

Article

Refining Spatial and Temporal XCO₂ Characteristics Observed by Orbiting Carbon Observatory-2 and Orbiting Carbon Observatory-3 Using Sentinel-5P Tropospheric Monitoring Instrument NO₂ Observations in China

Kaiyuan Guo ^{1,2}, Liping Lei ^{1,2}, Mengya Sheng ^{3,4} , Zhanghui Ji ^{1,2} and Hao Song ^{5,*}

¹ Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; guokaiyuan20@mails.ucas.ac.cn (K.G.); leilp@aircas.ac.cn (L.L.); jizhanghui21@mails.ucas.ac.cn (Z.J.)

² International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China

³ School of Earth and Space Sciences, Peking University, Beijing 100871, China; shengmy@radi.ac.cn

⁴ China Highway Engineering Consultants Corporation, Beijing 100089, China

⁵ Guangxi Key Laboratory of Culture and Tourism Smart Technology, Guilin Tourism University, Guilin 541006, China

* Correspondence: 3001200117@email.cugb.edu.cn

Abstract: The spatial and temporal variations in the atmospheric CO₂ concentrations evidently respond to anthropogenic CO₂ emission activities. NO₂, a pollutant gas emitted from fossil fuel combustion, comes from the same emission sources as CO₂. Exploiting the simultaneous emissions characteristics of NO₂ and CO₂, we proposed an XCO₂ prediction approach to reconstruct XCO₂ data based on the data-driven machine learning algorithm using multiple predictors, including satellite observation of atmospheric NO₂, to resolve the issue of data gaps in satellite observation of XCO₂. The prediction model showed good predictive performance in revealing CO₂ concentrations in space and time, with a total deviation of 0.17 ± 1.17 ppm in the cross-validation and 1.03 ± 1.15 ppm compared to ground-based XCO₂ measurements. As a result, the introduction of NO₂ obtained better improvements in the CO₂ concentration responding to the anthropogenic emissions in space. The reconstructed XCO₂ data not only filled the gaps but also enhanced the signals of anthropogenic CO₂ emissions by using NO₂ data, as NO₂ strongly responds to anthropogenic CO₂ emissions ($R^2 = 0.92$). Moreover, the predicted XCO₂ data preferred to correct the abnormally low XCO₂ retrievals at satellite observing footprints, where the XCO₂-uncertainty field in the OCO-2 and OCO-3 products indicated a larger uncertainty in the inversion algorithm.

Keywords: XCO₂; gaps; NO₂; homologous emission; machine learning



Citation: Guo, K.; Lei, L.; Sheng, M.; Ji, Z.; Song, H. Refining Spatial and Temporal XCO₂ Characteristics Observed by Orbiting Carbon Observatory-2 and Orbiting Carbon Observatory-3 Using Sentinel-5P Tropospheric Monitoring Instrument NO₂ Observations in China. *Remote Sens.* **2024**, *16*, 2456. <https://doi.org/10.3390/rs16132456>

Academic Editor: Pavel Kishcha

Received: 28 March 2024

Revised: 27 May 2024

Accepted: 1 July 2024

Published: 4 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The increase in atmospheric carbon dioxide (CO₂) has significantly enhanced global climate warming [1], mainly caused by carbon emissions from human activities [2]. In order to mitigate global warming, countries are adopting measures to reduce and control their anthropogenic carbon emissions. Changes in atmospheric CO₂ concentrations can be used as an indicator to monitor the effects of anthropogenic CO₂ emission reductions and control [3,4], as well as to evaluate the impacts on natural ecological CO₂ emissions due to extreme climate change [5,6]. Thus, accurate data on atmospheric CO₂ concentrations in space and time can help us better understand the changing mechanisms of the atmospheric CO₂ concentration induced by anthropogenic CO₂ emissions and natural effects to support government decision-making policies for carbon emission reduction and control.

Satellite observations of atmospheric CO₂ have been used to obtain global CO₂ data effectively, with the advantage of global and period coverage. Currently, the XCO₂ (columnar average of molar fractions of carbon dioxide in dry air) retrievals derived from CO₂

observation satellites such as the Greenhouse Gases Observing Satellite (GOSAT), GOSAT-2, Orbiting Carbon Observatory-2 (OCO-2), and OCO-3 are available from 2009 [7–10]. These XCO₂ retrievals, however, have numerous gaps in space and time due to the satellite observation mode, cloudy or aerosol-contaminated conditions, data quality filtering, etc. These gaps result in missing the fine-tuning changes of CO₂ in space and time, making it difficult to correlate the correspondence of CO₂ changes with anthropogenic emissions. A data-driven methodology, which is different from the model-based data assimilation method, was developed in this research to fill the gaps and reconstruct satellite-observed XCO₂ data.

Applying the geo-statistics of satellite XCO₂ retrievals in space and time, some studies have developed a gap-filling method [11,12]. Global spatio-temporal XCO₂ datasets in a 1-degree grid over three-day or monthly periods from 2009 have been generated using GOSAT, GOSAT-2, OCO-2, and OCO-3 [13] and have been released in the HARVARD Dataverse [14]. The feasibility and reliability of these datasets have been verified by comparing them with the ground-based measures (The Total Atmospheric Carbon Column Observation Network, TCCON), cross-validation methods, and application analyses to reveal changes in CO₂ concentrations corresponding to anthropogenic emissions [15] and extreme climate change. The spatial resolutions of data generated by this method greatly depend on the number of XCO₂ retrievals; the more XCO₂ retrievals, the higher the spatial resolution. It has been found that this method can only generate data with a resolution of 0.5 degrees when using the available observation data from a single GOSAT, or a combination of GOSAT, OCO-2, and OCO-3, in order to guarantee the accuracy of the filled gaps and the reliability of the generated spatio-temporal data [16].

Recently, machine learning (ML) and deep learning methodologies for fusing multi-source data have been developed to reconstruct high-resolution XCO₂ data in space and time, driven by the increase in satellite observations that directly or indirectly capture changes in CO₂. Algorithms such as Light Gradient-Boosting Machine (LGB), GWNN-GWT, CatBoost (CatB), Random Forest (RF), and Multi-Layer Perceptron (MLP) have been applied to develop XCO₂ predictive models using satellite-observed CO₂ as the target variable and the multiple parameter data from multisource observations and measurements as predictor variables [17–20]. The choice of predictor variables strongly influences the accuracy and effectiveness of ML methodology. Previous research has used predictor variables such as NDVI, land use, population, meteorological reanalysis data, and model-simulated CO₂ data. However, parameters such as land use and population, suggested to describe the anthropogenic emissions, may not effectively interpret dynamic CO₂ emission activity, especially from constantly changing point sources. Furthermore, previous research lacks clarity on the extent to which raw XCO₂ retrievals at the footprints are modified after being recomputed based on the ML model and whether the reconstructed XCO₂ accurately reflects responses to anthropogenic CO₂ emissions.

Atmospheric NO₂ concentration data, obtained from observations of Sentinel-5 Precursor/Tropospheric Monitoring Instrument (S5P/TROPOMI), have been widely used in recent years to quantify anthropogenic CO₂ emissions, as both NO₂ and CO₂ emitted by human activity such as fossil fuel combustion originate from the same sources [21–24]. It has been shown that NO₂ observations have the potential to provide much stronger constraints to total annual fossil CO₂ emissions [25–27]. NO₂ emitted from a point source can be sensitively detected with minimal background disturbance, as NO₂ is a gas with a short lifetime of only a few hours [28–30]. CO₂, however, is a long-lived gas [31]. CO₂ emissions from sources with weak signals of 1–4 ppm are overshadowed by background levels exceeding 400 ppm, making it difficult to detect enhancements in CO₂ emissions with the existing signal-to-noise ratios of sensors in satellites. The magnitude of NO₂ columns is generally much larger than background levels around the emitting source, making it a suitable tracer for CO₂ emissions and capable of characterizing anthropogenic emissions. Therefore, NO₂ can be used as a tracer for CO₂ emissions, and previous research has demonstrated the effectiveness of simultaneously using satellite observations of NO₂

and XCO₂ to detect anthropogenic CO₂ emissions [32]. Using NO₂ as a proxy for fossil CO₂ allows for the quantification of anthropogenic emissions and distinguishes them from biogenic sources of CO₂ [33,34].

China is a country with high CO₂ emissions and heavy air pollution due to rapid economic development. Anthropogenic CO₂ emissions in China are spatially and inhomogeneously distributed, with a significant concentration in the eastern region characterized by high population density, numerous large cities, and industrial enterprises. Evidence of satellite-based CO₂ observations strongly correlating with satellite-based NO₂ observations have been demonstrated in this area [14]. The Chinese government has been taking measures to reduce carbon emissions and put forward clear time programs for emissions reduction and has outlined clear timelines for emission reduction and control to contribute to the mitigation of global climate warming [35–37]. Synergistic analysis of satellite-based CO₂ and NO₂ could enhance our ability to better monitor the effects of emissions reduction.

Focusing on the effectiveness of predictors and reconstructed XCO₂ based on ML, as described above, we introduced NO₂ data as one of the predictor variables. Our study area was the Chinese mainland, and we developed an ML-based method to reconstruct XCO₂ data from the satellite observations using multiple sources of data. We assessed the performance of reconstructed XCO₂ and the contribution and effect of NO₂ in reconstructing XCO₂ for the quantification of anthropogenic emissions and differentiation from biogenic sources of CO₂. Additionally, we analyzed the co-variation of NO₂ and CO₂ in response to the dramatic reduction in anthropogenic CO₂ emissions in a special scenario.

2. Materials and Methods

2.1. Data Used for Reconstructing XCO₂

The signals of atmospheric CO₂ observed by satellites mainly originate from anthropogenic CO₂ emissions, CO₂ uptake and emissions (fluxes) from land ecosystems, and atmospheric transmission fluxes over the target area. We collected parameter data significantly related to these sources CO₂ [38–41], along with XCO₂ retrievals derived from OCO-2 and OCO-3 observations from January 2019 to December 2022 in the study area, as shown in Table 1.

Table 1. Summary table of the multisource data used.

Acronym	Parameter	Source	Resolution		Product
			Space	Time	
fp-XCO ₂	XCO ₂ retrievals	OCO-2	2.25 km × 1.29 km	16 days	OCO-2 _L2_Lite_FP_11r
		OCO-3	2.25 km × 1.29 km	16 days	OCO-3 _L2_Lite_FP_10.4r
NO ₂	Atmospheric NO ₂ column	TROPOMI-S5P	0.01°	Monthly	Sentinel-5P OFFL NO ₂ : Offline Nitrogen Dioxide
NDVI	Normalized-difference vegetation index	MODIS	0.05°	Monthly	MOD13C2
D2M T2M U10 V10	2 m dewpoint temperature 2 m temperature 10 m U-wind component 10 m V-wind component	ERA5— fifth-generation ECMWF atmospheric reanalysis	0.1°	Monthly	Complete ERA5 global atmospheric reanalysis
MXCO ₂	Mapping XCO ₂	Mapped geostatistical method using XCO ₂ retrievals	0.5°	Monthly	Global land 0.5° mapping XCO ₂ dataset using satellite observations of GOSAT, OCO-2, and OCO-3 from 2009 to 2022
T-XCO ₂	Ground-based XCO ₂ data	TCCON	Point	-	TCCON data from Hefei (PRC)

The XCO₂ retrieval data were collected from the OCO-2 XCO₂ (OCO-2_L2_Lite_FP 11r) and OCO-3 XCO₂ (OCO-3_L2_Lite_FP 10.4r) data products from August 2019 to December 2022. According to the threshold settings of the variables `land_fraction < 90` and `XCO2_quality_flag = 0` in the raw satellite observation data file, the land area and data quality were filtered, respectively, and only land area and high-quality XCO₂ data were used for the analysis. The XCO₂ retrievals were derived from CO₂ sounder observations on the Orbiting Carbon Observatory-2 (OCO-2) satellite (launched in 2014) and the Orbiting

Carbon Observatory-3 (OCO-3) satellite (launched in 2019). The XCO₂ data products were generated by the Atmospheric CO₂ Observations from Space (ACOS) XCO₂ retrieval algorithm, which is an all-physical algorithm and has been filtered through recommended data screening and bias-corrected XCO₂ in the Lite files. The retrieval algorithm exhibited a significant regional bias, i.e., the surface pressure was incorrectly estimated in areas of high topographic variability, leading to a more significant bias (~1 ppm) [42]. Cloud-disturbed errors persist in the OCO-2 data, as noted by Massie et al. [43]. The deviation between the bias-corrected XCO₂ data and the Total Carbon Column Observing Network (TCCON) is 0.78 ± 1.14 ppm. Additionally, the temporal and spatial coverage of the observation data was not continuous, and there were many blank areas due to limitations of clouds and observation modes.

