atmosphere

Article

A Transfer Learning–CNN Framework for Marine Atmospheric Pollutant Inversion Using Multi-Source Data Fusion

Xiaoling Li [1,2], Xiaoyu Liu [1], Xiaohuan Liu [3], Zhengyang Zhu [1], Yunhui Xiong [1], Jingfei Hu [1] and Xiang Gong [1,2,*]

[1] School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China; 2023091009@mails.qust.edu.cn (X.L.); 2024091013@mails.qust.edu.cn (X.L.); 2406030124@mails.qust.edu.cn (Z.Z.); 2206030124@mails.qust.edu.cn (Y.X.); 2025092002@mails.qust.edu.cn (J.H.)

[2] Shandong Engineering Research Center for Marine Scenarized Application of Artificial Intelligence, Qingdao University of Science and Technology, Qingdao 266061, China

[3] Key Laboratory of Marine Environment and Ecology, Ministry of Education, Ocean University of China, Qingdao 266100, China; liuxh1983@ouc.edu.cn

[*] Correspondence: gongxiang@qust.edu.cn

Abstract

The concentration characteristics of SO_2, NO_2, O_3, and CO in the marine atmosphere are of great significance for understanding air–sea interactions and regional atmospheric chemical processes. However, due to the challenging conditions of marine monitoring, long-term continuous observational data remain scarce. To address this gap, this study proposes a Transfer Learning–Convolutional Neural Network (TL-CNN) model that integrates ERA5 meteorological data, EAC4 atmospheric composition reanalysis data, and ground-based observations through multi-source data fusion. During data preprocessing, the Data Interpolating Empirical Orthogonal Function (DINEOF), inverse distance weighting (IDW) spatial interpolation, and Gaussian filtering methods were employed to improve data continuity and consistency. Using ERA5 meteorological variables as inputs and EAC4 pollutant concentrations as training targets, a CNN-based inversion framework was constructed. Results show that the CNN model achieved an average coefficient of determination (R^2) exceeding 0.80 on the pretraining test set, significantly outperforming random forest and deep neural networks, particularly in reproducing nearshore gradients and regional spatial distributions. After incorporating transfer learning and fine-tuning with station observations, the model inversion results reached an average R^2 of 0.72 against site measurements, effectively correcting systematic biases in the reanalysis data. Among the pollutants, the inversion of SO_2 performed relatively poorly, mainly because emission reduction trends from anthropogenic sources were not sufficiently represented in the reanalysis dataset. Overall, the TL-CNN model provides more accurate pollutant concentration fields for offshore regions with limited observations, offering strong support for marine atmospheric environment studies and assessments of marine ecological effects. It also demonstrates the potential of combining deep learning and transfer learning in atmospheric chemistry research.

Keywords: marine atmosphere; ERA5/EAC4 reanalysis; Convolutional Neural Network; transfer learning

check for updates

Received: 31 August 2025
Revised: 4 October 2025
Accepted: 6 October 2025
Published: 8 October 2025

Citation: Li, X.; Liu, X.; Liu, X.; Zhu, Z.; Xiong, Y.; Hu, J.; Gong, X. A Transfer Learning–CNN Framework for Marine Atmospheric Pollutant Inversion Using Multi-Source Data Fusion. *Atmosphere* **2025**, *16*, 1168. https://doi.org/10.3390/atmos16101168

1. Introduction

The concentration variations of SO_2, NO_2, O_3, and CO in the marine atmosphere are of great significance for understanding air–sea interactions and atmospheric chemical processes [1]. These gases are not only typical atmospheric pollutants but also key factors in regional atmospheric chemical cycles [2,3]. For instance, SO_2 and NO_2 can be converted into sulfate and nitrate aerosols, which contribute to acid deposition, thereby influencing ocean acidification and nutrient inputs in coastal waters. In this study, we focus specifically on NO_2 from the NO_x family. Compared with nitric oxide (NO), which has a short atmospheric lifetime and limited observational coverage, NO_2 is more stable and extensively monitored, making it a reliable metric for model training and long-term environmental assessment [4]. O_3, as a strong oxidant, regulates photochemical processes at the air–sea interface and directly affects phytoplankton photosynthesis as well as regional atmospheric oxidizing capacity. CO, by reacting with hydroxyl radicals (OH), modulates the atmospheric oxidation capacity and indirectly affects the lifetime of greenhouse gases. These processes interact with one another and jointly influence regional climate and marine ecosystems. Hence, revealing the spatiotemporal variability of these gaseous pollutants is of significant scientific and practical importance.

Despite their environmental and climatic relevance, observational data of these pollutants over offshore regions remain highly limited. This scarcity primarily arises from the challenges of marine monitoring: long-term, stable observational sites are sparse, and the deployment and maintenance of monitoring instruments under harsh marine conditions are difficult. As a result, existing studies largely rely on limited observations, which makes it difficult to comprehensively characterize the spatiotemporal variability of marine atmospheric pollutants. This data gap significantly constrains in-depth understanding of the transport, transformation, and deposition of pollutants in marine environments.

To compensate for the lack of observations, regional atmospheric chemical transport models have been widely applied to simulate and predict offshore air pollution. Models such as WRF-Chem and CMAQ can simultaneously account for meteorological processes, atmospheric chemical reactions, and pollutant transport mechanisms, thereby generating physically constrained, three-dimensional, consistent datasets [5,6]. However, their accuracy strongly depends on the quality of input data, parameterizations, and chemical mechanisms, which introduces considerable uncertainty. Moreover, their high computational demand makes them less feasible for large-scale applications when computing resources are limited.

Since the 1990s, artificial intelligence (AI) methods have been gradually applied to air pollution forecasting [7–9]. With the rapid advancement of deep learning in recent years, the capacity of models for estimating and predicting air pollutants has improved markedly. Deep learning has been demonstrated to outperform traditional methods in reproducing spatiotemporal patterns, short-term forecasts, and long-term trend analyses across multiple regions and cities, and it is increasingly recognized as a powerful tool in environmental monitoring and management. For example, Li et al. [10] used an LSTM model to predict hourly $PM_{2.5}$ concentrations in Beijing, while Reddy et al. [11] extended the forecast horizon to 5–10 h using time series of pollutants and meteorological variables. Given that Convolutional Neural Network (CNN) can effectively extract temporal features from air quality data, researchers have developed hybrid CNN–LSTM models that significantly improve prediction accuracy compared to standalone CNN or LSTM approaches [12]. In addition, graph neural networks (GNNs) have been widely applied to prediction problems with strong spatial dependencies. For instance, Qi et al. [13] proposed a GC-LSTM model combining graph convolution and LSTM to capture spatiotemporal variations of $PM_{2.5}$ concentrations, while Iskandaryan et al. [14] developed a model incorporating attention mechanisms, gated recurrent units, and graph convolution to predict hourly NO_2 concentrations in Madrid. Nevertheless, most existing studies rely on single-source data and lack integration of physical prior knowledge. Moreover, systematic evaluations of model generalization and extrapolation ability across different surface conditions remain scarce.

Against this background, the present study integrates multi-source datasets—including ground-based observations and reanalysis products—to develop a Transfer Learning–Convolutional Neural Network (TL-CNN) model for estimating the concentrations of major gaseous pollutants (SO_2, NO_2, O_3, and CO) over the Bohai and Yellow Seas (35.25° N–38° N, 119° E–123° E). Deep learning has already been shown to be effective in spatiotemporal simulation, short-term forecasting, and long-term trend analysis of air pollutants across cities and regions, making it an increasingly important tool in environmental monitoring and management. However, its application to estimating marine atmospheric pollutants remains constrained by sparse observational data, limiting comprehensive assessments of their variability. Transfer learning (TL), an important approach in machine learning and deep learning, is particularly valuable in such contexts: it leverages knowledge learned from a source task to improve performance on a target task, especially when the target dataset is small or difficult to obtain [15,16]. By combining the spatial feature extraction capability of CNNs with the adaptability of transfer learning, this study aims to overcome the limitations of sparse offshore observations and provide a novel application case for atmospheric chemistry and marine environment research.

2. Materials and Methods

2.1. Data Resource

The study area is defined as 35.25° N–42.75° N and 116.25° E–123.75° E (Figure 1), covering the Bohai Sea and the Yellow Sea. The Bohai Sea, China's largest inland sea, is enclosed by the Shandong and Liaodong Peninsulas, and its atmospheric circulation and moisture transport exert significant influence on regional meteorology and pollutant dispersion. The climate of this region is controlled by the East Asian monsoon, characterized by cold and dry winters, warm and humid summers, and distinct seasonal variations. These features strongly determine the transport and diffusion of atmospheric pollutants. Therefore, when delineating the study area, both the meteorological conditions and the pollution characteristics of the Bohai Sea and the northeastern Liaodong Peninsula were comprehensively considered.

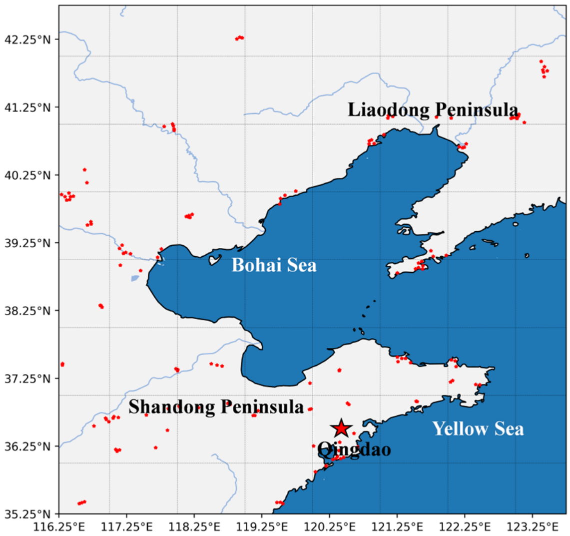

Figure 1. Study area (35.25° N–42.75° N, 116.25° E–123.75° E) showing the locations of 150 air monitoring sites (red dots). The red star denotes the Qingdao station, which was used for comparing observational and reanalysis data and for validating the model results.

