remote sensing

MDPI

Reconstruction of Daily Sea Surface Temperature Based on Radial Basis Function Networks

Zhihong Liao [1,2,3], Qing Dong [1,*], Cunjin Xue [1] , Jingwu Bi [1,2] and Guangtong Wan [1]

[1] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; liaozh@radi.ac.cn (Z.L.); xuecj@radi.ac.cn (C.X.); jingwubi@163.com (J.B.); wanguangtong@126.com (G.W.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] National Meteorological Information Center, China Meteorological Administration, Beijing 100081, China
* Correspondence: qdong@radi.ac.cn; Tel.: +86-10-8217-8121

Abstract: A radial basis function network (RBFN) method is proposed to reconstruct daily Sea surface temperatures (SSTs) with limited SST samples. For the purpose of evaluating the SSTs using this method, non-biased SST samples in the Pacific Ocean (10°N–30°N, 115°E–135°E) are selected when the tropical storm Hagibis arrived in June 2014, and these SST samples are obtained from the Reynolds optimum interpolation (OI) v2 daily 0.25° SST (OISST) products according to the distribution of AVHRR L2p SST and in-situ SST data. Furthermore, an improved nearest neighbor cluster (INNC) algorithm is designed to search for the optimal hidden knots for RBFNs from both the SST samples and the background fields. Then, the reconstructed SSTs from the RBFN method are compared with the results from the OI method. The statistical results show that the RBFN method has a better performance of reconstructing SST than the OI method in the study, and that the average RMSE is 0.48 °C for the RBFN method, which is quite smaller than the value of 0.69 °C for the OI method. Additionally, the RBFN methods with different basis functions and clustering algorithms are tested, and we discover that the INNC algorithm with multi-quadric function is quite suitable for the RBFN method to reconstruct SSTs when the SST samples are sparsely distributed.

Keywords: sea surface temperature (SST); radial basis function network (RBFN); improved nearest neighbor cluster (INNC) algorithm

1. Introduction

High-quality sea surface temperature (SST) data play an important role in many applications, including climate predictions [1], ocean data assimilation [2], and global change research [3]. However, raw SST data can vary significantly between different types of measurements [4], such as in-situ measurements (e.g., ships or buoys) and satellite sensors (e.g., infrared thermal radiometers or microwave radiometers on satellite platforms). In-situ SST measurements are accurate but sparse, while satellite-retrieved SST measurements have poor accuracy but provide a dense coverage globally. Therefore, one method to solve these problems is to combine different sources of SSTs to reconstruct new SST fields [4,5].

Even through many different sources of SSTs are combined, due to the limitations of satellite orbits and the field of view (FOV) of sensors, global ocean coverage remains difficult to achieve [6]. Additionally, high-quality SST samples usually have irregular distributions as some data are discarded, such as in-situ SST data from questionable drifters and satellite-retrieved SST data contaminated by clouds [6,7]. Especially for the SST data from infrared radiometers, the SSTs in some regions are largely missing since they can not get the information through clouds. Thus, it is critical to reconstruct SST fields from incomplete SST samples.

Several alternative methods, such as empirical orthogonal function (EOF) [8,9], data interpolating orthogonal function (DINEOF) [10,11], and optimum interpolation (OI) [7], have previously been used for SST reconstruction. Currently, the most popular method to reconstruct SST is with the OI algorithm, and many high-quality SST analysis products using the OI method have been designed with different kinds of temporal and spatial resolutions for various applications [5,6,12–17]. However, when the distribution of SST samples are very sparse and the local correlation scales are not large enough in the OI method, the SSTs are often replaced by the background fields and the SST accuracy in these regions is relatively poor [6].

A SST field can generally be described as a nonlinear pattern in space and time, which can be expressed as a set of complex functions. Radial basis function network (RBFN) is an artificial neural network that uses radial basis functions (RBFs) as activation functions in the field of mathematical modeling. It is a good function approximator when RBFNs have enough hidden neurons and samples [18–20]. RBFNs have been employed in many applications [21–23]. A RBFN-based response model for SST has recently been developed by Ryu et al. [24]. However, this model does not optimize the quantity and distribution of hidden knots, which are sensitive to the accuracy of the model and should be adjusted with different distributions of SST samples.

Thus, we introduce a RBFN method for reconstructing daily SST fields when the SST samples are limited and sparsely distributed, and design an improved nearest neighbor cluster (INNC) algorithm for optimizing the quantity and distribution of hidden knots for RBFNs in this study. The remainder of this paper is organized as follows. Section 2 provides an outline of the SST samples that are used in this study. Then, the RBFN method and the INNC algorithm are described in Section 3. Section 4 evaluates the performance of the RBFN method by using different basis functions and clustering methods for RBFNs, and analyze its SST accuracy with the OI method. Then, some characteristics of this method are discussed in Section 5. Finally, the conclusions are summarized in Section 6.

2. Data Description

Since the satellite-retrieved SST data from infrared radiometers are often contaminated by clouds, the SST samples that we can obtain is very limited, especially when typhoon or tropical storm comes with heavy clouds. This study considers data from the Pacific Ocean (10°N–30°N, 115°E–135°E) (see Figure 1a). The SST samples in the study are selected from both the in-situ SST and the AVHRR SST L2p data, but the coverage of SST samples is still very low when the tropical storm Hagibis arrived from 12 June to 18 June 2014 (see Figure 1b). The AVHRR SST L2p data have been provided by the group for high resolution sea surface temperature (GHRSST) (ftp://data.nodc.noaa.gov/pub/data.nodc/ghrsst/) and have been widely used in SST studies [4]. The in-situ SSTs are from the in-situ SST quality monitor (iQuam) system by using the highest quality flag value of 5 that is designed for high-accuracy applications (http://www.star.nesdis.noaa.gov/sod/sst/iquam/index.html) [25]. These SST samples combining the AVHRR L2p data and in-situ SST data have been averaged onto a 0.25° grid. To evaluate the SST errors only from RBFNs, the OISST product was used as a reference, and non-biased SST samples were selected from the OISST products according to the distributions of the SST samples. The OISST products are the level-4 data in GHRSST with a grid resolution of 0.25°, where in-situ data have been combined with AVHRR data through the analyses system of National Oceanic and Atmospheric Administration (NOAA)'s national climatic data center (NCDC) [6]. Thus, the distributions of SST samples are consistent with the AVHRR L2p data and in-situ SST data, while the values of SST samples were assigned by the OISST products. Accordingly, the global SST samples from 13 June 2014 are also used to test the performance of RBFN method on the global SST model. In addition, the reconstructed SSTs from this RBFN method are validated with the reference SSTs in the study.
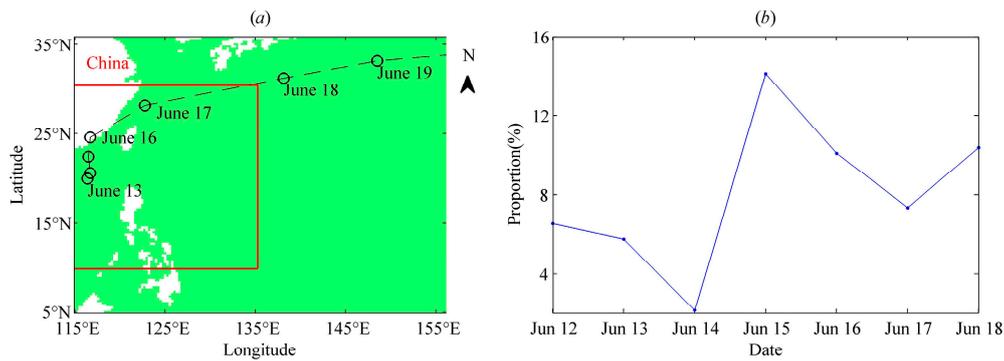
Figure 1. (**a**) The study area (10°N–30°N, 115°E–135°E) in red rectangle region and the track of the tropical storm Hagibis labeled by day. The white and green areas in (**a**) show land and ocean, respectively; (**b**) The coverage of sea surface temperature (SST) samples in the study region during the period from 12 June to 18 June 2014.