We collected three-level OFFL NO₂ data products in the study area from January 2019 to December 2022 through the GEE platform. The raw NO₂ observations were obtained from the Copernicus Ecosystem's Sentinel-5 Precursor observation satellite, a global atmospheric pollution monitoring satellite launched on 13 October 2017, which carries the Tropospheric Monitoring Raster Spectrometer (TROPOspheric Monitoring Instrument (TROPOMI)). Compared with previous remote-sensing instruments for monitoring atmospheric composition, the technical characteristics of the TROPOMI sensor have been greatly improved, with the signal-to-noise ratio increased by a factor of 1 to 5 [44]. The high temporal and spatial resolution and high-precision data observed by TROPOMI can be used to study urban-scale air pollution conditions and air quality monitoring for major events, providing a scientific data basis for policy decisions related to pollution emissions [23,45–47]. TROPOMI inverts atmospheric NO₂ data using Differential Optical Absorption Spectroscopy (DOAS) in the UV–Vis spectral band. The data processing system is based on the DOMINO-2 product and the Ozone Monitoring Instrument (OMI) EU QA4ECV NO₂ reprocessing dataset, and the algorithm is further optimized with the inversion–assimilation–modeling algorithm that uses the TM5-MP chemical transport model at a $1^\circ \times 1^\circ$ resolution in the global three dimensions as the basic element [44]. Since October 2018, the Copernicus ecosystem has provided NO₂ data through three Level 2 data products including Near Real-Time (NRTI), Offline (OFFL), and Reprocessing (RPRO). Some of these products may have undergone multiple processing iterations, potentially resulting in improved data quality but delayed availability. Since Level 2 data products are generated on a per-orbit basis, further processing is required for areas where scanned orbits are repeatedly observed. Google Earth Engine (GEE) provides Level 3 data products with a spatial resolution of 1.1132 km, derived from Level 2 data that undergo screening for quality and grid-based reprocessing. The satellite-observed NO₂ data are the total vertical column concentration in mol/m². Since the effect of data observation noise can lead to negative observations in the clean region, only anomalous observations below -0.001 mol/m² need to be removed, as recommended in the data description document. In this study, monthly-averaged atmospheric NO₂ data were calculated for the study area using the GEE platform for online processing, converted to common units of molec/cm² by a multiplication factor of 6.02214×10^{19} , and the spatial resolution was standardized to a 0.01° grid for use in this study.

We collected normalized vegetation index (NDVI) and ECMWF Reanalysis v5 (ERA5) data during the same period as the XCO₂ data, which were used as predictors to account for the effects of the vegetation ecological system and atmospheric transport on the atmospheric CO₂ concentrations. NDVI, obtained by the Moderate Resolution Imaging Spectroradiometer (MODIS), can account for the CO₂ uptake and release of the vegetation ecological system. The seasonal variation in NDVI, caused by vegetation CO₂ uptake and release, results in seasonal variations in the CO₂ concentration [48,49]. Meteorological data, including T2M (2 m temperature), D2M (dew point temperature), U10 (10 m U-wind fraction), and V10 (10 m V-wind fraction) were obtained from the ERA5 data. These data were derived from the fifth generation of the reanalysis climatic dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF).

Mapping-XCO₂ (MXCO₂), was used to constrain the global variation in XCO₂ as one of the predictors in modeling XCO₂ prediction using ML. MXCO₂, which represents spatio-temporally continuous XCO₂ data, was generated through geo-statistical analysis of the XCO₂ retrievals in space and time obtained from multiple satellite observations, including GOSAT (April 2009 to August 2014), OCO-2 (September 2014 to December 2020), and OCO-3 (August 2019 to December 2022). This dataset is available on the HARVARD Dataverse (<https://dataverse.harvard.edu/>, accessed on 17 August 2021) [14]. The Mapping-XCO₂ dataset, which exhibited a -0.29 ± 1.04 ppm deviation compared to the TCCON, enabled examination of temporal and spatial changes at global and regional scales [50,51].

The ground-based XCO₂ data released from 2019 to 2022 from Hefei station (117.17°E, 31.9°N) in China, which is in the Total Atmospheric Carbon Column Observation Network (TCCON) [32,52] (<https://tcon-wiki.caltech.edu/>, accessed on 12 May 2021), were used to validate the accuracy of the reconstructed data by the predictive model. TCCON data, an atmospheric CO₂ column concentration derived by ground-based Fourier Transform Spectrometer (FTSE) observations, have been widely used to validate greenhouse gas products from space-based observations.

2.2. Methodology

The modeling XCO₂ prediction at spatio-temporal location *i* can be expressed as follows using the NO₂ data in addition to the other parameter variables.

$$XCO_{2[i]} = f(NO_{2[i]}, NDVI_{[i]}, T2M_{[i]}, D2M_{[i]}, U10_{[i]}, V10_{[i]}, MXCO_{2[i]}, T_{[i]}) \quad (1)$$

where *f* represents the function obtained by the ML algorithm based on the training data. XCO_{2[i]} is the predicted XCO₂ by model *f* at spatio-temporal location *i*; the variables, including NO_{2[i]}, NDVI_[i], T2M_[i], D2M_[i], U10_[i], V10_[i], MXCO_{2[i]}, and T_[i], are the predictors of the spatio-temporal constrain fields for the atmospheric CO₂ concentration. Among the predictive parameters, D2M and T2M denote the 2 m dew point temperature and 2 m air temperature, respectively, elucidating the impact of meteorological temperature conditions on the atmospheric CO₂ concentration [53]. NO₂ serves as an indicator of anthropogenic CO₂ emissions, stemming from the same sources [24]. TimeIndex represents timestamps expressed in 1–12 corresponding to the current month, which was used to capture the temporal seasonal variation in the CO₂ concentration. NDVI represents physical quantities of the normalized-difference vegetation index, which is used to characterize CO₂ uptake and emissions by surface vegetation ecosystems [54]. U10 and V10 represent the physical quantities of the 10 m U-wind component and V-wind component, which are used to characterize the transport effects of the atmospheric wind field [53]. Additionally, MXCO₂ represents the spatio-temporal continuum of atmospheric XCO₂, serving as a constraint on the overall spatial distribution [14].

The framework for refining XCO₂ characteristics in space and time using simultaneously emitted NO₂ alongside CO₂, as shown in Figure 1, includes ML-based reconstruction of XCO₂ data and assessment of the NO₂ field constraint to modeling CO₂.

2.2.1. Modeling XCO₂ Prediction and Reconstructing XCO₂ in Space and Time

We built an XCO₂ prediction model by co-locating the fp-XCO₂ as the Y variable with the predictor variables as the X variable, including NO₂, NDVI, T2M, D2M, U10, V10, MXCO₂, and TimeIndex in Equation (1).

The training samples used for modeling XCO₂ predictions are especially critical, as the type of predictor variables and their spatial scale greatly influence the modeling accuracy and the performance of the predicted XCO₂. In addition to incorporating NO₂, we applied MXCO₂ data generated through mathematical geostatistics of the XCO₂ retrievals rather than simulated XCO₂ such as CarbonTracker, as carried out by other researchers [17]. This approach aims to constrain the spatio-temporal relationship of XCO₂, relying on a data-driven method that identifies a mathematical (constraining) relationship between observations of XCO₂ from OCO satellites and multiple parameters. MXCO₂ demonstrates

the spatio-temporal relationships among the XCO₂ retrievals and addresses gaps in satellite observations through mathematical geostatistical methods [13,14].

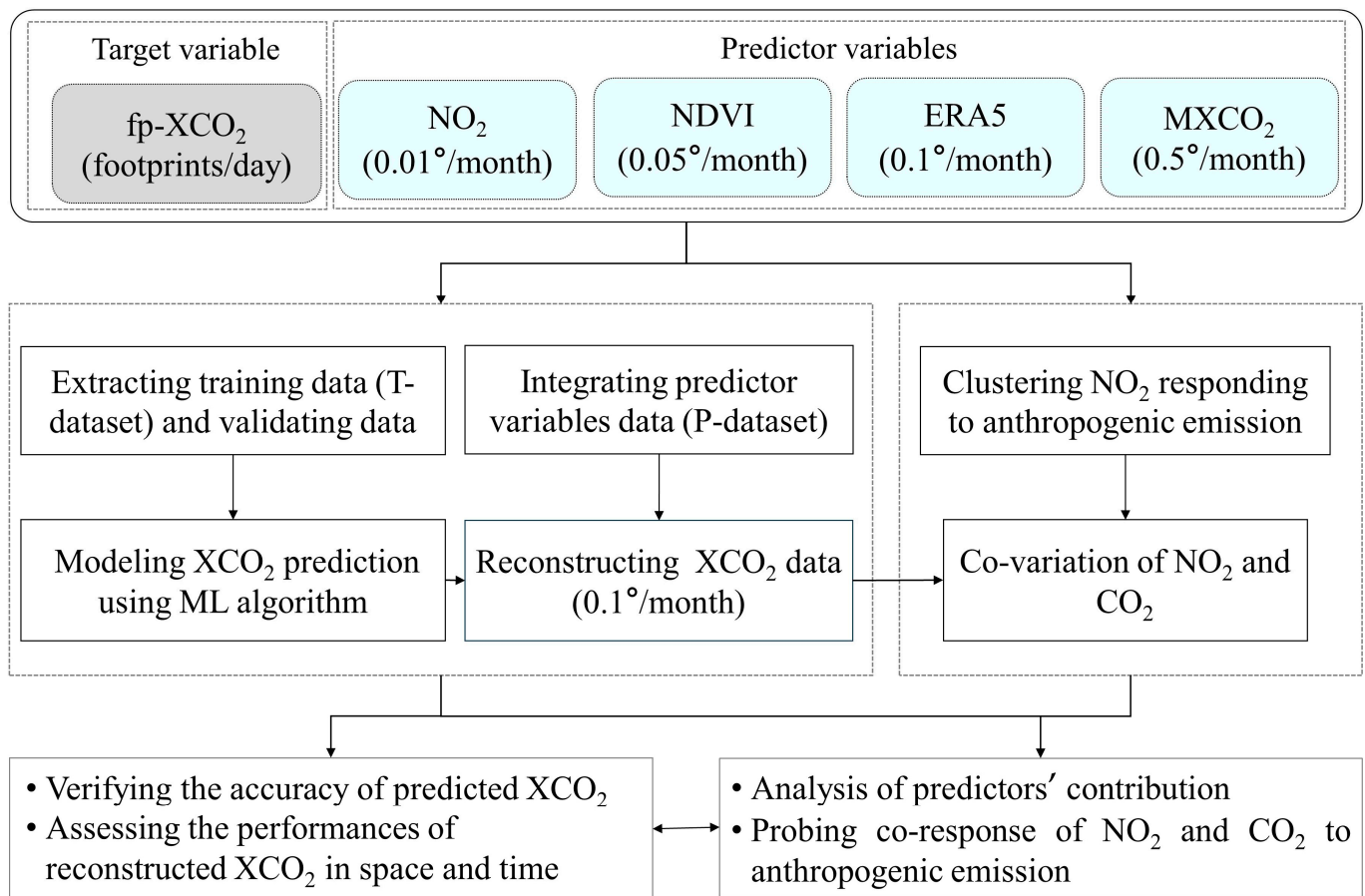


Figure 1. Framework for predicting XCO₂ based on ML, involving reconstruction XCO₂ data and analysis of the NO₂ field to constrain XCO₂ predictions.

We assessed the predictive effects of model training data at different scales of a 0.5° grid that is the same as MXCO₂ and a 0.1° grid that is the same as the reconstructed XCO₂ grid using four ML algorithms, CatB, LGB, MLP, and LSTM, respectively. The results showed that both 0.5° and 0.1° training data without MXCO₂ demonstrated much lower prediction accuracy than did those with MXCO₂ via cross-validation for all ML algorithms, indicating that MXCO₂ can help constrain the spatial relationship of XCO₂ during model training. Furthermore, XCO₂ predictions from the 0.5°-grid training data demonstrated higher accuracy than those from the 0.1°-grid training data, because resampling MXCO₂ to a 0.1° grid within a 0.5° grid could result in the loss of spatio-temporal characteristic information due to the smoothing of scale-up resampling.

As a result of the training data analysis above, we generated a dataset consisting of a set of co-located fp-XCO₂ and predictors (NO₂, NDVI, T2M, D2M, U10, V10, and MXCO₂) from the years 2019 to 2022 in a 0.5° grid/month unit. We then split it into two parts; one is 10% of this dataset, which was randomly extracted as a cross-validation dataset, while the remaining data were used as the training dataset for modeling XCO₂ predictions (hereinafter referred to as the T-dataset).

