Article

Physics-Informed Deep Learning for Reconstruction of Spatial Missing Climate Information in the Antarctic

Ziqiang Yao [1], Tao Zhang [2,*], Li Wu [1,*], Xiaoying Wang [1] and Jianqiang Huang [1]

1   Department of Computer Technology and Applications, Qinghai University, Xining 810016, China
2   Brookhaven National Laboratory, Upton, NY 11973, USA
*   Correspondence: tzhang@bnl.gov (T.Z.); wuli777@qhu.edu.cn (L.W.)

Abstract: Understanding the influence of the Antarctic on the global climate is crucial for the prediction of global warming. However, due to very few observation sites, it is difficult to reconstruct the rational spatial pattern by filling in the missing values from the limited site observations. To tackle this challenge, regional spatial gap-filling methods, such as Kriging and inverse distance weighted (IDW), are regularly used in geoscience. Nevertheless, the reconstructing credibility of these methods is undesirable when the spatial structure has massive missing pieces. Inspired by image inpainting, we propose a novel deep learning method that demonstrates a good effect by embedding the physics-aware initialization of deep learning methods for rapid learning and capturing the spatial dependence for the high-fidelity imputation of missing areas. We create the benchmark dataset that artificially masks the Antarctic region with ratios of 30%, 50% and 70%. The reconstructing monthly mean surface temperature using the deep learning image inpainting method RFR (Recurrent Feature Reasoning) exhibits an average of 63% and 71% improvement of accuracy over Kriging and IDW under different missing rates. With regard to wind speed, there are still 36% and 50% improvements. In particular, the achieved improvement is even better for the larger missing ratio, such as under the 70% missing rate, where the accuracy of RFR is 68% and 74% higher than Kriging and IDW for temperature and also 38% and 46% higher for wind speed. In addition, the PI-RFR (Physics-Informed Recurrent Feature Reasoning) method we proposed is initialized using the spatial pattern data simulated by the numerical climate model instead of the unified average. Compared with RFR, PI-RFR has an average accuracy improvement of 10% for temperature and 9% for wind speed. When applied to reconstruct the spatial pattern based on the Antarctic site observations, where the missing rate is over 90%, the proposed method exhibits more spatial characteristics than Kriging and IDW.

Keywords: deep learning; missing value reconstruction; numerical climate model; ERA5; Antarctica

1. Introduction

The intimate interaction between Antarctica and the rest of the world can have significant impacts on the global climate system [1]. With the effects of global warming, the changing climate in Antarctica could bring catastrophic consequences. The amount of sea level rise largely depends on the melting Antarctic ice sheet and threatens coastal cites all the world [2,3]. However, the contribution of Antarctica to global sea level rise remains uncertain due to the competing processes of increasing ice loss and snowfall accumulation in a warming climate, as indicated by climate system models [4]. The previous studies examining Antarctica's climate variations were mainly based on observational data from a limited number of Antarctic stations [5]. Although the reanalysis data, such as European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 [6], provide the global spatial distribution, there are still significant uncertainties at high latitudes [7], particularly in the Antarctic, due to the few meteorological stations and the small coverage area of observation data.

Due to the sparseness of the observation network, the distance between stations in Antarctica can reach hundreds of kilometers, posing significant challenges for reconstructing the spatial pattern of meteorological properties. Kriging [8] and Inverse Distance Weighted (IDW) [9] are traditional spatial interpolation methods that are commonly used in climate science for reconstructing distributions. Kriging is a regression algorithm for the spatial modeling and prediction (interpolation) of random processes and random fields based on covariance functions. IDW is based on the first law of geography and assumes that the attribute value of a point to be interpolated is the inverse distance weighted average of a group of known sample point attribute values in the neighborhood. Kriging and IDW have been employed to fill in the mass balance data of glaciers, precipitation and temperature data, among other applications [10–12]. However, the accuracy of Kriging depends heavily on parameters such as the variogram, which is a statistical value that describes the spatial correlation of random field and random processes. The efficiency of IDW is very sensitive to the outliers and sampling radius due to its simple computational means [13].

Recently, deep learning-based image inpainting methods have been creatively utilized in the reconstruction of climate spatial data [14–17]. Compared to traditional interpolation methods that primarily utilize distribution characteristics and limited parameters of existing data, such as distance, to estimate missing parts, the powerful nonlinear expression capability of deep learning methods can better analyze complex and chaotic climate data. Generative Adversarial Network (GAN)-based [18] and Convolution Neural Network (CNN)-based [19] methods are the two most commonly used approaches.

GAN-based methods belong to the generative model that learns to generate realistic data in an adversarial manner. It consists of a generator and a discriminator. The generator captures the data distribution and generates real data, while the discriminator estimates the probability that the data come from the real data space. Dong et al. [20] and Shibata et al. [21] utilized GAN to restore sea surface temperature (SST) satellite images to deal with the problem of cloud occlusion. Dewi et al. [22] reconstructed cloud vertical structure with the GAN, demonstrating the feasibility of GAN to solve problems in atmospheric remote sensing.However, it is difficult to guarantee the optimal convergence of the generator and discriminator at the same time because their optimization goals are different. As a result, the GAN model can have difficulty maintaining balance, leading to instability during training and potentially resulting in large errors and meaningless outputs [23].

CNN-based methods have made significant progress in the field of image inpainting by scanning the entire spatial information to obtain sufficient spatial features, and they use them to gradually restore missing parts. Li et al. [24] proposed a progressive image inpainting method, RFR, which involved the partial convolutions, encoder–decoder and attention mechanism. The algorithm achieved better reconstruction results in the face of large-scale irregular deletions, with its success benefitting from the following parts: first, the partial convolution operator learns the missing part according to the surrounding area. Second, the encoder–decoder structure prevents the deep learning models from becoming overly complex, thereby avoiding overfitting. Finally, the attention mechanism captures the spatial dependence to further improve the imputation accuracy [25]. Kadow et al. [26] absorbed the idea of such progressive image inpainting and used it with 20CR (20th Century Reanalysis) and CMIP5 (Coupled Model Intercomparison Project Phase 5) data, achieving better results than Kriging interpolation and principal component analysis-based infilling. However, the limitation of this method is that the initial values of all spatially missing parts have no physical meaning, resulting in a low learning efficiency of neural networks and low accuracy of final reconstruction.

To address this challenge, we propose a novel deep learning method, PI-RFR, which sets the initial values with spatial numerical model data in the RFR model to restore the spatial structure of Antarctic climate station data. PI-RFR leverages the current state-of-the-art deep learning image inpainting algorithm for climate spatial data reconstruction, while also using spatial-pattern initial values of missing parts in the Antarctic region to provide the deep learning method with a better basis for feature inference.

The experiments aim to quantitatively evaluate the spatial missing reconstruction capabilities of deep learning methods and traditional interpolation methods at different missing rates to prove that our method is robust and outperforms other approaches at higher missing rates. Taking the reconstruction of two climate variables, surface temperature and surface wind speed at Antarctic climate stations, as the examples, the spatial data reconstruction capabilities of Kriging, IDW, RFR and PI-RFR under 30%, 50% and 70% masking rates were compared, respectively. The reconstruction models are evaluated and trained with the restoration of artificially missing ERA (ECMWF Reanalysis) data, and the spatial structure of the entire Antarctic region is constructed by using the Antarctic climate station data input into the optimal trained reconstruction model.

In the remainder of this paper, Section 2 presents the new deep learning method. Section 3 presents the data and experiments. Section 4 illustrates the detailed experimental evaluation results together with the corresponding analysis. Section 5 summarizes this study and discusses about the future work.

