https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

Open Access Earth System Science Data Discussions

A full-coverage daily XCO_2 dataset in China from 2015 to 2020 based on DSC-DF-LGB

Xinfeng Huang[1], Hui Yang[1,*], Qingzhou Lv[1], Huaiwei Fan[1], Liu Cui[1], Yina Qiao[1], Yuejing Yao[1], Gefei Feng[2]

[1]School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China
[2]School of Linguistic Sciences and Arts, Jiangsu Normal University, Xuzhou 221009, China

Correspondence to: Hui Yang (yanghui@cumt.edu.cn)

Abstract. Carbon dioxide (CO_2), as a major greenhouse gas, is one of the important causes of global warming. In recent years, the atmospheric CO_2 concentration in China has been increasing year by year. Satellite observation is the main means of obtaining atmospheric CO_2 concentration. However, the current onboard sensors used for measuring atmospheric CO_2 have a narrow observation range and cannot obtain spatiotemporal continuous atmospheric CO_2 concentrations. Therefore, this paper proposes a daily full-coverage XCO_2 dataset generation method based on the DSC-DF-LGB (Deep Separable Convolutional Neural Network and Deep Forest concatenated with LightGBM) model to obtain the spatiotemporal distribution of atmospheric CO_2 in China. The DSC-DF-LGB model was established to train the mapping relationship between OCO-2 XCO_2 retrieval and related variables (reanalysis XCO_2, vegetation parameters, human factors, elevation, and meteorological parameters). The model was used to generate a daily 0.1° full-coverage XCO_2 dataset for China from 2015 to 2020. The cross validation (CV) result indicates that the model has strong performance in estimating XCO_2, with R^2 and RMSE of 0.9633 and 0.9761 ppm. The TCCON independent site validation result indicates that the estimated XCO_2 has high consistency with in-situ measurements, with R^2 and RMSE of 0.8786 and 1.5452 ppm. The full-coverage and high-resolution XCO_2 dataset can provide data support for research on carbon sources and sinks. The dataset is available at https://zenodo.org/doi/10.5281/zenodo.12696674 (Huang, 2024).

1 Introduction

The increase in the amount of CO_2 is generally considered the main cause of climate change (Kump, 2000). Since the Industrial Revolution, global carbon dioxide concentration has been continuously increasing. At present, the global CO_2 concentration has increased from 280 ppm in the pre-industrial to 419.31 ppm in 2023 (Lan et al., 2024; Petit and Raynaud, 2020). Under this background, nearly 200 countries and regions signed the Paris Climate Agreement in 2016, making the goal of controlling temperatures by 1.5℃ possible (IPCC, 2018; Voigt, 2016). Subsequently, at the 75th United Nations General Assembly, the Chinese government announced that it will strive to peak carbon dioxide emissions before 2030 and achieve carbon neutrality target before 2060. In order to achieve the "dual carbon" target, it is urgent to carry out

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

spatiotemporal continuous CO_2 concentration mapping to understand the spatiotemporal variation pattern of CO_2 concentration in China.

So far, the atmospheric CO_2 concentration mainly measured by in situ stations, aircraft flights and satellite instruments (Yang et al., 2020). However, the spatial distributions of in situ measurements and aerial measurements are sparse, making it impossible to obtain the distribution of CO_2 concentration over a large-scale region (He et al., 2023). Satellite-based observations have made up for this deficiency. Even though satellite-based observations have low temporal resolution, their coverage is wide which makes them the most effective way to measure regional and global atmospheric CO_2 (Mustafa et al., 2021; Wang et al., 2021).Satellite-based observations use near-infrared (NIR) and shortwave infrared (SWIR) spectra to monitor atmospheric CO_2 from space, which can provide high-precision global distribution of CO_2 and help improve our understanding of CO_2 flux (Liu et al., 2018; Yue et al., 2016). The current carbon satellites in orbit mainly include Japanese Greenhouse gases Observing Satellites (GOSAT) (Yoshida et al., 2011), the United States' the Orbiting Carbon Observatory-2 (OCO-2)/Orbiting Carbon Observatory-3 (OCO-3) (Miller et al., 2014), and Chinese Carbon Dioxide Observation Satellite (TanSat) (Ran and Li, 2019). The GOSAT satellite was launched into space by Japan in 2009, which is the first satellite in the word specifically designed to measure the distribution of greenhouse gas concentrations. OCO-2 satellite was launched by NASA in 2014, which is dedicated to monitor carbon dioxide emissions. And Tansat is the first global atmospheric CO_2 observation satellite developed by China, which was successfully launched in 2016. The XCO_2 products retrieved by GOSAT and OCO-2 have been widely used in various studies. These satellites can all provide XCO_2 products. However, due to the presence of clouds and aerosols, and the narrow satellite orbiting tracks, the retrieved atmospheric CO_2 concentration data is not spatially continuous. Besides, the satellites have long revisiting intervals, which results in satellite data not being temporally continuous. The spatiotemporal discontinuity limits our research on regional carbon sources and sinks. Therefore, it is of great significance to map spatiotemporally continuous XCO_2 concentration in China.

