An Interpolation and Prediction Algorithm for XCO₂ Based on Multi-Source Time Series Data

Kai Hu 1,2*, Qi Zhang 1, Xinyan Feng 1, Ziran Liu 1, Pengfei Shao 1, Min Xia 1,2 and Xiaoling Ye 1,2

Abstract: Carbon satellites are an important observation tool for analyzing ground carbon emission. From the perspective of the Earth’s scale, the spatiotemporal sparse characteristics of raw data observed from carbon satellite requires the accurate interpolation of data, and based on only this work, people predict future carbon emission trends and formulate appropriate management and conservation strategies. The existing research work has not fully considered the close correlation between data and seasons, as well as the characteristics accumulated over a long time scale. In this paper, firstly, by employing extreme random forests and auxiliary data, we reconstruct a daily average CO₂ dataset at a resolution of 0.25°, and achieve a validated determination coefficient of 0.92. Secondly, introducing technologies such as Time Convolutional Networks (TCN), Channel Attention Mechanism (CAM), and Long Short-Term Memory networks (LSTM), we conduct atmospheric CO₂ concentration interpolation and predictions. When conducting predictive analysis for the Yangtze River Delta region, we train the model by using quarterly data from 2016 to 2020; the correlation coefficient in summer is 0.94, and in winter it is 0.91. These experimental data indicate that compared to other algorithms, this algorithm has a significantly better performance.

Keywords: carbon satellite; interpolation; prediction; Yangtze River Delta region

1. Introduction

Carbon dioxide (CO₂) is one of the most significant greenhouse gases in the atmosphere, constituting 0.04% of the total atmospheric composition [1]. Due to human activities, its concentration has risen from 280 ppm before the Industrial Revolution to the current level of 414 ppm. This increase, coupled with other greenhouse gas emissions, has resulted in a global average temperature rise of approximately 1.09 °C over the past century, causing irreversible damage to ecosystems [2]. The United Nations Framework Convention on Climate Change and the Paris Agreement aim to control and reduce atmospheric CO₂ concentration [3], making climate change an integral part of the United Nations' Sustainable Development Goals with profound implications for global health and sustainable development [4]. As one important step in technology, the accurate prediction of atmospheric CO₂ concentration is crucial for formulating emission reduction plans to achieve the “net-zero” target by 2050, aligning with both international and national emission reduction goals [5]. The study aims to establish an impartial carbon emission monitoring system by utilizing environmental variables, with the goal of providing crucial references and support for future anthropogenic economic activity carbon emissions.

Ground-based observations and satellite monitoring are commonly used methods for estimating carbon dioxide concentrations in the atmosphere [6]. Ground-based CO₂ concentration observations provide long-term, high-precision data but are sparsely distributed...
with limited spatial coverage. In contrast, satellite observations overcome the limitations of ground stations by covering extensive spatial ranges [7].

Satellites such as the Greenhouse Gases Observing Satellite (GOSAT) and the Orbiting Carbon Observatory-2 (OCO-2) can accurately detect global atmospheric CO$_2$ concentrations [8]. These satellite monitors utilize near-infrared solar radiation reflected from the Earth’s surface in the CO$_2$ spectral and O$_2$ A bands to generate XCO$_2$, aiming to enhance estimates of the spatial distribution of carbon sources and sinks [9]. Despite the numerous advantages of using carbon satellites for monitoring CO$_2$ concentrations, there are inevitably two challenges.

1. From a global perspective, the monitoring range is still limited by satellite observation methods, and satellites are susceptible to the influence of cloud cover and aerosols [10, 11].

2. Due to insufficient satellite data coverage, the acquisition of long-term time series data is limited, thus making the accurate prediction of future CO$_2$ concentrations more challenging.

For instance, even after quality control, OCO-2 satellite’s effective observation amounts to only about 10% of the total observations [12]. Currently, the satellite monitoring of atmospheric XCO$_2$ has a relatively low coverage. This low coverage of XCO$_2$ concentration data adversely impacts the accurate estimation of carbon sources and sinks [13]. Therefore, filling the gaps in XCO$_2$ data is crucial for subsequent predictions.

In previous research, three main approaches have been developed to reconstruct high-coverage XCO$_2$ (carbon dioxide column-averaged dry air mole fraction) data [14], including spatial interpolation [15], multisensor fusion [16], and modeling based on machine learning [8].

In recent years, with the abundance of data and sufficient computational power, machine learning methods have introduced a novel perspective for data fusion. A popular strategy involves utilizing machine learning to establish relationships between auxiliary factors and XCO$_2$ data, followed by the reconstruction of CO$_2$ concentrations in regional or global atmospheres. For instance, Siabi et al. [8] employed a multilayer perceptron model to construct a nonlinear correspondence between OCO-2 satellite XCO$_2$ data and multiple data sources, effectively filling gaps in satellite observations. He et al. [17] and colleagues utilized elevation, meteorological conditions, and CarbonTracker XCO$_2$ data, employed LightGBM to achieve comprehensive XCO$_2$ data coverage for China. Using extreme random forest and random forest models, Li et al. [18] and Wang et al. [19] generated continuous spatiotemporal atmospheric CO$_2$ concentration data at both global and regional scales. However, most studies are limited to constructing datasets, without delving into the subsequent prediction of CO$_2$ concentration changes.

Currently, only a few pieces of the literature have attempted to forecast CO$_2$ column concentrations. For example, Zheng et al. [20] utilized the GOSAT dataset and applied differential moving autoregressive models and long short-term memory (LSTM) neural network models to predict the trend of CO$_2$ concentration changes in the near-surface region of China. However, this experiment did not consider meteorological and vegetation factors related to CO$_2$, and the data resolution was relatively low, posing challenges for regional predictions, with less-than-satisfactory prediction accuracy. Meng et al. [21] employed a heterogeneous spatiotemporal dataset obtained from OCO-2, GOSAT, and self-built wireless carbon sensors, attempting to use the LSTM model for prediction. However, they tested only one location, lacking a more comprehensive validation. On the other hand, Li et al. [22] selected OCO-2 satellite spectral data from 2019 and used five machine learning models, considering various meteorological, surface, and vegetation factors for estimation. But, they did not adequately account for regional seasonal variations in CO$_2$ and long-term trends. Moreover, there is currently no publicly available dataset.

It is noteworthy that, currently, there is a lack of relevant research employing deep learning algorithms, particularly based on OCO-2 data, for the accurate estimation of CO$_2$ column concentrations. The primary advantage of deep learning methods lies in
their powerful ability to automatically learn advanced features from extensive datasets, a crucial step in bridging the gap between data patterns at different feature levels. Given the outstanding feature extraction performance of deep learning neural networks, they hold significant potential in fusing multisource data to extract crucial spatial information [23,24].

