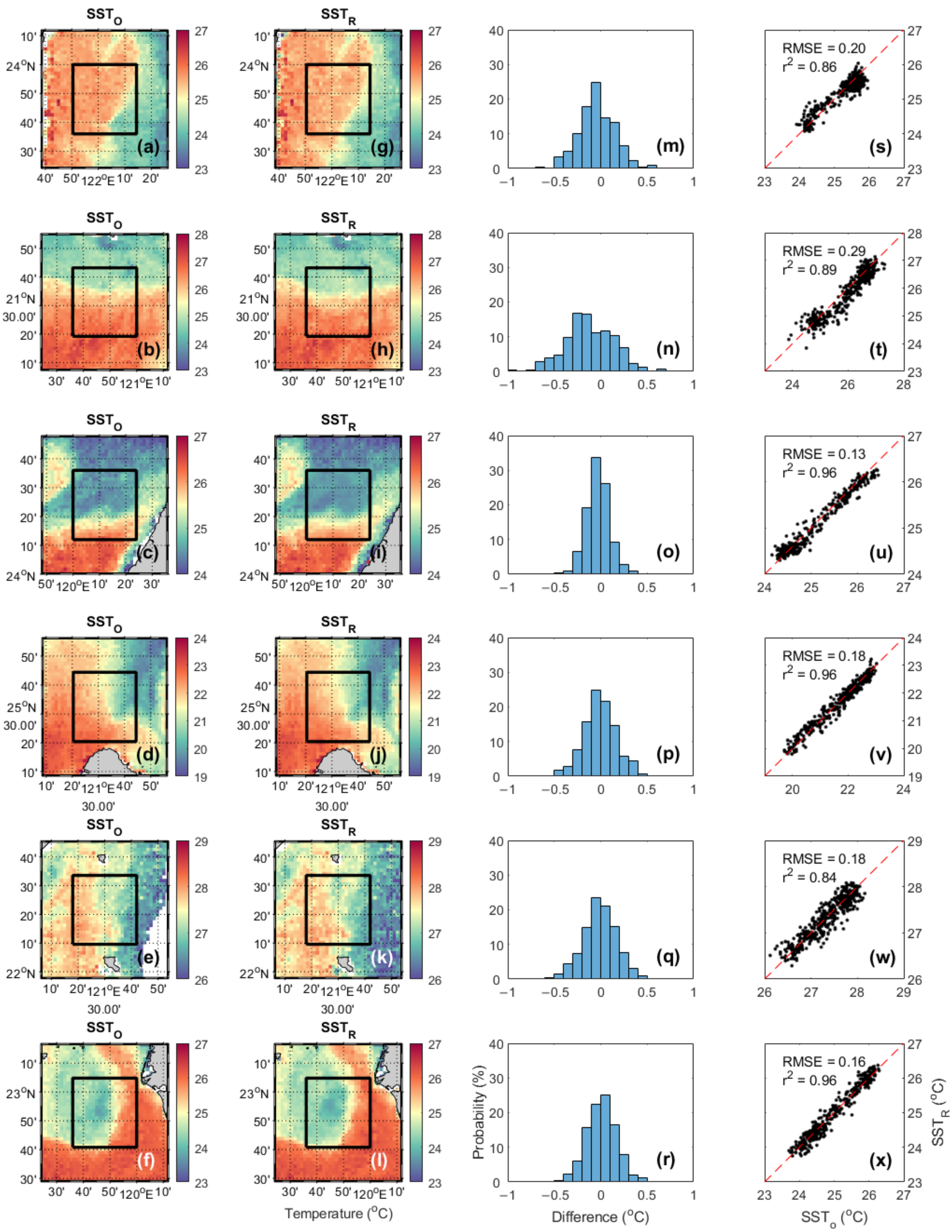

Figure 6. The best results of DINEOF SST$_R$ in the six study areas and the areas from top to bottom are east, south, west, north, southeast, and southwest. (**a**–**f**) SST$_O$ images. (**g**–**l**) SST$_R$ images using the DINEOF method. (**m**–**r**) Histograms of the difference between the SST$_O$ data and the SST$_R$ data. (**s**–**x**) Scatter plots of the SST$_O$ data and the SST$_R$ data.