2. Methodology

The method proposed in this paper is derived from the field of image inpainting and has been adapted for the specific application of reconstructing missing meteorological values. Image inpainting and missing value reconstruction in geoscience share a similarity in that they both seek to explore the mapping relationship between existing data and missing data. Therefore, the method of image inpainting has important value for designing the method of missing meteorological data reconstruction. The image inpainting algorithm RFR based on deep learning can be used to reconstruct the missing meteorological data. The RFR model used in this study consists of four components: partial convolution (PConv) layers, a CNN encoder, attention mechanism and a transposed convolution decoder, as shown in Figure 1. Partial convolution is used to update the mask in iterations and learn the missing parts. The CNN encoder is used to extract high-level features of input meteorological data in order to fill the missing area with the spatial information. The newly designed attention mechanism accumulates the attention scores of the two adjacent iterations in proportion and uses the attention scores to control the important of information to find the features most relevant to the missing data for full fusion. The transposed convolution decoder is finally used to transform the high-level features to complete meteorological data. Based on this, PI-RFR innovatively introduces the physics-spatial-aware initial values for each property to accelerate the convergency of deep learning and achieve high-fidelity imputation.
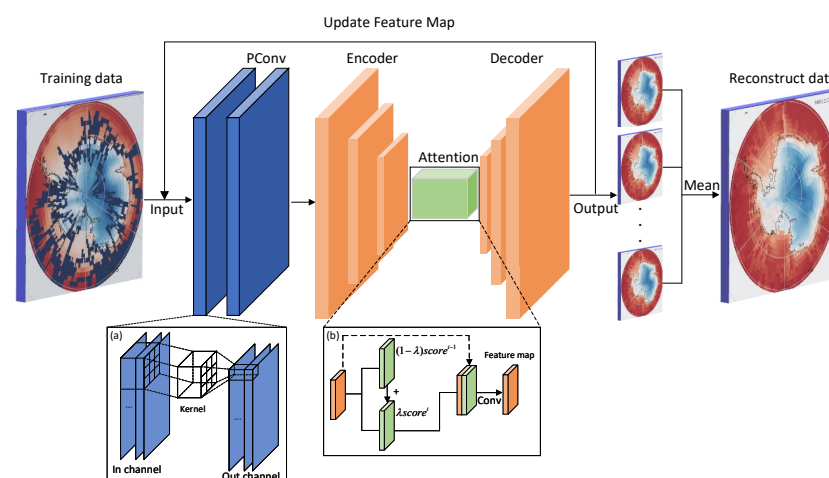

Figure 1. The overall network structure consists of input, partial convolution, encoder, attention mechanism, decoder and output modules. The masked reanalysis data and mask are fed into the network. The convolution operation in partial convolution is shown in (**a**). The unfolded structure of the attention mechanism is shown in (**b**), where λ is a learnable parameter, $score^i$ represents the attention score of step i, and $score^{i-1}$ is the attention score of the previous step.

2.1. Initialized by Spatial Pattern from Climate Model

The original RFR algorithm does not discuss the selection of the default initial value of the mask area and instead uses 0 as the initial value. However, there is a significant difference between the manually masked data and the ground truth, which can slow down the model's convergence and negatively impact the reconstruction results. To address this issue, the RFR method in this paper is chosen to directly change the initial value of the missing part to the average climate value calculated from the reanalysis data. However, due to the huge difference in the spatial distribution of climate, this method is suboptimal.

Therefore, we propose a new method called PI-RFR, which utilizes climatology data with spatial information from climate system models as initial values. Hence, PI-RFR starts from a relatively reliable foundation and uses the data analysis and fusion capabilities of the deep learning model to reconstruct the missing data.

The initial state of Antarctica's skin temperature in PI-RFR under a 30% missing rate is shown in Figure 2c. It can be seen intuitively that the initial value after filling the mask with the numerical model space data is more consistent with the spatial distribution of Antarctic reanalysis data (Figure 2a) than the initial value without any filling (Figure 2b).
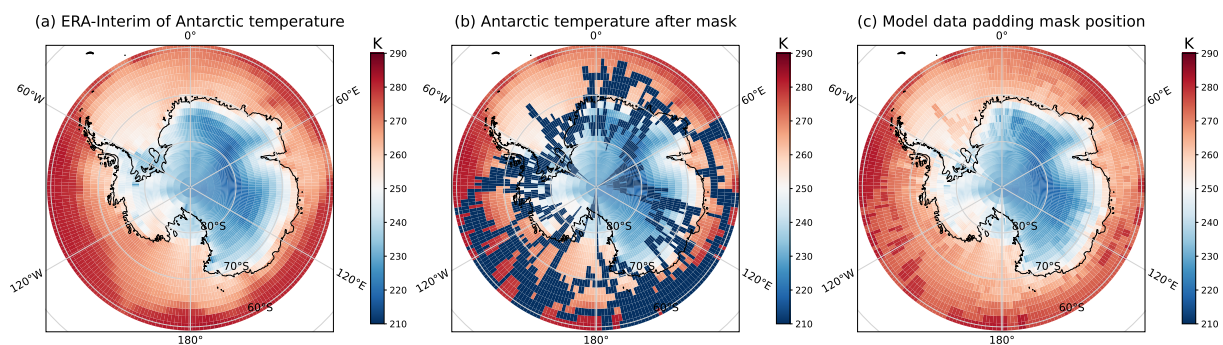
Figure 2. ERA-interim of Antarctic temperature as ground truth (**a**), ERA-interim of Antarctic temperature with artificial random 30% mask rate (**b**) and E3SM model temperature data as initial value of the mask area (**c**).

2.2. Recurrent Feature Reasoning

This paper innovatively applies the image inpainting algorithm RFR to the reconstruction of missing meteorological data. The following are the components of the RFR model and the important techniques used in this study.

(1) Partial convolutional network

The partial convolutional network (PConv) proposed by Liu et al. [17]. was developed based on CNN. Both CNN and PConv can obtain the feature map of data, but PCnov can synchronously update the mask to learn the missing area. As shown in Figure 3, the PConv network performs convolution operations on the spatial meteorological data with the missing parts and the corresponding size of the pure mask composed of the missing and the data identifiers, such as 0 and 1 here [27]. As the number of networks increases, the missing parts of the mask becomes smaller, and the effective area in the output result becomes larger.

The PConv layer consists of a partial convolution operation and mask update function. Let W be the weight of the convolution filter and b be its corresponding bias. X denotes the feature values for the current convolution sliding window, and M is the corresponding mask. The partial convolution at every location can be defined as Equation (1).

$$x = \begin{cases} W^T(X \odot M)\frac{sum(m)}{sum(M)} + b & if \ sum(M) > 0 \\ 0 & else \end{cases} \tag{1}$$

where $\odot$ denotes element-wise multiplication, and m is the convolution kernel of the mask for convolution operation, which has the same shape as M, and all the elements are 1.

The $sum(m)/sum(M)$ as a coefficient adjusts the unmasked inputs. It can be seen that output values are more affected by the mask edge (A smaller $sum(M)$ will lead to a larger $sum(m)/sum(M)$).
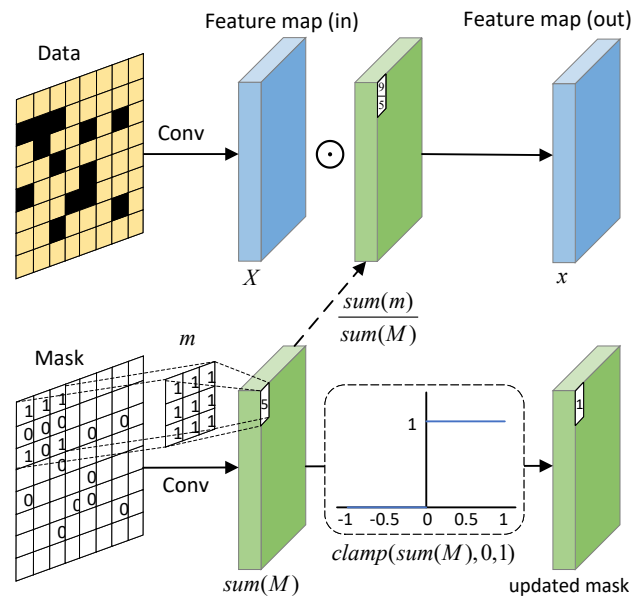

Figure 3. Partial convolution. Input the masked meteorological data and corresponding mask into PCnov to obtain the feature map and the updated mask. Partial convolution operation is divided into convolution operation on the masked data and updating the mask. The feature map (in) obtained by convolution operation is affected by the mask to obtain the final feature map (out), where $\odot$ is the element-wise multiplication and function $clamp(M, min, max)$ compresses the values in the matrix M into the interval $[min, max]$.