The primary physical parameters used in this study include reanalysis datasets (meteorological variables and major pollutant concentrations) and observational data of key atmospheric gaseous pollutants from monitoring stations. The reanalysis data were applied for pretraining the deep learning network model, while the station data were used for transfer learning. The data sources are described as follows:

(1) Ground-based air quality observation data: These data were obtained from the national urban air quality real-time platform released by the China National Environmental Monitoring Center (CNEMC). Relying on a widely distributed monitoring network, this platform provides stable and continuous air quality observations. Hourly concentration data of SO_2, NO_2, O_3, and CO from 150 sites within the study area were collected for the period 1 January 2014 to 20 September 2020. These data were mainly used for model transfer learning and independent validation.

(2) ERA5 meteorological reanalysis data: ERA5 is the fifth-generation global atmospheric reanalysis product provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), and it is widely used in climate research and environmental simulations. It has a spatial resolution of 0.25° × 0.25° and a temporal resolution of 1 h, providing a reliable representation of regional meteorological processes. The selected variables in this study include 10 m wind speed (U and V components), 2 m temperature, 2 m dew point temperature, sea surface temperature (SST), mean sea level pressure, surface pressure, precipitation rate, and total precipitation (Table 1). These variables were used as input features for the CNN model to characterize the influence of meteorological conditions on pollutant concentrations.

Table 1. Data variables and descriptions.

Types	Variables	Abbr.	Unit	Resolution
ERA5	10 Meter U Wind Component	U	$M\cdot s^{-1}$	Spatial resolution: 0.25° × 0.25°; Time span: 1 January 2014–31 December 2021; Temporal resolution: 1 h.
	10 Meter V Wind Component	V	$m\cdot s^{-1}$	
	2 Meter Dewpoint Temperature	DPT	°C	
	2 Meter Temperature	Temp	°C	
	Sea Surface Temperature	SST	°C	
	Mean Sea Level Pressure	MSL	Pa	
	Surface Pressure	SP	Pa	
	Mean Total Precipitation Rate	MTP	$kg\cdot m^{-2}\cdot s^{-1}$	
	Total Precipitation	TP	m	
EAC4	NO_2	-	$\mu g\cdot m^{-3}$	Spatial resolution: 0.75° × 0.75°; Time span: 1 January 2014–31 December 2021; Temporal resolution: 3 h.
	SO_2	-	$\mu g\cdot m^{-3}$	
	O_3	-	$\mu g\cdot m^{-3}$	
	CO	-	$\mu g\cdot m^{-3}$	

(3) EAC4 atmospheric composition reanalysis data: The EAC4 dataset is the fourth-generation atmospheric composition reanalysis product provided by the Copernicus Atmosphere Monitoring Service (CAMS). It integrates multiple observational datasets with model simulations and is one of the most widely used atmospheric composition reanalysis products. It has a spatial resolution of 0.75° and a temporal resolution of 3 h, covering the period 2014–2021. In this study, reanalysis concentrations of SO_2, NO_2, O_3, and CO from EAC4 were downloaded and used as the target output data for CNN model pretraining (Table 1).

To evaluate the applicability of the EAC4 dataset, observational data from the Qingdao site were selected for comparison (Figure 2). The results show that EAC4 is able to reasonably reproduce the interannual and seasonal variations of the major pollutants, with its periodic fluctuations largely consistent with observations. However, systematic biases are evident in terms of numerical accuracy, with SO_2 concentrations being significantly overestimated. This indicates that, although EAC4 provides high spatiotemporal resolution, its simulated values cannot be directly used as a reliable basis for pollutant studies in offshore regions. Therefore, in the subsequent analysis, deep learning methods are introduced to correct these biases and obtain pollutant concentration fields that are more consistent with actual observations.

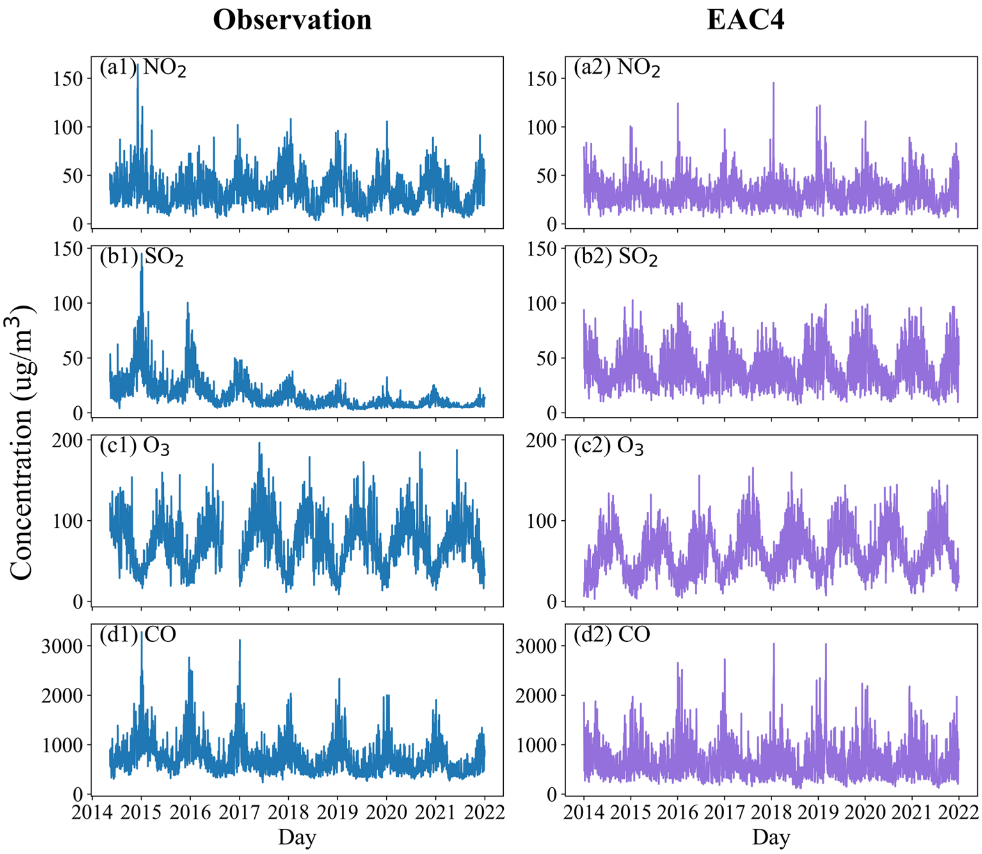

Figure 2. (**a1**–**d1**) Observed concentrations of SO_2, NO_2, O_3, and CO at the Qingdao site, compared with (**a2**–**d2**) corresponding EAC4 reanalysis data. Each pair of subplots illustrates the temporal profiles of the observed and modeled quantities for the same pollutant, highlighting agreements and discrepancies between measurements and reanalysis.

2.2. Data Preprocessing

Before model training, multi-source datasets need to be preprocessed to ensure spatiotemporal consistency and numerical stability. The main steps include missing data reconstruction, multi-source data matching and fusion, and normalization.

2.2.1. Missing Data Reconstruction

Analysis of ERA5 meteorological data revealed that the missing rates were 86.5% for sea surface temperature (SST), 1.5% for dew point temperature (DPT), and 1.8% for both mean sea level pressure (MSLP) and surface pressure (SP). When the missing rate of a given grid point along the temporal dimension exceeded 90%, it was identified as a land area and masked accordingly. Large missing areas of SST mainly corresponded to land; therefore, interpolation was conducted after masking land grids. The interpolation method adopted was DINEOF (Data Interpolating Empirical Orthogonal Function), which extracts the dominant spatiotemporal modes through empirical orthogonal decomposition and optimizes the reconstruction using cross-validation [17]. This approach effectively preserves both spatial continuity and temporal characteristics of the data (Figure 3). Validation results show that the root mean square errors (RMSE) of the reconstructed variables are all much smaller than their standard deviations (Table 2), indicating high reliability of the interpolation results.