3. Methodology

3.1. RBFN Method

The daily SST can be considered as two parts: the background field and the SST increment [6,26]. The previous SST field is used as the background field. In particular, the values of climatology change are added to the background field. The reconstructed SST T_t can be defined as

$$\begin{cases} T_t = Y_t + T_t^b \\ T_t^b = T_{t-1} + (T_t^c - T_{t-1}^c) \end{cases} \tag{1}$$

where Y_t is the SST increment, T_t^b is the background field, and T_{t-1} is the reconstructed SST for the previous time. For convenience, we use the OISST product as the a priori SST, and to minimize the effects of this, the procedure was run for 10 days in advance of the first date. The daily SST climatology, T_t^c, was obtained from 30 years (1982–2011) of OISST data [27]. $(T_t^c - T_{t-1}^c)$ is the SST climatology that changes between time t and $t - 1$. The SST increment Y_t is calculated from the RBFN proposed by Ryu et al. [24]. It should be noted that this RBFN is used to estimate SST increments, not SSTs as in Ryu's et al. study. The RBFN can be given by

$$\begin{cases} Y_t = f(X_t) + \varepsilon_t \\ f(X_t) = \beta_{t,0} + X_t^d \beta_{t,m}^T + \sum_{k=1}^{K_t} \phi(z_{t,k}) \beta_{t,k+5} \end{cases} \tag{2}$$

where $f(X_t)$ is a RBFN function, $X_t = (X_{t1}, X_{t2})$ is a 2-dimensional matrix, and X_{t1}, X_{t2} are the longitude and latitude at time t, respectively. X_t^d denotes the set of all possible power with d degree. Here, $d = 2$ and the 2-degree polynomial can be expressed as $X_t^d = (1, X_{t1}, X_{t2}, X_{t1}X_{t2}, X_{t1}^2, X_{t2}^2)$. $z_{t,k} = \|X_t - \mu_{t,k}\|$ denotes the distance between X_t and hidden knots $u_{t,k}$ for $k = 1, ..., K_t$, and $\phi(z) = z$. $\beta_{t,0}, \beta_{t,m}, m = 1, ..., 5$; and $\beta_{t,k+5}$ denote regression coefficients for the intercept, polynomials, and basis functions, respectively. ε_t is an independent Gaussian random error with mean 0 and variance $\sigma_{y_t}^2$.

Using this RBFN method, the number of hidden knots K_t, the position, and error variance are unknown, and are strongly related to the precision of the reconstructed SST. The error variance is estimated by the squared deviation of the difference between $f(X_t)$ and SST increments Y_t. Since the number of hidden knots and the position can be affected by the distribution of SST samples, the INNC algorithm was designed to determine the hidden knots in the SST fields. Figure 2 shows the schematic diagram of this RBFN method for reconstructing SST.
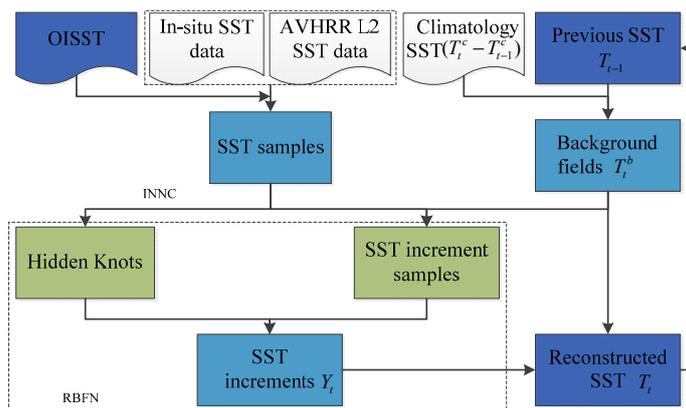
Figure 2. Schematic diagram of the RBFN method for reconstructing SST.

3.2. INNC Algorithm

The nearest neighbor cluster (NNC) algorithm is used to search the optimal hidden knots only from the SST samples for RBFNs [28,29], while the INNC algorithm can choose the hidden knots from both SST samples and background field, and consider the values of SST samples. The INNC algorithm is described as follows.

(1) Standardizing the original SST data, and make sure each variable of (x, y, z) in the SST matrix, with the mean of 0 and standard deviation of 1, where z is the value of the SST at the position of (x, y) in the SST matrix.

(2) Define a minimal distance D and set the first SST sample (x_1, y_1, z_1) as the first center c_1.

(3) For the second SST sample (x_2, y_2, z_2), the Euclidean distance s to the center c_1 is calculated. If $s > D$, then the position (x_2, y_2, z_2) is the next center c_2, otherwise the algorithm searches for the next SST sample (x_3, y_3, z_3).

(4) For the i-th SST sample (x_i, y_i, z_i), the Euclidean distance $\{s_k\}$ to each center $\{c_k\}$ is calculated, $k = 1, \ldots, K$. K is the number of center. If the minimal distance $s_m > D$, then the position (x_i, y_i, z_i) is the next center c_{K+1}, otherwise the algorithm searches for the next SST sample, until the last one is found.

(5) The values from the background field T_t^b are used to fill the positions without SST samples, before repeating step (3) for each position to select the centers from the background field. This continues until all of the positions are processed in the SST matrix, and the hidden knots u_t are obtained by using the positions of centers $\{c_k\}$ in the SST matrix.

In terms of the INNC algorithm, the parameters K_t and u_t are determined when the minimal distance D is defined. The value of D ranges from 0.2 to 1.5 in this study, and the optimal D is selected with the minimal error variance $\sigma_{y_t}^2$ to obtain the optimal K_t and u_t for RBFN. Thus, the optimal K_t and u_t may vary with different quantities and distributions of SST samples. In addition, the SST values at the positions of the optimal hidden knots from the background fields are added to the SST samples, so that there is at least one SST sample in the SST field within the minimal distance of D.