Additionally, we detrended the impact of background on XCO₂ (hereafter referred to as dXCO₂) by subtracting the monthly average values of the entire XCO₂ dataset in the study area for the XCO₂ in values in both the T-dataset and P-dataset. This step removed effects from the background due to the atmospheric transportation and the accumulations

of CO₂ over a long lifetime. Lastly, the XCO₂ predictions were calculated for each grid by adding the predicted dXCO₂ value to the background trend value used in the same month.

We assessed the performance of the ML algorithms, including CatB, LGB, MLP, and Long Short-Term Memory (LSTM), which have been proven more effective for nonlinear prediction. CatB and LGB are decision tree models, while MLP and LSTM are neural network models. After parameter testing and tuning, the optimal parameter settings for each model are shown below. The CatB model had 10,000 iterations and a learning rate of 0.001. The LGB model had 500 leaf nodes, a model depth of 8, and a learning rate of 0.001. The MLP model had (64, 32, 8) hidden neurons and a learning rate of 0.001. The LSTM model had 50 hidden neurons and a learning rate of 0.001. As a result, the predictions by CatB significantly outperformed those of LGB, MLP, and LSTM through model cross-validation, demonstrating reasonable performance in predicting XCO₂ in both space and time. CatB showed the smallest deviation from fp-XCO₂, which is 0.17 ± 1.17 ppm ($R^2 = 0.81$), while LGB, MLP, and LSTM showed deviations of 0.21 ± 1.18 ppm ($R^2 = 0.79$), 1.49 ± 3.18 ppm ($R^2 = 0.47$), and 1.03 ± 2.16 ppm ($R^2 = 0.58$), respectively (see Figure A1). Specific comparative validation screenings of the model results are outlined in the Discussion section. Therefore, we applied CatB for modeling XCO₂ predictions to reconstruct XCO₂.

2.2.2. Evaluation of Reconstructed XCO₂ and Effects of NO₂ Constraints

The multiple predictor variable data (NO₂, NDVI, T2M, D2M, U10, V10, and MXCO₂) in Equation (1), each with different spatial and temporal resolutions as shown in Table 1, were integrated into the dataset at a resolution of 0.1° grid/month unit for each variable from the years 2019 to 2022 (hereinafter referred to as the P-dataset). This integration was achieved through resampling, where those variables with a resolution of less than 0.1° in the grid were averaged within each grid for each variable, and MXCO₂ was resampled using the nearest-neighbor method.

XCO₂ was predicted using the XCO₂ prediction model built previously with the P-dataset and reconstructed XCO₂ data at a resolution of 0.1°/month from 2019 to 2022, as shown in Figure 1.

The accuracy and performance of the reconstructed XCO₂ were implemented through three methods, namely, model cross-validation, ground-based observed XCO₂, and analysis of the predictor's contribution and co-response of NO₂ and CO₂ to anthropogenic emissions, as well as the effects of NO₂ for disentanglement from biogenic sources of CO₂.

We extracted ground-based XCO₂ data from the TCCON site in Hefei, obtained between 12:00 and 14:00 around the local observation time of OCO series satellites (13:30), and calculated monthly averaged values (hereafter referred to as T-XCO₂). The paired data of the co-located predicted XCO₂ and T-XCO₂ were used to calculate three metrics, the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute percentage error (MAPE), using the following equations to evaluate the accuracy of the predicted XCO₂.

The coefficient of determination, R-squared (R^2), was calculated by Equation (2), where a value closer to 1 indicates a better fit of the model.

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y})}{\sum(Y_i - \bar{Y})} \quad (2)$$

The formula for the root mean square error (RMSE) and mean absolute prediction error (MAPE) are as follows, respectively:

$$X_{RMSE} = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}} \quad (3)$$

$$X_{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{Y_i - \hat{Y}_i}{Y} \right| \quad (4)$$

where Y_i is the observed value, \hat{Y} is the model predicted value, \bar{Y} is the mean of the observations, and N is the number of samples.

We applied the SHAP (Shapley Additive exPlanations) method to assess the contribution of predictor variables to the XCO_2 predictions and to explore the contribution of NO_2 co-emitted with fossil CO_2 during combustion, along with the co-variation of NO_2 and CO_2 in response to anthropogenic CO_2 emissions through clustering analysis. Clustering analysis of the spatio-temporal characteristics of NO_2 can demonstrate the emission intensity in clustered area units, thereby integrating different chemical processes of NO_2 and CO_2 in the atmosphere. This approach allows us to investigate the co-variation of NO_2 and XCO_2 in response to anthropogenic CO_2 emissions by statistically analyzing clustered areas, which helps avoid the effects of chemical processes in NO_2 and CO_2 . The clustering analysis of NO_2 data at a resolution of 0.1° /month from the years 2019 to 2022 classified the study area into 14 categories using the K-means method.

3. Results

3.1. Model Prediction Accuracy and Performance of the Reconstructed XCO_2

3.1.1. Accuracy of the XCO_2 Predictions

Figure 2 demonstrates the predicted XCO_2 from the CatB-based built predictive model compared to both the satellite-based fp- XCO_2 and the ground-based observed XCO_2 .

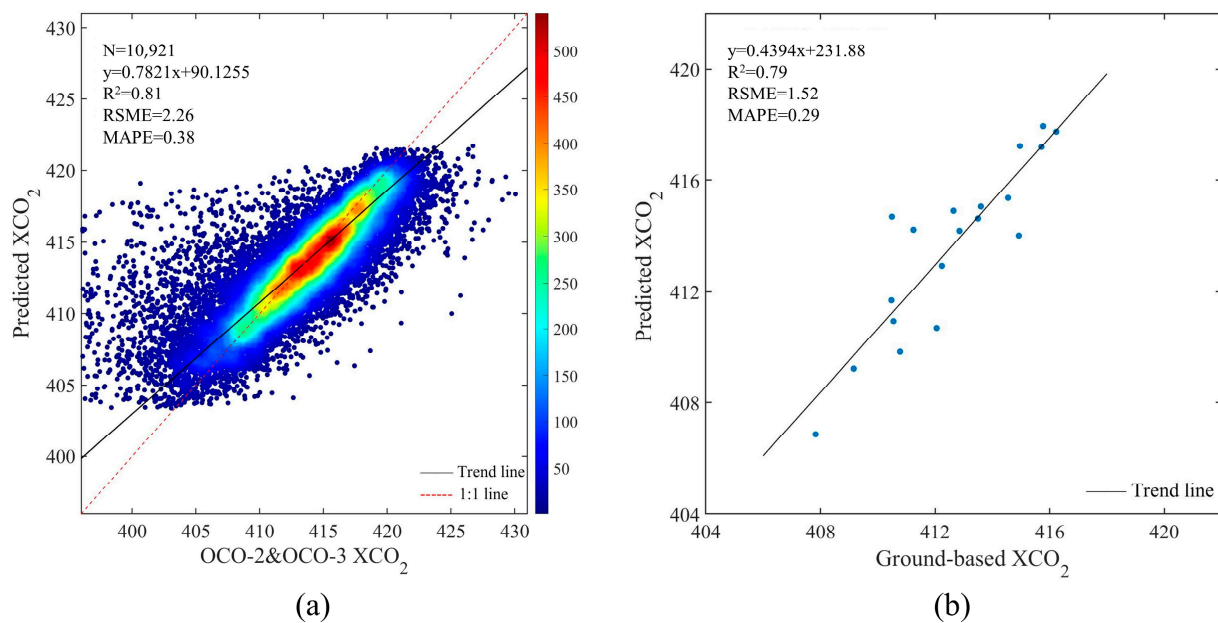


Figure 2. Validations of the model predicting XCO_2 through (a) cross-validation and (b) comparison with T- XCO_2 .

The results from the cross-validation, as shown in Figure 2a, demonstrate that the R^2 , RMSE, and MAPE of the XCO_2 predictions compared with the fp- XCO_2 are 0.81, 2.26 ppm, and 0.38, respectively, with a total deviation of 0.17 ± 1.17 ppm. The model's predicted XCO_2 tended to be higher than the satellite-observed XCO_2 in the low-value range of fp- XCO_2 , which implies that the model predictions increase those low XCO_2 retrievals mostly in the areas without anthropogenic emissions or irregular values. The validation of the predicted XCO_2 against T- XCO_2 , shown in Figure 2b, demonstrates R^2 , RMSE, and MAPE values of 0.79, 1.52 ppm, and 0.29, respectively. The total deviation of the predicted XCO_2 was 1.03 ± 1.15 ppm, similar to the deviation of the co-located fp- XCO_2 and TCCON site, which was 0.89 ± 1.23 ppm.

3.1.2. Performance of the Model Predictions

We applied the XCO₂ prediction model constructed by the CatB algorithm and T-dataset to generate an XCO₂ dataset at a resolution of 0.1° grid/month, denoted as Cm-XCO₂, from 2019 to 2022 using the predictor dataset from 2019 to 2022. Figure 3 shows the averaged values for each grid calculated from the model-predicted XCO₂, as well as the original satellite-observed XCO₂ values within the 0.1° grid. The model-predicted XCO₂ exhibited significantly finer spatial features and appeared more reasonable than the original satellite XCO₂ retrievals, particularly evident in the resolution of anomalous high XCO₂ in the southern region, and the western Taklamakan Desert in the original satellite observations in the study area.

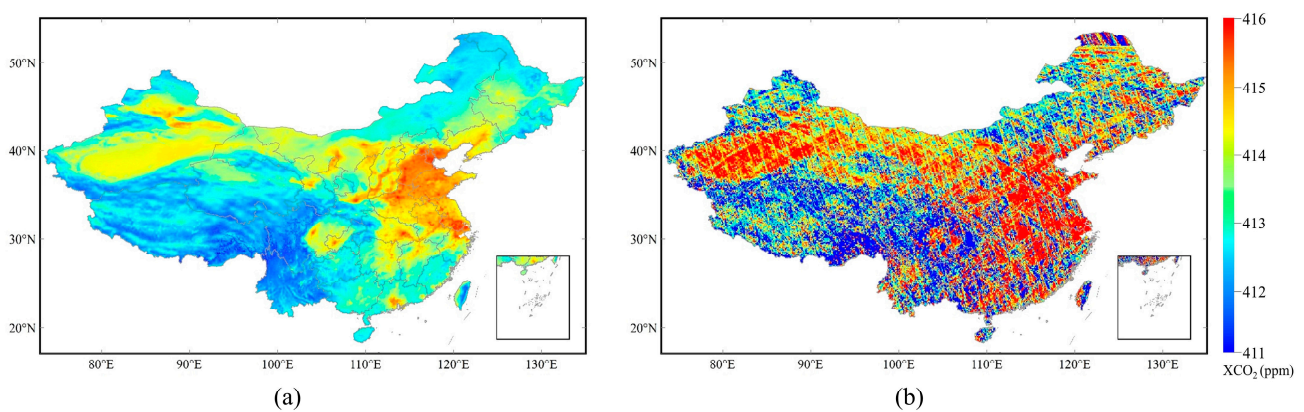


Figure 3. Spatial distribution of the mean XCO₂ values in the study area. (a) Predicted XCO₂ (Cm-XCO₂) and (b) fp-XCO₂ for each 0.1° grid.

We assessed the impact of model prediction on the original satellite-observed XCO₂ (fp-XCO₂) that was re-calculated by the predicting model comparing the co-located fp-XCO₂ and Cm-XCO₂. Figure 4 shows a diagram, histogram, and the difference of the co-located Cm-XCO₂ and fp-XCO₂, along with a comparison with the posterior uncertainty in the XCO₂ algorithm released in L2 product data [55].

It can be found from Figure 4a,b that the model predicting XCO₂ narrowed the range of fp-XCO₂ values from 399.06–428.67 ppm to 403.04–424.14 ppm within the cumulative frequency range of 5–95%. The low fp-XCO₂ values of 395–402 ppm were enlarged to 403–417 ppm using the predictive model, occurring mostly in mountainous areas with highly undulating terrain along the edge of the Tibetan Plateau, as shown in Figure 4c. Here, the differences between Cm-XCO₂ and fp-XCO₂ ranged from 1 to 4 ppm. It is known that a CO₂ concentration of less than 400 ppm is unreasonable since 2019 according to the global CO₂ background concentration (see Figure A2), where the minimum of CO₂ concentration was 408 ppm in 2019, as observed from ground-based XCO₂ measurements from the TCCON station in Hefei shown in Figure 2b.