4. Discussion

4.1. Comparison of SST_R with Other SST Data

The accuracy of SST_R has been assessed with SST_O. However, when using DINEOF to reconstruct missing values in geophysical data, missing values should be compared with other data, such as higher resolution data, to check if they are correct. Thus, SST_R data were further assessed with in situ SST Quality Monitor (iQuam) SST measurements and the moderate resolution imaging spectroradiometer (MODIS) SST data onboard the Terra and Aqua satellites.

4.1.1. iQuam SST

The iQuam SST dataset is an in situ calibrated monitoring system for SST developed by NOAA. The main purpose is to ensure the quality of the in situ SST provided by different countries, agencies, and platforms [24]. The iQuam SST dataset has a quality level indicator. The suggested use quality level for high-accuracy applications is level 5 and for general applications is level 4. This study used only quality level 5 data to compare with SST_O and SST_R.

Figure 7 shows the time series comparison of iQuam SST (SST_Q) with SST_O and SST_R in the six areas. The selection of the SST_Q is based on the data of level 5 in the study area and then compared with SST_O and SST_R of the closest time and the nearest pixel of the SST_Q. The most available iQuam data for comparison is in the south area with 14 points, while the least is in the west area with four points. The statistical result of the comparison between SST_Q and SST_O/SST_R is listed in Table 2. The RMSE between SST_Q and SST_R was from 0.44 to 1.29 °C. It seems that the RMSE is large, but the r^2 values were all above 0.77, showing a good correlation. Overall, the statistical results of SST_R and SST_O with SST_Q were similar, although some regional SST_R results had lower RMSE and higher r^2 compared to SST_O. However, the statistical confidence may not be sufficient due to fewer matching points in some areas. Comparing the data of the six areas, the results are shown in Figure 8. The r^2 was both 0.86 and the RMSE was between 0.83 and 0.85 °C, indicating that the error in the results of SST_R was acceptable.