After performing partial convolution, the mask of the corresponding region of the convolution window contains at least one value of 1 ($sum(M) \neq 0$), and the mask at the corresponding position is updated to 1. In this way, after each partial convolution, the mask will eventually be all ones, if the input has any valid data. Passing though the partial convolution layers, the feature maps are processed by a normalization layer and an activation function before being sent to the encoder–decoder structure.

(2) Encoder and Decoder

An encoder–decoder structure is used to extract features and reconstruct data. The encoder downsamples the data through convolution, pooling and normalization to extract critical feature information [28]. The decoder upsamples the encoded feature map to recover the data. The encoder–decoder architecture is shown in Figure 4.
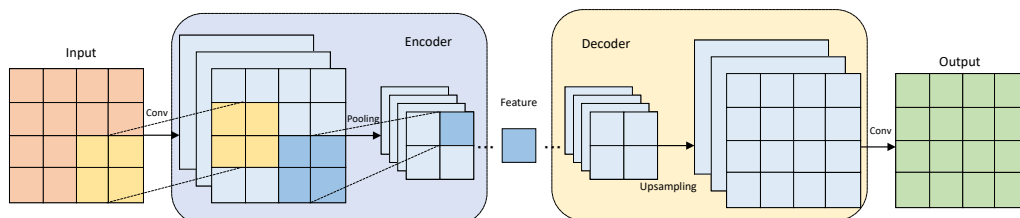
Figure 4. Encoder and decoder. The input data through PConv extracts higher-level spatial features through the encoder and decoder structures.

For this study, the encoder learns high-level feature maps from the missing meteorological data by convolutional neural networks (CNNs). The stride of each convolution layer is 2. The spatial sizes of the feature were simultaneously reduced, meaning that the model requires less GPU memory usage. Furthermore, a larger receptive field was obtained

for the encoder compared to the original inputs. Therefore, this will help the model to learn a wide range of spatial features. Then, the decoder reconstructs missing data and recovers the data size by deconvolution. The rectified linear unit (ReLU) activity layer was used for all layers to prevent the gradient from disappearing.

Furthermore, the feature spatial pattern after the encoder and decoder are fed into the partial convolution again, and the missing parts are gradually filled from the regional edge to the regional center. A higher quality spatial pattern can be obtained through the encoder–decoder structure.

(3) Attention Mechanism

The attention mechanism can make a neural network recognize important features by weighting input values. For this study, the attention mechanism measures the importance of meteorological data features in different regions by calculating attention scores, meaning that the model can focus on important regions to obtain better spatial patterns. The attention score is used as the weight of the input meteorological data. Through weights, we can focus the attention on the more valuable part to train the model more efficiently. In this method, in the face of spatially missing meteorological data, the neural network will focus on the parts that are not missing and try to reconstruct the missing data from the features of the existing data. The RFR uses an attention mechanism, which can adaptively fuse the attention scores proportionally accumulated from previous recurrences, both in the encoder and decoder. As a result, the mechanism helps the neural network to better extract features and search for possible characteristics in the background. The computational details of the attention mechanism are shown below.

We first calculate the spatial correlation coefficient of each grid point by Equation (2):

$$\widehat{sim}^i_{lat,lon,lat',lon'} = \frac{v_{lat,lon} \cdot v_{lat',lon'}}{\left\| v_{lat,lon} \right\| \left\| v_{lat',lon'} \right\|} \tag{2}$$

where $sim^i_{lat,lon,lat',lon'}$ indicates the similarity between the feature at location (lat, lon) and that at location (lat', lon'), and the attention scores are smoothed by averaging the correlation of a target point data in an adjacent area (side length is k).

$$sim^i_{lat,lon,lat',lon'} = \frac{\sum\limits_{p,q \in \{-k,\ldots,k\}} \widehat{sim}_{lat+p,lon+q,lat',lon'}}{k \times k} \tag{3}$$

Then, the result of similarity is input into the softmax function to calculate the score for the (lat, lon) position. The value range of the softmax function is $(0, 1)$, and the data processed by this function is suitable as a weight. When calculating the final attention score of a position, the score of the position in the last iteration may be referred to. If the value of the current position (lat, lon) in the last cycle is valid (the mask value $m^{i-1}_{lat,lon}$ is 1), the current attention score will be calculated as the weighted sum of the current iteration and the previous iteration attention score, respectively. The calculation of the weighted sum of the scores of adjacent cycles is done to allow the feature maps calculated by different iterations to be better fused, where λ is a learnable parameter:

$$score^i_{lat,lon,lat',lon'} = \lambda score^i_{lat,lon,lat',lon'} + (1 - \lambda)score^{i-1}_{lat,lon,lat',lon'} \tag{4}$$

The current attention scores are multiplied with the feature maps to obtain new feature maps.

3. Data and Experiments

In order to evaluate the reconstruction effects of different methods and carry out spatial data inpainting in the real Antarctic region with a large missing rate, this paper designs ideal experiments and real experiments, respectively. Among them, the ideal experiment compares the performance of the new deep learning model and other traditional methods under different missing rates to verify the advancement of our proposed method beyond

the interpolation methods and its robustness in different scenarios. Using the Antarctic station data as input in the real experiment, the spatial structure of the entire Antarctic is reconstructed by the model trained under the high missing rate of the ideal experiment. The workflows of ideal and real experiments for the RFR and PI-RFR models are shown in Figure 5. The following sections describe the process and data details used in the two experiments.
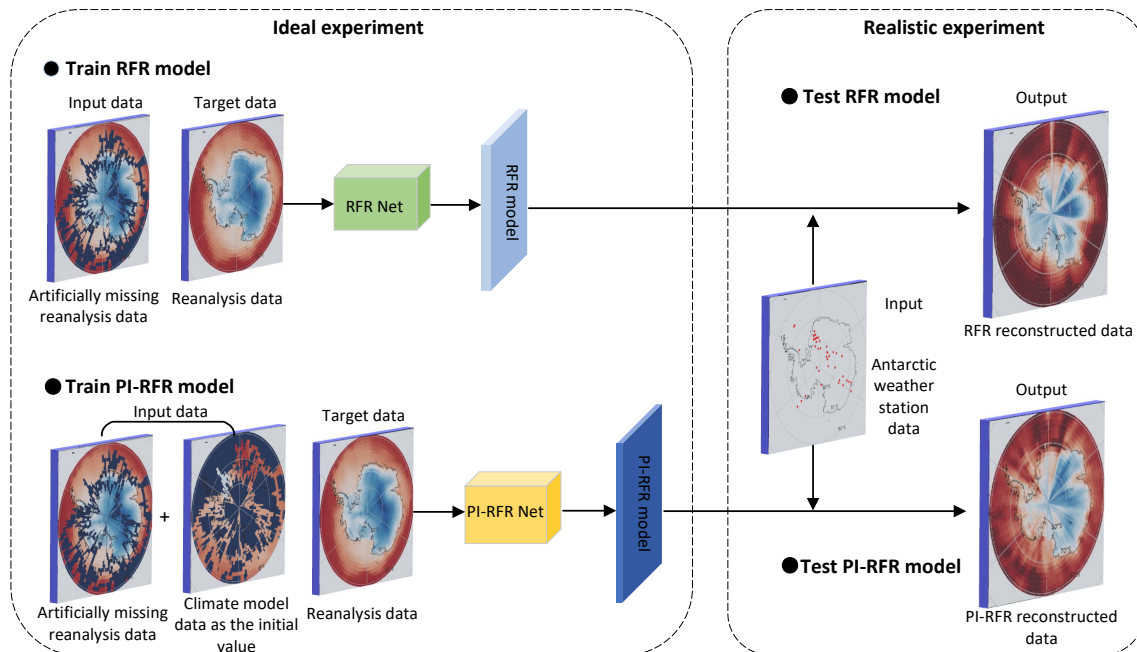