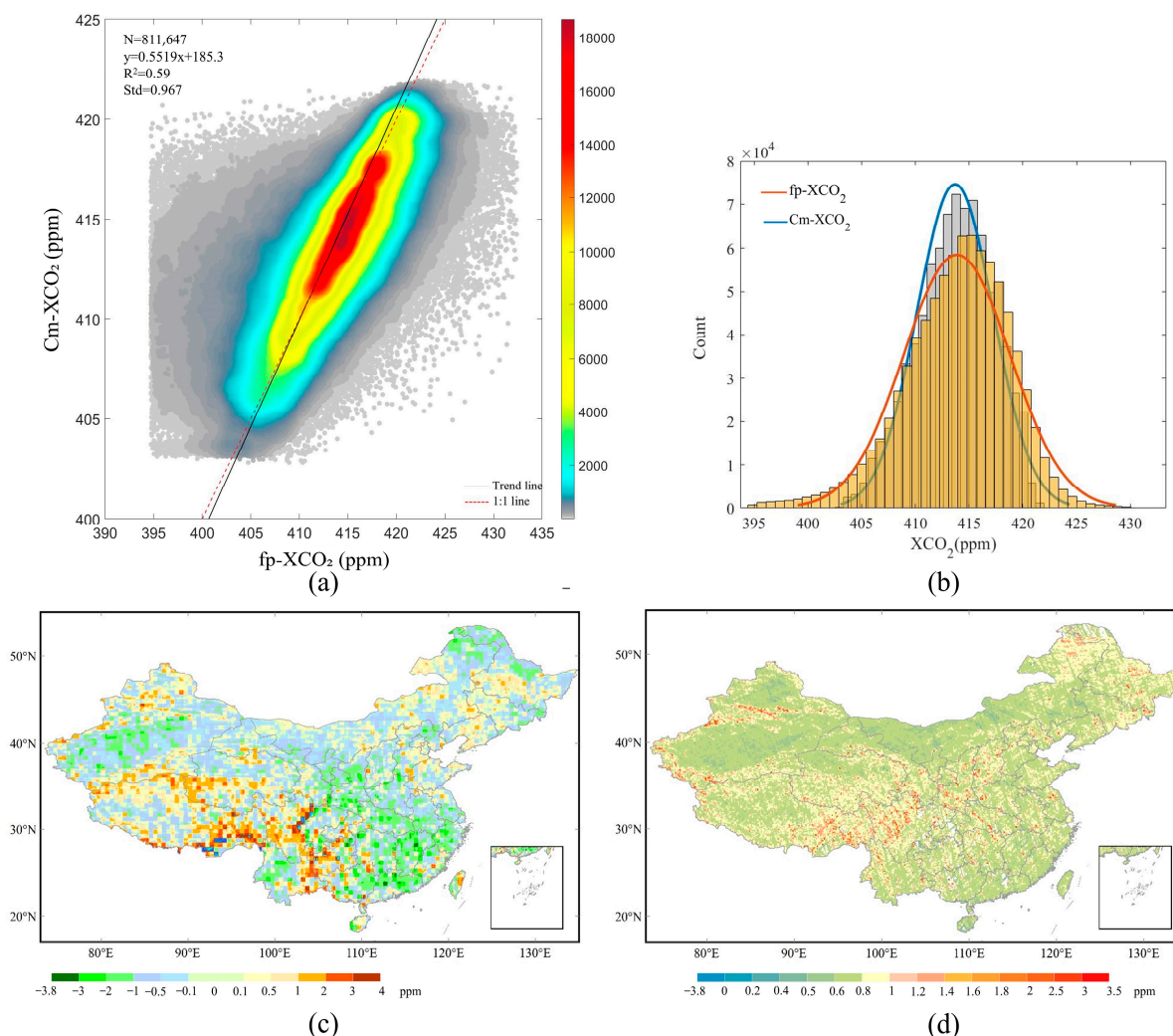


Figure 4. Effects of the predictive model for the original satellite-observed XCO₂ (fp-XCO₂) that was re-calculated using this model. (a) Comparison of the co-located fp-XCO₂ and Cm-XCO₂ based on the monthly averages of the XCO₂ retrievals within a 0.1° grid from 2019 to 2022. (b) Histograms of the co-located fp-XCO₂ and Cm-XCO₂. (c) Mean difference between the co-located Cm-XCO₂ and fp-XCO₂ between 2019 and 2022 in the study area. (d) Mean posterior uncertainty in XCO₂ from the L2 product algorithm.

Additionally, comparing Figures 4c and 4d reveals that the large differences between Cm-XCO₂ and fp-XCO₂ correspond to large posterior uncertainty in the XCO₂ retrieval data products derived by the OCO L2 algorithm [56] over the edge of the Tibetan Plateau, respectively, which has higher elevation and topographic variations. This is likely because the sharp variations in elevation in the undulating terrain result in anomalous values of fp-XCO₂. Currently, the remaining uncertainties in XCO₂ retrievals may indicate large biases, especially in regions with large topographic variations [54]. Therefore, the reconstructed XCO₂ for these abnormally low values of fp-XCO₂, by constraint computation of XCO₂ through the ML of the multiple variables, rectified these XCO₂ retrievals with large uncertainties, which could be induced by the uncertainty of input parameters in the algorithms of XCO₂ retrievals [55].

The high values of fp-XCO₂ ranged from 422 to 430 ppm, which are currently unreasonable, as the Global Atmosphere Watch Programme (GAW) global background station located in Waliguan, Qinghai, China, observed that the atmospheric CO₂ concentrations peaked at 419.3 ppm in 2022 [57]. These high values are predicted to be reduced to low

values ranging from 415 to 420 ppm in $C_m\text{-XCO}_2$, mostly in eastern areas with differences ranging from -2 ppm to -3 ppm.

3.2. Co-Variation of CO_2 and NO_2 to Anthropogenic CO_2 Emissions

3.2.1. NO_2 Constraints Enhancing the XCO_2 Response to Anthropogenic Emissions

The larger the SHAP value of a predictor variable, the greater its contribution to the model prediction. The contribution of an individual variable is ranked in order of contributions by the average SHAP value, as shown in Figure 5. This is defined as the average of the absolute SHAP value (in ppm) and the standard deviation calculated from the training data from 2019 to 2022.

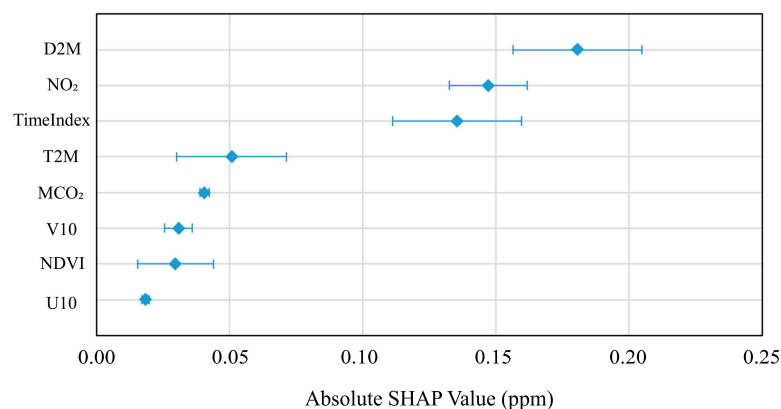


Figure 5. Contribution of the predictor variables in modeling XCO_2 using the CatB algorithm.

The results of the SHAP values show that D2M, NO_2 , and TimeIndex significantly contributed to modeling XCO_2 predictions much more than the other variables. The largest contributor, D2M, and third-largest, TimeIndex, are likely significant because the distinct seasonal variation in CO_2 is strongly impacted by CO_2 uptake and release in terrestrial ecosystems, which generally depend on temperature and water vapor, as explained by D2M and monthly time. The contribution of the vegetation index NDVI was low, ranking second from the bottom. The spatial NDVI demonstrated the different seasonal variations in CO_2 absorption capacity over various land surfaces with different vegetation and non-vegetation densities. Wind can result in CO_2 flowing in space, diminishing the CO_2 signals in the atmosphere from CO_2 fluxes of land ecosystem surfaces [19].

NO_2 ranked second in the contribution of multiple variables to modeling XCO_2 prediction. NO_2 , a pollutant gas emitted from fossil fuel combustion, strongly affects the modeling of XCO_2 prediction due to its co-emission with CO_2 . NO_2 showed the strongest linear relationship with $C_m\text{-XCO}_2$ ($R^2 = 0.57$) among the predictor variables (see Figure A3), while D2M and NDVI showed a nonlinear relationship with $C_m\text{-XCO}_2$. Furthermore, we investigated the co-variation of NO_2 and XCO_2 in relationship to anthropogenic CO_2 emissions.

NO_2 , as a tracer of CO_2 emissions from fossil fuel combustion, can constrain the predictive model to enhance the information on CO_2 from the anthropogenic emissions. The spatial distribution of NO_2 , as shown in Figure 6a, generally aligned with the anthropogenic CO_2 emissions (Figure 6b), and the hotspots of emissions demonstrated the spread of high NO_2 concentrations. NO_2 showed a strong response to the anthropogenic CO_2 emissions with an R^2 of 0.92, as shown in Figure 6c, when calculating their correlation between NO_2 and EDGAR emissions using the averaged values of the grids within each clustered area derived from the spatio-temporal clustering analysis of NO_2 data. This result implies that NO_2 , as a predictor variable, could strengthen the response to anthropogenic CO_2 emissions in modeling XCO_2 predictions. The predicted XCO_2 for those locations of the satellite-observed XCO_2 ($fp\text{-XCO}_2$) showed an R^2 of 0.79 with the co-located anthropogenic emission, which is higher than the R^2 of $fp\text{-XCO}_2$ ($R^2 = 0.66$). Here, the CO_2 response to anthropogenic emissions (66–79%) was lower than that of NO_2 (92%), reasonably indicating

that atmospheric CO₂ still includes natural emissions from land ecosystems and transported CO₂ fluxes.

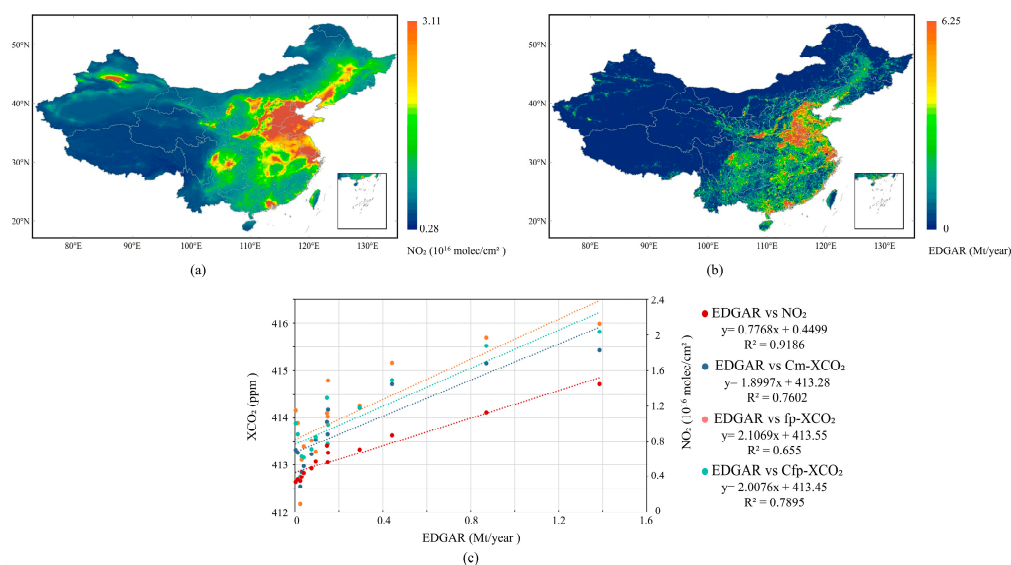


Figure 6. The relationship between NO₂ and anthropogenic CO₂ emissions from EDGAR from 2019 to 2022. (a) Annual mean atmospheric NO₂ concentration. (b) Annual mean anthropogenic CO₂ emissions (2019–2022). (c) NO₂ and XCO₂ response to anthropogenic CO₂ emissions from EDGAR calculated based on the averaged values of the grids within each clustered area derived from the spatio-temporal clustering analysis of NO₂ data.

The relationship between the co-located NO₂ and fp-XCO₂ reached up to 0.74 based on the clustered areas, compared with the calculations based on the grids ($R^2 = 0.57$ in Figure A3b). The clustered areas derived from the spatio-temporal clustering analysis of NO₂ data smoothed the impacts of different spreads and lifetimes between NO₂ and CO₂ in the grid by averaging said grids for each clustered area. The strong correlations between the predicted XCO₂ (Cfp-XCO₂ and Cm-XCO₂) and NO₂, with R^2 values of 0.86 and 0.92, respectively, as shown in Figure 7, indicated that NO₂, as one of the predictors, can constrain XCO₂ in modeling XCO₂ predictions. This is consistent with Emily G. Yang et al.'s suggestion that the satellite-based NO₂ can serve as a proxy species to constrain CO₂ [34].