3.3. Evaluating the Performance of the RBFN Method

As described above, a simplified multi-quadric function $\sqrt{z^2 + s_1^2}$ (s_1 equal 0 here) is chosen as a basis function, and the INNC algorithm is designed to search for the optimal hidden knots for RBFNs. In order to evaluate the performance of this RBFN method, a Gaussian function $\exp(-s_2 z^2)$ is selected for comparison and $s_2 = K_t/d_{\max}$ [30], because $d_{\max}$ is the maximum distance between the hidden knots, thus $d_{\max} = D$ in the study. K-means algorithm and Kohonen-map algorithm are also tested for selecting hidden knots for RBFNs. While the number of hidden knots K_t should be assigned for these

two clustering algorithms, so the value of K_t ranges from 20 to 120 for K-means algorithm, and the Kohonen-maps with regular array of $m \times m$ hidden knots are tested and the value of m ranges from 5 to 15, then the optimal values of K_t and m are separately selected with the minimal error variance $\sigma^2_{y_t}$. More Details about K-means algorithm and Kohonen-map algorithm are described in [31–33].

Additionally, the OI method is used to compare with this RBFN method. Since the values of SST samples are assigned by the reference SST in the study, we select the noise-to-signal standard deviation ratio of AVHRR data for SST samples (the value is 0.5) and the average correlation scales for OI method (zonal and meridional correlation scales are 151 km and 155 km, respectively). Details of the OI method are described by Reynolds et al. [6].

To validate the accuracy of SST, the reconstructed SSTs are compared with the reference SSTs, and four commonly used error metrics are calculated: root mean square error (RMSE), mean absolute error (MAE), Pearson correlation coefficient (R), and signal-to-noise ratio (SNR). The SNR is the ratio of the standard deviation of the SST results to the standard deviation of the errors [34].

4. Results

In this section, the RBFN methods with different basis functions and clustering methods are tested, and the performance of the RBFN method for reconstructing daily SSTs is evaluated by comparing with the OI method.

4.1. Results from Different Basis Functions

Using the INNC algorithm, the optimal hidden knots can be chosen from SST samples and background fields. But the distributions of SST samples are very sparse in this study, the reconstructed SSTs from the RBFN method may be influenced by the type of the basis function.

Figures 3 and 4 display the SST increments that are estimated by using Equation (2) with Gaussian function and multi-quadric function on 13 June and 18 June, respectively. It is clear that the SST increments obtained from Gaussian function have a lot of anomaly circular regions, as shown in Figures 3a and 4a, where the SST increments are significantly different to the values from the regions behind them. While the SST increment fields using multi-quadric function in Figures 3b and 4b are quite close to their neighborhood regions. Besides, due to the characteristics of the Gaussian function, the effective areas of the hidden knots are limited, so the SST increments are strongly affected by the hidden knots that are selected from the background fields when there are few SST samples nearby, which may cause large errors on SST increments.
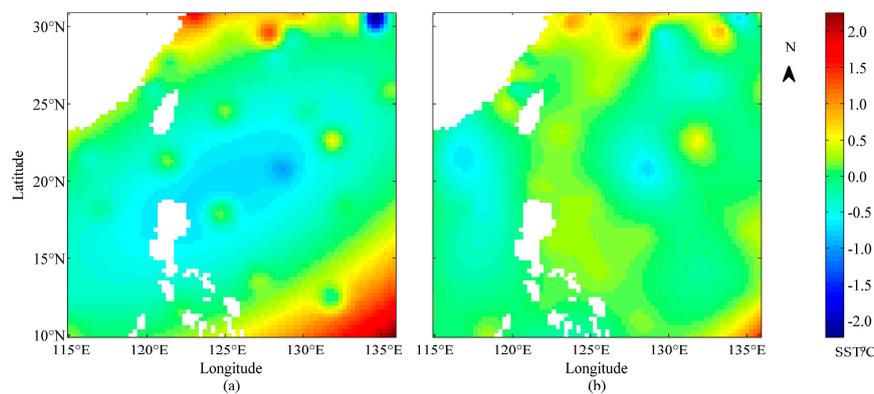

Figure 3. Reconstructed SST increments on 13 June 2014 obtained by using (**a**) Gaussian function and (**b**) multi-quadric function. White areas are lands in the images.
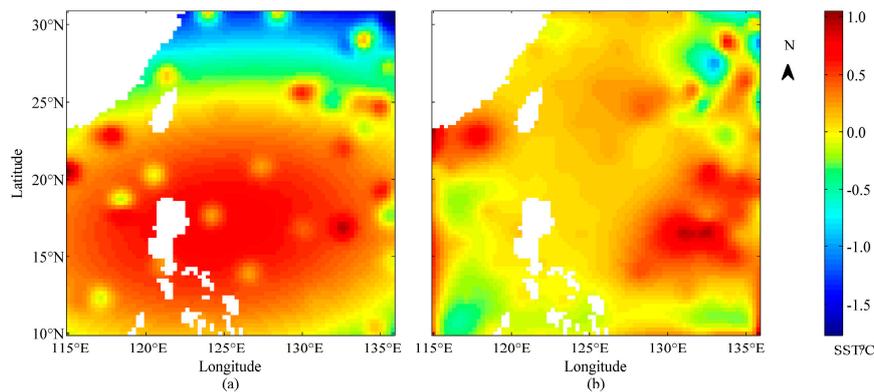
Figure 4. Reconstructed SST increments on 18 June 2014 obtained by using (**a**) Gaussian function and (**b**) multi-quadric function. White areas are lands in the images.

The statistical results from Tables 1 and 2 demonstrate that the multi-quadric function is significantly better than the Gaussian function as the basis function for the RBFN method, especially from the results on 18 June, the RMSE from the multi-quadric function is 0.47 °C, which is quite smaller than the value of 0.73 °C from the Gaussian function.

Table 1. Statistical results from different basis functions and clustering methods on 13 June 2014. The subscripts G and M stand for Gaussian function and multi-quadric function, respectively.

	RMSE (°C)	MAE (°C)	R	SNR
$INNC_G$	0.66	0.96	0.49	3.43
$INNC_M$	0.51	0.98	0.37	4.61
$K-means_M$	0.66	0.97	0.49	4.02
$Kohonen-map_M$	0.55	0.97	0.43	4.18

Table 2. Statistical results from different basis functions and clustering methods on 18 June 2014. The subscripts G and M stand for Gaussian function and multi-quadric function, respectively.