The objective of this study is to fill the CO$_2$ data gaps, enhance the high spatiotemporal resolution of the data, and use a deep learning neural network for prediction, with a specific focus on estimating medium- to long-term fully covered daily scale CO$_2$ data. Due to the advantages of Temporal Convolutional Networks (TCN), Channel Attention mechanism (CA), and Long Short-Term Memory networks (LSTM), this paper intends to combine them to alleviate the problem of interpolation and prediction of XCO$_2$ data. The contributions of this study are as follows:

1. Augmenting the existing multisource data with ground semantic information has been incorporated, enhancing the predictive capabilities of the model.
2. A daily dataset of seamless XCO$_2$ in the Yangtze River Delta region with a spatial resolution of 0.25°, derived from the fusion of multisource data spanning from 2016 to 2020, has been established.
3. The adoption of the TCN-Attention module has improved the quality and efficiency of feature aggregation, enabling the better capture of both local and global spatial features.
4. Leveraging the LSTM structure, long-term trends in multisource spatiotemporal data are effectively modeled, facilitating the integration of features across multiple time steps.

The workflow of this study is outlined as follows: Section 2 introduces the data utilized, encompassing XCO$_2$ data from satellite observations, and auxiliary data, and details the data processing and analysis procedures. This section delves into the prediction methodology, encompassing the deep learning approach and the model’s schematic diagram. Section 3 encompasses the model evaluation, along with a detailed discussion of the spatiotemporal distribution. Section 4 gives a conclusion and future prospects. Figure 1 provides an overview of the whole workflow.

Figure 1. Flowchart of our work.
2. Materials and Methods

2.1. Study Area

The study area encompasses the Yangtze River Delta (YRD) region in China (latitude 29°20′–32°34′N, longitude 115°46′–123°25′E), with Shanghai as its central point, including the Jiangsu, Anhui, and Zhejiang provinces. The total area is 358,000 square kilometers, situated in the alluvial plain formed before the Yangtze River enters the sea. The terrain is low-lying, with elevations ranging from 200 to 300 m. The region is crisscrossed by rivers, characterized by developed agriculture, a dense population, and numerous cities. The predominant land cover types include farmland, forests, and water bodies. It is the area with the highest river density in China, with over 200 lakes on the plain. The Yangtze River Delta experiences a subtropical monsoon climate. The vegetation cover data for the year 2020 is shown in Figure 2. Figure 2a displays the vegetation coverage map of China during the summer of 2016. Figure 2b shows the Yangtze River Delta region under study, while Figure 2c presents the trend of XCO₂ (https://earthdata.nasa.gov/, accessed on 10 May 2024) growth from 2016 to 2020. The red line depicts the CO₂ concentration dynamics, showing lower levels in summer, higher levels in winter, and an overall upward trend. The data are sourced from the MOD13C2 product, available for download from (https://modis.gsfc.nasa.gov/, accessed on 10 May 2024).

![Vegetation Cover in China](image1)

**Figure 2.** Study area: (a) the NDVI coverage map of China, (b) the study area, (c) the growth trend of XCO₂ in the YRD from 2016 to 2020.

Economic development has a significant impact on carbon emissions [25]. The YRD region, as the most economically powerful center in China, aggregates factors such as economy, technology, and talent, leading to high intensity of pollutant emissions. The region is prominently affected by regional atmospheric pollution and is a key area for air pollution prevention and control in China. Therefore, it is essential to predict high concentrations of CO₂ in the YRD region [26]. This prediction serves as a scientific basis for regional ecological environment quality monitoring, environmental health assessment, and decision-making management to achieve pollution reduction, carbon reduction, and coordinated efficiency enhancement. In the bottom right corner, the trend chart illustrates the CO₂ concentration in the YRD region from 2016 to 2020. It reveals a seasonal cyclic variation in CO₂, with concentrations continuously increasing.
2.2. Multisource Data

Table 1 summarizes the data used to estimate XCO\(_2\) from multiple sources, covering seven main categories, including OCO-2 XCO\(_2\), CAMS XCO\(_2\), vegetation data, the Fifth Generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5) meteorological variables, land cover data, elevation data, and Total Carbon Column Observing Network (TCCON) station XCO\(_2\) measurement data.

Table 1. Data Used.

<table>
<thead>
<tr>
<th>Data</th>
<th>Variables</th>
<th>Spatial Resolution</th>
<th>Temporal Resolution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite Data</td>
<td>XCO(_2)</td>
<td>1.29 × 2.25 km</td>
<td>16 day</td>
<td><a href="https://earthdata.nasa.gov/">https://earthdata.nasa.gov/</a></td>
</tr>
<tr>
<td>Reanalysis Data</td>
<td>XCO(_2)</td>
<td>0.75°</td>
<td>3 h</td>
<td><a href="https://ads.atmosphere.copernicus.eu/">https://ads.atmosphere.copernicus.eu/</a></td>
</tr>
<tr>
<td>Meteorological Data</td>
<td>Relative Humidity (Rh)</td>
<td>0.25°</td>
<td>3 h</td>
<td><a href="https://cds.climet.copernicus.eu/">https://cds.climet.copernicus.eu/</a></td>
</tr>
<tr>
<td></td>
<td>10-m U Component of Wind (U)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-m V Component of Wind (V)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 m Temperature (T2M)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation Data</td>
<td>DEM</td>
<td>90 m × 90 m</td>
<td>-</td>
<td><a href="https://engine-aiearth.aliyun.com/">https://engine-aiearth.aliyun.com/</a></td>
</tr>
<tr>
<td>CLCD</td>
<td>Land, Forest, Grassland, Water Body, Shrubland and so on</td>
<td>30 m</td>
<td>-</td>
<td><a href="https://engineaiearth.aliyun.com/">https://engineaiearth.aliyun.com/</a></td>
</tr>
<tr>
<td>Station Data</td>
<td>XCO(_2) Point</td>
<td>∼2 m</td>
<td></td>
<td><a href="https://tccondata.org/">https://tccondata.org/</a></td>
</tr>
</tbody>
</table>

2.2.1. OCO-2 XCO\(_2\) Data

The CO\(_2\) column concentration data used in this study are sourced from the OCO-2 satellite product (OCO\(_2\)_L2_Lite_FP). OCO-2, launched by NASA in July 2014, is the first dedicated carbon observation satellite designed for measuring XCO\(_2\) and monitoring near-surface carbon sources and sinks. The satellite observes the Earth around 13:30 local time, with a spatial resolution of 2.25 km × 1.29 km (∼0.02°) and a revisit cycle of 16 days [27]. In comparison to other CO\(_2\) observation satellites, OCO-2 satellite data offers superior spatial resolution and monitoring accuracy [12]. The XCO\(_2\) data utilized in this study cover the period from 1 January 2018, to 31 December 2020. Figure 3 depicts the mean OCO-2 XCO\(_2\) values for the year 2016 in the Chinese region.
2.2.2. Cams XCO\textsubscript{2} Data

CAMS reanalysis is the latest global atmospheric composition reanalysis dataset, encompassing aerosols, chemical substances, and greenhouse gases [28]. The CAMS Global Greenhouse Gas Reanalysis, which includes CO\textsubscript{2} and CH\textsubscript{4}, currently spans from 2003 to 2020, with temporal and spatial resolutions of 3 h and 0.75\degree, respectively. In the generation process of CAMS XCO\textsubscript{2}, OCO-2 data are not assimilated. Therefore, the fusion of CAMS XCO\textsubscript{2} with OCO-2 XCO\textsubscript{2} data holds the potential to integrate advantages from multiple data sources [29]. Verification indicates that CAMS XCO\textsubscript{2} data demonstrates potential and feasibility for atmospheric CO\textsubscript{2} analysis. Generated by the Integrated Forecast System (IFS) model and the 4DVar data assimilation system at ECMWF, CAMS XCO\textsubscript{2} data are derived from atmospheric data storage, utilizing the “Column-Averaged Mole Fraction of CO\textsubscript{2}” variable in this study.