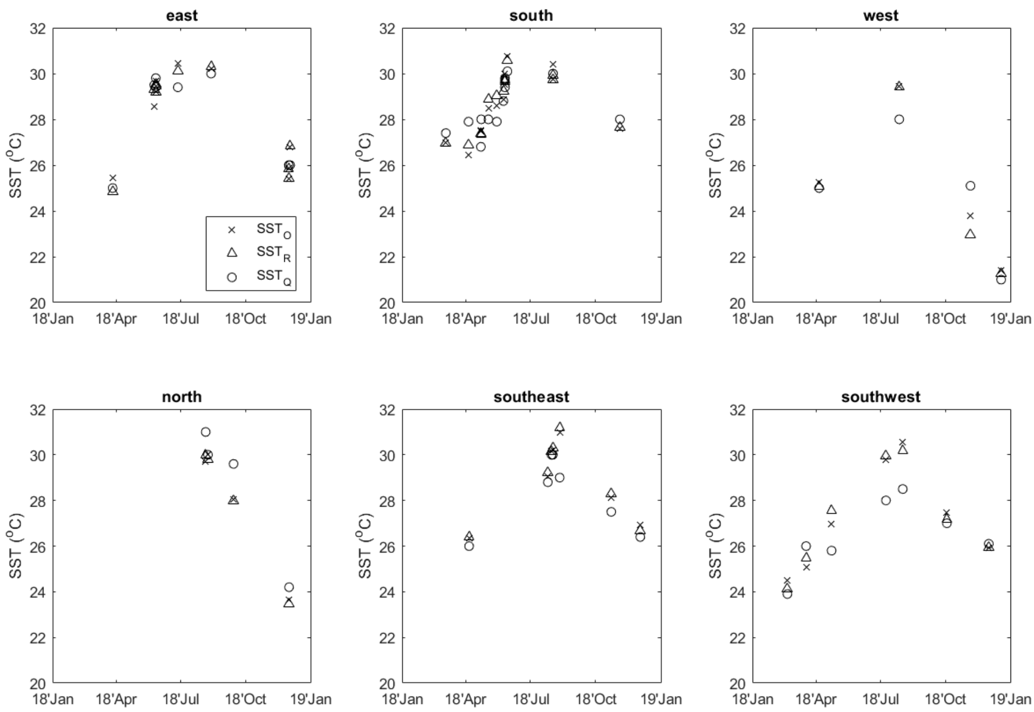

Figure 7. Time series comparison of SST_Q with SST_O and SST_R in the six areas.

Table 2. Statistic results of the comparison between SST_Q and SST_R/SST_O in the six areas.

Sea Area	Number of Matching Points	SST_Q with SST_R			SST_Q with SST_O		
		RMSE (°C)	r^2	*p*-Value	RMSE (°C)	r^2	*p*-Value
East (E)	10	0.44	0.95	<0.001	0.58	0.91	<0.001
South (S)	14	0.58	0.77	<0.001	0.60	0.80	<0.001
West (W)	4	1.29	0.83	0.0895	1.01	0.89	0.0546
North (N)	4	1.02	0.96	0.0187	1.05	0.94	0.0279
Southeast (SE)	7	0.91	0.86	0.0029	0.83	0.87	0.0022
Southwest (SW)	7	1.19	0.85	0.0029	1.20	0.85	0.0030

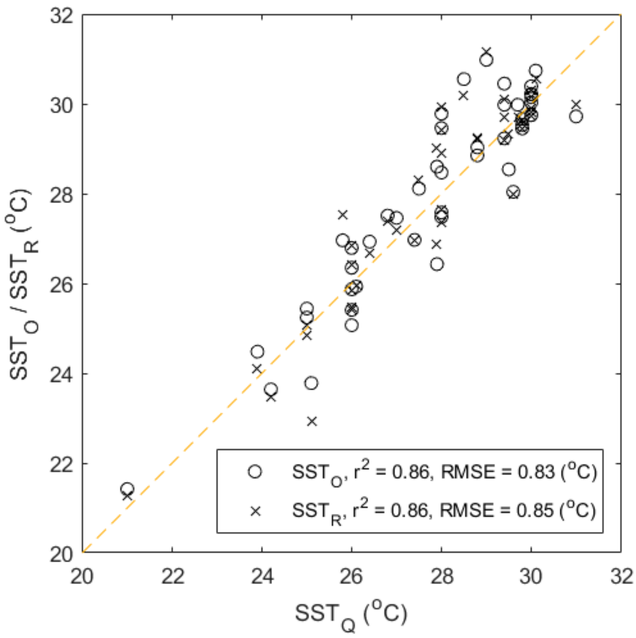

Figure 8. Scatter plot of SST_O and SST_R with SST_Q for all matching points in six areas.

4.1.2. MODIS SST

The DINEOF reconstruction results were compared with MODIS SST (SST_M) data in addition to iQuam SST. Because MODIS is onboard both Terra and Aqua satellites, the SST_R was compared with Terra MODIS SST and Aqua MODIS SST. The MODIS SST used in this study was L3 daytime and nighttime data. They are available from the MODIS website. Since the grid size of MODIS data is 4 km × 4 km, the Himawari data were also processed to 4 km × 4 km for comparison. The comparison results between SST_R and SST_M and between SST_O and SST_M are listed in Table 3. The RMSE and r^2 of the reconstructed data were between 0.58 and 0.84 °C and 0.67 and 0.97, respectively. However, in terms of RMSE, the nighttime data are better than the daytime data, but in terms of r^2, the opposite is true. Overall, the comparisons were about the same whether Aqua or Terra or daytime or nighttime data. It is simply that the reconstructed data were slightly worse than the original data.