Figure 5. Overall framework with the RFR and PI-RFR methods for reconstructing the climatological spatial missing information. The left part is the ideal experiments, training the RFR and PI-RFR methods separately. The right part is the realistic experiments, with spatial missing reconstruction from the Antarctic site data.

3.1. Ideal Experiments

The ideal experiment evaluates the performance of deep learning-based RFR, PI-RFR and traditional Kriging and IDW methods under three different missing rates of 30%, 50% and 70%.

The artificially missing reanalysis data are fed into the reconstruction model to restore the spatial pattern as far as possible. The reanalysis data are selected from the monthly average of the ERA-Interim derived from European Centre for Medium-Range Weather Forecasts (ECMWF) with the period of 40 years from 1979 to 2018 and 1.5° latitude × 1.5° longitude resolution. The region of the data is the Antarctic (longitude (0°–360°), latitude (60° S–90° S)) [29–31]. Data of the first 30 years are used for model training, and the data of the last 10 years are used to test and validate the performance. The variables such as skin temperature (K, Kelvins) and surface wind speed (m/s, meter per second) are selected for experiments.

To ensure the randomness of the experiments and the credibility of the results, the mask is generated in a completely random manner. The initial mask has the same spatial size as the full reanalysis data. The artificial masking input data is obtained by multiplying the mask with the full reanalysis data. As shown in Figure 6, the update method of the mask randomly selects a beginning position and starts random movement from this position. There are four moving directions of up, down, left and right, and the mask value at the passing position is set to missing in probability. The missing rate is controlled by computing the sum of all elements in the mask divided by the number of mask elements. Figure 6 shows the mask generated by the above method.
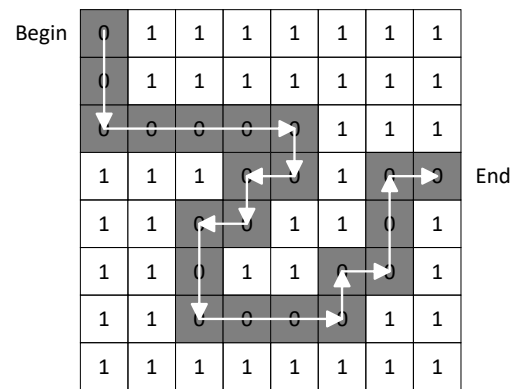
Figure 6. A randomly generated mask, where "0" represents mask area and "1" represents reserved data.

The default RFR sets the initial data of all mask parts to 0. However, the initialization with the low value is far from the actual missing variable values, and the deep learning model needs to be trained too many times to fill in the data similar to the real values. Instead, the global average of the 1979–2008 reanalysis data in the Antarctic region is used as the initial values to improve the convergence of the RFR model in this study. In the PI-RFR method, in order to further reflect the obvious changes in the spatial characteristics of the Antarctic region, the climatology properties from the pre-industry simulation of the Energy Exascale Earth System Model (E3SM [32]) are used as the spatial-aware initial values.

3.2. Realistic Experiments

The real experiment approximately maps real station data into the grid-by-grid spatial pattern in the Antarctic with 1.5° latitude × 1.5° longitude grid due to the resolution of training dataset. However, the resolution can be increased by a higher resolution of the training dataset or by applying super-resolution methods. After this map processing, the obtained spatial data forms are similar to the input data of the training set, and the areas for which there are no site data are regarded as missing parts. It has been calculated that the missing rate in this Antarctic region exceeds 90%. The reconstruction model of the Antarctic spatial data with missing regions obtained from this mapping is the PI-RFR model with the best performance of a 70% missing rate in the ideal experiment, because this missing rate is the closest to the real scene. In addition, the missing spatial grid point data is input into this reconstruction model to recover the data of the Antarctic region in the entire grid space. Details about site data and mapping techniques are as follows.

The Antarctic weather station data are derived from the Antarctic Meteorological Research Center (AMRC) and Automatic Weather Station (AWS) program from 1980 to 2019. The method of mapping site data to grid points uses site data as the data of the nearest grid point to each site. The mapping process is shown in Figure 7.


Antarctic weather station data
(high resolution)

Mapping

Antarctic weather station data
(low resolution)
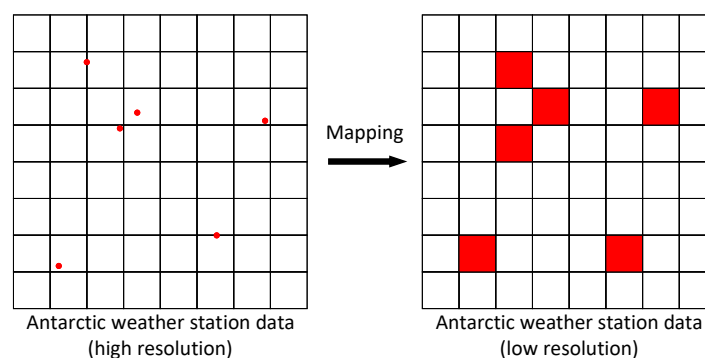
Figure 7. Map Antarctic station data into the grid-by-grid spatial pattern with 1.5° latitude × 1.5° longitude grid.