2.2.3. Vegetation Data

The NDVI, as a component of carbon sink, can characterize vegetation growth status and has been demonstrated to be closely related to CO\textsubscript{2} concentration [30]. Therefore, in the reconstruction process, NDVI is employed as one of the auxiliary predictive factors. The MODIS instrument (https://modis.gsfc.nasa.gov/, accessed on 10 May 2024), a crucial tool on the Terra and Aqua satellites, is widely utilized for vegetation growth monitoring due to its large observation coverage (approximately 2330 km) and high data quality. Thus, monthly MOD13C2 products were obtained at a resolution of 0.05\degree for this study [31,32].

2.2.4. Meteorological Data

In addition to considering natural vegetation factors, this study also takes into account the influence of meteorological parameters on atmospheric CO\textsubscript{2} concentration. Given the significant impact of meteorological factors on the temporal and spatial variations of CO\textsubscript{2} concentration, key meteorological factors affecting concentration include wind speed, temperature, and humidity [33,34]. ERA5, the fifth-generation ECMWF global climate and weather reanalysis dataset, features a spatial resolution of 0.25\degree \times 0.25\degree and a temporal resolution of 1 h, distributed on a grid. ERA5 incorporates more historical observational data, particularly satellite data, into advanced data assimilation and modeling systems to estimate more accurate atmospheric conditions. In this context, wind speed (wspd) and wind direction (wdir) are calculated based on the U-component (UW, m/s) and V-component (VW, m/s) of wind velocity, employing the following formula. Additionally, temperature (TEM, K) and relative humidity (RH, %) are introduced for modeling CO\textsubscript{2} concentration estimation. All meteorological data used here are from the time interval between 13:00 and 14:00 during satellite overpasses [35].

\[
\text{wspd} = \sqrt{(u^2 + v^2)},
\]

\[
\text{wdir} = \arctan\left(\frac{v}{u}\right),
\]

2.2.5. Elevation Data

The Space shuttle Radar Topography Mission (SRTM) is an international project initiated by the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA). It spanned 11 days and aimed to acquire and generate high-resolution global terrain elevation products. The dataset employed in this study is SRTM3, featuring a spatial resolution of 90 m.

2.2.6. Land Cover Data

Ground semantic information provides insights into different regional ecosystems and land uses, impacting the processes of atmospheric CO\textsubscript{2} absorption and emission. Therefore, this study incorporates the Chinese Land Cover Dataset (CLCD). Created by the team at
Wuhan University, this dataset, based on Landsat imagery, characterizes land use and land cover across various regions in China. It typically includes multiple categories such as forests, grasslands, water bodies, wetlands, and farmland. The spatial resolution of the CLCD used in this study is 30 m.

2.2.7. Tccon XCO2 Data

The TCCON employs ground-based Fourier-transform spectrometers to record near-infrared spectra, subsequently retrieving column-averaged carbon dioxide concentrations. Due to its high precision in CO2 detection, TCCON station data is widely utilized for validating satellite-derived CO2 products [36,37]. Hence, in this study, TCCON data are utilized as ground-based in situ CO2 data to assess the reconstruction performance. The research region includes a ground monitoring station, the Hefei station (located at 117.17°E, 31.9°N), with data collection spanning from January 2016 to December 2020 [38].

2.2.8. Data Preprocessing

This study collect multi-source data covering China and processed it to harmonize these data in the spatiotemporal dimension [39].

Initially, to ensure data quality, refinement is performed on the collected OCO-2 XCO2 data based on quality flags, eliminating pixels with poor quality (where the xco2_quality_flag parameters of 0 and 1 represent good and poor quality, respectively) [40]. Subsequently, daily data passing through 13:00 are selected as CAMS daily data, and the ERA5 meteorological data’s average values were used to represent different pressure level data [19]. Taking into account spatial heterogeneity and assessment criteria for different factors, these variable factors are resampled to a spatial resolution of 0.25° to construct a temporally consistent dataset. For CAMS and ERA5 data with spatial resolutions greater than 0.25°, inverse distance weighting interpolation was applied, while bilinear interpolation is used for vegetation and meteorological data with resolutions of less than 0.25°. DEM and land cover data were transformed into CSV format through batch cropping and processing of remote sensing images. The resampling process ensures a consistent spatial resolution of 0.25° for various factors.

Next, an Extreme Random Forest (ERF) regression model is trained [17], ensuring full utilization of the entire dataset to train a single decision tree. The parameters of the Extreme Random Forest are set as follows: n_estimators is set as 200, the random seed is set as 42, max_depth is set as 10, and max_features is set as 0.8. The model was evaluated through 10-fold cross-validation, achieving a fitting degree of 92%.

After constructing the complete dataset, long-term observations from ground stations are crucial for the performance of reconstructing XCO2 results. Although CO2 observation stations in the YRD region are limited, the TCCON Hefei station’s data covers the period from 2015 to 2020, along with some climate background station observations of near-surface CO2 concentrations [41,42]. Comparing the XCO2 results from the Hefei station with the reconstructed XCO2 model data, the average deviation was approximately 0.4 ppm, the Standard Deviation (SD) was about 0.75 ppm, and the Root Mean Square Error (RMSE) is around 1.01 ppm. As shown in Table 2, Li et al. estimated an RMSE of 1.71 ppm for XCO2 from 2015 to 2020 compared to ground-based TCCON data [43]. Zhang et al. validated XCO2 from the Hefei TCCON site against ML results, showing an average deviation of −0.60 ppm, an SD of 0.99 ppm, and an RMSE of 1.18 ppm [44]. He et al. validated XCO2 from the Hefei TCCON site against ML results, showing an average deviation of −0.60 ppm, an SD of 0.99 ppm, and an RMSE of 1.18 ppm [44]. These results are consistent with our analysis, further supporting the reliability and validity of our findings [45]. The error of the validation results is depicted in Figure 4. The x-axis is based on time, the left y-axis (XCO2/ppm) represents the concentration of XCO2, and the right y-axis (bias) shows the difference between the actual station data and the reconstructed data. Clearly, the results from the Hefei station closely align with TCCON observations, indicating the good performance of the model data in simulating XCO2.
Therefore, this dataset is named Yangtze River Delta \textit{XCO}_2 (YRD\textunderscore XCO\textsubscript{2}) and serves as the research dataset in this paper.