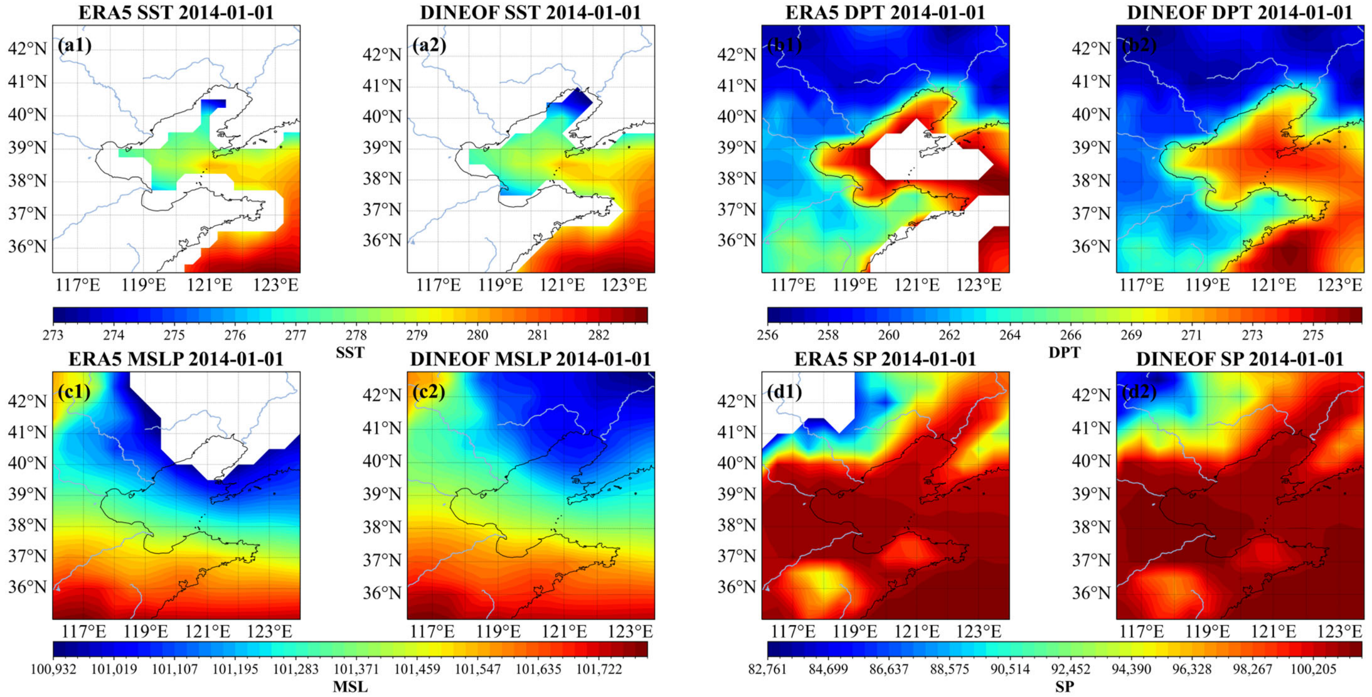

Figure 3. Comparison of ERA5 data for sea surface temperature (SST), dew point temperature (DPT), mean sea level pressure (MSLP), and surface pressure (SP) (**a1**–**d1**) with corresponding DINEOF interpolation results (**a2**–**d2**).

Table 2. Results of the cross-validation dataset using the DINEOF method.

	SST	DPT	MSLP	SP
Convergence level	0.01	0.01	0.01	0.01
Optimal EOF modes	4	4	5	5
Optimal iterations	96	125	20	27
CV error	0.77	0.75	19.44	15.53
Std. dev.	5.16	6.22	301.39	256.98

2.2.2. Data Matching and Fusion

To ensure consistency in the spatiotemporal resolution of the input and output data, the ERA5 and EAC4 datasets were unified. The EAC4 atmospheric composition data, originally at a resolution of 0.75° and 3 h, were first interpolated using a two-dimensional spatial interpolation method to a resolution of 0.25°. Subsequently, these data were temporally and spatially matched with the ERA5 meteorological data (which were downsampled to a 3 h resolution). The resulting ERA5–EAC4 matched dataset has a spatial resolution of 0.25° and a temporal resolution of 3 h.

For the station observation data used in the transfer learning stage, the measurements from 150 sites were first interpolated onto a 0.25° grid using the inverse distance weighting (IDW) method and then unified to a 3 h resolution (Figure 4). IDW estimates values at unsampled locations by assigning higher weights to observations closer to the target point and lower weights to those farther away, ensuring that nearby measurements have greater influence on the interpolated results. Previous studies have shown that IDW generally achieves higher estimation accuracy for pollutant concentrations than linear interpolation and Kriging [18–21], making it a suitable choice for this analysis.

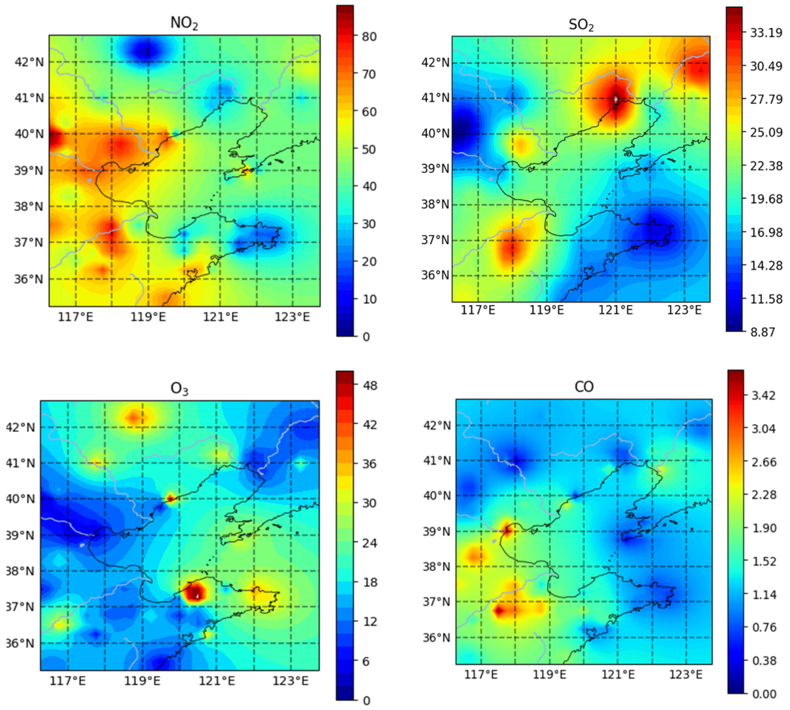

Figure 4. Inverse distance weighting (IDW) visualization of the four air pollutants.

Since IDW interpolation may introduce local anomalies, Gaussian filtering was further applied to enhance spatial continuity (Figure 5). The filtered fields exhibit improved spatial coherence and better agreement with expected pollutant distribution patterns.

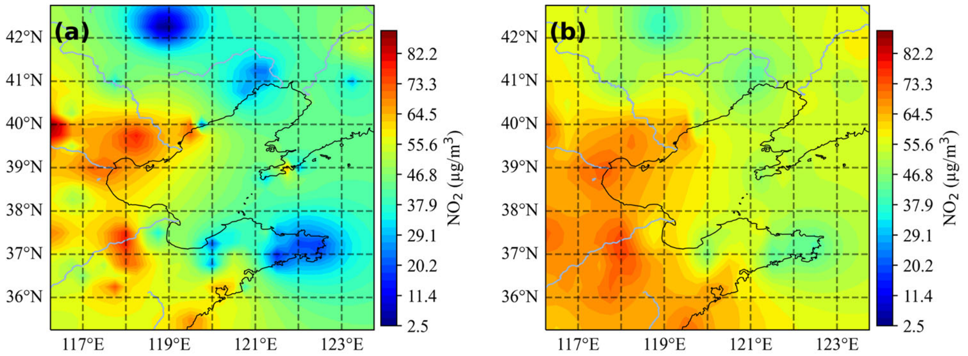

Figure 5. Comparison of NO_2 spatial distribution before (**a**) and after (**b**) Gaussian low-pass filtering at 00:00 on 1 January 2020.

2.2.3. Data Normalization

To eliminate dimensional differences among variables and ensure balanced contributions of each input feature during model training, all variables were standardized prior to training. In addition, relative humidity was calculated from ERA5 temperature and dew point temperature and incorporated into the model as an additional input feature.

2.3. TL-CNN Model Design and Configuration

2.3.1. CNN Inversion Model Design

In this study, a pollutant inversion model based on a CNN was constructed, with ERA5 meteorological variables as input features and EAC4 reanalysis pollutant concentrations as output labels (Figure 6). Traditional CNNs can effectively extract spatial features, but their practical applications face three major challenges: (1) when the input data are limited, the model is easily affected by noise and lacks generalization capability; (2) convolution and pooling operations may cause the loss of spatial distribution information, making it difficult to reconstruct nearshore concentration gradients; and (3) the fully connected layers contain a large number of parameters, which increases the risk of overfitting. To address these issues, the CNN model was improved as follows:

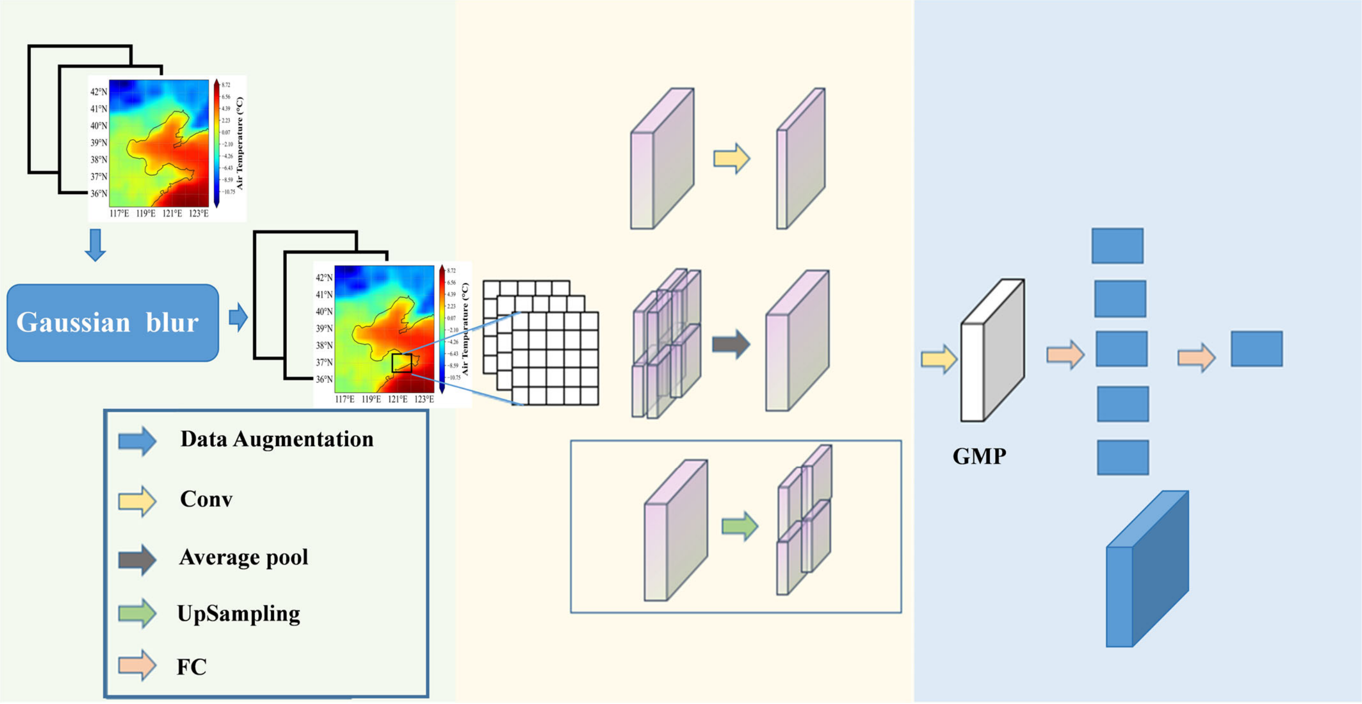

Figure 6. Structure of the CNN inversion model.

(1) Data augmentation (Gaussian noise): At the input stage, Gaussian noise with a mean of 0 and a standard deviation of $\sigma = 0.1$ was added to the meteorological variables (Figure 7) to simulate uncertainty and enhance model robustness. Compared with the original input data, the noise-augmented model achieved an approximately 5% improvement in R^2 on the test set, indicating stronger adaptability to small perturbations.

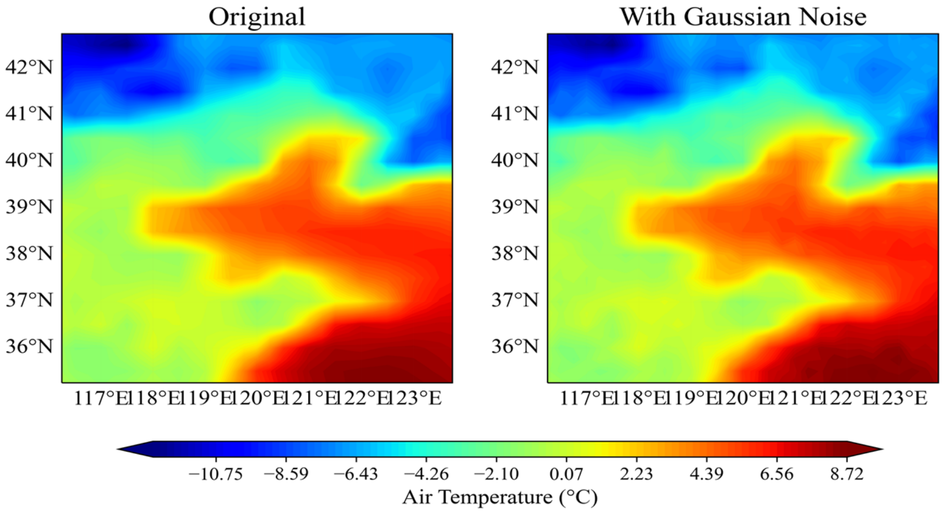

Figure 7. Comparison of Gaussian noise processing results.

(2) UpSampling layer: During feature extraction, a deconvolution-based upsampling module was introduced to restore the resolution lost during convolution–pooling operations to a level close to the original. This design improved the reconstruction of pollutant spatial distributions, particularly enabling more accurate reproduction of concentration gradients along coastal boundaries.

(3) Global Max Pooling: In the estimation stage, a global max pooling layer was used instead of traditional fully connected layers, mapping outputs by extracting only the maximum value from the feature maps. This significantly reduced the number of parameters and lowered the risk of overfitting. Experimental results demonstrate that this method outperforms fully connected layers in capturing extreme values, such as high-concentration pollution events.

Compared with conventional CNNs, the proposed improved model shows clear advantages in spatial distribution reconstruction and extreme-value detection: it better preserves concentration gradient continuity in coastal nearshore regions, while substantially reducing errors in identifying high-pollution episodes. These improvements make the model more suitable for application in offshore atmospheric environments, where observations are limited and spatial distributions are complex.

2.3.2. Input Variables

The input features include ERA5 meteorological variables (10 m wind speed U/V components, 2 m temperature, 2 m dew point temperature, sea surface temperature, mean sea level pressure, surface pressure, precipitation rate, and total precipitation), as well as relative humidity calculated from temperature and dew point temperature (Equation (1)). Relative humidity has been shown to play an important role in the simulation of SO_2 and NO_2 [22,23], and was therefore explicitly included as an input variable. To fully exploit spatial information, latitude and longitude coordinates were also added to the inputs.

$$\begin{cases} \mathrm{RH} = \frac{e}{e_s} \\ e = e_{s0} \exp\left(\frac{a t_d}{273.15 + t_d - b}\right) \\ e_s = e_{s0} \exp\left(\frac{a t}{273.15 + t - b}\right) \end{cases} \tag{1}$$

In the equation, RH denotes relative humidity, e_s represents the saturation vapor pressure, e is the actual vapor pressure, e_{s0} is the saturation vapor pressure at 0 °C, t denotes temperature (°C), and t_d denotes dew point temperature (°C). When t or $t_d > -15$ °C, $a = 17.269$ and $b = 35.86$; when t or $t_d < -40$ °C, $a = 21.8746$ and $b = 7.66$.

2.3.3. Transfer Learning Strategy

Given the scarcity of offshore observations and the limited numerical accuracy of EAC4 despite its relatively high resolution, this study further introduces TL technology. The core idea of transfer learning is to leverage pretrained model parameters from a source task (ERA5–EAC4) and transfer them to a target task (ERA5–IDW station observations), thereby improving modeling capability under data-sparse conditions.

The matched ERA5-EAC4 dataset, consisting of 23,376 consecutive 3-hourly intervals, was divided into training and testing sets according to time. To maintain temporal consistency, 22,560 records from 00:00 on 1 January 2014 to 21:00 on 20 September 2021 were used as the training set, with 20% randomly selected for validation. The remaining 816 records from 00:00 on 21 September 2021 to 21:00 on 31 December 2021 were designated as the test set. Note that the model was trained on a continuous 7-year dataset, while testing was conducted on a three-month period to avoid data leakage and overfitting. This partitioning ensures that the model is evaluated on data independent of the training period while preserving the temporal structure of the dataset.

The specific procedure is as follows: (1) Pretraining stage: A CNN model was trained using ERA5–EAC4 matched data from 2014 to 2020 (22,560 samples, with 20% used as the validation set) to obtain baseline weights. (2) Transfer stage: The pretrained weights were loaded and fine-tuned with ERA5–IDW observational data to correct systematic biases in EAC4. (3) Validation stage: Independent test data from 2021 (816 samples) were used to evaluate model performance, which was also compared against conventional Random Forest (RF) and Deep Neural Network (DNN) models.

Through transfer learning, the model not only retains the high spatiotemporal resolution features of EAC4 but also incorporates observational information to correct biases, thereby producing pollutant concentration fields closer to reality. This strategy proves highly applicable to offshore regions where large-scale in situ measurements are lacking.

After pretraining and fine-tuning through transfer learning, the final model configuration is summarized in Table 3. In the table, "Conv" denotes convolutional layers, all with a kernel size of 2 × 2. The average pooling (Average Pool) and upsampling (UpSampling) layers also employ a 2 × 2 grid size. ReLU is used as the activation function in all nonlinear layers except the output layer. Except for the output layer, each layer contains 32 neurons. The model was trained using the Adam optimizer with an initial learning rate of 0.01, a batch size of 20,440, and 800 training iterations. The output layer uses a Sigmoid activation function to map predicted pollutant concentrations to the [0, 1] range.

Table 3. TL-CNN network architecture.

Layer	Kernel/Size	Output Shape	Activation	Output Shape Explanation
Conv.	2 × 2	(Autoset, 12, 2, 32)	ReLU	Feature maps of height 12, width 2, 32 channels
Average Pool	2 × 2	(Autoset, 6, 1, 32)	-	Downsampled feature maps, height 6, width 1, 32 channels
Conv.	2 × 2	(Autoset, 3, 1, 32)	ReLU	Extracted features, height 3, width 1, 32 channels
UpSampling	2 × 2	(Autoset, 6, 2, 32)	-	Upsampled feature maps, height 6, width 2, 32 channels
Conv.	2 × 2	(Autoset, 6, 2, 32)	ReLU	Processed features, height 6, width 2, 32 channels
Global Max Pooling	-	(Autoset, 32)	-	Collapsed spatial dimensions, 32-length vector
Dense	-	(Autoset, 32)	Sigmoid	Fully connected layer with 32 neurons
Dense (Output)	-	(Autoset, 1)	Sigmoid	Output layer, single predicted value

Model performance was evaluated using the coefficient of determination (R^2) and the root mean square error (RMSE) (Equations (2) and (3)).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}, \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}, \tag{3}$$

where $\hat{y}$ denotes the predicted value obtained from the machine learning model, while $\overline{y}$ represents the mean value of the observed samples (y).

3. Results and Discussion

3.1. Evaluation of CNN Inversion Performance

The evaluation metrics of the CNN inversion model, pretrained with reanalysis data from 00:00 on 1 January 2014 to 21:00 on 20 September 2021, are presented in Table 4.