4. Results

4.1. Comparison of the RFR with Traditional Methods

The performance of each method is evaluated according to the grid-to-grid RMSE. Figures 8 and 9 and Table 1 show the performance of traditional methods, including Kriging and IDW, and the RFR method under 30%, 50% and 70% missing rates. Taking skin temperature as an example, the RFR method performs the best with 1.72, 2.46 and 3.07 in terms of RMSE at the corresponding missing rates. In comparison, the errors of RFR reconstruction temperature are about 63% and 71% smaller than Kriging and IDW averaging at different missing rates. It is worth noting that the performances of Kriging and IDW drastically decrease when the missing rate increases. For instance, the errors of Kriging and IDW are 309% and 391% errors of RFR at a 70% missing rate. In the Antarctic area, due to the limited number of observation sites, the missing ratio is over 90%. These traditional methods are incompetent at reconstructing the spatial pattern. In contrast, the deep learning method still keeps good accuracy even though the missing ratio is 70%; for example, the accuracy of RFR is 68% and 74% higher than Kriging and IDW at temperature.
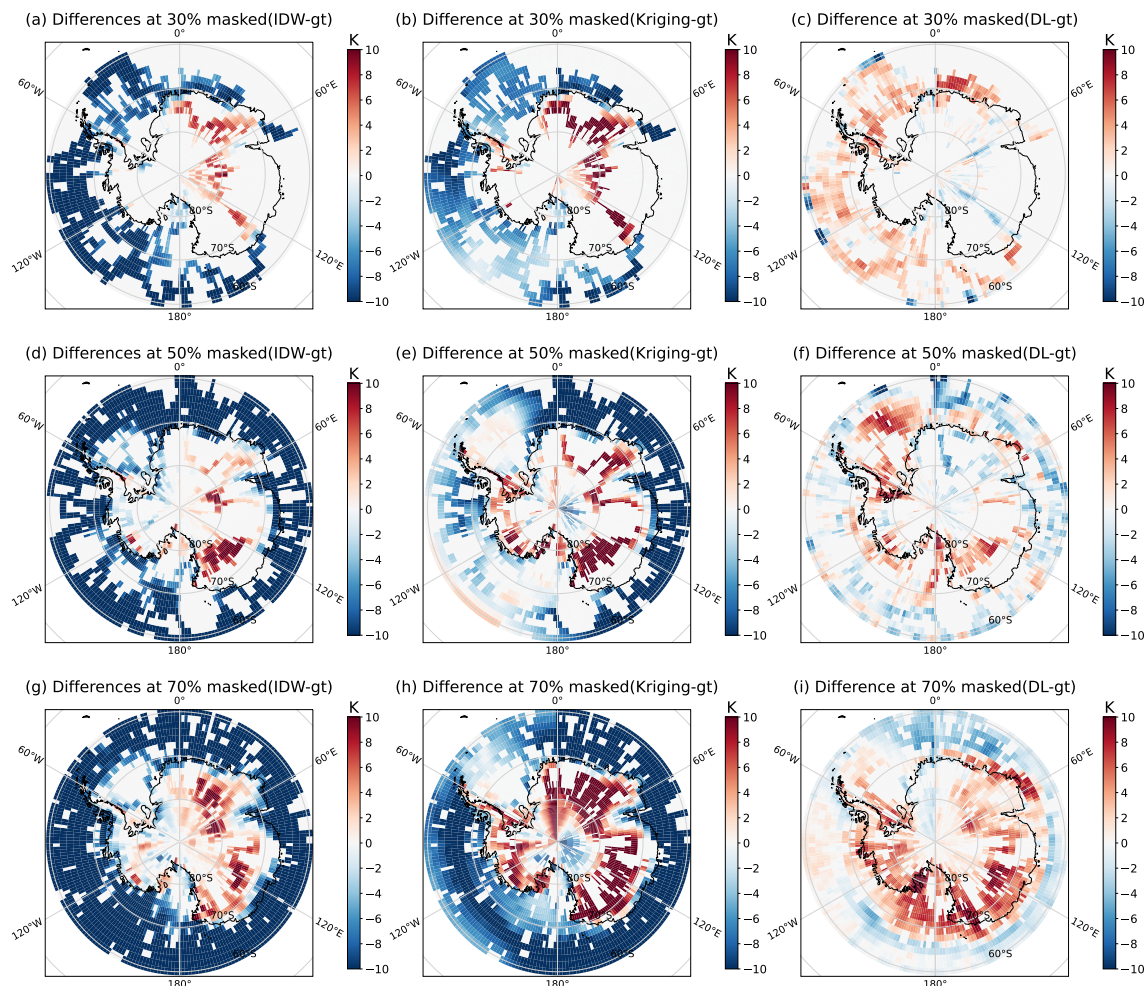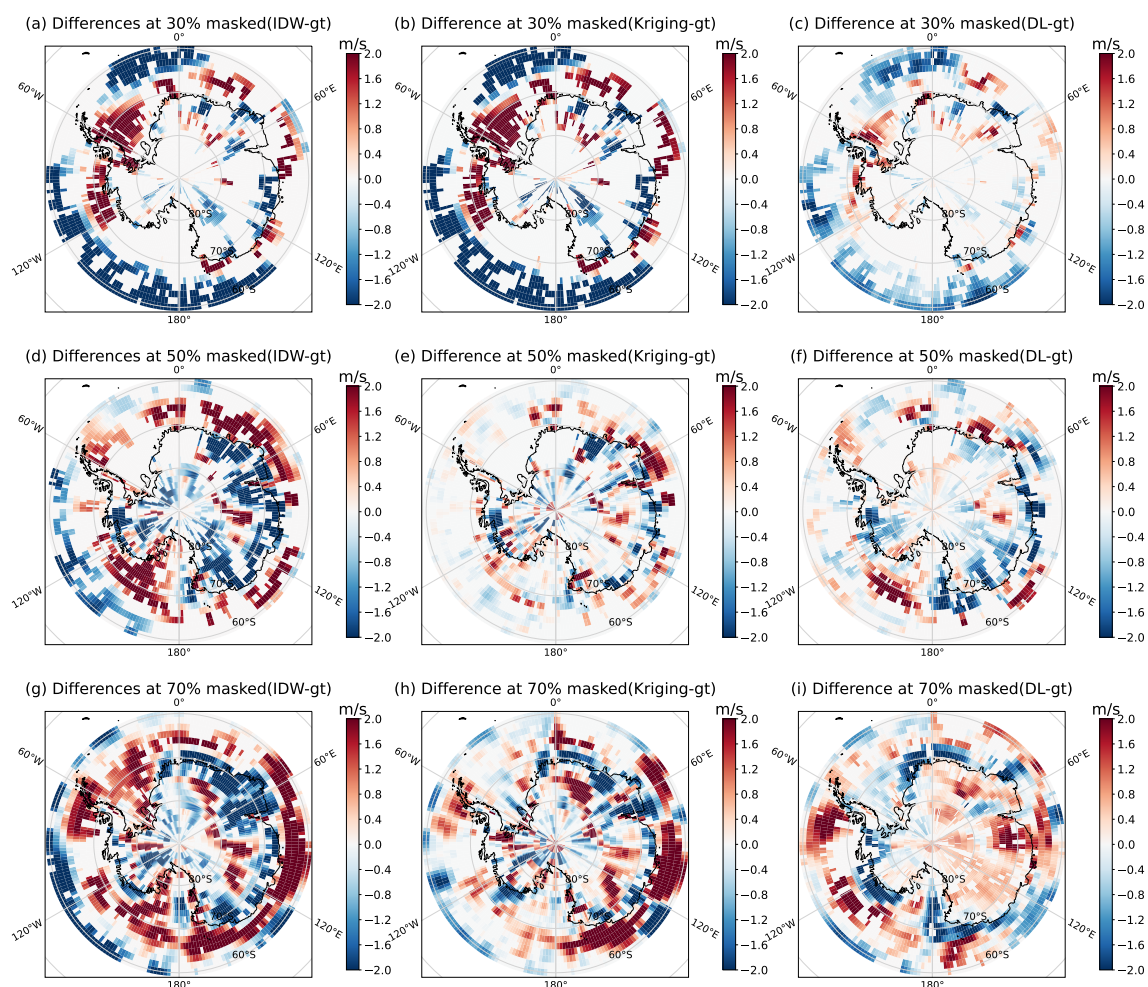

Figure 8. Differences between the reconstructed skin temperature of IDW (**a**,**d**,**g**), Kriging (**b**,**e**,**h**) and RFR (**c**,**f**,**i**) and the ground truth with 30%, 50% and 70% missing rates, respectively.

Table 1. Reconstruction performance (RMSE) of skin temperature and wind speed by IDW, Kriging and RFR methods under different missing rates.

	Skin Temperature			Wind Speed		
Missing Rate	**IDW**	**Kriging**	**RFR**	**IDW**	**Kriging**	**RFR**
30%	5.08	3.96	1.72	1.06	0.83	0.48
50%	8.94	6.77	2.46	1.25	0.84	0.62
70%	12.01	9.48	3.07	1.46	1.28	0.79

Figure 8 indicates that the spatial distribution of skin temperature is characterized by the warmer regions at the outer ocean and colder at the inner land. The reconstructions at the ocean achieved by Kriging and IDW methods are cooler than the true values, while the results on land are warmer. The reason could be that these interpolation methods impute the missing values by approximating the average of the surrounding data. When the missing rate is large or the spatial pattern has a continuum missing, the surrounding data are inadequate to achieve good accuracy. On the other hand, the deep learning method learns the global spatial dependence, which employs more global valid data to impute the missing information instead of the limited surrounding data. Similar to the skin temperature, the deep learning method significantly outperforms the Kriging and IDW for surface wind speed, with improvements of 36% and 50%, respectively, on average for the three missing rates. (Table 1 and Figure 9). Especially at a high missing ratio (70%), the RFR reduces the error by 38% and 46% for Kriging and IDW, respectively.