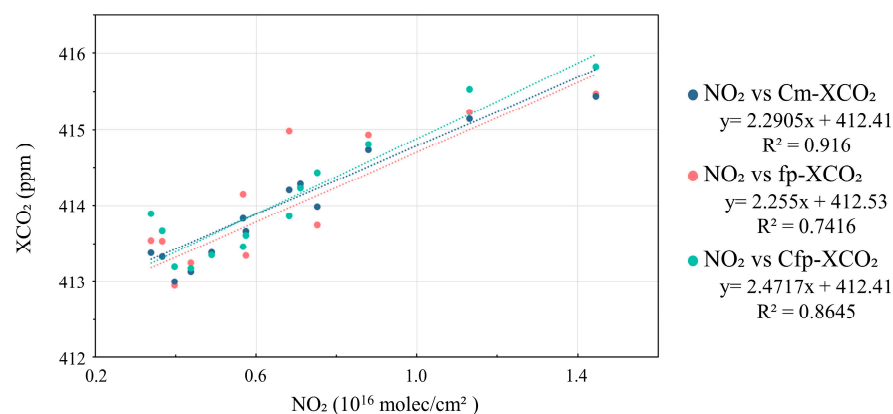


Figure 7. The relationship between NO₂ and XCO₂ calculated based on the averaged values of the grids within each clustered area derived from the spatio-temporal clustering analysis of NO₂ data.

3.2.2. Co-Response of NO₂ and CO₂ Concentrations to Anthropogenic Emissions under Special Scenarios of Human Activity

The significant reduction in human activities during the 2020–2022 COVID-19 period resulted in decreases in NO₂ and CO₂ from anthropogenic emissions [14]. We analyzed the co-variation of CO₂ and NO₂ under this particular reduction in anthropogenic emissions from 2019 to 2022 to find how NO₂ and XCO₂ synchronized in response to an abnormal reduction in anthropogenic emissions. Based on the clustering results of the spatio-temporal characteristics of the NO₂ concentration and anthropogenic CO₂ emissions in China, we selected 13 regions of interest (ROIs) with high NO₂ values and anthropogenic emissions, as shown in Figure 8a. ROI1–ROI13 (referred to as R1–R13 in Figure 8) indicate Wuhan (R1-Wuhan), Shanghai (R2-Shanghai), the Yangtze River Delta (R3-YRD), Beijing–Tianjin–Hebei (R4-BTH), Xi’an (R5-Xian), Shanxi Coal Mining Belt (R6-ShanxiCoal), Jinan (R7-Jinan), Chengdu–Chongqing (R8-ChengduC), Guangzhou (R9-Guangzhou), the strip west-northwest of Urumqi toward Shihezi (R10-Wulumuqi), Yinchuan (R11-Yinchuan), the strip south of Hohhot (R12-Huhehaote), and Shenyang (R13-Shenyang).

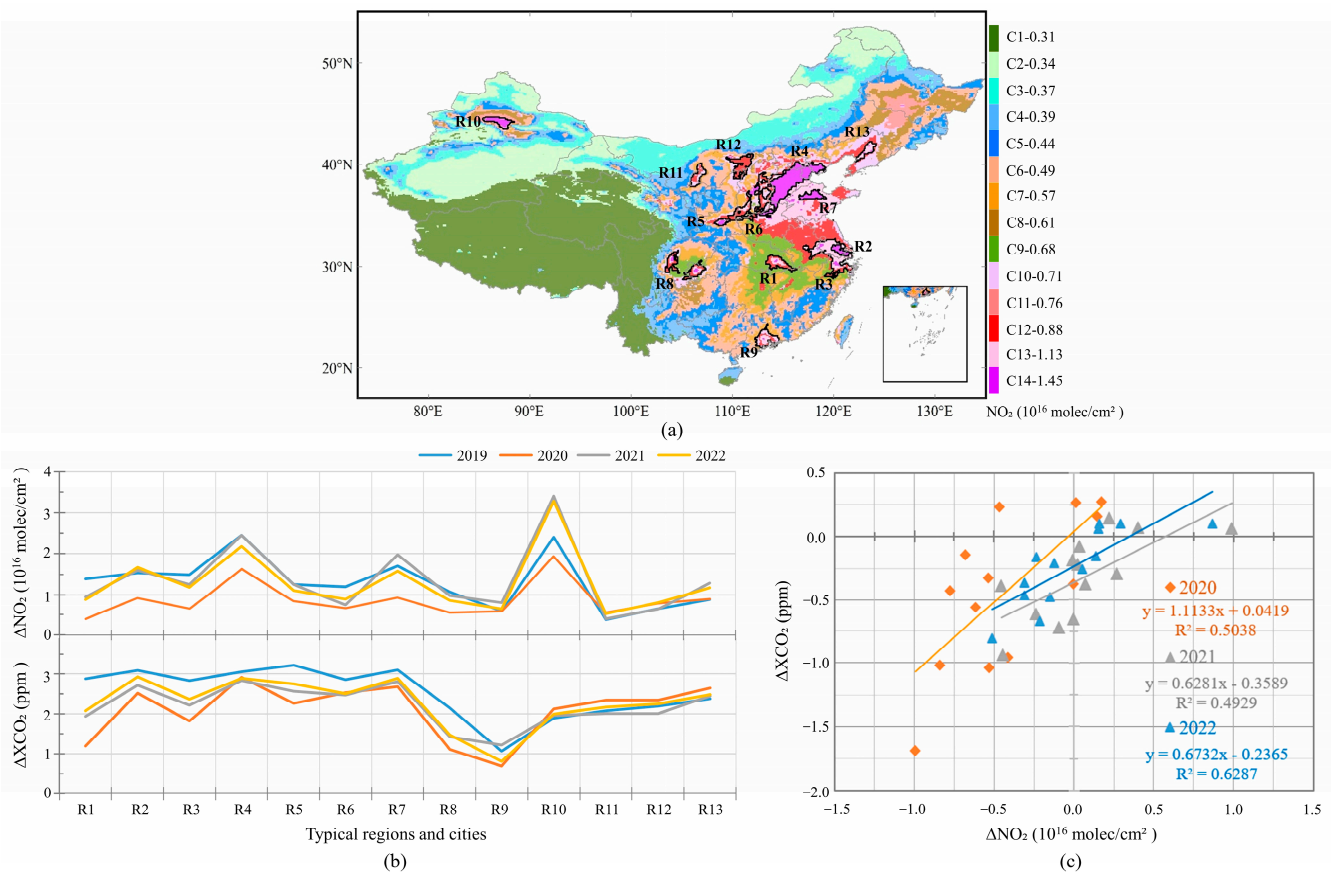


Figure 8. Analysis of NO₂ and XCO₂ in regions of interest (ROIs) with (a) ROI locations where the background map shows the clustering results of NO₂ data from 2019 to 2022 and the legend presents the 14 classes and averaged values of NO₂ from 2019 to 2022 for each class to the left and right of symbol, (b) yearly average values of ΔNO₂ and ΔXCO₂ (y-axis) for each ROI (x-axis) in the winter (December–February) for the four years of 2019–2022, and (c) the relationship between NO₂ and XCO₂, which is calculated by using 2019 yearly averages as contrasting values for each ROI in the three years 2020, 2021, and 2022, respectively.

It is known that the maximum anthropogenic emissions generally occur in the winter season when the XCO₂ shows the largest enhancements [14]. Therefore, we calculated the enhancement relative to the background value in the winter (December–February) for each ROI (winter value in ROIs minus the winter average of the overall study area in the same

year) for the four years from 2019 to 2022 (hereafter referred to as ΔXCO_2 and ΔNO_2), as shown in Figure 8b. Figure 8b shows that both ΔNO_2 and ΔXCO_2 , after 2019, ahead of the COVID-19 outbreak and control, were lower than the values in the normal year of 2019 from R1 to R7, especially in R1-Wuhan in the eastern part of China, where ΔXCO_2 was around 3 ppm. These yearly changes clearly respond to the unexpected reduction of human activity in 2020 and 2021 during the COVID-19 outbreak and control, with human economic activity not returning to 2019 levels in 2022. The reduction in anthropogenic activity during the COVID-19 period resulted in a decrease in ΔNO_2 of $0.4\text{--}1.0 \times 10^{16}$ molec/cm² and in ΔCO_2 of 0.5–1.7 ppm compared to 2019. The largest decrease in ΔCO_2 , 1.7 ppm, occurred when emissions from human activity were minimized in Wuhan.

The ΔCO_2 in R8–R13 were less than those of R1–R7, and the interannual variations were also small due to a lesser impact from COVID-19. It should be noted that ΔNO_2 in R10-Urumqi presented the largest value among the ROIs abnormally, with the maximum value, 3.4×10^{16} molec/cm², occurring in 2021, not in the normal year of 2019. The maximum interannual difference was up to 75% during 2019–2022, while ΔCO_2 did not present similar changes to ΔNO_2 correspondingly, with the maximum interannual difference being only 12%. This abnormal ΔNO_2 was due to high emissions from many air-polluting enterprises (mainly the coal and chemical industry and non-ferrous metals) that have been rapidly increasing in recent years in this region, resulting in it having the worst air pollution in China in recent years. The discrepancy, where the change in CO_2 did not correspond to the change in NO_2 , was likely caused by two reasons. One reason is probably due to the fossil fuel combustion processes in these air-polluting enterprises, of coal chemical industry and non-ferrous metal, in which NO_2 , the pollutant gas emitted by companies, has been rising, while CO_2 has reached its maximum. That is, the emitted CO_2 does not respond to an increase in emitted NO_2 from these sources when NO_2 emissions increase beyond a certain level. Another reason is probably due to the uncertainties of XCO_2 data; the aerosols, which are one of the key input parameters in the XCO_2 retrieval algorithm, likely introduced large biases due to the heavy air pollution (see Figure 4d), even if the ML predictions (see Figure 4c) modified some of their biases.

The extreme reductions in anthropogenic emissions in 2020 and 2021 impacted the corresponding variation in NO_2 and XCO_2 when comparing NO_2 and XCO_2 in these two years to the normal year 2019, as shown in Figure 8c. We found that the R^2 in 2020 and 2021 (0.49–0.50) was less than in 2022 (0.63) when anthropogenic emissions had begun to recover but not yet returned to 2019 levels. This implies that CO_2 may inaccurately respond to NO_2 changes in these extreme scenarios and during extreme NO_2 events like R10-Urumqi.

Further, in contrast to 2019, the monthly differences between monthly values from 2020 to 2022 and the same month in 2019 showed that XCO_2 inaccurately responded to NO_2 as well. Figure 9 shows an example of these differences for R1-Wuhan, R2-Shanghai, R3-YRD, and R4-BTH, where the fluctuations in human economic activities were the highest during the COVID-19 period. Figure 9 shows a co-decrease in NO_2 and XCO_2 during the first period of COVID-19, covering January–March of 2020, especially in R1-Wuhan, which experienced an extreme reduction in human activities due to COVID-19. NO_2 showed a decrease of 1.3×10^{16} molec/cm² in February 2020, corresponding to a reduction of 1.4 ppm in XCO_2 . The faint co-decrease in NO_2 and XCO_2 during the COVID-19 controlling period in April 2022 in R2-Shanghai is shown as well, with a decrease of 0.5×10^{16} molec/cm² in NO_2 , corresponding to a reduction of 0.5 ppm in XCO_2 .

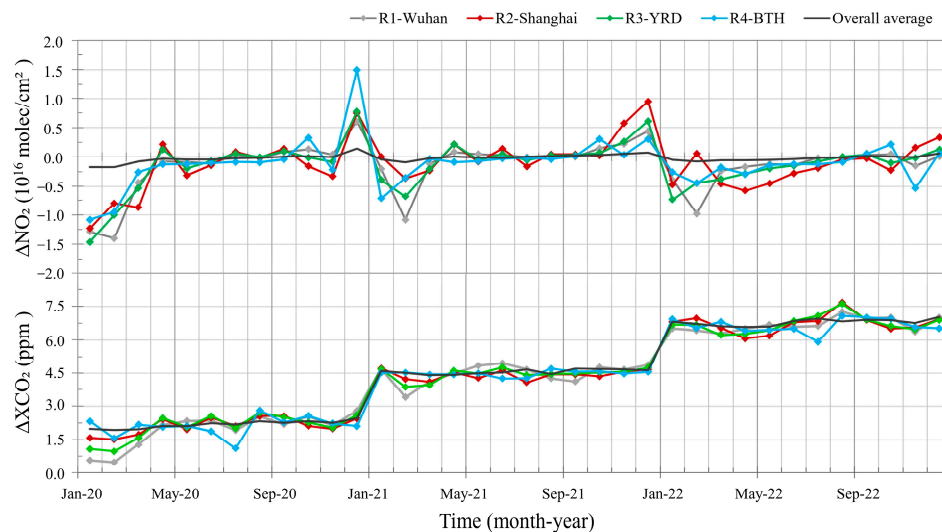


Figure 9. An example of differences between the monthly XCO₂ and NO₂ values from 2020 to 2022 and the same month in 2019 for R1-Wuhan, R2-Shngahia, R3-YRD, and 4-BTH, as well as the overall study area.