**Table 2.** Comparison with previous studies.

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>SD</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li [43]</td>
<td>1.71 ppm</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zhang [44]</td>
<td>1.18 ppm</td>
<td>0.99 ppm</td>
<td>-0.6</td>
</tr>
<tr>
<td>He [45]</td>
<td>1.123 ppm</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ours</td>
<td>1.01 ppm</td>
<td>0.75</td>
<td>0.4</td>
</tr>
</tbody>
</table>

**Figure 4.** Validation chart of TCCON and reconstructed XCO\textsubscript{2} at Hefei site from 2016 to 2020 (publicly available sites in the Yangtze River Delta region).

The prediction of CO\textsubscript{2} concentration requires a parameterized model, with each parameter or variable having different scales in the dataset. To prevent parameters with large value ranges from exerting excessive influence, feature normalization is performed to scale all features equally. This normalization eliminates the influence of absolute values across different units, enabling fair comparisons among indicators. The Max – Min normalization method is employed to ensure that all features are normalized to the same range, transforming the original data of each feature into the range [0, 1].

\[
X = \frac{(X - X_{\text{min}})}{(X_{\text{max}} - X_{\text{min}})} \quad (3)
\]

where \(X\) represents the original value, \(X_{\text{min}}\) is the minimum value, and \(X_{\text{max}}\) is the maximum value.

2.3. **Data Analysis**

2.3.1. Seasonal Analysis

CO\textsubscript{2} concentration is influenced by seasonal variations, and Figure 5a illustrates a schematic representation of original satellite data using the Local Polynomial Regression and Scatterplot Smoothing (LOESS) method for Seasonal-Trend decomposition using LOESS (STL) [46,47]. This method decomposes the original time series into secular trend (Figure 5b), seasonal variation (Figure 5c), and residual terms (Figure 5d).
Figure 5. Time series after STL decomposition (a) original data, (b) trend component, (c) seasonal component, (d) residual component.

Concurrently, the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) are used to examine the seasonality of the original data. The $p$-values in Table 3 provide sufficient evidence to reject the null hypothesis for autocorrelation. If the $p$-values for the first-, second-, and third-order autocorrelations are greater than 0.05, there is no autocorrelation. However, if the $p$-values for the fourth-order or higher-order autocorrelations are less than 0.05, autocorrelation is present.

Table 3. Results of ACF and PACF.

<table>
<thead>
<tr>
<th>Lag Order</th>
<th>AC Value</th>
<th>PAC Value</th>
<th>Q Statistic</th>
<th>$p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>0.938</td>
<td>0.938</td>
<td>290,163.434</td>
<td>0.030</td>
</tr>
<tr>
<td>2nd</td>
<td>0.888</td>
<td>0.069</td>
<td>550,369.730</td>
<td>0.020</td>
</tr>
<tr>
<td>3rd</td>
<td>0.843</td>
<td>0.022</td>
<td>784,843.896</td>
<td>0.001</td>
</tr>
<tr>
<td>4th</td>
<td>0.801</td>
<td>0.011</td>
<td>996,667.104</td>
<td>0.000</td>
</tr>
<tr>
<td>5th</td>
<td>0.763</td>
<td>0.014</td>
<td>1,188,779.593</td>
<td>0.000</td>
</tr>
<tr>
<td>6th</td>
<td>0.728</td>
<td>0.018</td>
<td>1,363,936.930</td>
<td>0.000</td>
</tr>
<tr>
<td>7th</td>
<td>0.697</td>
<td>0.014</td>
<td>1,524,238.478</td>
<td>0.000</td>
</tr>
<tr>
<td>8th</td>
<td>0.669</td>
<td>0.027</td>
<td>1,672,195.034</td>
<td>0.000</td>
</tr>
<tr>
<td>9th</td>
<td>0.645</td>
<td>0.026</td>
<td>1,809,783.745</td>
<td>0.000</td>
</tr>
<tr>
<td>10th</td>
<td>0.624</td>
<td>0.022</td>
<td>1,938,508.098</td>
<td>0.000</td>
</tr>
</tbody>
</table>

YRD_XCO$_2$ concentrations exhibit clear periodic patterns and strong autocorrelation in Figure 5, confirming that the collected CO$_2$ concentration data from past time points can be used for subsequent predictions. This supports the rationale for using time series to construct the Temporal Convolutional Network (TCN) model in the study. However, on 11 December 2018, the carbon dioxide concentration was 419 ppmv, which is 8.69 ppmv higher than the expected value of 409 ppmv, resulting in a residual. This indicated an abnormally
increased concentration compared to the expected value, possibly influenced by non-periodic meteorological factors, which may be related to extreme weather conditions, posing a challenge for accurate predictions of YRD_XCO₂ concentrations [48,49]. Considering that CO₂ concentration is influenced by various factors, including time, weather, vegetation, elevation, and semantic information, we selected time information, meteorological input parameters, vegetation parameters, elevation information, and semantic data as important variables for model prediction. Additionally, CAMS XCO₂ reanalysis data were considered to enhance spatiotemporal resolution and improve the spatiotemporal resolution of satellite XCO₂ data. This parameter was used as an auxiliary variable input.

Figure 6 displays descriptive statistical data for CO₂ concentrations in the YRD region from 2016 to 2020, showing a yearly increase in the mean CO₂ concentration. The annual average growth of YRD_XCO₂ falls within the range of 2.8 ± 0.8 ppm/yr. Differences in CO₂ concentrations are observed among the four seasons, with noticeably higher average concentrations in spring and winter compared to summer and autumn. Specifically, variations in CO₂ concentrations are observed in April (spring) and September (summer), with fluctuations occurring mainly between spring, summer, and the arrival of winter to the next spring. Changes between summer and autumn are relatively small. According to Falahatkar et al.’s study [50], the rise in spring temperatures and vegetation recovery accelerate soil microbial activity, leading to increased CO₂ release. Simultaneously, the combustion of fossil fuels during winter releases a substantial amount of CO₂, contributing to the rise in atmospheric CO₂ concentrations during spring. Subsequently, enhanced vegetation growth and photosynthesis during spring and summer gradually reduce CO₂ concentrations. In autumn and winter, when vegetation growth ceases and photosynthesis weakens, coupled with fossil fuel combustion for heating during winter, CO₂ concentrations gradually increase. In the following sections, this paper will explore the statistical relationships between various variables, as depicted in Figure 6.

Figure 6. Seasonal and annual changes in CO₂ concentrations in the Yangtze River Delta from 2016 to 2020.
2.3.2. Statistical Relationship between Variables

XCO\textsubscript{2} denotes column-averaged carbon dioxide data, CAMS refers to reanalysis data, r represents relative humidity, ndvi indicates vegetation coverage, t denotes temperature, u signifies horizontal wind speed, v stands for vertical wind speed, classid represents surface semantic data, and dem refers to elevation data. In Figure 7, the statistical relationships and importance between variables are illustrated, with correlation coefficients ($r$) used to indicate their correlations. The formula is shown below.

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}}$$

(4)

where $X_i$ and $Y_i$ represent the $i$th observations for two variables, $\bar{X}$ and $\bar{Y}$ are the respective means of all observations, and $n$ is the number of observations.