Figure 9. Differences between the reconstructed wind speed of IDW (**a**,**d**,**g**), Kriging (**b**,**e**,**h**) and RFR (**c**,**f**,**i**) and the ground truth with 30%, 50% and 70% missing rates, respectively.

4.2. Comparison of the PI-RFR with RFR

For RFR, the temperature reconstruction of the Antarctic land presents a significant warm bias, while that of the Antarctic ocean shows a cold bias (Figure 10a–c. The reason is that the masked areas are initialized by the unifying spatial average in RFR. As a result, there is no spatial structure in the masked areas, leading to slow convergence and low accuracy. In this study, we propose the new method, PI-RFR, which is improved by initializing the missing value with the spatial pattern from climate models. Table 2 indicates that PI-RFR outperforms RFR at any missing rates for temperature and surface wind speed, and on average, PI-RFR improves accuracy by 10% and 9%, respectively. In particular, the new method improves by 23% at a 50% missing rate for temperature data and 13% at a 30% missing rate for wind speed data. Figure 11d,e shows that the reconstruction of PI-RFR is colder than RFR for Antarctic land and warmer than RFR for the Antarctic ocean. This implies that the PI-RFR can reduce the error both for the Antarctic land and ocean. In the same way as wind speed, shown in Figure 11, PI-RFR reduces the warm/cold biases in comparison with RFR.
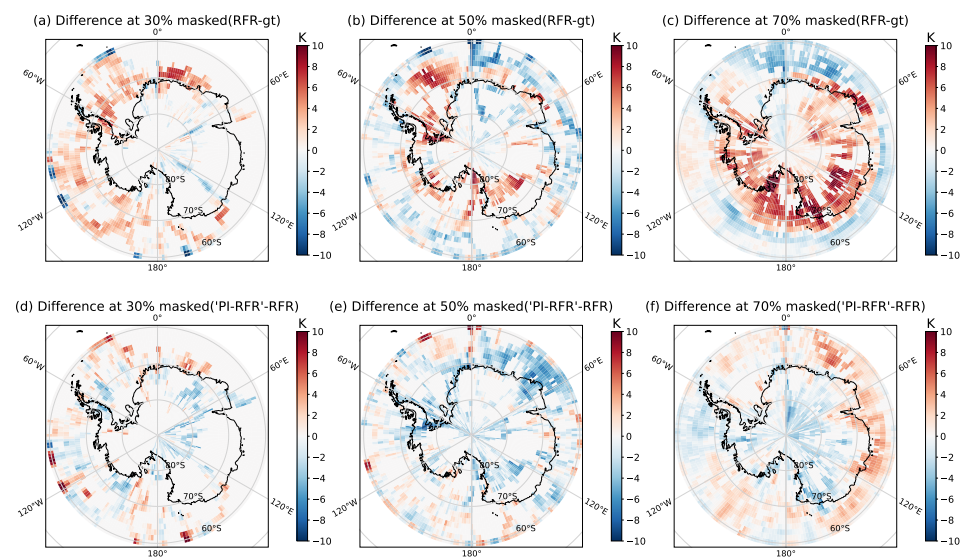
Figure 10. Differences between the reconstructed skin temperature of RFR (**a**–**c**) and PI-RFR (**d**–**f**) and the ground truth with 30%, 50% and 70% missing rates, respectively.
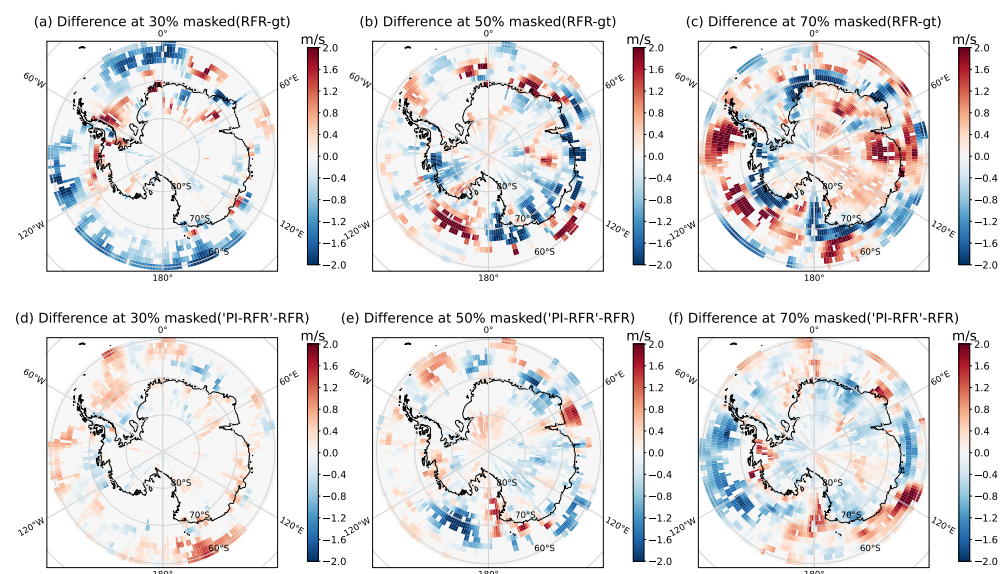

Figure 11. Differences between the reconstructed wind speed of RFR (**a**–**c**) and PI-RFR (**d**–**f**) and the ground truth with 30%, 50% and 70% missing rates, respectively.

Table 2. Reconstruction performance (RMSE) of skin temperature and wind speed by RFR and PI-RFR methods under different missing rates.

Missing Rate	Skin Temperature		Wind Speed	
	RFR	PI-RFR	RFR	PI-RFR
30%	1.72	1.68	0.48	0.42
50%	2.46	1.89	0.62	0.59
70%	3.07	2.94	0.79	0.71