	RMSE (°C)	MAE (°C)	R	SNR
$INNC_G$	0.73	0.97	0.52	3.60
$INNC_M$	0.47	0.98	0.34	4.64
$K-means_M$	0.59	0.97	0.43	4.06
$Kohonen-map_M$	0.69	0.96	0.51	3.39

4.2. Results from Different Clustering Algorithms

A RBFN with too many or too few hidden knots is not conductive to simulating SST fields. Additionally, the distribution of hidden knots has a strong influence on the construction of a RBFN. Therefore, the quality of reconstructed SSTs from the RBFN method is highly related to the quantity and distribution of hidden knots, and it is important to select the suitable hidden knots for RBFNs.

Three clustering algorithms, including K-means algorithm, Kohonen-map algorithm, and INNC algorithm are compared to selecting the optimal hidden knots for RBFNs on 13 June and 18 June, respectively. Since the RBFN is quite sensitive to the hidden knots, the values of hidden knots are very important to the reconstructed SSTs. As a result, if a hidden knot is obtained from the background field, the value of which is close to 0 in the SST increment field in Equation (2), and the SST accuracy

will decrease in this situation. Thus, it is conductive to select the hidden knots from the SST samples rather than from the background field.

Figures 5 and 6 show the distributions of hidden knots from these three algorithms. It is easy to discover that there are more hidden knots that are selected from the SST samples (green points) by the INNC algorithm than those by the other two algorithms, because the clustering centers of the K-means algorithm and the Kohonen-map algorithm are chosen by training the input samples, while the INNC algorithm directly selects the hidden knots from the SST samples and make sure that as many hidden knots as possible are obtained from SST samples within the minimal distance *D*. As the statistical results shown in Tables 1 and 2, the INNC algorithm in this study with the highest accuracy of reconstructed SSTs, is more suitable than the K-means algorithm and the Kohonen-map algorithm to select the optimal hidden knots for the RBFN method.
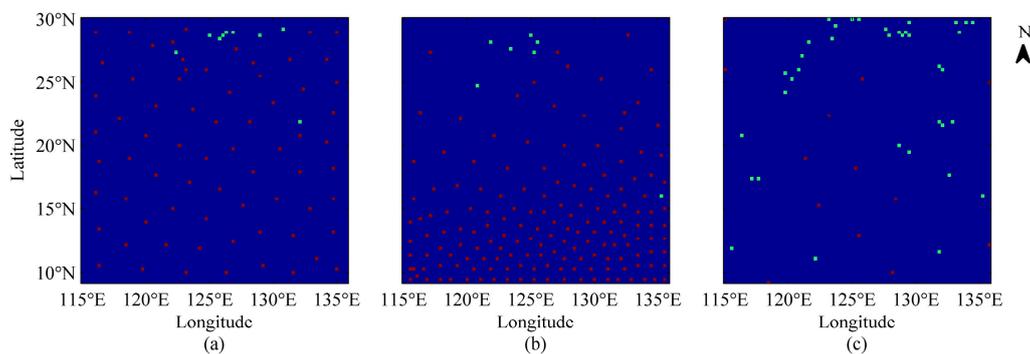

Figure 5. The distributions of hidden knots for radial basis function network (RBFN) on 13 June 2014 from (**a**) K-means algorithm; (**b**) Kohonen-map algorithm and (**c**) improved nearest neighbor cluster (INNC) algorithm respectively. The red points and green points are the locations of hidden knots for RBFNs separately selected from the background field and the SST samples.
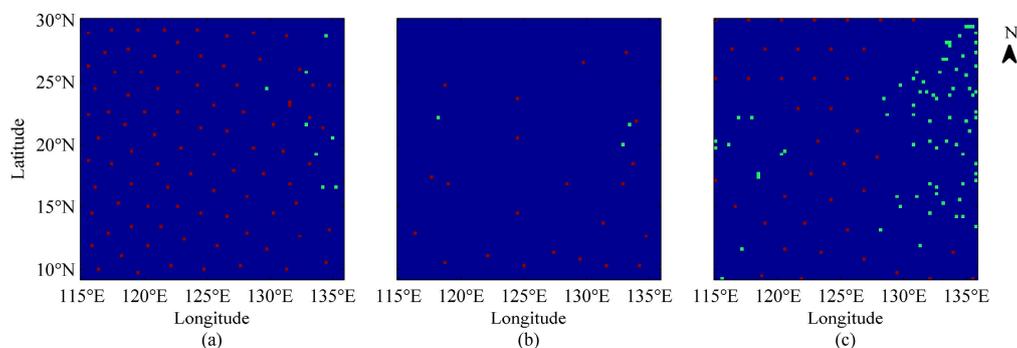

Figure 6. The distributions of hidden knots for RBFN on 18 June 2014 from (**a**) K-means algorithm; (**b**) Kohonen-map algorithm and (**c**) INNC algorithm respectively. The red points and green points are the locations of hidden knots for RBFNs separately selected from the background field and the SST samples.

4.3. Comparison with the OI Method

Due to the limitation of the correlation scales for the OI method, the SST increments will close to 0 when there are no SST samples behind them, while the SST increments from the RBFN method are obtained based on the tendency of the whole SST increment field. This is the key difference between the OI method and the RBFN method for reconstructing SST.

Figure 7 displays the time-series of each error metric from the OI and RBFN methods during the period from 12 June to 18 June 2014. The performance of each error metric from the RBFN method is much better than that from the OI method. The average values of these error metrics

in Table 3 indicate that the RBFN method increases R and SNR from 0.96 and 3.82 to 0.98 and 4.94, respectively, and decreases RMSE and MAE from 0.69 °C and 0.46 °C to 0.48 °C and 0.35 °C, respectively. Although the RBFN method requires more computation time than the OI method, due to the selection of optimal hidden knots by using the INNC algorithm, the RBFN method has a better accuracy of reconstruction than the OI method.
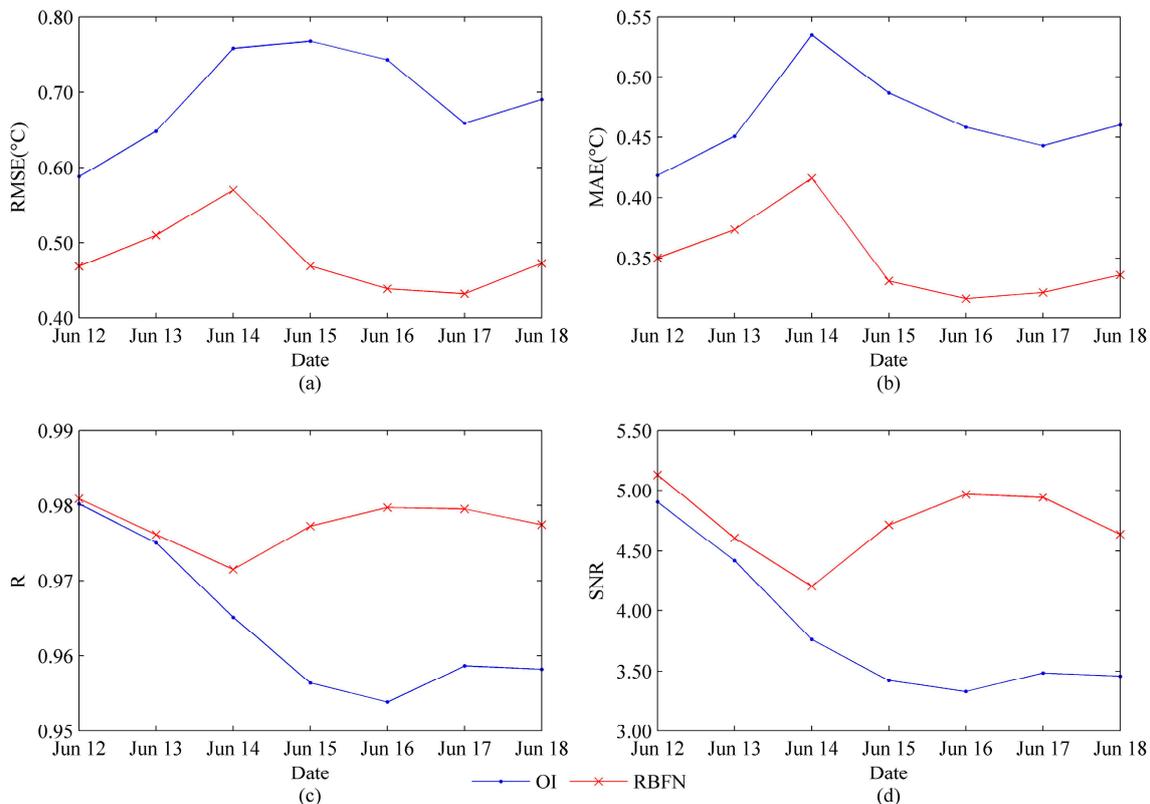