The correlation coefficient $r$ between satellite XCO\textsubscript{2} and CAMS\textsubscript{XCO\textsubscript{2}} is 0.71, while it shows a negative correlation with vegetation data $r = -0.20$. Regarding emissions, there is a significant correlation with meteorological factors, such as a negative correlation with temperature ($r = -0.52$) and a $r$ value of $-0.28$ with sea-level pressure. The correlation with elevation and ground semantic information is weaker, with negative correlations of $-0.04$ and $-0.02$, respectively. The interrelation among these variables is intricate. Despite some correlations not being highly pronounced, the model constructed in this study is capable of extracting valuable information from these complex relationships. Therefore, meteorological input parameters, vegetation parameters, elevation information, and semantic data were chosen as auxiliary data for training in this study.

2.4. Prediction Models

This study introduces an innovative CO\textsubscript{2} concentration prediction model based on feature fusion. By incorporating Time Convolutional Network (TCN), the model is able to effectively extract mid-term and periodic variations in the CO\textsubscript{2} concentration sequence. The use of a Channel Attention Mechanism aids in learning the relationships between different features, and the Long Short-Term Memory (LSTM) network is employed to capture the long-term dependencies in the time series. The research objective is to comprehensively predict the variation trends of CO\textsubscript{2} concentration in the Yangtze River Delta across different seasons in 2020. To thoroughly assess the model’s performance, this study employs evaluation metrics for time series regression models, providing an in-depth analysis of the model’s performance on the test data.

The unique one-dimensional causal convolution structure of TCN ensures the preservation of temporal sequence characteristics in the data. Residual connection units expedite the network’s convergence speed, and dilated convolutions guarantee the extraction of all data features. CAM is an additional module designed to learn the weights of each feature channel. It enables the model to better comprehend which features are more crucial.
for the task, guiding the model to focus its attention on those more important channels. The LSTM model, as a variant of classical RNN, possesses outstanding non-linear fitting capabilities, making it suitable for handling sequence modeling problems. The prediction of YRD_XCO$_2$ concentration involves a time series forecasting problem with non-linear features. Factors influencing YRD_XCO$_2$ concentration include meteorological conditions, vegetation, and semantic information from the ground. In this study, TCN and CAM modules are fused, combined with the LSTM model to construct a CATCN-LSTM model for non-linear feature atmospheric CO$_2$ concentration prediction using multi-source data. The structure of CATCN-LSTM is illustrated in Figure 8, and the primary process for predicting YRD_XCO$_2$ concentration is described as follows.

![Model architecture](image.png)

**Figure 8.** Model architecture: (a) CATCN-LSTM, (b) Dilated Causal Convolution, (c) CAM, (d) LSTM, (e) TCN-LSTM.

### 2.4.1. Tcn Module

The Temporal Convolutional Network (TCN) is employed as the predictive model for YRD_XCO$_2$ concentration. TCN is a simple and versatile convolutional neural network architecture designed for addressing time-series problems, primarily composed of multiple stacked residual units [51]. The residual module comprises two convolutional units and a non-linear mapping unit. TCN exhibits several advantages in time-series prediction tasks: (a) it can address the issues of gradient vanishing and exploding; (b) it can compute convolutions in parallel, thereby accelerating training speed; (c) TCN possesses a highly effective historical length, making it capable of capturing temporal correlations for discontinuous and widely spaced historical time series data.
1. Causal Convolution

The causal convolution imparts a strict temporal constraint on the TCN module with respect to the input XCO₂ sequence \( x_0, x_1, \ldots, x_{t-1}, x_t, \ldots \). The output \( y_t \) at time \( t \) is expressed such that it is only related to the inputs up to and including time \( t \). As illustrated in Figure 8b, its mathematical representation is as follows:

\[
y_t = f(x_1, x_2, \ldots, x_t).
\]  
(5)

Here, \( x_t \) is a one-dimensional vector containing \( n \) features, and \( y_t \) is the variable to be predicted. There exists some relationship between \( x_t \) and \( y_t \), denoted by the function \( f \). To ensure that the output tensor and input tensor have the same length, a strategy of zero-padding on the left side of the input tensor is employed. Causal convolution is a unidirectional structure that processes the value at time \( t \) and uses only data before time \( t \) to ensure the temporal nature of data processing. However, to obtain longer and complete historical information, as the network depth increases, issues such as gradient vanishing, computational complexity, and poor fitting effects may arise. Therefore, dilated convolution is introduced.

2. Dilated Convolution

Dilated convolution allows exponentially increasing the receptive field without increasing parameters and model complexity. As shown in Figure 8c, the network structure of dilated convolution is presented. Unlike traditional convolutional neural networks (CNN), dilated convolution permits the input of convolution to have interval sampling controlled by the dilation factor, denoted as \( d \). In the bottom layer, \( d = 1 \) means that the input is sampled at each time point, and in the hidden layers, \( d = 2 \) means that the input is sampled every 2 time points as one input. For a one-dimensional XCO₂ concentration sequence \( X = (x_0, x_1, \ldots, x_t, \ldots) \), the definition of dilated convolution \( F(S) \) with a filter \( f \) on \( 0, \ldots, k-1 \) is given as follows:

\[
F(S) = \sum_{i=0}^{k-1} f(i) x(s-d \cdot i),
\]  
(6)

where \( S \) is the input sequence information, \( d \) is the dilation factor, \( k \) is the filter size, \( f(i) \) represents the weight of the convolutional kernel, \( d \cdot i \) is the total displacement on the input sequence, and \( (s-d \cdot i) \) denotes the position in the historical information of the sequence. The dilation factor \( d = (1, 2, 4) \) is used, and as \( d \) increases, the receptive field \( \omega \) of TCN expands, ensuring that the convolutional kernel can flexibly choose the length of historical data information. The receptive field \( \omega \) of TCN is expressed as:

\[
\omega = 1 + (k-1) \cdot \frac{b^n - 1}{b - 1}.
\]  
(7)

Here, \( n \) is the number of layers, and \( b \) is the base of the dilatation convolution (dilation factor \( d = b^i, i = 1, 2, \ldots, n \)). It can be observed that when the filter size is 3 and the dilation factors are \([1, 2, 4]\), the output \( y_t \) at time \( t \) is determined by the inputs \((x_1, x_2, \ldots, x_t)\), indicating that the receptive field can cover all values in the input sequence.