Table 4. Evaluation metrics of the pretrained model on the test set.

Pollutant	Mean Concentration	RMSE	R^2
NO_2 ($\mu g \cdot m^{-3}$)	42.6	8.68	0.82
SO_2 ($\mu g \cdot m^{-3}$)	38.2	9.39	0.84
O_3 ($\mu g \cdot m^{-3}$)	73.5	17.21	0.81
CO ($mg \cdot m^{-3}$)	0.93	0.25	0.80

Note: All R^2 values are statistically significant with $p < 0.05$.

As shown in Table 4, R^2 for all pollutants on the test set exceeded 0.8, and RMSE values were relatively small, indicating that the CNN model can effectively invert gaseous pollutant concentrations. Among the four pollutants, SO_2 exhibited the highest coefficient of determination ($R^2 = 0.84$), while O_3 showed the lowest ($R^2 = 0.81$), suggesting that the CNN model performed slightly better for SO_2 than for O_3 on the test set.

NO_2 and SO_2 showed intermediate errors (RMSE = 8.68 and 9.39 $\mu g/m^3$, respectively), whereas O_3 exhibited the largest RMSE (17.21 $\mu g/m^3$), likely due to its strong diurnal cycle and secondary formation processes, which increase prediction complexity. It should also be noted that the unit of CO concentration (mg/m^3) differs from that of NO_2, SO_2, and O_3 ($\mu g/m^3$). Therefore, the RMSE value of CO (0.25 mg/m^3) is not directly comparable to those of the other pollutants. However, relative to its concentration scale, this value still reflects high prediction accuracy. These RMSE values correspond to the average concentrations of 42.6 $\mu g \cdot m^{-3}$ for NO_2, 38.2 $\mu g \cdot m^{-3}$ for SO_2, 73.5 $\mu g \cdot m^{-3}$ for O_3, and 0.93 $mg \cdot m^{-3}$ for CO, indicating that the prediction errors are moderate relative to typical pollutant levels. Overall, the model achieved robust inversion performance across pollutants, with particularly strong performance for SO_2 and CO, while O_3 remained the most challenging to predict.

In addition to the overall evaluation metrics, residual diagnostics were conducted to further assess model performance. Figure 8 displays the histograms of residuals with Gaussian fitting curves and the standardized Q–Q plots for NO_2, SO_2, O_3, and CO. For all pollutants, the residuals were approximately normally distributed, although NO_2 and O_3 showed mild skewness and heavier tails. The Shapiro–Wilk test indicated deviations from strict normality ($p < 0.05$), but these deviations were minor. The residuals remained centered around zero without systematic bias, and predictive errors were relatively small.

Overall, the model achieved robust inversion performance across pollutants on the test set, with particularly strong performance for SO_2 and CO, while O_3 remained the most challenging to predict.

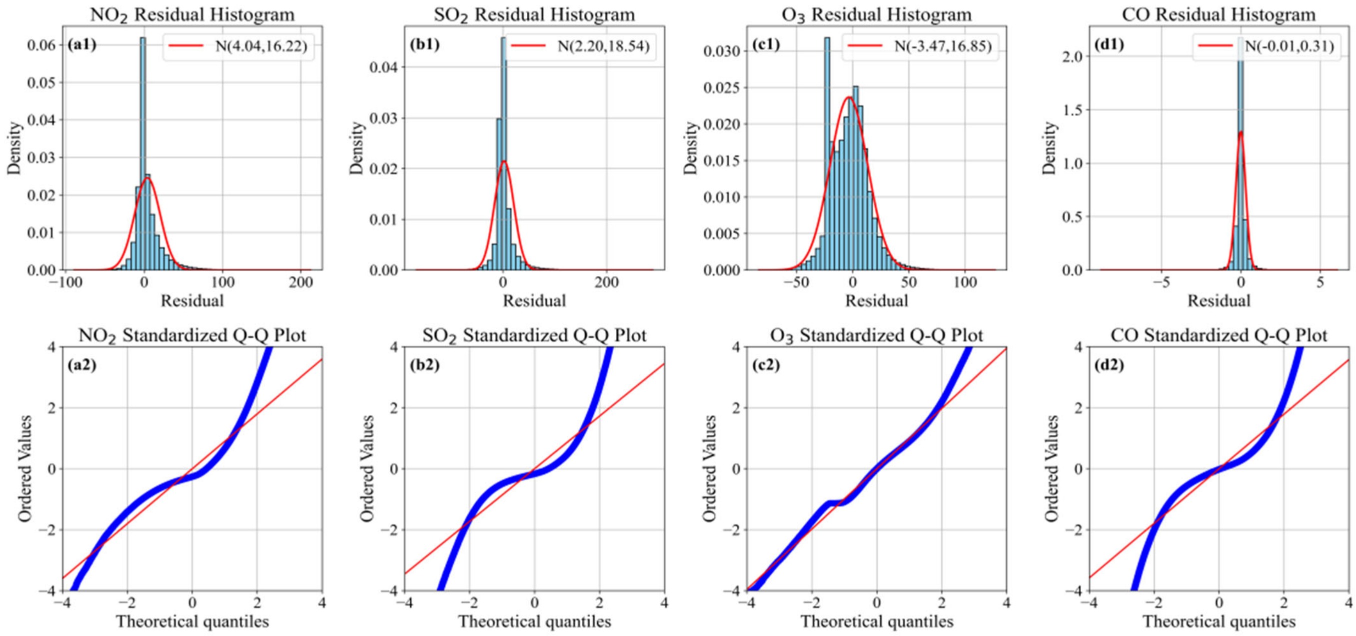

Figure 8. Residual analysis for NO_2, SO_2, O_3, and CO. Top row (**a1**–**d1**) shows residual histograms with fitted normal curves; bottom row (**a2**–**d2**) presents the corresponding standardized Q–Q plots.

3.2. Evaluation Against Other Modeling Approaches

To further assess the accuracy of the CNN inversion model on the ERA5–EAC4 dataset, two benchmark models—RF and DNN—were also constructed. Their optimal parameter configurations were determined through parameter tuning, and the corresponding test set performance is summarized in Table 5. Compared with the RF and DNN models (Table 5), the CNN demonstrated overall superior performance. For example, in the inversion of NO_2, the RMSE of CNN was reduced by 48.5% compared with RF and by 29.4% compared with DNN; in the case of O_3, the R^2 of CNN improved by more than 20%. These results highlight the inherent advantage of CNNs in capturing spatial features of pollutants. By contrast, RF relies mainly on decision tree partitioning and struggles to extract spatial continuity, while DNNs, though capable of learning complex nonlinear relationships, involve a large number of parameters and are prone to overfitting under limited sample conditions. Therefore, the CNN achieved the best generalization performance in this study.

Table 5. Random Forest (RF) model and Deep Neural Network (DNN) model evaluation metrics.

Statistical Metrics	**DNN**				**RF**			
	NO_2	**SO_2**	**O_3**	**CO**	**NO_2**	**SO_2**	**O_3**	**CO**
RMSE	12.3	13.65	26.35	0.75	16.85	14.35	21.26	0.86
R^2	0.71	0.75	0.58	0.65	0.53	0.71	0.72	0.51

Note: All R^2 values are statistically significant with $p < 0.05$.

Figure 9 compares the average inversion results of four atmospheric pollutants by the three models over a continuous 30-day test period (from 00:00 on 21 September 2021 to 21:00 on 20 October 2021). Figure 9a (upper left) shows the inversion results for NO_2 concentrations. Panel a1 presents the EAC4 data, while panels a2, a3, and a4 display the results from the RF, DNN, and CNN models, respectively. The three inversion models all captured the spatial distribution reasonably well, but for high-value regions, both RF and DNN were less accurate. In particular, RF and DNN erroneously overestimated concentrations over the Liaodong Peninsula and the western and northern coastal areas of the Bohai Sea.