Figure 7. (**a**) root mean square error (RMSEs); (**b**) mean absolute error (MAEs); (**c**) Pearson correlation coefficient (Rs) and (**d**) signal-to-noise ratio (SNRs) of the reconstructed SSTs from the optimum interpolation (OI) and RBFN methods in the study region during the period from 12 June to 18 June 2014.

Table 3. Statistical results from the OI and RBFN methods.

	RMSE (°C)	MAE (°C)	R	SNR
OI	0.69	0.46	0.96	3.82
RBFN	0.48	0.35	0.98	4.94

In Figures 8 and 9, the reconstructed SSTs from 13 June and 18 June are separately displayed as examples. The results show that the reconstructed SSTs from the OI method (shown in Figures 8b and 9b) and the RBFN method (shown in Figures 8d and 9d) are quite similar to their reference SSTs (see Figures 8c and 9c). But, the SSTs in the region outlined by the square from the OI method have a significantly high SST anomaly, where these phenomenon are not obvious in the SSTs from the RBFN method and their reference data. This indicates that the RBFN method has a relatively better performance than the OI method to reconstruct SSTs when the original SST samples is quite sparse, as shown in Figures 8a and 9a.
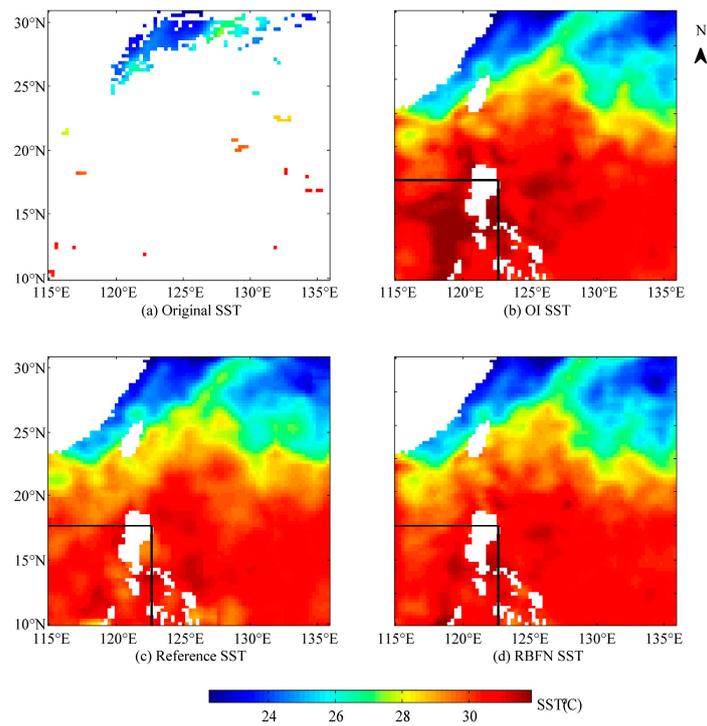
Figure 8. The distributions of (**a**) original SST samples; (**b**) OI SST; (**c**) reference SST; and, (**d**) RBFN SST on 13 June 2014. The region outlined by the square is used for comparison in detail. White areas in the images indicate no data or land.
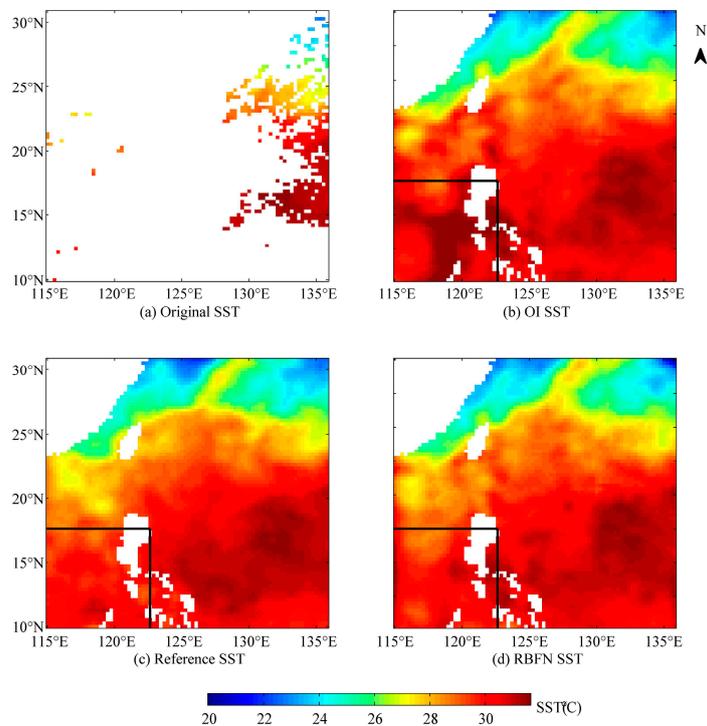

Figure 9. The distributions of (**a**) original SST samples; (**b**) OI SST; (**c**) reference SST; and, (**d**) RBFN SST on 18 June 2014. The region outlined by the square is used for comparison in detail. White areas in the images indicate no data or land.