4.3. Reconstruction from Antarctic Site Data

In the realistic case, we apply the deep learning techniques in the reconstruction from Antarctic site observation. This brings huge challenges due to the few observation sites, as shown in Figure 12b. The spatial missing ratio exceeds 90% in the Antarctic region. The deep learning models RFR/PI-RFR for reconstruction from sites use the trained models with a 70% missing ratio.
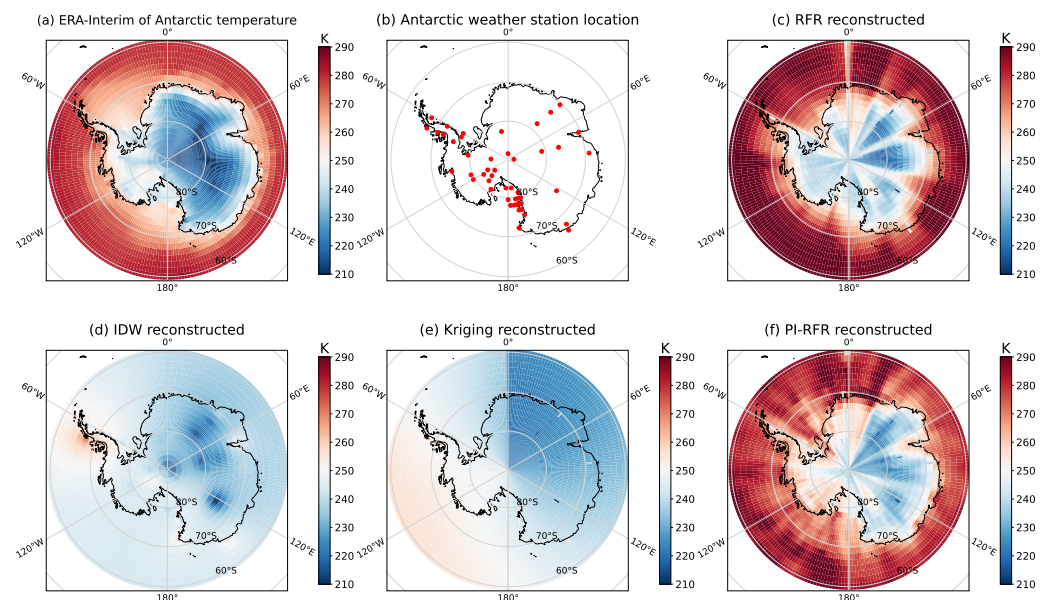
Figure 12. Reconstruction results of Antarctic station skin temperature data by IDW, Kriging, RFR and PI-RFR methods. The subfigure (**a**) is the data of the reanalysis data, and we take it as the ground truth. The subfigure (**b**) represents the location of the Antarctic weather station. The subfigures (**c–f**) show the reconstruction results of RFR, IDW, Kriging and PI-RFR, respectively.

Figures 12 and 13 compare the spatial reconstructions from site observation among IDW, Kriging and deep learning RFR/PI-RFR in terms of skin temperature and surface wind speed. The Kriging method cannot recover any spatial signal compared with the ERA-Interim reanalysis data. The IDW only demonstrates some features around the observation sites and loses most spatial structures. Both methods only depend on the neighboring points. The bad performance is caused by the small number of observation sites. In contrast, both deep learning methods, i.e., RFR/PI-RFR, present the characteristics of a warm ocean and cold land for skin temperature. In particular, for wind speed, the deep learning methods display a similar structure to ERA-Interim. Figure 14 shows that the reconstruction of PI-RFR has good correlation with the ERA-Interim data and also displays the results of significance testing. It also shows that the PI-RFR method achieves the smallest RMSE in reconstructing Antarctic station data into grid data.
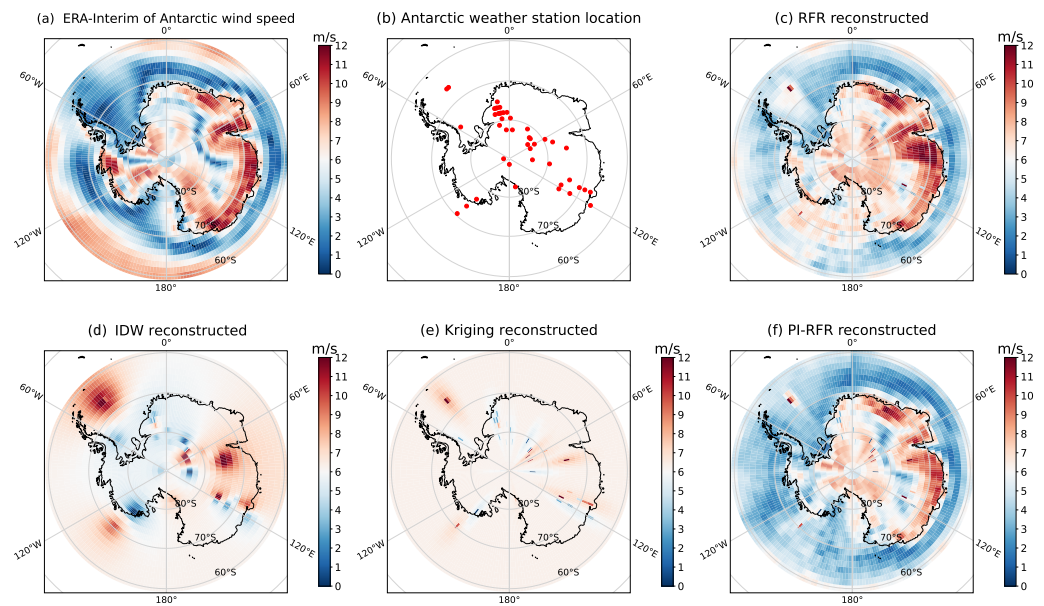
Figure 13. Reconstruction results of Antarctic station wind speed data by IDW, Kriging, RFR and PI-RFR methods. The subfigure (**a**) is the data of the reanalysis data, and we take it as the ground truth. The subfigure (**b**) represents the location of the Antarctic weather station. The subfigures (**c**–**f**) show the reconstruction results of RFR, IDW, Kriging and PI-RFR, respectively.
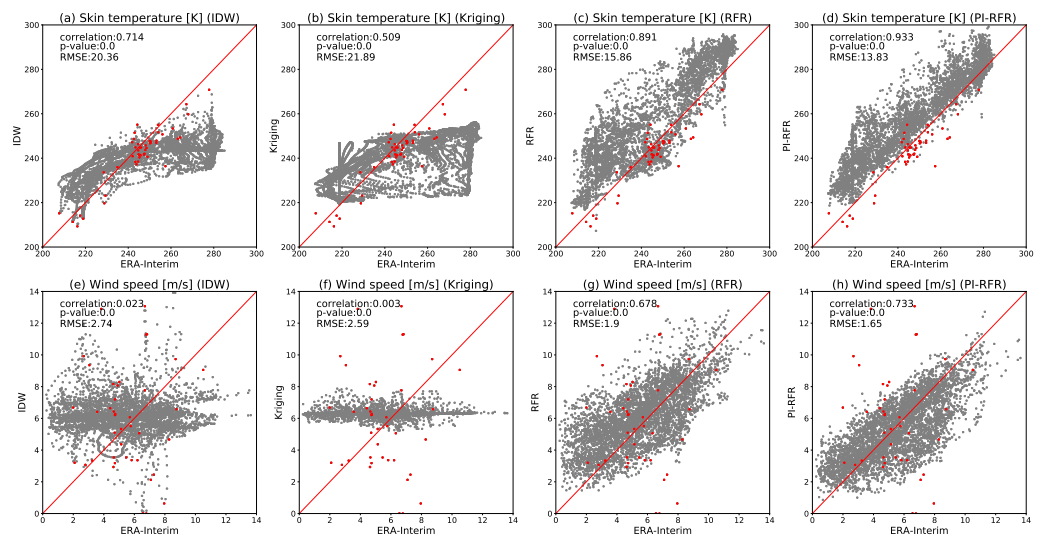
Figure 14. Correlation between the reconstruction results of IDW (**a**,**e**), Kriging (**b**,**f**), RFR (**c**,**g**) and PI-RFR (**d**,**h**) and ERA-Interim reanalysis data for skin temperature and wind speed. The red points represent the values of observation sites. Gray points represent reconstructed values.

5. Conclusions and Future Work

In this paper, we propose a physics-informed deep learning method, called PI-RFR, for meteorological missing value reconstruction, based on an advanced image inpainting algorithm, RFR. To verify the effectiveness and practicability of the deep learning reconstruction model, ideal experiments and realistic experiments are designed, respectively. In the ideal experiment, the PI-RFR method is evaluated based on the missing reanalysis data, using the physical properties of the numerical model data together with the powerful feature reasoning method of RFR to restore the missing parts. Then, the trained model is also applied to the actual scenarios of recovering the spatial structure of the entire Antarctic from the realistic Antarctic site data in the realistic experiment, realizing the function of restoring from sparse sites to uniform data.