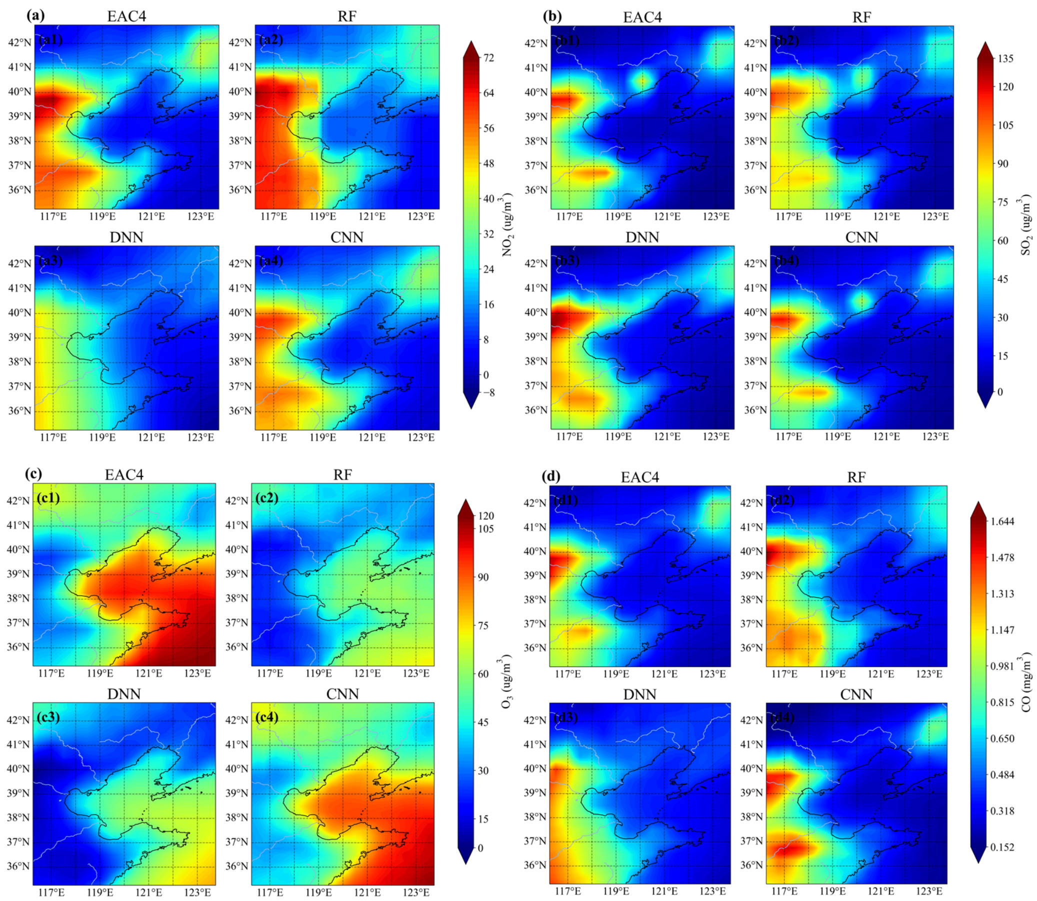

Figure 9. Comparison of EAC4 reanalysis with CNN, RF, and DNN models in retrieving NO_2, SO_2, O_3, and CO concentrations over the study area: (**a**) NO_2, (**b**) SO_2, (**c**) O_3, and (**d**) CO.

Figure 9b (upper right) shows the inversion results for SO_2 concentrations. Compared with the EAC4 data (b1), both RF (b2) and DNN (b3) failed to accurately capture the high SO_2 concentrations at northern Bohai coastal sites, while simultaneously overestimating SO_2 concentrations over the western Shandong Peninsula and northern Bohai. By contrast, the CNN model (b4) more accurately reproduced the spatial distribution patterns.

Figure 9c (lower left) presents the inversion results for O_3 concentrations. Compared with other pollutants, O_3 (c1) is primarily distributed over the Bohai and Yellow Seas, showing clear coastal boundaries and higher concentrations farther from shore. All three models reproduced this spatial feature. However, the DNN model (c3) severely underestimated O_3 concentrations along the Bohai–Yellow Sea coastline and adjacent areas, whereas the RF model (c2) overestimated them. The CNN model (c4) accurately captured these gradients.

Figure 9d (lower right) shows the inversion results for CO concentrations. Panel d1 presents the EAC4 data, while panels d2, d3, and d4 show the RF, DNN, and CNN results, respectively. CO is mainly distributed over inland regions west of the Bohai and Yellow Seas and over the Liaodong Peninsula. The DNN model (d3) failed to reproduce the high CO concentrations over the Liaodong Peninsula, and the RF model (d2) underestimated them, whereas the CNN model (d4) successfully captured the distribution in this region. Two major high-CO areas were identified over the Tangshan region and parts of inland Shandong. While all models reproduced the Tangshan high-concentration area reasonably well, all three overestimated the spatial extent of high CO concentrations over inland Shandong.

Overall, the comparison in Figure 8 demonstrates that the CNN not only reproduces the high-pollution features over the Liaodong Peninsula and the western and northern Bohai coast more accurately, but also avoids the systematic overestimation or underestimation of high-value regions observed in RF and DNN. For example, in the case of O_3, the CNN accurately captured the “low nearshore, high offshore” concentration gradient, while the DNN severely underestimated and the RF overestimated it. This further illustrates that the convolutional structure can effectively extract local spatial patterns and integrate them into global features, thereby significantly improving prediction accuracy.

The overall daily variations in pollutant concentrations simulated by the three models on the test set are shown in Figure 10. As illustrated, the daily mean concentrations of all pollutants exhibit pronounced fluctuations, with an overall increasing trend observed for all pollutants except O_3. In the inversion of NO_2 and CO, the RF model (green line) exaggerated these variations, resulting in a much larger range. For O_3, the DNN model captured the trend and fluctuations reasonably well, but the overall values were consistently underestimated, consistent with the results shown in Figure 9(c3). Among the four pollutants, SO_2 exhibited the best inversion performance across all models; however, for extreme values, the CNN model outperformed the other two.

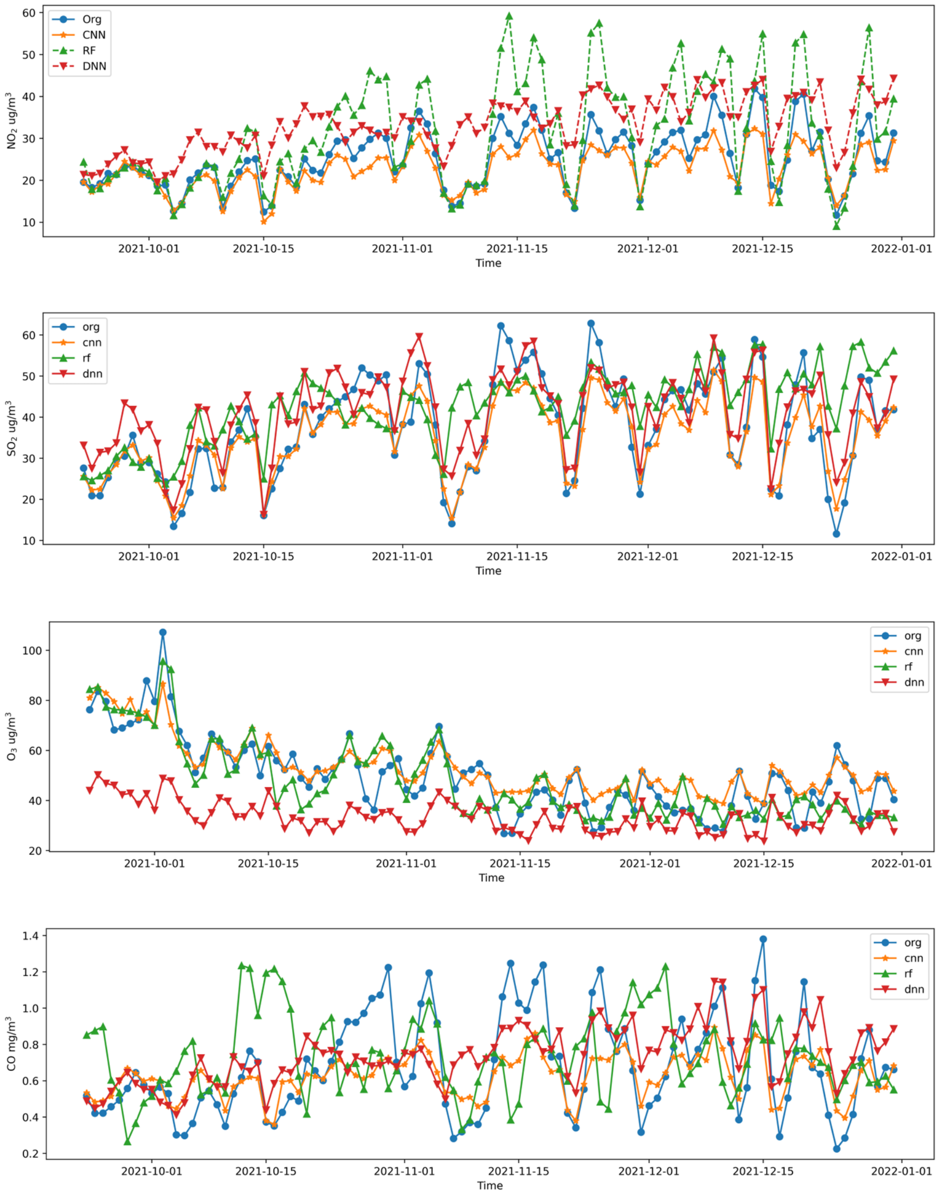

Figure 10. Time-series comparison of predicted concentrations (NO_2, SO_2, O_3, and CO) from CNN, RF, and DNN models against the original EAC4 on the test sets.

Overall, the CNN inversion results were more consistent with the EAC4 reanalysis data than those of RF and DNN. For daily mean pollutant concentrations, the CNN demonstrated the smallest deviations from the reference data in terms of temporal fluctuations, and it achieved the best performance in capturing both the overall trends and the mean variations. Nevertheless, for certain extreme cases—such as NO_2 during 21–25 October 2021, SO_2 during 14–16 November and 28 November 2021, and CO in particular—the CNN showed limited accuracy. This was mainly due to its difficulty in reproducing abrupt high pollutant concentrations at specific inland locations in western Bohai and on the Liaodong Peninsula.

3.3. Performance of the TL-CNN Model After Transfer Learning

After transferring the pretrained model weights to the ERA5–IDW dataset, the direct testing results are shown in Table 6. It can be seen that the performance of direct transfer was not ideal, with the R^2 for SO_2 being only 0.42. The main reason is that SO_2 is strongly influenced by anthropogenic emission controls. Since 2014, station observations have shown a pronounced decreasing trend (Figure 2), whereas the EAC4 reanalysis data failed to adequately capture this trend, leading to poor transfer performance of the model. The RMSE values were 13.86 $\mu g \cdot m^{-3}$ for NO_2, 22.36 $\mu g \cdot m^{-3}$ for SO_2, 23.36 $\mu g \cdot m^{-3}$ for O_3, and 0.36 $mg \cdot m^{-3}$ for CO. Compared to the average concentrations of 42.6 $\mu g \cdot m^{-3}$ for NO_2, 38.2 $\mu g \cdot m^{-3}$ for SO_2, 73.5 $\mu g \cdot m^{-3}$ for O_3, and 0.93 $mg \cdot m^{-3}$ for CO (Table 4), these RMSE values indicate moderate prediction errors relative to typical pollutant levels. Overall, the model captures general pollutant patterns reasonably well; however, further fine-tuning is necessary to improve accuracy, particularly for pollutants exhibiting strong temporal trends.