However, the global distribution of original SST samples is much more complex than the local regions. As shown in Figure 10, the distribution of them is very irregular on 13 June 2014 in the global ocean, and the coverage of these SST data is not complete, especially in the equatorial regions, where the data is relatively limited. Therefore, in order to test the performance of RBFN method, we run the RBFN using the global SST samples and compare the results with the OI method. It should be noted that both the OI method and the RBFN method in this study choose the previous OISST product as the background field, and the global SST field from the RBFN method are combined by each 20° × 20° boxes of reconstructed SST samples that using the same scheme as the local experiment in Section 3.
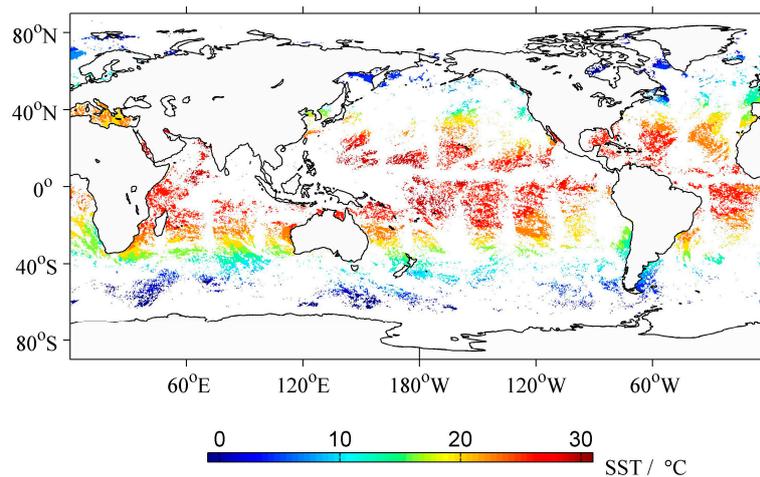

Figure 10. The global distribution of original SST samples on 13 June 2014. White areas in the images indicate no data or land, and the coastlines are highlighted with back contours.

Figure 11 displays the global distributions of SST biases separately from the OI method and the RBFN method on 13 June 2014 according to the corresponding reference SST. The results show that the variations of SST biases from these two methods are quite similar, but the SST biases from the RBFN method are relatively greater than those from the OI method when the SST samples in Figure 10 are not sparsely distributed. It indicates that this RBFN method might not be as good as the OI method when the coverage of SST samples is relatively high. Additionally, when compared with the global statistical results from the OI method in Table 4, the values of RMSE and MAD from the RBFN method are slightly larger, but its SNR value is relatively smaller. Thus the performance of RBFN method in the study is not very stable with different distributions of SST samples, and it still needs to be improved when applied in the global region.

Table 4. Global Statistical results from the OI and RBFN methods on 13 June 2014.

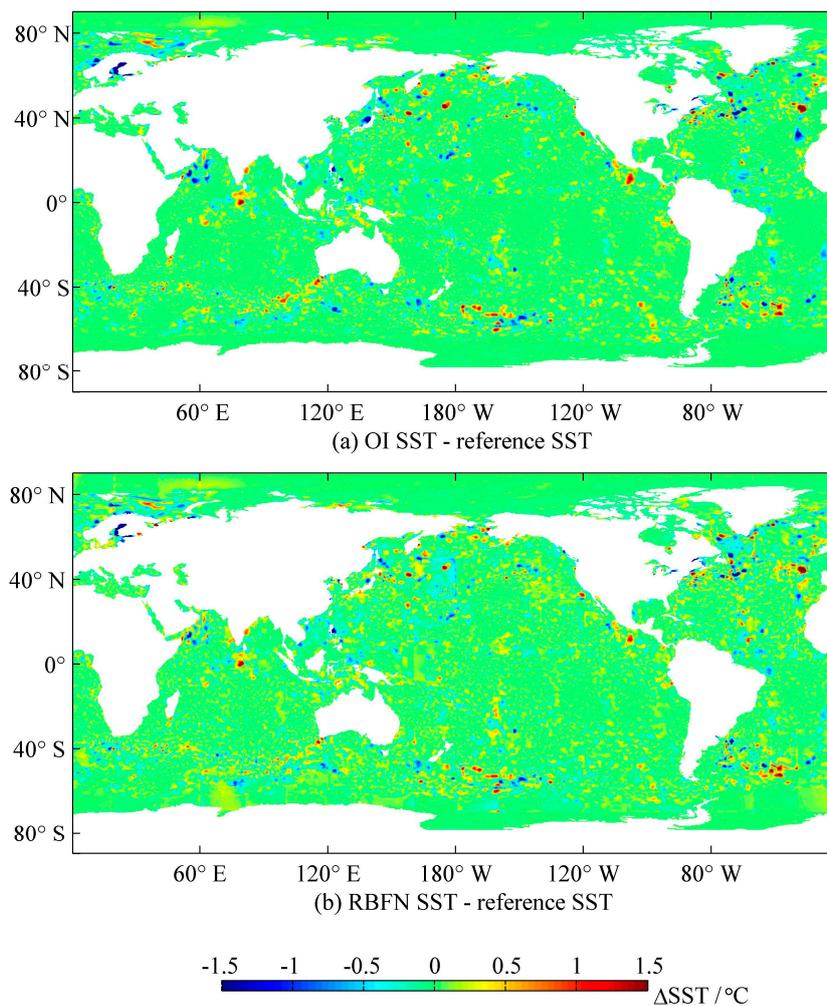|	RMSE (°C)	MAE (°C)	R	SNR
OI	0.18	0.08	0.99	65.38
RBFN	0.19	0.10	0.99	62.21

Figure 11. The global distributions of (**a**) OI SST—reference SST and (**b**) RBFN SST—reference SST on 13 June 2014. White areas in the images indicate no data or land.

5. Discussion

5.1. The INCC Algorithm for RBFNs

The hidden knots for RBFNs are commonly selected by clustering algorithms [28,29,35] or random sampling [24]. Since the hidden knots with different quantities and distributions should be tested in RBFNs, huge computational resources are required when random sampling is used to select the optimal hidden knots from SST fields. Commonly used clustering algorithms for RBFNs, such as K-means algorithm and Kohonen-map algorithm, are only used to determine the distribution of hidden knots when the number of hidden knots is given [32,35]. The K-means algorithm should be executed hundreds of times, which also takes a lot of time even though the algorithm has a fast convergence speed. For the Kohonen-map algorithm, an input vector is not only related to the nearest hidden knot, but also to its neighbor hidden knots, so the computation is also huge and the training step for the clustering centers consume a lot of time before the iteration ends. While using the INNC algorithm with a parameter D, both the quantity and the distribution of hidden knots could be directly determined without any iterative computation, and the values of D in this study were chosen from 0.2 to 1.5. Thus, the INNC algorithm is more efficient in selecting the optimal hidden knots for RBFNs.