4.2. Influence of SST Variations on SST_R

As mentioned in the previous section, the DINEOF method is based on EOF theory. If the SST has obvious characteristics in the time or space domain, that is, if the SST varies greatly, then the DINEOF method can be used to reconstruct missing data to obtain more accurate results. To further examine this conclusion, the correlation between the r^2 of SST_O and SST_R and the SST deviation of SST_O was plotted in Figure 9. Figure 9a shows the relationship between r^2 and the SST_O deviation for the data of the six areas. The r^2 appears to increase with SST deviation. To present it more clearly, we averaged the SST deviation per 0.1 interval of r^2 (Figure 9b). The result clearly shows that r^2 increases with increasing SST deviation, that is, the larger the SST_O variation, the more accurate the SST_R. The quality of the reconstructed data caused by the differences in the SST characteristics of the six areas can also be given in a scatter plot (Figure 9c). It shows the relationship between the r^2 of SST_O and SST_R and SST_O deviation in the 0.4 × 0.4 degree box of the six areas. The highest SST_O deviation with the highest r^2 is found in the western sea area. The eastern, southeastern, and southern sea areas have lower SST_O deviation and thus lower r^2 of SST_O and SST_R. These results confirm that the higher variation in SST_O results in a more accurate SST_R using the DINEOF method.

Table 3. Statistical results between SST_M and SST_R/SST_O.

Sea Area	Terra Daytime SST_M				Terra Nighttime SST_M				Aqua Daytime SST_M				Aqua Nighttime SST_M			
	SST_R		SST_O		SST_R		SST_O		SST_R		SST_O		SST_R		SST_O	
	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2
E	0.79	0.88	0.73	0.91	0.63	0.92	0.69	0.91	0.78	0.85	0.78	0.86	0.59	0.91	0.55	0.92
S	0.67	0.83	0.63	0.86	0.65	0.74	0.63	0.76	0.68	0.82	0.68	0.82	0.84	0.67	0.85	0.69
W	0.69	0.95	0.68	0.95	0.58	0.94	0.55	0.95	0.67	0.93	0.67	0.93	0.67	0.94	0.64	0.95
N	0.59	0.97	0.60	0.97	0.67	0.93	0.67	0.93	0.78	0.95	0.75	0.96	0.65	0.94	0.62	0.94
SE	0.76	0.86	0.74	0.86	0.72	0.81	0.73	0.81	0.67	0.86	0.66	0.86	0.69	0.80	0.69	0.80
SW	0.79	0.87	0.71	0.90	0.76	0.82	0.73	0.84	0.67	0.92	0.65	0.92	0.74	0.85	0.73	0.86

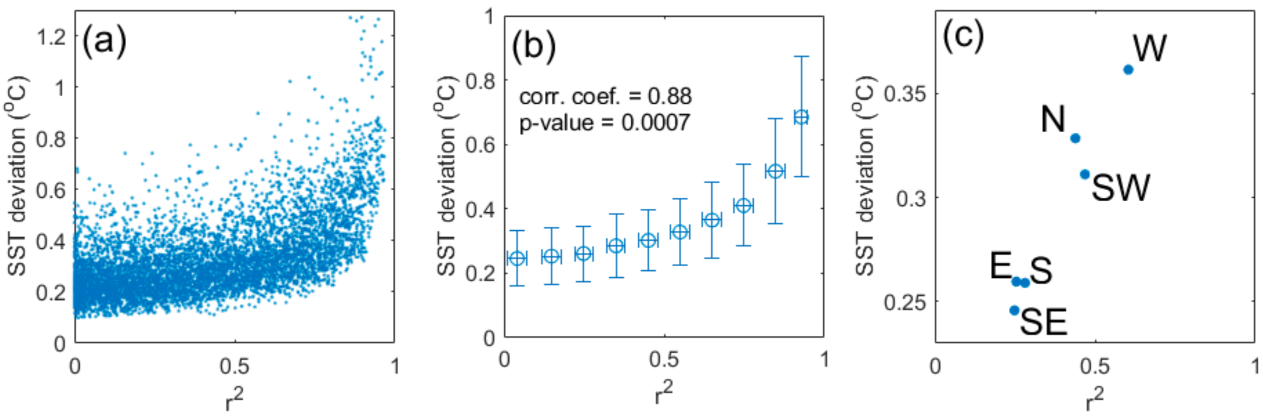

Figure 9. (**a**) Scatter plot of r^2 and SST_O deviation. (**b**) Scatter plot of r^2 and averaged SST_O deviation per 0.1 interval of r^2, bars are error bars of one standard deviation. (**c**) Scatter plot of r^2 and SST_O deviation for the six areas.