XCO₂ still demonstrated a yearly increase of, on average, 2.3 ppm from 2020 to 2022, which implies that regional reductions in anthropogenic emission activities, such as in China, cannot effectively mitigate the increase in CO₂ globally. This means that to mitigate further increase in CO₂, global action is required.

4. Discussion

Considering that the prediction results of different machine learning algorithms differ, we further compared the prediction performance of LGB, MLP, and LSTM. The results showed that the prediction results of all three models were worse than CatB. The R² and total deviation of the cross-validation for the four models (CatB, LGB, MLP, and LSTM) were 0.81, 0.79, 0.47, and 0.58, and 0.17 ± 1.17 ppm, 0.21 ± 1.18 ppm, 1.49 ± 3.18 ppm, and 1.03 ± 2.16 ppm (see Figures 2 and A1), respectively. The predictions of CatB and LGB, based on decision trees, were better than those of MLP and LSTM, based on neural networks. LGB and LSTM generally presented overestimates, while MLP presented severe underestimates (see Figures 3 and 10).

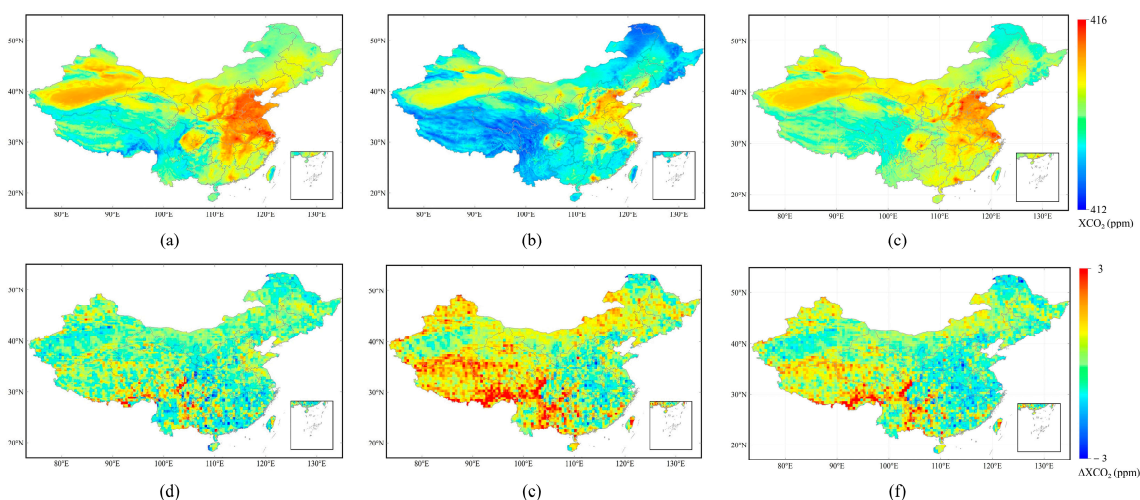


Figure 10. The means and the differences of fp-XCO₂ from 2019 to 2022 derived from (a,d) LGB, (b,e) MLP, and (c,f) LSTM.

These results are likely due to the fact that different machine learning algorithms have different abilities to capture the features of multisource satellite observations. CatB, and LGB are optimized algorithms based on the Gradient-Boosting Decision Tree (GBDT) algorithm, which is the most widely used ML algorithm. GBDT applies weak classifiers (decision trees) for training to obtain an optimal model iteratively, which has the advantages of a good training effect and being hard to overfit. MLP and LSTM are forward-structured artificial neural networks containing an input layer, an output layer, and several hidden layers. The error backward propagation technique was employed to train the parameters of the neural network model and is capable of handling nonlinear separable problems.

The total contribution of the model predictor parameters to the predictions (see Figures 5 and 11) showed that the contribution of the predictor variables in MLP and LSTM were 24.7%, 13.6%, 13.6%, 10.8%, 10.7%, 10.4%, 8.5%, and 7.7% (NDVI, NO₂, MXCO₂, T2M, V10, U10, D2M, and TimeIndex) and 22.7%, 19%, 16.5%, 11.9%, 11.2%, 8%, 6.3%, and 4.5% (TimeIndex, MXCO₂, NO₂, D2M, NDVI, T2M, V10, and U10), respectively. It can be seen that NDVI and MXCO₂ had unusually high contributions in MLP, as well as TimeIndex and MXCO₂ in LSTM, which led to significantly worse predictions in both models. In addition, TimeIndex contributed a higher percentage in LGB and LSTM than in CatB. As described in Section 2.2, for the physical significance and impact of the predicted parameters, overconsideration of the temporal increase in the atmospheric CO₂ concentration leads to an overall overestimation of its predictions.

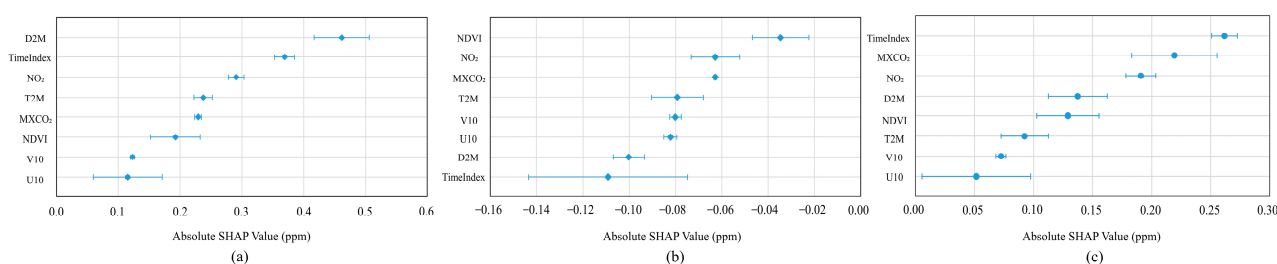


Figure 11. Machine learning model SHAP value swarm graphs, (a) LGB, (b) MLP, and (c) LSTM.

5. Conclusions

The spatio-temporal discontinuities in the original satellite XCO₂ observations and the large number of data gaps make it difficult to accurately reveal the spatial and temporal characteristics of the atmospheric CO₂ concentration. We reconstructed satellite-observed XCO₂ data using multiple predictors, including satellite-observed NO₂ with co-emissions from the same sources based on a machine learning algorithm to generate spatially and temporally continuous monthly XCO₂ data from January 2019 to December 2022 in Chinese regions. The accuracy of the ML predictive model and the effectiveness of introducing NO₂ to enhance the XCO₂ response to anthropogenic emission activities were assessed through the cross-validation, validation against TCCON observations, and comparison with original satellite XCO₂ retrievals.

The results showed that the predictive model had good predictive performance in terms of revealing the spatial and temporal concentrations of CO₂, with an R² and total deviations of 0.79, 0.17 ± 1.17 ppm, and 0.81, 1.03 ± 1.15 ppm, respectively, compared to ground-based XCO₂ measurements and model cross-validation. The reconstructed XCO₂ data not only fills in the gaps but also corrects the large uncertainties of the original observations over areas with high elevation and steep terrain. The strong correlation of NO₂ with the anthropogenic CO₂ emissions (R² = 0.92), due to simultaneous anthropogenic CO₂ emissions, implies that the introduction of NO₂ can enhance the anthropogenic emissions information in the model-predicted XCO₂ with R² = 0.76, which is higher than the R² = 0.66 found for the raw satellite-observed XCO₂.

The findings, by analyzing co-response variations of XCO₂ and NO₂ under special anthropogenic emission scenarios during the COVID-19 control period of 2020–2022 in the

study area, indicate a synchronized CO₂ and NO₂ response to the dramatic reduction in anthropogenic activities; however, CO₂ could incompletely respond to NO₂ changes in the case of overly heavy gas polluting scenarios. This XCO₂ and NO₂ co-response variation mechanism to anthropogenic CO₂ emissions needs further verification via practical CO₂ and NO₂ emissions measures.

Author Contributions: Conceptualization, K.G., L.L. and M.S.; Methodology, K.G. and H.S.; Software, K.G., Z.J. and H.S.; Validation, K.G.; Formal analysis, K.G.; Data curation, K.G.; Writing—original draft, K.G.; Writing—review & editing, K.G. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (grant no. 2022YFC3800700 and 2020YFA0607503).

Data Availability Statement: The research data presented in this study are available on request from the corresponding author. The data are not publicly available as this project is still in the research phase.

Acknowledgments: We are grateful for the OCO-2 v11r data and OCO-3 v10.4r data, which were provided by the OCO-2/OCO-3 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2/OCO-3 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. We thank the European Space Agency (ESA) and Google Earth Engine for providing Sentinel-5P NO₂ products and the World Data Centre for Greenhouse Gases (WDCGG) for providing global atmospheric CO₂ data. We also acknowledge the Land Processes Distributed Active Archive Center (LP DAAC) at the National Aeronautics and Space Administration (NASA) for sharing land cover type and NDVI data derived from MODIS. We thank the Total Carbon Column Observing Network (TCCON) for providing XCH₄ observed data products.

Conflicts of Interest: The authors declare no conflicts of interest. Author Mengya Sheng was employed by the company China Highway Engineering Consultants Corporation. The company China Highway Engineering Consultants Corporation had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

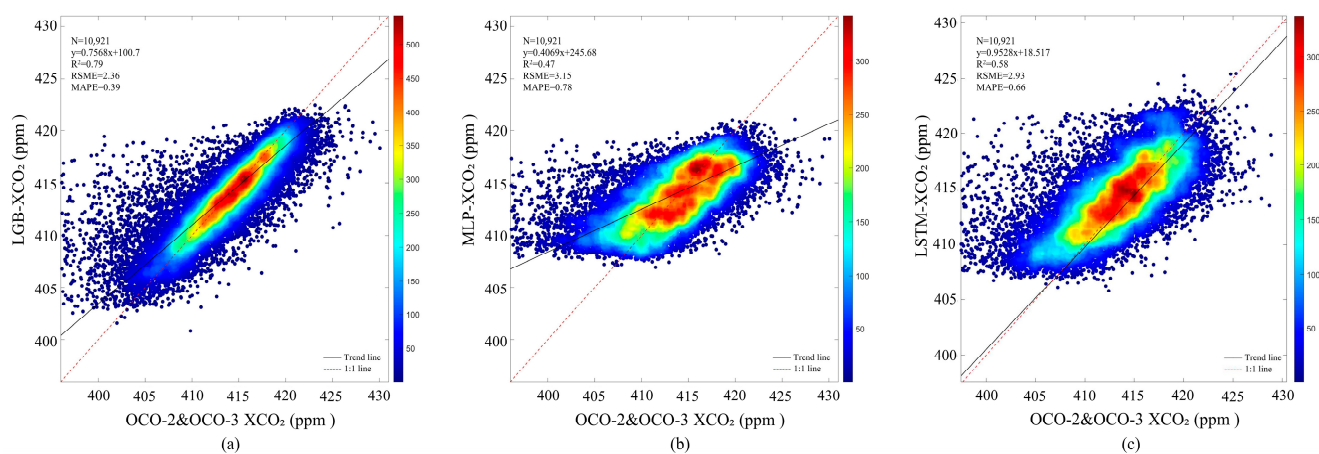


Figure A1. Plot of cross-validation results for models, (a) LGB, (b) MLP, and (c) LSTM.