Additionally, the INNC algorithm select the hidden knots from SST samples and background fields within the minimal distance D, and then the values of hidden knots from the background

fields are added to the SST samples, which makes sure that the distributions of the hidden knots and the SST samples are close to uniform in the SST field. While the multi-quadric basis function for RBFNs considers large related domains in SST field, so the values from both the SST samples and the backgrounds fields will contribute to the reconstructed SSTs when there are few SST samples nearby.

5.2. SST Samples

The coverage of SST samples that we selected in the Pacific Ocean is very low because the tropical storm Hagibis come with heavy clouds, which makes poor quality of satellite-retrieved SSTs, and the SSTs we can used are very limited. Theoretically, the errors of reconstructed SSTs are from both background fields and SST increments. The SST increments are estimated by using current SST samples, and the quality of background fields is strongly influenced by the previous SST samples, thus the distributions of SST samples at both the previous time and the current time play a significant role in reconstructing SST. It can well explain the fact that the minimal RMSE did not happen on 15 June in Figure 7a, which has the highest coverage of SST samples, as shown in Figure 1b. Because the 14 June with the lowest SST coverage in the study has a low accuracy of reconstructed SST, and make a poor quality of the background field for 15 June, so the accuracy of reconstructed SST on 15 June is not the best even though it has the highest coverage of SST samples.

The quality of SST samples is not only relevant to the coverage of SST samples, but it also depends on their accuracy. Since the SST sample values that were used were assigned by the reference SSTs, which can be considered as non-biased SST data, the reconstructed SSTs are of relatively high quality. However, the satellite-retrieved SST is often contaminated by the atmosphere and the SST data from different sources may contain errors from their respective surveying systems, meaning that it is not easy to obtain non-biased SST data, and the accuracy will decrease if large errors exist in the SST samples.

The SST samples with high coverage are conductive to acquiring high-quality daily SSTs. While acquiring larger quantities of SST samples often needs the combination of different sources of SSTs, more SST errors may come from these different kinds of sensors, and this has a negative influence on the quality of the SST samples overall. For the purpose of obtaining high-accuracy SST data from the RBFN method, more SST data should be used in reconstructing the SST. Additionally, strict quality control is required for SSTs from various measurements, and bias corrections for satellite-retrieved SSTs are necessary to eliminate the errors in SST samples.

5.3. The Performance of the RBFN Method

Since the OI method estimates SSTs using only the information within a limited distance, the estimated SSTs in some areas rely on background fields, which create large errors in the SST fields. The RBFNs with enough hidden knots use all the SST samples to approximate the SST field, so the SSTs from the RBFN method contain more information on the whole SST field, not only the information about SST samples nearby. Thus, the accuracy of SST that is obtained using the RBFN method is better than that using the OI method when the distribution of SST samples is quite sparse.

As shown in Figure 7a, the RMSE on 15 June from the RBFN method is significantly smaller than that from the OI method when the quality of the background field is quite poor. This shows that when compared with the SSTs from the OI method, the quality of SSTs from the RBFN method is more stable when the SST errors are associated with the background field.

On the other hand, the OI method is effective in estimating SST using neighboring samples, which was conductive to obtaining a higher accuracy of SSTs when the coverage of the SST samples is very high. While the quantity and distribution of hidden knots in RBFNs are determined by the minimal error variance, which is a global variable of the SST field, and the distribution of hidden knots may not be suitable in some local regions, and the SST errors of RBFNs may occur even in some areas with large amount of SST samples nearby. This is may be the reason why the RBFN method is not as effective as the OI method in the global region. Because both the OI method and the RBFN method is designed to minimize the errors of the SST field, and if the SST samples is sparsely distributed,

the distribution of expected errors is relatively uniform in the SST field. In this case, a global parameter of RBFN is suitable for the whole SST field. But when the coverage of SST samples is high, the expected errors of RBFN in the regions with SST samples is much smaller than those without SST samples nearby, then the expected errors are distributed unevenly in the SST field, and a global parameter may not adapt for the RBFN to simulate the high quality of SST, which may make the errors much larger than those of the OI method, as shown in Figure 11.

In addition, although the INNC algorithm with *D* is very efficient in searching for hidden knots in SST fields, the optimal *D* of the INNC algorithm should be selected from the values between 0.2 and 1.2, which makes the RBFN method more computationally expensive than the OI method. Overall, the OI and RBFN methods have their own characteristics in terms of reconstructing SST, but the RBFN method is more effective than the OI method when the SST samples are sparsely distributed.

6. Conclusions

Reconstructing SST fields from a limited number of SST samples is important for data applications. In this study, a RBFN method is proposed and a INNC algorithm is designed to search for the optimal hidden knots for RBFNs. When compared with the other clustering algorithms as K-means algorithm and Kohonen-map algorithm, the INNC algorithm can obtain more hidden knots from the SST samples, which will contribute to acquiring a high quality of reconstructed SST. Using the multi-quadric function as the basis function is more effective than the Gaussian function to avoid the SST increment anomaly on the regions of hidden knots. Thus, the INNC algorithm with multi-quadric function is quite suitable for the RBFN method to reconstruct SSTs in the study.

To evaluate the accuracy of this RBFN method, SST samples with low coverage are reconstructed by using the RBFN and OI methods, respectively. The results show that when compared with the SSTs from the OI method, the quality of SST from the RBFN method is more stable when the SST errors are associated with the background field. Moreover, the RBFN method is less affected by missing values in SST samples, and the SSTs from the RBFN method have a higher accuracy than those from the OI method when SST samples are sparsely distributed. According to this characteristic of the RBFN method, it is quite suitable to be used for SST reconstruction that only combine the in situ data and the satellite retrieved SSTs from infrared radiometers.

When considering the efficiency of RBFNs, hidden knots with different quantities and distributions are determined simultaneously using various *D* values in the INCC algorithm, and the optimal *D* for RBFN is selected with the minimal error variance. The step-length of *D* is set as 0.02 in this study. We believe that if the step-length of *D* can be set smaller, then the accuracy may be improved.

However, the accuracy of reconstructed SSTs is strongly influenced by the quality of SST samples from both the present and current time, and the biases in SST samples will increase the errors in SSTs, thus the strict quality control and bias corrections to SST samples are required in practice. Though the RBFN method has a better performance than the OI method to reconstruct SST field with the limited SST samples, the advantage is not obvious when the coverage of SST samples is high (the details are described in the Supplemental Material), so this RBFN method is appropriate to be used in some local regions, not to the global region at the present stage, and we will improve this RBFN method in our future work.

Author Contributions: Zhihong Liao proposed the idea and the algorithm, wrote the introduction and the algorithm sections, and integrated the contributions from all authors as a whole. Qing Dong and Cunjin Xue designed the experiments and the structure of this manuscript, and collaborated with Zhihong Liao in writing the discussion and conclusion sections. Jingwu Bi and Guangtong Wan contributed the experiment, and collaborated with Zhihong Liao in writing the results section.