4.3. Influence of Missing Data Rate on SST_R

To examine whether the missing data rate of SST_O affects the accuracy of SST_R, the same procedure was performed for processing SST_O variation. Figure 10a shows the relationship between r^2 and the missing data rate of SST_O for the data of the six areas. It seems that there is no relationship between the two. This is consistent with the results of a previous study [6], where the authors concluded that the reconstructed result was robust even when there was a very high amount of missing data. However, it can be seen from Figure 10 of that study [6] that there is a higher root mean square (RMS) for around 80% missing data. It seems that the higher the missing data rate, the higher the RMS. To further understand whether a similar phenomenon occurs in our research data, we averaged the missing data rate per 0.1 interval of r^2, as shown in Figure 10b. The r^2 decreases with increasing missing data rate; that is, the larger the missing data in SST_O, the more inaccurate the SST_R result. This result also seems to reflect the results of the previous study [6]. However, it should be noted that although the change trend of the mean value is significant ($p < 0.01$), the standard deviation is quite large. To better understand the spatial difference between r^2 and the missing data rate, a scatter plot was plotted, as shown in Figure 10c. The missing data rates in the six areas were almost the same, except for the eastern sea area, which has a higher missing data rate. The r^2 varies from 0.2 to 0.6 in the five areas with nearly the same missing data rate. Therefore, the change in r^2 is unrelated to the missing data rate. A closer look reveals that the change in r^2 is related to the variation in SST_O.

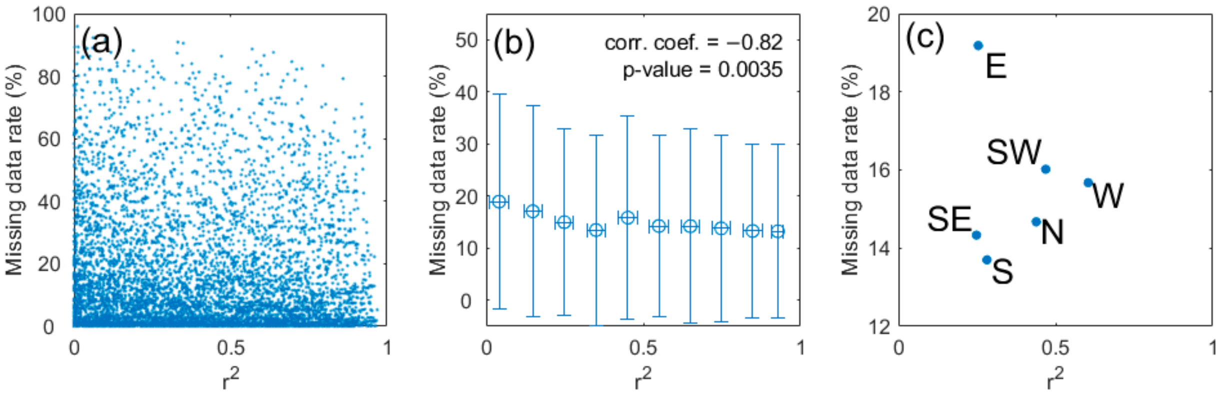

Figure 10. (**a**) Scatter plot of r^2 and SST_O missing data rate. (**b**) Scatter plot of r^2 and averaged SST_O missing data rate per 0.1 interval of r^2, bars are error bars of one standard deviation. (**c**) Scatter plot of r^2 and SST_O missing data rate for the six areas.