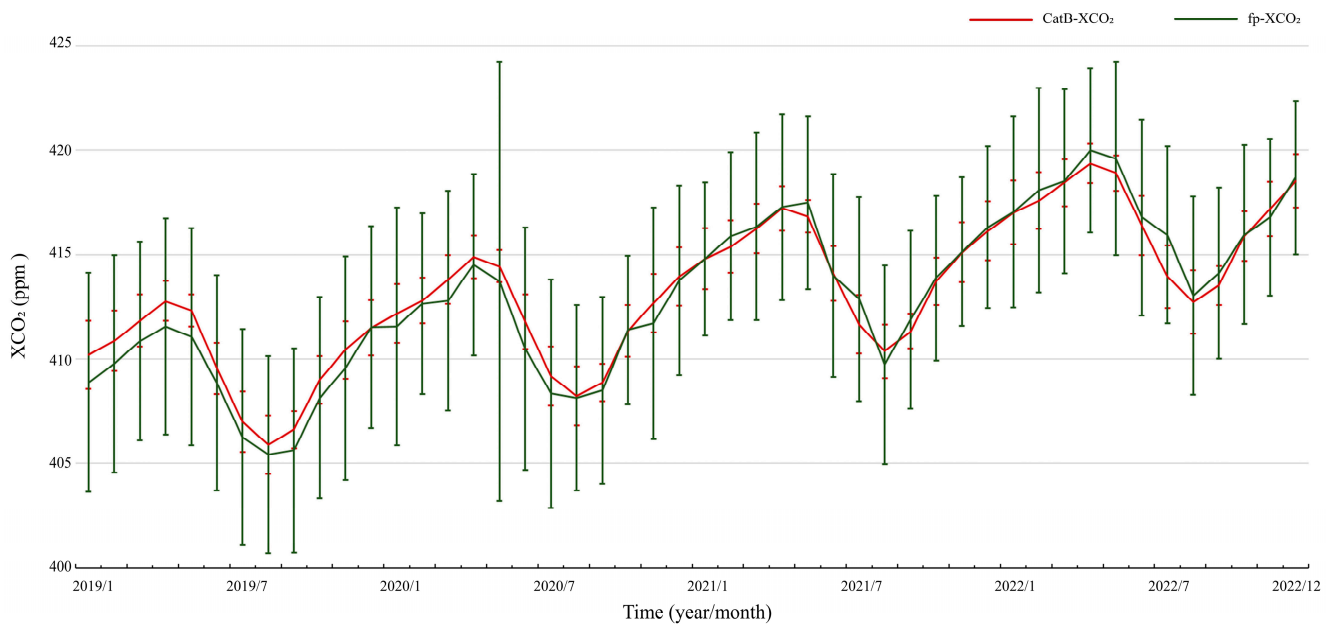


Figure A2. Time series of monthly mean XCO₂ values from CatB and satellite raw observations.

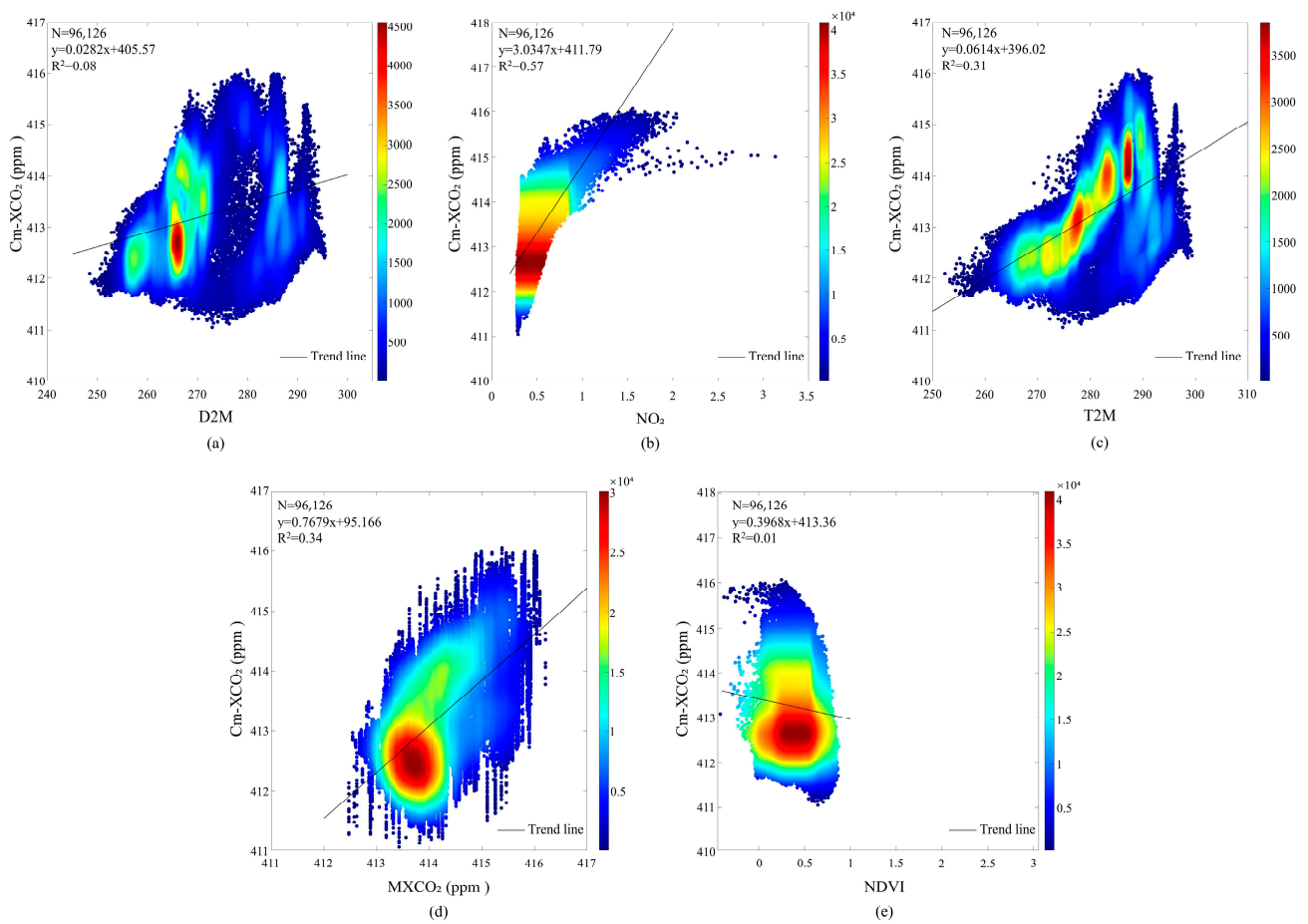


Figure A3. CatB model predictions versus parameters (D2M, NO₂, T2M, MXCO₂, and NDVI on the x-axis of (a–e), respectively).

References

1. Wang, H.; Liu, Z.; Zhang, Y.; Yu, Z.; Chen, C. Impact of Different Urban Canopy Models on Air Quality Simulation in Chengdu, Southwestern China. *Atmos. Environ.* **2021**, *267*, 118775. [[CrossRef](#)]
2. Hsueh, Y.-H.; Li, K.-F.; Lin, L.-C.; Bhattacharya, S.K.; Laskar, A.H.; Liang, M.-C. East Asian CO₂ Level Change Caused by Pacific Decadal Oscillation. *Remote Sens. Environ.* **2021**, *264*, 112624. [[CrossRef](#)]
3. Liang, M.; Zhang, Y.; Ma, Q.; Yu, D.; Chen, X.; Cohen, J.B. Dramatic Decline of Observed Atmospheric CO₂ and CH₄ during the COVID-19 Lockdown over the Yangtze River Delta of China. *J. Environ. Sci.* **2023**, *124*, 712–722. [[CrossRef](#)]
4. Fang, S.; Du, R.; Qi, B.; Ma, Q.; Zhang, G.; Chen, B.; Li, J. Variation of Carbon Dioxide Mole Fraction at a Typical Urban Area in the Yangtze River Delta, China. *Atmos. Res.* **2022**, *265*, 105884. [[CrossRef](#)]
5. Fu, Y.; Sun, W.; Luo, F.; Zhang, Y.; Zhang, X. Variation Patterns and Driving Factors of Regional Atmospheric CO₂ Anomalies in China. *Environ. Sci. Pollut. Res.* **2022**, *29*, 19390–19403. [[CrossRef](#)]
6. Wang, W.; He, J.; Feng, H.; Jin, Z. High-Coverage Reconstruction of XCO₂ Using Multisource Satellite Remote Sensing Data in Beijing–Tianjin–Hebei Region. *Int. J. Environ. Res. Public Health* **2022**, *19*, 10853. [[CrossRef](#)]
7. Crisp, D.; Fisher, B.M.; O'Dell, C.; Frankenberg, C.; Basilio, R.; Bösch, H.; Brown, L.R.; Castano, R.; Connor, B.; Deutscher, N.M.; et al. The ACOS CO₂ Retrieval Algorithm—Part II: Global XCO₂ Data Characterization. *Atmos. Meas. Tech.* **2012**, *5*, 687–707. [[CrossRef](#)]
8. Nakajima, M.; Kuze, A.; Suto, H. The Current Status of GOSAT and the Concept of GOSAT-2. In Proceedings of the Sensors, Systems, and Next-Generation Satellites XVI, Edinburgh, UK, 24–27 September 2012; Meynart, R., Neeck, S.P., Shimoda, H., Eds.; SPIE: Bellingham, WA, USA, 2012; p. 853306.
9. Kataoka, F.; Crisp, D.; Taylor, T.; O'Dell, C.; Kuze, A.; Shiomi, K.; Suto, H.; Bruegge, C.; Schwandner, F.; Rosenberg, R.; et al. The Cross-Calibration of Spectral Radiances and Cross-Validation of CO₂ Estimates from GOSAT and OCO-2. *Remote Sens.* **2017**, *9*, 1158. [[CrossRef](#)]
10. Jin, C.; Xue, Y.; Jiang, X.; Zhao, L.; Yuan, T.; Sun, Y.; Wu, S.; Wang, X. A Long-Term Global XCO₂ Dataset: Ensemble of Satellite Products. *Atmos. Res.* **2022**, *279*, 106385. [[CrossRef](#)]
11. Guo, X.; Zhang, Z.; Cai, Z.; Wang, L.; Gu, Z.; Xu, Y.; Zhao, J. Analysis of the Spatial–Temporal Distribution Characteristics of NO₂ and Their Influencing Factors in the Yangtze River Delta Based on Sentinel-5P Satellite Data. *Atmosphere* **2022**, *13*, 1923. [[CrossRef](#)]
12. Zhang, L.; Li, T.; Wu, J. Deriving Gapless CO₂ Concentrations Using a Geographically Weighted Neural Network: China, 2014–2020. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103063. [[CrossRef](#)]
13. Zeng, Z.; Lei, L.; Hou, S.; Ru, F.; Guan, X.; Zhang, B. A Regional Gap-Filling Method Based on Spatiotemporal Variogram Model of CO₂ Columns. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3594–3603. [[CrossRef](#)]
14. Sheng, M.; Lei, L.; Zeng, Z.-C.; Rao, W.; Zhang, S. Detecting the Responses of CO₂ Column Abundances to Anthropogenic Emissions from Satellite Observations of GOSAT and OCO-2. *Remote Sens.* **2021**, *13*, 3524. [[CrossRef](#)]
15. Zhang, S.; Lei, L.; Sheng, M.; Song, H.; Li, L.; Guo, K.; Ma, C.; Liu, L.; Zeng, Z. Evaluating Anthropogenic CO₂ Bottom-Up Emission Inventories Using Satellite Observations from GOSAT and OCO-2. *Remote Sens.* **2022**, *14*, 5024. [[CrossRef](#)]
16. Uddin, M.S.; Czajkowski, K.P. Performance Assessment of Spatial Interpolation Methods for the Estimation of Atmospheric Carbon Dioxide in the Wider Geographic Extent. *J. Geovis. Spat. Anal.* **2022**, *6*, 10. [[CrossRef](#)]
17. He, C.; Ji, M.; Li, T.; Liu, X.; Tang, D.; Zhang, S.; Luo, Y.; Grieneisen, M.L.; Zhou, Z.; Zhan, Y. Deriving Full-Coverage and Fine-Scale XCO₂ Across China Based on OCO-2 Satellite Retrievals and CarbonTracker Output. *Geophys. Res. Lett.* **2022**, *49*, e2022GL098435. [[CrossRef](#)]
18. Wu, C.; Ju, Y.; Yang, S.; Zhang, Z.; Chen, Y. Reconstructing Annual XCO₂ at a 1 Km × 1 Km Spatial Resolution across China from 2012 to 2019 Based on a Spatial CatBoost Method. *Environ. Res.* **2023**, *236*, 116866. [[CrossRef](#)]
19. Kallio, J.; Tervonen, J.; Räsänen, P.; Mäkynen, R.; Koivusaari, J.; Peltola, J. Forecasting Office Indoor CO₂ Concentration Using Machine Learning with a One-Year Dataset. *Build. Environ.* **2021**, *187*, 107409. [[CrossRef](#)]
20. Zhao, Z.; Xie, F.; Ren, T.; Zhao, C. Atmospheric CO₂ Retrieval from Satellite Spectral Measurements by a Two-Step Machine Learning Approach. *J. Quant. Spectrosc. Radiat. Transf.* **2022**, *278*, 108006. [[CrossRef](#)]
21. Finch, D.; Palmer, P.; Zhang, T. Automated Detection of Atmospheric NO₂ Plumes from Satellite Data: A Tool to Help Infer Anthropogenic Combustion Emissions. *Atmos. Meas. Tech.* **2021**, *15*, 721–733. [[CrossRef](#)]
22. Hakkarainen, J.; Ialongo, I.; Oda, T.; Szélag, M.E.; O'Dell, C.W.; Eldering, A.; Crisp, D. Building a Bridge: Characterizing Major Anthropogenic Point Sources in the South African Highveld Region Using OCO-3 Carbon Dioxide Snapshot Area Maps and Sentinel-5P/TROPOMI Nitrogen Dioxide Columns. *Environ. Res. Lett.* **2023**, *18*, 035003. [[CrossRef](#)]
23. Saw, G.K.; Dey, S.; Kaushal, H.; Lal, K. Tracking NO₂ Emission from Thermal Power Plants in North India Using TROPOMI Data. *Atmos. Environ.* **2021**, *259*, 118514. [[CrossRef](#)]
24. Fuentes Andrade, B.; Buchwitz, M.; Reuter, M.; Bovensmann, H.; Richter, A.; Boesch, H.; Burrows, J.P. A Method for Estimating Localized CO₂ Emissions from Co-Located Satellite XCO₂ and NO₂ Images. *Atmos. Meas. Tech.* **2024**, *17*, 1145–1173. [[CrossRef](#)]
25. Goldberg, D.L.; Lu, Z.; Oda, T.; Lamsal, L.N.; Liu, F.; Griffin, D.; McLinden, C.A.; Krotkov, N.A.; Duncan, B.N.; Streets, D.G. Exploiting OMI NO₂ Satellite Observations to Infer Fossil-Fuel CO₂ Emissions from U.S. Megacities. *Sci. Total Environ.* **2019**, *695*, 133805. [[CrossRef](#)]
26. Kononov, I.B.; Berezin, E.V.; Ciais, P.; Broquet, G.; Zhuravlev, R.V.; Janssens-Maenhout, G. Estimation of Fossil-Fuel CO₂ Emissions Using Satellite Measurements of “Proxy” Species. *Atmos. Chem. Phys.* **2016**, *16*, 13509–13540. [[CrossRef](#)]