Table 6. TL-CNN model evaluation metrics on the ERA5–IDW dataset.

Evaluation Metrics	NO_2	SO_2	O_3	CO
RMSE	13.86	22.36	23.36	0.36
R^2	0.65	0.42	0.63	0.66

Note: All R^2 values are statistically significant with $p < 0.05$.

By reloading the complete CNN inversion model, including both the network architecture and pretrained weights, and subsequently retraining it with the ERA5–IDW dataset, the model was fine-tuned to optimize weights and parameter settings. As a result, the performance for NO_2, SO_2, O_3, and CO improved substantially, achieving R^2 values of 0.72, 0.69, 0.75, and 0.74, respectively, with overall trends closely aligned with the observations. The mean R^2 across all four pollutants reached 0.72, demonstrating that transfer learning effectively strengthened the model's inversion capability for pollutants in offshore regions. Compared with transferring pretrained weights only, the R^2 improvements were approximately 11%, 64%, 19%, and 12% for NO_2, SO_2, O_3, and CO, respectively, highlighting the substantial contribution of architecture transfer combined with fine-tuning.

Figure 11 presents the monthly average NO_2 concentrations for January 2020, obtained from the TL-CNN model. As shown, the spatial distribution after transfer learning (Figure 11c) closely resembles that of EAC4 (Figure 11a). Figure 11b shows the IDW interpolation results. Compared with EAC4, the IDW dataset exhibits generally lower values and a relatively uniform spatial distribution, ranging from 35 to 55 $\mu g \cdot m^{-3}$, with the main differences occurring over marine regions. Transfer learning enabled the model to capture this feature, resulting in pollutant concentration ranges narrower than those of the EAC4 dataset.

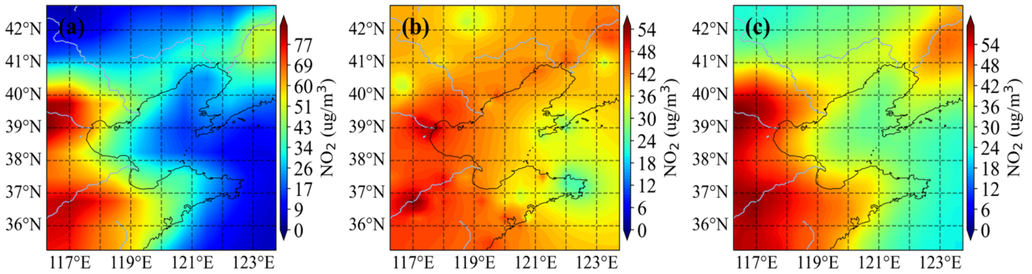

Figure 11. Spatial comparison of monthly average NO_2: (**a**) EAC4 reanalysis data, (**b**) IDW interpolation, and (**c**) TL-CNN predictions.

To comprehensively evaluate the stability and generalization performance of the TL-CNN model after fine-tuning, daily average pollutant concentrations from 1 January to 30 June 2020 were validated for major coastal cities on the Shandong Peninsula, with results for Qingdao presented in Figure 12. The transfer learning results for the four pollutants indicate that the TL-CNN model captures the overall trends and variability well for NO_2, O_3, and CO. In contrast, for SO_2, the transfer learning results tend to overemphasize the declining trend, suggesting that the transfer performance is not yet optimal and warrants further investigation. Moreover, the model's ability to capture underlying mechanisms appears limited, which may be attributed to two main factors: (1) systematic errors introduced when extracting station data using nearest-grid substitution, and (2) the low proportion of SO_2 extreme values in the original dataset, which limited the model's learning.

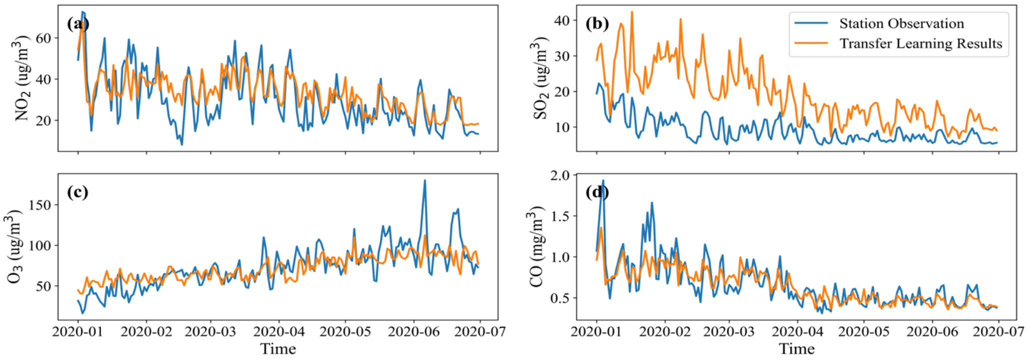

Figure 12. Time-series comparison of TL-CNN predictions and observations for Qingdao City: (**a**) NO_2, (**b**) SO_2, (**c**) O_3, and (**d**) CO.

Figure 13 shows density plots of TL-CNN predictions versus IDW-interpolated concentrations from 150 randomly selected test points between 1 January and 20 September 2020. Predicted concentrations are primarily distributed along the 1:1 diagonal line, indicating good overall model performance. Higher point densities occur at lower concentration ranges (~20 $\mu g \cdot m^{-3}$ for NO_2, ~15 $\mu g \cdot m^{-3}$ for SO_2, ~60 $\mu g \cdot m^{-3}$ for O_3, and ~0.5 $mg \cdot m^{-3}$ for CO), while scatter disperses at higher values due to the limited number of extreme samples, suggesting that the model's robustness under high-pollution events could be further improved.

By integrating CNN with transfer learning, the model not only outperformed RF and DNN in terms of accuracy but also corrected systematic biases in the reanalysis data through the incorporation of station observations. This improved its ability to capture spatiotemporal variations, making it particularly suitable for offshore regions where observational data are scarce.

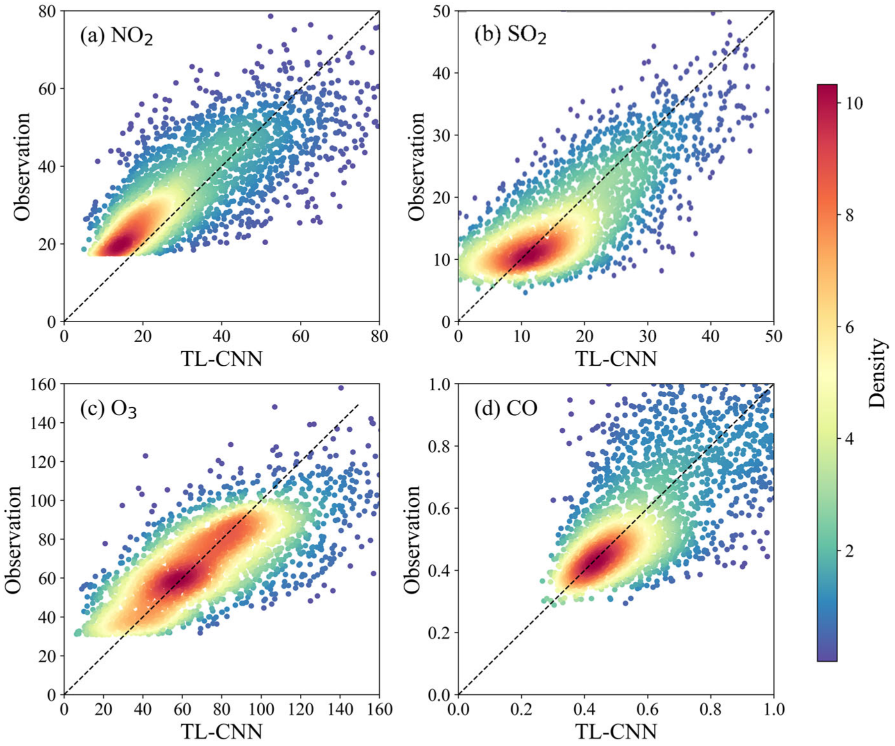

Figure 13. Density plots of TL-CNN predictions and observations for (**a**) NO_2, (**b**) SO_2, (**c**) O_3, and (**d**) CO, Warmer colors (yellow to red) indicate regions with higher data density, while cooler colors (blue to green) indicate lower density.

3.4. Implications and Limitations of This Study

This study highlights the potential of combining deep learning with transfer learning for the inversion of marine atmospheric pollutants in regions with sparse observational data. By fine-tuning a pretrained CNN model with available station observations, the proposed TL-CNN framework effectively improved retrieval accuracy and reduced systematic biases in reanalysis data. The high R^2 values achieved demonstrate that this approach can reliably reproduce the spatial distributions of NO_2, SO_2, O_3, and CO over the Bohai and Yellow Seas, providing an important technical foundation for atmospheric monitoring, pollutant transport studies, and policy evaluation in data-scarce marine regions.