3. Residual block

The residual structure of TCN is illustrated in Figure 8a. The output of different layers is added to the input data, forming a residual block. After passing through an activation function, the output is obtained. The residual block connection mechanism enhances the network’s feedback and convergence, and helps avoid issues like gradient vanishing and exploding commonly found in traditional neural networks. Each residual unit consists of two one-dimensional dilated causal convolutional layers and a non-linear mapping. Initially, the input data \( h_{t-1} \) undergoes a one-dimensional...
dilated causal convolution, followed by weight normalization to address gradient explosion and accelerate network training. Subsequently, a ReLU activation function is applied for non-linear operations. Dropout is added after each dilated convolution to prevent overfitting. Additionally, a $1 \times 1$ convolution is introduced to return to the original number of channels. Finally, the obtained result is summed with the input to generate the output vector

$$f_i = \text{conv}(w_i \times F_j + b_i),$$  \hfill (8)

$$\{f_0, f_1, \ldots, f_{t-1}, f_t\} = \text{weightnorm}(\{f_0, f_1, \ldots, f_{t-1}, f_t\}),$$  \hfill (9)

$$h_t = \{f_0, f_1, \ldots, f_{t-1}, f_t\} = \text{Relu}(\{f_0, f_1, \ldots, f_{t-1}, f_t\}),$$  \hfill (10)

where $f_i$ represents the feature vector obtained through convolution at time $i$, $w_i$ denotes the weights of the convolution calculation at time $i$, $F_j$ represents the convolutional kernel of the $j$-th layer, $b_i$ is the bias vector, weightnorm$(x) = \frac{\|w_i\|}{\|v\|} v$, $\|w_i\|$ represents the magnitude of the weight $w$ in the Relu$(x) = \max(0,x)$ operation, and $\frac{v}{\|v\|}$ indicates the unit vector in the same direction as $w$. $h_t$ represents the feature map obtained after the complete convolution of the $j$-th layer.

The TCN model performs feature extraction on input information. After the TCN model extracts features from the data, impurities in the data are significantly reduced, and features are more pronounced, aiding the subsequent CAM module in obtaining higher weights and capturing crucial relationships between features.

2.4.2. Tcn-Cam Module

The attention mechanism simulates human attention by weighting different features, highlighting key features, and enhancing model performance [52–54]. It has been widely applied in the machine translation and computer vision fields. In order to better learn the importance of each feature from the XCO$_2$ time series, calculate their attention scores, and further capture temporal relationships, this study designs a channel attention module suitable for TCN. The attention mechanism weights and sums the feature vectors input to the TCN network, as shown in Figure 8c. Two pooling layers, global average pooling and global maximum pooling, are used to obtain the importance of these features. The input is the hidden layer output vector $h_t$ (with shape $N \times C \times T$) from the TCN layer, where $C$ is the number of features or channels, $T$ is the time sequence length, and $N$ is the number of samples. After passing through the two global pooling layers, a channel feature of size $C \times 1 \times 1$ is obtained. Then, channel dimension reduction is performed through a $1 \times 1$ convolutional layer. This process is expressed as

$$\text{GAP} = \frac{1}{T \times N} \sum_{m=1}^{T} \sum_{n=1}^{N} V(m,n),$$  \hfill (11)

$$\text{GMP} = \max \{V(m,n,o)\},$$  \hfill (12)

where $m = 1, 2, \ldots, T$; $n = 1, 2, \ldots, N$.

$$x_{\text{GAP}} = \text{conv}(\text{GAP}(X)),$$  \hfill (13)

$$x_{\text{GMP}} = \text{conv}(\text{GMP}(X)),$$  \hfill (14)

In the context, GAP and GMP represent global average pooling and global max pooling, respectively. The variables “$m$” and “$n$” denote positions along the dimensions $T$ and $N$. The term “conv” refers to a convolutional layer with a kernel size of 1. GAP and GMP represent the extracted features through global average pooling and global max
pooling, respectively. Subsequently, the output vectors of these two pooling operations are concatenated and fed into a convolutional layer with a kernel size of 1. By applying the sigmoid function, attention weights “a” are computed \([55,56]\). The input vector is then multiplied by the attention scores to obtain the weighted new feature. The calculation formula is as follows:

\[
a = \text{sigmoid}(\text{conv(cat}(\text{Gap, Gmp}))),
\]

\[
y_i = \sum_{i=0}^1 a \cdot h_i,
\]

where “cat” represents the concatenation operation, “sigmoid” is the activation function, and “conv” denotes a convolutional layer with a kernel size of 1. The input \(h_i\) is subjected to channel attention, resulting in the new feature \(y_i\). Subsequently, the weighted new feature \(y_i\) is input into an LSTM module for further predictions.

### 2.4.3. LSTM Module

The LSTM model is employed for processing time-series data. The LSTM model features memory cells with self-connections to store temporal states, as illustrated in Figure 8d. The LSTM model comprises three gates: the forget gate, the input gate, and the output gate. At each time step \(t\), the input sequence vector, the hidden layer output, and the cell state are considered. The outputs include the LSTM hidden layer output and the cell state \([57,58]\). The formulas for the forget gate, input gate, and output gate are as follows:

\[
f_t = \sigma(w_f \cdot [x_t, h_{t-1}] + b_f),
\]

\[
i_t = \sigma(w_i \cdot [x_t, h_{t-1}] + b_i),
\]

\[
o_t = \sigma(w_o \cdot [x_t, h_{t-1}] + b_o),
\]

The input gate and forget gate respectively determine the proportion of information carried over from \(c_{t-1}\) and contributed by \(\tilde{c}_t\) in the current cell state \(c_t\). The current cell state is determined by

\[
c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t
\]

The output formula for the hidden layer is

\[
h_t = o_t \odot \tanh(c_t),
\]

where \(f_t\), \(i_t\), and \(o_t\) represent the forget gate, input gate, and output gate, respectively; \(\sigma\) and \(\tanh\) denote the sigmoid function and hyperbolic tangent function; \(w_f\), \(w_i\), \(w_o\) and \(w_c\) are the weight matrices of the LSTM model; \(h_{t-1}\) is the state information passed from the previous time step; \(b_f\), \(b_i\), \(b_o\) and \(b_c\) are the bias matrices of the LSTM; \(\tilde{c}_t\) represents the candidate memory cell; \(c_t\) denotes the current cell state. The symbol \(\odot\) represents element-wise multiplication of two matrices.

The features extracted by the TCN model are input into the LSTM model to enable the model to handle long-term sequential data and accurately predict the concentration of YRD_XCO2 in the next time step. The output vector learned by the LSTM layer is then fed into a fully connected network. Through iterative training, the final estimate of YRD_XCO2 is obtained.

### 2.5. Model Evaluation Metrics

Prediction involves speculating on future trends based on existing data using specific methods and rules. To assess the quality of the prediction results, it is essential to introduce a dedicated error evaluation system to represent the discrepancies between predicted
values and actual values. A smaller error between predicted and actual values indicates better prediction results, reflecting a more effective predictive model. Otherwise, a larger error suggests the poorer predictive performance of the model.

In this study, we introduce the coefficient of determination ($R^2$), RMSE, Mean Absolute Error (MAE), and mean absolute percentage error (MAPE) as effective metrics for evaluating the disparities between predicted and actual values. The expressions for each metric are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i}((y_i - \bar{y})^2)},$$  \hspace{1cm} (23)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2},$$  \hspace{1cm} (24)

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|,$$  \hspace{1cm} (25)

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%,$$  \hspace{1cm} (26)

In the equations, $y_i$ represents the actual value, $\hat{y}_i$ represents the predicted value, $\bar{y}$ represents the mean of the actual values, and $N$ is the total number of samples.

3. Results and Discussion
3.1. Experimental Environment

This section outlines the configuration and settings of the experiments, as well as introduces the dataset and evaluation criteria.