5. Conclusions

The main purpose of this study was to use the high temporal resolution Himawari-8 SST data to explore the phenomenon of missing data in the adjacent waters of Taiwan and its reconstruction using the DINEOF method. Then, the influence of the SST characteristics of different sea areas on the reconstruction of the SST using the DINEOF method was analyzed through comparisons. The conclusions of the analysis are as follows:

1. The DINEOF method is affected by the magnitude of SST variation in the study region. If there are more obvious SST characteristics, that is, higher variation in the region, the results of the reconstruction are better and the RMSE is smaller.
2. The missing rate of the original data does not substantially affect the accuracy of the reconstructed data. However, from a statistical point of view, the higher the missing data rate of the original data, the less accurate the reconstructed data may be.

Author Contributions: Conceptualization, C.-R.H., Y.-C.C. and T.-Z.Y.; methodology, Y.-C.Y.; software, Y.-C.Y. and C.-Y.L.; validation, S.-J.H. and C.-R.H.; formal analysis, Y.-C.Y. and C.-Y.L.; investigation, Y.-C.Y., C.-Y.L. and C.-R.H.; resources, Y.-C.C. and T.-Z.Y.; data curation, Y.-C.Y. and C.-Y.L.; writing—original draft preparation, Y.-C.Y.; writing—review and editing, C.-R.H. and S.-J.H.; visualization, Y.-C.Y.; supervision, Y.-C.C.; project administration, C.-R.H.; funding acquisition, C.-R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Central Weather Bureau of Taiwan, grant number 1102084A, and the APC was funded by the Ministry of Science and Technology of Taiwan, grant number MOST110-2611-M-019-001.

Data Availability Statement: The research data of SST (produced from Himawari-8) that was used in this paper was supplied by the P-Tree System, Japan Aerospace Exploration Agency (JAXA). The data were obtained from https://www.eorc.jaxa.jp/ptree/terms.html, accessed on 20 March 2020. The iQuam SST data were downloaded from https://www.star.nesdis.noaa.gov/socd/sst/iquam/data.html, accessed on 12 May 2022, and the MODIS SST data were downloaded from https://oceandata.sci.gsfc.nasa.gov/directaccess/MODIS-Aqua/Mapped/Daily/4km/sst/2018/, https://doi.org/10.5067/MODST-1D4D9, accessed on 12 May 2022.

Acknowledgments: The authors thank JAXA for providing Himawari-8 SST data, NOAA for the iQuam data, and NASA for the MODIS data. Many thanks also to GeoHydrodynamics and Environment Research for providing the DINEOF source code.