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

References

1. Saha, S.; Moorthi, S.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Behringer, D.; Hou, Y.T.; Chuang, H.Y.; Iredell, M.; et al. The NCEP climate forecast system version 2. *J. Clim.* **2014**, *27*, 2185–2208. [CrossRef]
2. Shirvani, A.; Nazemosadat, S.; Kahya, E. Analyses of the Persian Gulf sea surface temperature: Prediction and detection of climate change signals. *Arabian J. Geosci.* **2015**, *8*, 2121–2130. [CrossRef]
3. Knutson, T.R.; McBride, J.L.; Chan, J.; Emanuel, K.; Holland, G.; Landsea, C.; Held, I.; Kossin, J.P.; Srivastava, A.K.; Sugi, M. Tropical cyclones and climate change. *Nat. Geosci.* **2010**, *3*, 157–163. [CrossRef]
4. Donlon, C.; Rayner, N.; Robinson, I.; Poulter, D.J.S.; Casey, K.S.; Vazquez-Cuervo, J.; Armstrong, E.; Bingham, A.; Arino, O.; Gentemann, C.; et al. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 1197–1213. [CrossRef]
5. Martin, M.; Dash, P.; Ignatov, A.; Banzon, V.; Beggs, H.; Brasnett, B.; Cayula, J.F.; Cummings, J.; Donlon, C.; Gentemann, C.; et al. Group for High Resolution Sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). *Deep Sea Res. Part II* **2012**, *77*, 21–30. [CrossRef]
6. Reynolds, R.W.; Smith, T.M.; Liu, C.; Chelton, D.B.; Casey, K.S.; Schlax, M.G. Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* **2007**, *20*, 5473–5496. [CrossRef]
7. Reynolds, R.W. A real-time global sea surface temperature analysis. *J. Clim.* **1988**, *1*, 75–87. [CrossRef]
8. Smith, T.M.; Reynolds, R.W.; Livezey, R.E.; Stokes, D.C. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Clim.* **1996**, *9*, 1403–1420. [CrossRef]
9. Smith, T.M.; Livezey, R.E.; Shen, S.S. An improved method for analyzing sparse and irregularly distributed SST data on a regular grid: The tropical Pacific Ocean. *J. Clim.* **1998**, *11*, 1717–1729. [CrossRef]
10. Ping, B.; Su, F.; Meng, Y. Reconstruction of satellite-derived sea surface temperature data based on an improved DINEOF algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4181–4188. [CrossRef]
11. Beckers, J.-M.; Rixen, M. EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Ocean. Technol.* **2003**, *20*, 1839–1856. [CrossRef]
12. Beggs, H.; Zhong, A.; Warren, G.; Alves, O.; Brassington, G.; Pugh, T. RAMSSA—An operational, high-resolution, Regional Australian Multi-Sensor Sea surface temperature Analysis over the Australian region. *Aust. Meteorol. Oceanogr. J.* **2011**, *61*, 1. [CrossRef]
13. Brasnett, B. The impact of satellite retrievals in a global sea-surface-temperature analysis. *Q. J. R. Meteorol. Soc.* **2008**, *134*, 1745–1760. [CrossRef]
14. Cummings, J.A. Operational multivariate ocean data assimilation. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 3583–3604. [CrossRef]
15. 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]
16. Gemmill, W.; Katz, B.; Li, X. *Daily Real-Time Global Sea Surface Temperature—High-Resolution Analysis: RTG_SST_HR*; NCEP No. 260; EMC Office Note: Maryland City, MD, USA, 2007.
17. Gentemann, C.L.; Wentz, F.J.; DeMaria, M. Near real time global optimum interpolated microwave SSTs: Applications to hurricane intensity forecasting. In Proceedings of the 27th Conference on Hurricanes and Tropical Meteorology, Monterey, CA, USA, 23–28 April 2006.
18. Park, J.; Sandberg, I.W. Universal approximation using radial-basis-function networks. *Neural Comput.* **1991**, *3*, 246–257. [CrossRef]
19. Holmes, C.; Mallick, B. Bayesian radial basis functions of variable dimension. *Neural Comput.* **1998**, *10*, 1217–1233. [CrossRef]
20. Powell, J. Radial basis function approximations to polynomials. In *Numerical Analysis 1987*; Longman Publishing Group: White Plains, NY, USA, 1987; pp. 223–241.
21. Liu, L.; Chua, L.; Ghista, D. Mesh-free radial basis function method for static, free vibration and buckling analysis of shear deformable composite laminates. *Compos. Struct.* **2007**, *78*, 58–69. [CrossRef]
22. Miazhynskaia, T.; Frühwirth-Schnatter, S.; Dorffner, G. Neural network models for conditional distribution under bayesian analysis. *Neural Comput.* **2008**, *20*, 504–522. [CrossRef] [PubMed]

23. Konishi, S.; Ando, T.; Imoto, S. Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika* **2004**, *91*, 27–43. [CrossRef]

24. Ryu, D.; Liang, F.; Mallick, B.K. Sea surface temperature modeling using radial basis function networks with a dynamically weighted particle filter. *J. Am. Stat. Assoc.* **2013**, *108*, 111–123. [CrossRef]

25. Xu, F.; Ignatov, A. In Situ SST quality monitor (i Quam). *J. Atmos. Ocean. Technol.* **2014**, *31*, 164–180. [CrossRef]

26. Reynolds, R.W.; Smith, T.M. Improved global sea surface temperature analyses using optimum interpolation. *J. Clim.* **1994**, *7*, 929–948. [CrossRef]

27. Banzon, V.F.; Reynolds, R.W.; Stokes, D.; Xue, Y. A 1/4°-spatial-resolution daily sea surface temperature climatology based on a blended satellite and In Situ analysis. *J. Clim.* **2014**, *27*, 8221–8228. [CrossRef]

28. Diao, X.H. Study on regional financial risk early warning system based on uniform design method and Nearest Neighbo-Clustering RBFNN. In Proceedings of the 2011 IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS), Guilin, China, 24–26 October 2011.

29. Zhu, M.; Zhang, D. Study on the algorithms of selecting the radial basis function center. *J. Anhui Univ.* **2000**, *3*, 72–78.

30. Lowe, D. Adaptive radial basis function nonlinearities, and the problem of generalisation. In Proceedings of the 1989 First IEE International Conference on Artificial Neural Networks, London, UK, 16–18 October 1989. No. 313.

31. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.D.; Silverman, R.; Wu, A.Y. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [CrossRef]

32. Latif, B.A.; Lecerf, R.; Mercier, G.; Hubert-Moy, L. Preprocessing of low-resolution time series contaminated by clouds and shadows. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2083–2096. [CrossRef]

33. Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. [CrossRef]

34. Zhang, H.; Qin, S.; Ma, J.; You, H. Using residual resampling and sensitivity analysis to improve particle filter data assimilation accuracy. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1404–1408. [CrossRef]

35. Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; McMaster University: Hamilton, ON, Canada, 2009.