The experiments in this paper are conducted on a system with the following specifications: a 64-bit Windows 10 operating system, TensorFlow framework (open-source) for software, and an Nvidia GeForce RTX 3060 GPU for hardware acceleration. The TCN model’s parameters are configured with three residual blocks, a convolutional kernel size of three, 128 convolutional kernels, and dilation factors set to [1,2,4]. All experiments employ the Stochastic Gradient Descent (SGD) algorithm with a batch size of 1024. The initial learning rate is set to 0.001, and the training process is conducted for 30 iterations. MSE is chosen as the loss function, and various model parameters were continuously adjusted during training. To ensure adaptive learning, the learning rate of the model is decayed during training, ensuring a prompt reduction when the model no longer exhibited a decrease in loss or an increase in accuracy. The loss curves of the models depict the trends in their training performance. From these curves, crucial information about model overfitting can be discerned. Figure 9 displays the loss curve of the models, indicating that the CATCN-LSTM model exhibited favorable learning outcomes with no signs of overfitting or underfitting. Figure 10 presents a comparison between predicted and actual values, while Figure 11 provides a comparison between the actual and predicted values for each model across the four seasons. Finally, Figure 12 illustrates the annual average CO$_2$ concentration values for the YRD region in 2020.
Figure 9. Loss plot of CATCN-LSTM.

Figure 10. Fitting plot of observed and estimated XCO₂.

Figure 11. Trends of observed and predicted values for different models across seasons.
Figure 12. Annual average CO$_2$ concentration map in the Yangtze River Delta region for the year 2020.

In this study, a CO$_2$ concentration prediction model based on the multi-input CATCN-LSTM architecture is developed. The prediction results are illustrated in Figure 10, demonstrating a strong positive correlation between the predicted values and the observed values with a fitting degree of 93%. Based on the left side of Figure 10, it is evident that the model performs well in predicting regions with high amplitude and frequency. Some outliers can be attributed to extreme weather conditions or industrial incidents, such as during the COVID-19 pandemic when certain regions experienced abnormal fluctuations in CO$_2$ concentration due to lockdowns and reduced economic activities. Hence, the occurrence of these outliers may be closely linked to environmental factors and human activities.

3.2. Sensitivity Analysis

The model proposed in this paper is primarily constructed from three core components: TCN, CAM, and LSTM. Utilizing a dataset spanning from January 2016 to December 2019 as the training set, the model predicts YRD XCO$_2$ data for January 2020 to December 2020. In order to validate the effectiveness of the proposed model, sensitivity analysis, involving five different combinations, is conducted:

1. LSTM, denoted as Model 1;
2. TCN, denoted as Model 2;
3. TCN-CAM, denoted as Model 3;
4. TCN-LSTM, denoted as Model 4;
5. CATCN-LSTM, representing the integrated model proposed in this paper.

Table 4 presents the results of the sensitivity analysis, revealing that the proposed model achieved optimal predictive performance across four evaluation metrics after multiple cross-validation cycles. Model 1, serving as the baseline model, exhibited relatively lower prediction accuracy. Model 2, which utilizes the TCN model to capture global temporal information, shows slight improvements in predictive accuracy, particularly in RMSE and MAE, compared to the baseline model. Model 3, which incorporates a channel attention mechanism on top of TCN, demonstrates slightly higher predictive accuracy metrics compared to the standalone TCN model. Model 3, which employs a decompose-and-integrate prediction strategy based on the divide-and-conquer principle, reduces prediction complexity, leading to improved performance compared to the single TCN model. Model 4, which incorporates LSTM on top of the TCN model, exhibits a modest improvement over Model 3.
Table 4. Experimental ablation results.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>0.75</td>
<td>0.77</td>
<td>1.25</td>
<td>0.010</td>
</tr>
<tr>
<td>TCN</td>
<td>0.85</td>
<td>0.58</td>
<td>0.92</td>
<td>0.014</td>
</tr>
<tr>
<td>TCN-CAM</td>
<td>0.86</td>
<td>0.54</td>
<td>0.90</td>
<td>0.014</td>
</tr>
<tr>
<td>TCN-LSTM</td>
<td>0.90</td>
<td>0.40</td>
<td>0.69</td>
<td>0.009</td>
</tr>
<tr>
<td>CATCN-LSTM</td>
<td>0.92</td>
<td>0.34</td>
<td>0.62</td>
<td>0.007</td>
</tr>
</tbody>
</table>

The proposed model integrates the strengths of residual network modules and channel attention units. With multi-scale components obtained through residual blocks and channel attention units focusing on distinctive features at different frequencies, the CATCN-LSTM model outperformed all testing models, achieving the highest value. Compared to a single LSTM model, CATCN-LSTM reduced RMSE by 70%, MAE by 33%, and improved $R^2$ accuracy by 23%. In comparison to the TCN-LSTM model, RMSE decreased by 13%, and MAE decreased by 6%.

3.3. Comparison of CATCN-LSTM with Other Models

The proposed model is being compared with other models in a comparative experiment, including SVR, XGBOOST, RNN, and CNN-LSTM. To ensure fairness, the prediction processes of these models are aligned with the proposed model. LSTM is used with 10 neurons in a single hidden layer, employing ReLU as the activation function and a sliding window length of 10. The output layer consisted of a single fully connected layer. Training parameters, such as the learning rate, are consistent with the proposed model. The SVR model utilizes default parameters with an RBF kernel from the sklearn library. Table 4 presents the results of the comparative experiment, demonstrating the predictive performance of each model. Figure 11 visualizes the prediction results across the models. To account for the strong seasonality of CO$_2$ concentration, the training data are divided into seasons: spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). Each data subset is used for model training. Since the test set data for winter only extended until December 2020, predictions are made solely for this month.

As shown in Figure 11, from top to bottom, each panel’s curves represent the true values of CO$_2$ concentration and the predicted values of each model. CATCN-LSTM consistently provides more accurate predictions across the entire forecast range compared to the other models. XGBOOST and SVR exhibited relatively weaker performances, while RNN and CNN-LSTM showed noticeable lags. The model achieves the best prediction performance during the summer season, considering that the Yangtze River Delta region experiences a subtropical monsoon climate during this period, typically characterized by higher temperatures. This season sees significant impacts on ecosystem activities and processes like plant photosynthesis, resulting in notable fluctuations in atmospheric CO$_2$ concentration. The model’s ability to accurately capture these seasonal variations contributes to its precision in predictions. In contrast, winter temperatures are generally lower, and the region experiences significant temperature fluctuations due to the convergence of cold and warm air masses. This can lead to phenomena such as snowfall, human activities related to heating facilities, and complex factors like emissions and energy consumption, introducing more noise and resulting in comparatively poorer model performance during this season.

The MAE of the CATCN-LSTM model is 25%, 18%, 15%, and 6% lower than SVR, XGBOOST, RNN, and CNN-LSTM, respectively. Compared to linear XGBOOST and SVR models, the RNN model achieves smaller errors, highlighting the ability of neural networks to model nonlinear relationships. The CNN-LSTM model outperforms the RNN model in terms of $R^2$, MAE, RMSE, and MAPE, indicating that the integration of CNN with LSTM preserved the LSTM encoder’s output for a given input sequence, selectively learning from the input sequence, and associating the output sequence with the input effectively to
discern the importance of information. However, CNN-based LSTM models may need to stack multiple convolutional layers to obtain a larger receptive field for extracting hidden unknown information.