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

References

1. IPCC. *Climate Change 2021: The Physical Science Basis*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press: Cambridge, UK, 2021.
2. Martin, S. *An Introduction to Ocean Remote Sensing*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2014.
3. Cochran, J.K.; Bokuniewicz, H.J.; Yager, P.L. *Encyclopedia of Ocean Sciences*, 3rd ed.; Elsevier Science: Amsterdam, The Netherlands, 2019.
4. Wallace, J.M.; Dickison, R.E. Empirical orthogonal representation of time series in the frequency domain. Part I: Theoretical considerations. *J. Appl. Meteorol. Climatol.* **1972**, *11*, 887–892. [CrossRef]
5. Beckers, J.M.; Rixen, M. EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Ocean. Technol.* **2003**, *20*, 1839–1856. [CrossRef]
6. Alvera-Azcárate, A.; Barth, A.; Rixen, M.; Beckers, J.M. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. *Ocean Model.* **2005**, *9*, 325–346. [CrossRef]
7. Daley, R. *Atmospheric Data Analysis*; Cambridge University Press: Cambridge, UK, 1991.
8. Beckers, J.M.; Barth, A.; Alvera-Azcárate, A. DINEOF reconstruction of clouded images including error maps-application to the Sea-Surface Temperature around Corsican Island. *Ocean Sci.* **2006**, *2*, 183–199. [CrossRef]
9. Elken, J.; Zujev, M.; She, J.; Lagemaa, P. Reconstruction of large-scale sea surface temperature and salinity fields using sub-regional EOF patterns from models. *Front. Earth Sci.* **2019**, *7*, 232. [CrossRef]
10. 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.* **2007**, *112*, C03008.
11. Hsu, P.C.; Lu, C.Y.; Hsu, T.W.; Ho, C.R. Diurnal to seasonal variations in ocean chlorophyll and ocean currents in the north of Taiwan observed by Geostationary Ocean Color Imager and coastal radar. *Remote Sens.* **2020**, *12*, 2853. [CrossRef]
12. Jan, S.; Chern, C.S.; Wang, J.; Chao, S.Y. The anomalous amplification of M2 tide in the Taiwan Strait. *Geophys. Res. Lett.* **2004**, *31*, L07308. [CrossRef]
13. Lan, K.W.; Kawamura, H.; Lee, M.A.; Chang, Y.; Chan, J.W.; Liao, C.H. Summertime sea surface temperature fronts associated with upwelling around the Taiwan Bank. *Cont. Shelf Res.* **2009**, *29*, 903–910. [CrossRef]
14. Chang, M.H.; Jan, S.; Mensah, V.; Andres, M.; Rainville, L.; Yang, Y.J.; Cheng, Y.H. Zonal migration and transport variations of the Kuroshio east of Taiwan induced by eddy impingements. *Deep-Sea Res. Part I Oceanogr. Res. Pap.* **2018**, *131*, 1–15. [CrossRef]
15. Hsu, P.C.; Ho, C.Y.; Lee, H.J.; Lu, C.Y.; Ho, C.R. Temporal variation and spatial structure of the Kuroshio-induced submesoscale island vortices observed from GCOM-C and Himawari-8 data. *Remote Sens.* **2019**, *12*, 883. [CrossRef]
16. Zheng, Q.; Hu, J.Y.; Ho, C.R.; Xie, L.L. Advances in research of regional oceanography of the South China Sea: Overview. In *Regional Oceanography of the South China Sea*; Hu, J.Y., Ho, C.R., Xie, L.L., Zheng, Q., Eds.; World Scientific: Singapore, 2019; pp. 1–18.
17. Weare, B.C.; Nasstrom1, J.S. Examples of extended empirical orthogonal function analyses. *Mon. Weather Rev.* **1982**, *110*, 481–485. [CrossRef]
18. Thomson, R.E.; Emery, W.J. *Data Analysis Methods in Physical Oceanography*, 3rd ed.; Elsevier Science: Amsterdam, The Netherlands, 2014.
19. Ping, B.; Su, F.; Meng, Y. An improved DINEOF algorithm for filling missing values in spatio-temporal sea surface temperature data. *PLoS ONE* **2016**, *11*, 0155928. [CrossRef]
20. Huynh, H.N.T.; Alvera-Azcárate, A.; Barth, A.; Beckers, J.M. Reconstruction and analysis of long-term satellite-derived sea surface temperature for the South China Sea. *J. Oceanogr.* **2016**, *72*, 707–726. [CrossRef]
21. Central Weather Bureau. Climate Monitoring: 2018 Annual Report. (In Chinese). Available online: https://www.cwb.gov.tw/Data/service/notice/download/publish_20191104141027.pdf (accessed on 27 April 2022).
22. Grodsky, S.A.; Bentamy, A.; Carton, J.A.; Pinker, R.T. Intraseasonal latent heat flux based on satellite observations. *J. Clim.* **2009**, *22*, 4539–4556. [CrossRef]
23. Rex, X.; Yang, X.-Q.; Hu, H. Subseasonal variations of wintertime North Pacific evaporation, cold air surges, and water vapor transport. *J. Clim.* **2017**, *30*, 9475–9491.
24. Xu, F.; Ignatov, A. In situ SST quality monitor (iQuam). *J. Atmos. Ocean. Technol.* **2014**, *31*, 164–180. [CrossRef]