27. Liu, F.; Duncan, B.N.; Krotkov, N.A.; Lamsal, L.N.; Beirle, S.; Griffin, D.; McLinden, C.A.; Goldberg, D.L.; Lu, Z. A Methodology to Constrain Carbon Dioxide Emissions from Coal-Fired Power Plants Using Satellite Observations of Co-Emitted Nitrogen Dioxide. *Atmos. Chem. Phys.* **2020**, *20*, 99–116. [[CrossRef](#)]
28. He, Q.; Ye, T.; Wang, W.; Luo, M.; Song, Y.; Zhang, M. Spatiotemporally Continuous Estimates of Daily 1-Km PM_{2.5} Concentrations and Their Long-Term Exposure in China from 2000 to 2020. *J. Environ. Manag.* **2023**, *342*, 118145. [[CrossRef](#)]
29. Park, H.; Jeong, S.; Park, H.; Labzovskii, L.D.; Bowman, K.W. An Assessment of Emission Characteristics of Northern Hemisphere Cities Using Spaceborne Observations of CO₂, CO, and NO₂. *Remote Sens. Environ.* **2021**, *254*, 112246. [[CrossRef](#)]
30. Hakkarainen, J.; Ialongo, I.; Maksyutov, S.; Crisp, D. Analysis of Four Years of Global XCO₂ Anomalies as Seen by Orbiting Carbon Observatory-2. *Remote Sens.* **2019**, *11*, 850. [[CrossRef](#)]
31. Liu, D.; Lei, L.; Guo, L.; Zeng, Z.-C. A Cluster of CO₂ Change Characteristics with GOSAT Observations for Viewing the Spatial Pattern of CO₂ Emission and Absorption. *Atmosphere* **2015**, *6*, 1695–1713. [[CrossRef](#)]
32. Wang, W.; Tian, Y.; Liu, C.; Sun, Y.; Liu, W.; Xie, P.; Liu, J.; Xu, J.; Morino, I.; Velazco, V.A.; et al. Investigating the Performance of a Greenhouse Gas Observatory in Hefei, China. *Atmos. Meas. Tech.* **2017**, *10*, 2627–2643. [[CrossRef](#)]
33. Ciais, P.; Dolman, A.J.; Bombelli, A.; Duren, R.; Peregon, A.; Rayner, P.J.; Miller, C.; Gobron, N.; Kinderman, G.; Marland, G.; et al. Current Systematic Carbon-Cycle Observations and the Need for Implementing a Policy-Relevant Carbon Observing System. *Biogeosciences* **2014**, *11*, 3547–3602. [[CrossRef](#)]
34. Yang, E.G.; Kort, E.A.; Ott, L.E.; Oda, T.; Lin, J.C. Using Space-Based CO₂ and NO₂ Observations to Estimate Urban CO₂ Emissions. *JGR Atmos.* **2023**, *128*, e2022JD037736. [[CrossRef](#)]
35. Ebi, K.L.; Anderson, C.L.; Hess, J.J.; Kim, S.-H.; Loladze, I.; Neumann, R.B.; Singh, D.; Ziska, L.; Wood, R. Nutritional Quality of Crops in a High CO₂ World: An Agenda for Research and Technology Development. *Environ. Res. Lett.* **2021**, *16*, 064045. [[CrossRef](#)]
36. Warren, J.M.; Jensen, A.M.; Ward, E.J.; Guha, A.; Childs, J.; Wullschleger, S.D.; Hanson, P.J. Divergent Species-specific Impacts of Whole Ecosystem Warming and Elevated CO₂ on Vegetation Water Relations in an Ombrotrophic Peatland. *Glob. Change Biol.* **2021**, *27*, 1820–1835. [[CrossRef](#)]
37. European Commission; Joint Research Centre. *Fossil CO₂ Emissions of All World Countries: 2018 Report*; Publications Office: Luxembourg, 2018.
38. He, Z.; Lei, L.; Welp, L.; Zeng, Z.-C.; Bie, N.; Yang, S.; Liu, L. Detection of Spatiotemporal Extreme Changes in Atmospheric CO₂ Concentration Based on Satellite Observations. *Remote Sens.* **2018**, *10*, 839. [[CrossRef](#)]
39. Chen, S.; Mihara, K.; Wen, J. Time Series Prediction of CO₂, TVOC and HCHO Based on Machine Learning at Different Sampling Points. *Build. Environ.* **2018**, *146*, 238–246. [[CrossRef](#)]
40. Kalra, S.; Lamba, R.; Sharma, M. Machine Learning Based Analysis for Relation between Global Temperature and Concentrations of Greenhouse Gases. *J. Inf. Optim. Sci.* **2020**, *41*, 73–84. [[CrossRef](#)]
41. Siabi, Z.; Falahatkar, S.; Alavi, S.J. Spatial Distribution of XCO₂ Using OCO-2 Data in Growing Seasons. *J. Environ. Manag.* **2019**, *244*, 110–118. [[CrossRef](#)]
42. O'Dell, C.W.; Eldering, A.; Wennberg, P.O.; Crisp, D.; Gunson, M.R.; Fisher, B.; Frankenberg, C.; Kiel, M.; Lindqvist, H.; Mandrake, L.; et al. Improved Retrievals of Carbon Dioxide from Orbiting Carbon Observatory-2 with the Version 8 ACOS Algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 6539–6576. [[CrossRef](#)]
43. Massie, S.T.; Sebastian Schmidt, K.; Eldering, A.; Crisp, D. Observational Evidence of 3-D Cloud Effects in OCO-2 CO₂ Retrievals. *JGR Atmos.* **2017**, *122*, 7064–7085. [[CrossRef](#)]
44. Van Geffen, J.; Boersma, K.F.; Eskes, H.; Sneep, M.; Ter Linden, M.; Zara, M.; Veefkind, J.P. S5P TROPOMI NO₂ Slant Column Retrieval: Method, Stability, Uncertainties and Comparisons with OMI. *Atmos. Meas. Tech.* **2020**, *13*, 1315–1335. [[CrossRef](#)]
45. Fan, C.; Li, Z.; Li, Y.; Dong, J.; Van Der, A.R.; De Leeuw, G. Variability of NO₂ Concentrations over China and Effect on Air Quality Derived from Satellite and Ground-Based Observations. *Atmos. Chem. Phys.* **2021**, *21*, 7723–7748. [[CrossRef](#)]
46. Fioletov, V.; McLinden, C.A.; Griffin, D.; Krotkov, N.; Liu, F.; Eskes, H. Quantifying Urban, Industrial, and Background Changes in NO₂ during the COVID-19 Lockdown Period Based on TROPOMI Satellite Observations. *Atmos. Chem. Phys.* **2022**, *22*, 4201–4236. [[CrossRef](#)]
47. Veefkind, J.P.; Aben, I.; McMullan, K.; Förster, H.; De Vries, J.; Otter, G.; Claas, J.; Eskes, H.J.; De Haan, J.F.; Kleipool, Q.; et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES Mission for Global Observations of the Atmospheric Composition for Climate, Air Quality and Ozone Layer Applications. *Remote Sens. Environ.* **2012**, *120*, 70–83. [[CrossRef](#)]
48. Guo, M.; Wang, X.; Li, J.; Wang, H.; Tani, H. Examining the Relationships between Land Cover and Greenhouse Gas Concentrations Using Remote-Sensing Data in East Asia. *Int. J. Remote Sens.* **2013**, *34*, 4281–4303. [[CrossRef](#)]
49. Yang, W.; Zhao, Y.; Wang, Q.; Guan, B. Climate, CO₂, and Anthropogenic Drivers of Accelerated Vegetation Greening in the Haihe River Basin. *Remote Sens.* **2022**, *14*, 268. [[CrossRef](#)]
50. Chen, X.; He, Q.; Ye, T.; Liang, Y.; Li, Y. Decoding Spatiotemporal Dynamics in Atmospheric CO₂ in Chinese Cities: Insights from Satellite Remote Sensing and Geographically and Temporally Weighted Regression Analysis. *Sci. Total Environ.* **2024**, *908*, 167917. [[CrossRef](#)]
51. Buchwitz, M.; Reuter, M.; Schneising, O.; Noël, S.; Gier, B.; Bovensmann, H.; Burrows, J.P.; Boesch, H.; Anand, J.; Parker, R.J.; et al. Computation and Analysis of Atmospheric Carbon Dioxide Annual Mean Growth Rates from Satellite Observations during 2003–2016. *Atmos. Chem. Phys.* **2018**, *18*, 17355–17370. [[CrossRef](#)]

52. Sun, Y.; Liu, C.; Zhang, L.; Palm, M.; Notholt, J.; Yin, H.; Vigouroux, C.; Lutsch, E.; Wang, W.; Shan, C.; et al. Fourier Transform Infrared Time Series of Tropospheric HCN in Eastern China: Seasonality, Interannual Variability, and Source Attribution. *Atmos. Chem. Phys.* **2020**, *20*, 5437–5456. [[CrossRef](#)]
53. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Quart. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
54. Jiang, Z.; Huete, A.R.; Chen, J.; Chen, Y.; Li, J.; Yan, G.; Zhang, X. Analysis of NDVI and Scaled Difference Vegetation Index Retrievals of Vegetation Fraction. *Remote Sens. Environ.* **2006**, *101*, 366–378. [[CrossRef](#)]
55. Chatterjee, A.; Payne, V.; Eldering, A.; Rosenberg, R.; Kiel, M.; Fisher, B.; Nelson, R.; Dang, L.; Rodrigues, G.K.; O'Dell, C.; et al. *Orbiting Carbon Observatory-3 (OCO-3) Data Quality Statement: Level 2 Forward and Retrospective Processing Data Release 10 (V10 and V10r), V10.4 Lite Files*; California Institute of Technology: Pasadena, CA, USA, 2022.
56. Keely, W.R.; Mauceri, S.; Crowell, S.; O'Dell, C.W. A Nonlinear Data-Driven Approach to Bias Correction of XCO₂ for NASA's OCO-2 ACOS Version 10. *Atmos. Meas. Tech.* **2023**, *16*, 5725–5748. [[CrossRef](#)]
57. Zhang, J. China Greenhouse Gas Bulletin 2022 post. *China Meteorological News*, 6 December 2023.

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.