The CATCN-LSTM model performs the best, demonstrating its capability to handle the periodic characteristics of CO$_2$ and the impact of extreme weather. Firstly, the robustness, memory capacity, nonlinear mapping ability, and self-learning capability of TCN make it more effective in predicting CO$_2$ concentration and capturing global information than other models. Secondly, despite the influence of periodic patterns and weather conditions on CO$_2$, the residual blocks of the TCN model add the input to the output of the convolutional layer, aiding in gradient propagation and model training. This mechanism enables better capture of local and short-term dependencies in the sequence. With the addition of the attention mechanism, the model can enhance its focus on different features. Lastly, LSTM is employed to handle the long-term dependencies of the entire sequence, further enhancing the accuracy of the final predictions. This method demonstrates practicality in capturing the atmospheric chemistry and physical nonlinearity. It can estimate CO$_2$ concentration trends for each season, providing essential data support for understanding and addressing climate change and environmental issues, contributing to the realization of carbon neutrality goals. Table 5 presents a comparison of the prediction errors between the proposed method and other typical machine learning methods.

Table 5. The comparison results of different models in different seasons in 2020.

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>$R^2$</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>CATCN-LSTM</td>
<td>0.917</td>
<td>0.403</td>
<td>0.681</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>CNN-LSTM</td>
<td>0.878</td>
<td>0.595</td>
<td>0.901</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>RNN</td>
<td>0.754</td>
<td>0.774</td>
<td>1.250</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>0.699</td>
<td>0.916</td>
<td>1.385</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>XGBOOST</td>
<td>0.602</td>
<td>1.027</td>
<td>1.594</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>CATCN-LSTM</td>
<td>0.941</td>
<td>0.344</td>
<td>0.559</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>CNN-LSTM</td>
<td>0.863</td>
<td>0.588</td>
<td>0.926</td>
<td>0.0014</td>
</tr>
<tr>
<td>Summer</td>
<td>RNN</td>
<td>0.748</td>
<td>0.821</td>
<td>1.279</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>0.685</td>
<td>1.074</td>
<td>1.390</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>XGBOOST</td>
<td>0.620</td>
<td>1.255</td>
<td>1.624</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>CATCN-LSTM</td>
<td>0.937</td>
<td>0.333</td>
<td>0.515</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>CNN-LSTM</td>
<td>0.855</td>
<td>0.604</td>
<td>1.006</td>
<td>0.0019</td>
</tr>
<tr>
<td>Autumn</td>
<td>RNN</td>
<td>0.721</td>
<td>0.871</td>
<td>1.483</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>0.682</td>
<td>0.916</td>
<td>1.425</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>XGBOOST</td>
<td>0.640</td>
<td>1.062</td>
<td>1.590</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>CATCN-LSTM</td>
<td>0.915</td>
<td>0.410</td>
<td>0.697</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>CNN-LSTM</td>
<td>0.880</td>
<td>0.567</td>
<td>0.992</td>
<td>0.0012</td>
</tr>
<tr>
<td>Winter</td>
<td>RNN</td>
<td>0.734</td>
<td>0.821</td>
<td>1.304</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>0.659</td>
<td>0.937</td>
<td>1.476</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>XGBOOST</td>
<td>0.582</td>
<td>0.860</td>
<td>1.534</td>
<td>0.0026</td>
</tr>
</tbody>
</table>

Based on the observations from Figure 11, the following conclusions can be drawn: the predicted average CO$_2$ concentrations for each season in 2020 were 415.11 ppm, 413.05 ppm, 413.18 ppm, and 414.71 ppm, with errors relative to the true values being 0.20 ppm, 0.13 ppm, 0.14 ppm, and 0.21 ppm, respectively. In the case of minimal concentration variation during spring, the model exhibits satisfactory performance in predicting values compared to the actual ones. However, during extreme increases or decreases in concentration in summer and winter, the proposed model demonstrates good fitting effects. In contrast, other models show noticeable lag and delay. This suggests that the proposed model has potential practical applications in addressing changes in CO$_2$ concentration in the field of carbon emissions. Figure 12 illustrates the annual average CO$_2$ values for YRD in 2020. It is evident that the estimated CO$_2$ values align well with the annual average XCO$_2$ values, showcasing the high consistency of these results. These findings provide robust
support for future climate and carbon emission management, highlighting the model’s applicability across different seasons and conditions.

4. Conclusions and Prospect

4.1. Conclusions

In order to address the challenges of carbon satellite data interpolation and prediction, this paper conducts the following works:

1. To address spatiotemporal sparse characteristics of data observed from carbon satellite raw data, this paper employs bilinear interpolation to resample multiple auxiliary datasets with XCO\textsubscript{2} data, achieving a daily data granularity of 0.25°. Subsequently, an Extreme Random Forest algorithm is utilized to reconstruct the data from 2016 to 2020. Through ten-fold cross-validation, the model’s robustness is verified, ensuring a high concordance of 92% with ground measurement station data.

2. CATCN-LSTM algorithm is proposed for predicting four seasons’ CO\textsubscript{2} concentrations in the Yangtze River Delta; it achieved higher predictive accuracy in summer and relatively weaker accuracy in winter. Compared to the LSTM model previously used by Meng and Li [21,22], this model effectively addresses the challenges posed by interdependent features in long sequences and provides a new approach for predicting CO\textsubscript{2} concentrations.

4.2. Prospective

In order to better integrate our work into several important areas, such as understanding the impact of climate change on ecosystems, predicting future trends, and formulating appropriate management and conservation strategies, we believe that further in-depth work is needed in the following three areas:

1. In terms of data, since satellite XCO\textsubscript{2} observational data are typically more accurate than reconstructed XCO\textsubscript{2} data, future studies can integrate more satellite data to enhance accuracy. For example, satellites like OCO-3 and GOSAT can be integrated, and deep learning techniques can be employed for interpolation when integrating high spatiotemporal resolution XCO\textsubscript{2} data. In addition, this study estimates XCO\textsubscript{2} data using environmental variables, but did not incorporate anthropogenic factors into the modeling process. Existing research has not adequately addressed this point [43–45], and in the future, incorporating social science factors into the model may improve our estimation accuracy.

2. In the model aspect, more advanced deep learning architectures or ensemble methods can be explored to further improve the predictive accuracy of CO\textsubscript{2} concentrations. Consideration can be given to incorporating technologies like Transformer and spatiotemporal attention mechanisms to better capture the complex spatiotemporal relationships of CO\textsubscript{2} concentrations in the atmosphere. Tuning model parameters and conducting sensitivity analyses are recommended to ensure model robustness and stability.

3. In terms of ground stations, it is advisable to increase the construction of CO\textsubscript{2} ground stations to enhance data reliability and coverage. Real-time monitoring data from ground stations can serve as crucial references for model validation and calibration, thereby increasing the credibility of the model in practical applications.

5. Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of this publication.

References


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