The ideal experiment results show that the reconstruction for the temperature and wind speed in the Antarctic region of the PI-RFR method are significantly better than Kriging and IDW methods at the three missing rates of 30%, 50% and 70%. The error of RFR is about 63% and 71% smaller than Kriging and IDW for skin temperature and 36% and 50% smaller for wind speed. Furthermore, the PI-RFR also outperforms RFR at any missing rates for temperature and surface wind speed, which improves by 23% at 50% missing rates for temperature data and 13% at 30% missing rates for wind speed data.

Traditional methods such as Kriging and IDW are weak at reconstructing the spatial pattern, especially with high missing rates. By comparison, the deep learning methods still maintain good accuracy when the missing rate is as high as 70%. The reason for the difference between the two kinds of methods is that the traditional method estimates the missing value by approximating the average value of the surrounding data. When the missing rate is large or the spatial pattern has continuous missing data, the surrounding data are not enough to achieve good accuracy. On the other hand, deep learning methods learn global spatial dependencies, which use more globally valid data to compute missing information than limited surrounding data.

In the realistic experiments, the reconstruction results of Kriging and IDW methods only contain some features around the sites, and most areas lose almost all spatial structure information through smooth transition. Both deep learning algorithms, i.e., RFR/PI-RFR, can make more comprehensive use of global data and are more effective in the case of large-scale missing sites; for example, the missing rate in this experiment is over 90%.

To further improve PI-RFR, our future work will focus on the design of multivariate neural networks, considering the influence of variables such as solar radiation and air humidity. These variables can be fed into the neural network together as covariates for training. The uncertainty will be calculated by Gaussian process regression and added to the calculation of the attention score to further improve the learning ability of the neural network.

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

Abbreviations

The following abbreviations are used in this manuscript:

IDW	Inverse Distance Weighted
RFR	Recurrent Feature Reasoning
PI-RFR	Physics-Informed Recurrent Feature Reasoning
ECMWF	European Centre for Medium-Range Weather Forecasts
AI	Artificial Intelligence
GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
SST	Sea Surface Temperature
E3SM	Energy Exascale Earth System Model
AMRC	Antarctic Meteorological Research Center
AWS	Automatic Weather Station
VGG	Visual Geometry Group
MDPI	Multidisciplinary Digital Publishing Institute

References

1. Rintoul, S.R.; Chown, S.L.; DeConto, R.M.; England, M.H.; Fricker, H.A.; Masson-Delmotte, V.; Naish, T.R.; Siegert, M.J.; Xavier, J.C. Choosing the future of Antarctica. *Nature* **2018**, *558*, 233–241. [CrossRef]
2. Liu, J.; Bromwich, D.; Chen, D.; Cordero, R.; Jung, T.; Raphael, M.; Turner, J.; Yang, Q. Preface to the Special Issue on Antarctic Meteorology and Climate: Past, Present and Future. *Adv. Atmos. Sci.* **2020**, *37*, 421–422. [CrossRef]
3. Parkinson, C.L. A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 14414–14423. [CrossRef] [PubMed]
4. Edwards, T.L.; Nowicki, S.; Marzeion, B.; Hock, R.; Goelzer, H.; Seroussi, H.; Jourdain, N.C.; Slater, D.A.; Turner, F.E.; Smith, C.J.; et al. Projected land ice contributions to twenty-first-century sea level rise. *Nature* **2021**, *593*, 74–82. [CrossRef] [PubMed]
5. Lazzara, M.A.; Weidner, G.A.; Keller, L.M.; Thom, J.E.; Cassano, J.J. Antarctic automatic weather station program: 30 years of polar observation. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1519–1537. [CrossRef]
6. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
7. Rodwell, M.; Palmer, T. Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.* **2007**, *133*, 129–146. [CrossRef]
8. Matheron, G. Principles of geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266. [CrossRef]
9. Shepard, D. *A Two-Dimensional Interpolation Function for Irregularly-Spaced Data*; Association for Computing Machinery: New York, NY, USA, 1968.
10. Hock, R.; Jensen, H. Application of kriging interpolation for glacier mass balance computations. *Geogr. Ann. Ser. Phys. Geogr.* **1999**, *81*, 611–619. [CrossRef]
11. Mair, A.; Fares, A. Throughfall characteristics in three non-native Hawaiian forest stands. *Agric. For. Meteorol.* **2010**, *150*, 1453–1466. [CrossRef]
12. Dhamodaran, S.; Lakshmi, M. Comparative analysis of spatial interpolation with climatic changes using inverse distance method. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 6725–6734. [CrossRef]
13. Bronowicka-Mielniczuk, U.; Mielniczuk, J.; Obroślak, R.; Przystupa, W. A comparison of some interpolation techniques for determining spatial distribution of nitrogen compounds in groundwater. *Int. J. Environ. Res.* **2019**, *13*, 679–687. [CrossRef]
14. Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv* **2014**, arXiv:1408.5882.
15. Zhang, T.; Lin, W.; Lin, Y.; Zhang, M.; Yu, H.; Cao, K.; Xue, W. Prediction of tropical cyclone genesis from mesoscale convective systems using machine learning. *Weather. Forecast.* **2019**, *34*, 1035–1049. [CrossRef]
16. Zhang, T.; Lin, W.; Vogelmann, A.M.; Zhang, M.; Xie, S.; Qin, Y.; Golaz, J.C. Improving convection trigger functions in deep convective parameterization schemes using machine learning. *J. Adv. Model. Earth Syst.* **2021**, *13*, e2020MS002365. [CrossRef]
17. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 85–100.
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 139–144.
19. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [CrossRef]

20. Dong, J.; Yin, R.; Sun, X.; Li, Q.; Yang, Y.; Qin, X. Inpainting of remote sensing SST images with deep convolutional generative adversarial network. *IEEE Geosci. Remote. Sens. Lett.* **2018**, *16*, 173–177. [CrossRef]

21. Shibata, S.; Iiyama, M.; Hashimoto, A.; Minoh, M. Restoration of sea surface temperature satellite images using a partially occluded training set. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2771–2776.

22. Leinonen, J.; Guillaume, A.; Yuan, T. Reconstruction of cloud vertical structure with a generative adversarial network. *Geophys. Res. Lett.* **2019**, *46*, 7035–7044. [CrossRef]

23. Dewi, C.; Chen, R.C.; Liu, Y.T.; Yu, H. Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation. *Appl. Sci.* **2021**, *11*, 2913. [CrossRef]

24. Li, J.; Wang, N.; Zhang, L.; Du, B.; Tao, D. Recurrent feature reasoning for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7760–7768.

25. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative image inpainting with contextual attention. In Proceedings of the PIEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5505–5514.

26. Kadow, C.; Hall, D.M.; Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **2020**, *13*, 408–413. [CrossRef]

27. Monteleoni, C.; Schmidt, G.A.; McQuade, S. Climate informatics: accelerating discovering in climate science with machine learning. *Comput. Sci. Eng.* **2013**, *15*, 32–40. [CrossRef]

28. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]

29. Simmons, A.; Willett, K.; Jones, P.; Thorne, P.; Dee, D. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res. Atmos.* **2010**, *115*, D01110. [CrossRef]

30. Uppala, S.; Dee, D.; Kobayashi, S.; Berrisford, P.; Simmons, A. Towards a climate data assimilation system: Status update of ERA-Interim. *ECMWF Newsl.* **2008**, *115*, 12–18.

31. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.; Balsamo, G.; Bauer, d.P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]

32. Golaz, J.C.; Van Roekel, L.P.; Zheng, X.; Roberts, A.F.; Wolfe, J.D.; Lin, W.; Bradley, A.M.; Tang, Q.; Maltrud, M.E.; Forsyth, R.M.; et al. The DOE E3SM Model Version 2: Overview of the physical model and initial model evaluation. *J. Adv. Model. Earth Syst.* **2022**, *14*, e2022MS003156. [CrossRef]