Despite these advantages, several limitations should be noted, including the potential impacts of residual model assumptions, temporal coverage constraints, and uncertainties in the input datasets. Specifically, the residual deviations in SO_2 inversion may arise not only from the limited representation of anthropogenic emission reduction trends and scarcity of extreme-value samples, but also from potential errors in the reanalysis meteorological fields that influence pollutant dispersion. Temporal coverage constraints, such as gaps in station measurements or seasonal biases in observations, could also affect the robustness of the TL-CNN model, particularly for pollutants with strong diurnal or seasonal cycles like O_3. Additionally, uncertainties in input datasets, including instrument calibration errors, representativeness of surface observations, and differences in spatial resolution between reanalysis and observational data, could propagate through the model and slightly affect the accuracy of predicted concentrations.

Nevertheless, our results indicate that the TL-CNN framework provides a reliable and flexible approach for reconstructing marine atmospheric pollutant distributions in regions with sparse observations. Future work could mitigate these limitations by integrating physical process modeling (e.g., ship emissions, sea-salt aerosols, photochemical reactions) and incorporating multi-source datasets such as satellite retrievals, high-resolution emission inventories, and chemical transport model outputs. Such enhancements would not only improve the accuracy and reliability of pollutant inversion but also extend the applicability of the model to other marine and coastal regions, different temporal periods, and extreme pollution events, thus providing a more comprehensive tool for atmospheric monitoring, pollutant transport studies, and policy evaluation in data-scarce marine environments.

4. Conclusions

In this study, a TL-CNN model was developed and validated to address the lack of SO_2, NO_2, O_3, and CO observations over the Bohai and Yellow Seas. The main findings are summarized as follows:

The pretrained CNN achieved an average $R^2 > 0.80$, outperforming Random Forest and standard DNN models, demonstrating strong capability in capturing spatial patterns of marine atmospheric pollutants.

Fine-tuning both network architecture and pretrained weights with station observations increased the overall mean R^2 across NO_2, SO_2, O_3, and CO to 0.72, with SO_2 showing a ~64% improvement over transferring pretrained weights alone. This highlights the TL-CNN's effectiveness in correcting systematic biases in reanalysis data.

Remaining deviations in SO_2 inversion reflect limited representation of anthropogenic emission reductions and scarcity of extreme-value samples, indicating the value of integrating diverse observational sources and physical processes.

Overall, this study demonstrates the advantages and applicability of combining deep learning with transfer learning for reconstructing marine atmospheric pollutant distributions in data-scarce regions. Future work could further enhance model performance by integrating physical processes (e.g., ship emissions and sea-salt aerosols) and multi-source datasets such as satellite retrievals and high-resolution emission inventories. Such improvements would strengthen the model's generalization across diverse marine and coastal regions, temporal scales, and extreme pollution events, providing a comprehensive tool for environmental assessment, pollutant transport studies, and policy evaluation.

Author Contributions: Conceptualization, X.G. and X.L. (Xiaohuan Liu); methodology, X.G. and X.L. (Xiaoling Li); software, X.L. (Xiaoling Li); validation, X.G. and Z.Z.; formal analysis, X.L. (Xiaoyu Liu) and Z.Z.; investigation, X.L. (Xiaoling Li), X.L. (Xiaoyu Liu) and Y.X.; resources, X.G.; data curation, X.L. (Xiaoling Li) and J.H.; writing—original draft preparation, X.L. (Xiaoling Li); writing—review and editing, X.G. and X.L. (Xiaohuan Liu); visualization, X.L. (Xiaoling Li) and Y.X.; supervision, X.G.; project administration, X.G.; funding acquisition, X.L. (Xiaohuan Liu). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Nature Science Foundation of China, grant number 42175129 and the Key Laboratory of Mathematics and Engineering Applications, Ministry of Education. The APC was funded by Xiaohuan Liu.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors sincerely thank Xingbin Jia and Guoju Wang for their valuable advice and technical support. We are also deeply grateful to the anonymous reviewers for their constructive comments and suggestions, which greatly helped improve the quality of this manuscript.

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

References

1. Zhang, Y.; Zhou, R.; Chen, J.; Rangel-Buitrago, N. The effectiveness of emission control policies in regulating air pollution over coastal ports of China: Spatiotemporal variations of NO_2 and SO_2. *Ocean Coast. Manag.* **2022**, *219*, 106064. [CrossRef]
2. Wei, J.; Li, Z.; Wang, J.; Li, C.; Gupta, P.; Cribb, M. Ground-level gaseous pollutants (NO_2, SO_2, and CO) in China: Daily seamless mapping and spatiotemporal variations. *Atmos. Chem. Phys.* **2023**, *23*, 1511–1532. [CrossRef]
3. Tan, W.; Liu, C.; Wang, S.S.; Xing, C.Z.; Su, W.J.; Zhang, C.X.; Xia, C.Z.; Liu, H.R.; Cai, Z.N.; Liu, J.G. Tropospheric NO_2, SO_2, and HCHO over the East China Sea, using ship-based MAX-DOAS observations and comparison with OMI and OMPS satellite data. *Atmos. Chem. Phys.* **2018**, *18*, 15387–15402. [CrossRef]
4. Nguyen, D.-H.; Lin, C.; Vu, C.-T.; Cheruiyot, N.K.; Nguyen, M.K.; Le, T.H.; Lukkhasorn, W.; Vo, T.-D.-H.; Bui, X.-T. Tropospheric ozone and NOx: A review of worldwide variation and meteorological influences. *Environ. Technol. Innov.* **2022**, *28*, 102809. [CrossRef]
5. Zhang, Q.; Xue, D.; Liu, X.H.; Gong, X.; Gao, H.W. Process analysis of PM2.5 pollution events in a coastal city of China using CMAQ. *J. Environ. Sci.* **2019**, *79*, 225–238. [CrossRef] [PubMed]
6. Chang, M.; Cao, J.; Ma, M.; Liu, Y.; Liu, Y.; Chen, W.; Fan, Q.; Liao, W.; Jia, S.; Wang, X. Dry deposition of reactive nitrogen to different ecosystems across eastern China: A comparison of three community models. *Sci. Total Environ.* **2020**, *720*, 137548. [CrossRef]
7. Brunelli, U.; Piazza, V.; Pignato, L.; Sorbello, F.; Vitabile, S. Two-days ahead prediction of daily maximum concentrations of SO_2, O_3, PM10, NO_2, CO in the urban area of Palermo, Italy. *Atmos. Environ.* **2007**, *41*, 2967–2995. [CrossRef]
8. Rahimi, A. Short-term prediction of NO_2 and NOx concentrations using multilayer perceptron neural network: A case study of Tabriz, Iran. *Ecol. Process.* **2017**, *6*, 4. [CrossRef]
9. Navares, R.; Aznarte, J.L. Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecol. Inform.* **2020**, *55*, 101019. [CrossRef]
10. Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* **2017**, *231*, 997–1004. [CrossRef]
11. Reddy, V.; Yedavalli, P.; Mohanty, S.; Nakhat, U. Deep air: Forecasting air pollution in Beijing, China. *Environ. Sci.* **2018**, *1564*, 1–8.
12. Sayeed, A.; Choi, Y.; Eslami, E.; Lops, Y.; Roy, A.; Jung, J. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Netw.* **2020**, *121*, 396–408. [CrossRef]
13. Qi, Y.; Li, Q.; Karimian, H.; Liu, D. A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* **2019**, *664*, 982–992. [CrossRef] [PubMed]
14. Iskandaryan, D.; Ramos, F.; Trilles, S. Graph neural network for air quality prediction: A case study in Madrid. *IEEE Access* **2023**, *11*, 2729–2742. [CrossRef]
15. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
16. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
17. Alvera-Azcárate, A.; Barth, A.; Sirjacobs, D.; Lenartz, F.; Beckers, J. Data Interpolating Empirical Orthogonal Functions (DINEOF): A tool for geophysical data analyses. *Mediterr. Mar. Sci.* **2011**, *12*, 5–11. [CrossRef]
18. Chen, S.; Zou, B.; Tang, J. Impact of spatial interpolation methods on identifying structure of heavy metal polluted soil. *Sci. Surv. Mapp.* **2015**, *40*, 5. [CrossRef]
19. Bai, Y.; Yang, J.; Chen, P.; Wen, Y.; Kuang, H.; He, Y.; Zhang, Y. Spatiotemporal characteristics and relationships of main air-pollutants during a typical heavy air pollution in Xi'an City based on a spatial interpolation method. *Res. Environ. Sci.* **2020**, *33*, 809–819. [CrossRef]
20. Ma, J.; Ding, Y.; Cheng, J.C.; Jiang, F.; Wan, Z. A temporal-spatial interpolation and extrapolation method based on geographic long short-term memory neural network for PM2.5. *J. Clean. Prod.* **2019**, *237*, 117729. [CrossRef]
21. Tan, S.; Wang, Y.; Yuan, Q.; Zheng, L.; Li, T.; Shen, H.; Zhang, L. Reconstructing global PM2.5 monitoring dataset from OpenAQ using a two-step spatio-temporal model based on SES-IDW and LSTM. *Environ. Res. Lett.* **2022**, *17*, 034014. [CrossRef]
22. Jia, X.; Gong, X.; Liu, X.; Zhao, X.; Meng, H.; Dong, Q.; Liu, G.; Gao, H. Deep sequence learning for prediction of daily NO_2 concentration in coastal cities of northern China. *Atmosphere* **2023**, *14*, 467. [CrossRef]
23. Wang, G.; Zhu, R.; Gong, X.; Li, X.; Gao, Y.; Yin, W.; Wang, R.; Li, H.; Gao, H.; Zou, T. A new hybrid deep sequence model for decomposing, interpreting, and predicting sulfur dioxide decline in coastal cities of northern China. *Sustainability* **2025**, *17*, 2546. [CrossRef]