At present, three kinds of methods for obtaining spatiotemporal continuous XCO_2 concentration dataset have been developed (He et al., 2022a), which include spatial interpolation (He et al., 2020), chemical transport model simulation (Fu et al., 2019) and multi-source data fusion (Liang et al., 2023; Wang et al., 2023). Spatial interpolation is a geostatistical method. The method cannot consider factors affecting XCO_2 concentration and capture the spatial gradients (Zhang and Liu, 2023). And the generated XCO_2 data through chemical transport model simulation has a low spatial resolution, making it difficult to gather some detailed information (Zhang et al., 2017). In recent years, with the increasing popularity of machine learning, various machine learning models have gradually been applied to multi-source data fusion. Although machine learning lacks interpretability, it can establish relationship models between multiple input variables and output variables. It allows us to train models and estimate atmospheric CO_2 concentrations by combining multiple relevant factors. More and more researches have proven its effectiveness in collaborative fusion of multi-source data. A 500 m spatially continuous normalized microwave reflection index (NMRI) product was obtained by integrating site data with other VWC-related products using Random Forest (RF) (Yuan et al., 2019). Various machine learning methods such as Extremely Randomized Trees (ERT) (Wei et al., 2020), RF (Tang et al., 2023; Yang et al., 2022; Zhao et al., 2019), and Deep Belief Network (DBN)

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

(Li et al., 2017) have been used to fuse AOD satellite data, meteorological data, and model data, obtaining spatiotemporal continuous $PM_{2.5}$ datasets. Besides, the DBN was utilized to downscale the 9 km SMAP SSM and 0.1° ERA5-Land SSM to 1 km (Huang et al., 2022). RF have also been used to plot the global distribution of XCH_4 concentrations, which has a higher accuracy and computing efficiency than the operational full-physics retrieval algorithms (Li et al., 2024). Zhang et al. (2024) generated a 1km high spatial resolution dataset of monthly moisture index collection in China by the light gradient boosting model (LightGBM).

As for CO_2, various neural networks and ensemble models based on decision tree have been widely used to generate global or regional atmospheric CO_2 concentration datasets. A common means is to use model to establish the relationship between ancillary factors and XCO_2 retrievals, and subsequently estimate full-coverage XCO_2 data (Li et al., 2023). To evaluate the spatial distribution of CO_2 concentration during the growing season in Iran in 2015, the XCO_2 data belonging to OCO-2 and eight environmental variables data were modeled by multi-layer perceptron (MLP) (Siabi et al., 2019). Wang et al. (2022) used RF to generate daily XCO_2 concentrations in the Beijing Tianjin Hebei region from 2015 to 2019. Li et al. (2022) proposed a reconstruction method using ERT to generate a spatiotemporally continuous CO_2 concentration dataset with high-spatiotemporal resolution based on satellite CO_2 retrievals and environmental factors. In addition, the improved ANN and GRWN also achieved good results in generating CO_2 concentration datasets in China (Zhang et al., 2022; Zhang and Liu, 2023).

Based on previous research, the decision tree based ensemble model has been proven to perform well in estimating atmospheric CO_2 concentration (He et al., 2023). Deep Forest (DF) is an advanced ensemble model that has demonstrated efficient performance in multi-source data fusion research in fields such as $PM_{2.5}$ (Song et al., 2022a; Song et al., 2022b), CO (Wang et al., 2022), O_3(Chen et al., 2023), surface soil moisture (Li et al., 2024). Therefore, the DF-LGB model was developed to generate high-precision and spatiotemporal continuous daily XCO_2 datasets with a resolution of 0.1° in China from 2015 to 2020. The concentration of XCO_2 is not only influenced by the local geographical environment, but also closely related to the surrounding carbon sources and sinks (Cui et al., 2024; He et al., 2022b). However, unlike Convolutional Neural Networks (CNN), general machine learning models cannot provide local receptive fields and simulate the impact of the surrounding environment on local CO_2 concentration. To address this problem, DSC is introduced to capture local features of vegetation parameters and anthropogenic factors. In addition, the inverse distance weighting (IDW) was used to preprocess OCO-2 XCO_2 retrievals to optimize the output dataset. After establishing the DSC-DF-LGB model, we conducted 10-fold cross validation (CV) and in situ site validation, and evaluated the ability of this model to generate earlier and latest datasets. Finally, the 0.1°×0.1° daily XCO_2 dataset from 2015 to 2020 was generated, and the spatiotemporal distribution pattern of XCO_2 in China was analyzed using the generated dataset.

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

2 Materials and methods

2.1 Datasets and preprocessing

The data used in this study includes satellite XCO_2 data, reanalyzed XCO_2 data, Leaf Area Index (LAI), Fractional Vegetation Coverage (FVC), CO_2 emissions (Emi), population (Pop), Digital Elevation Model (DEM), temperature (T), relative humidity (Rh), surface pressure (Sp), total precipitation (Tp), boundary layer height (Blh), and in situ XCO_2 data from 2015 to 2020. Table 1 summarizes the basic information of these data.

Table 1 Summary of research data

Source	Variable	Spatial resolution	Temporal resolution
OCO-2	XCO_2	2.25km×1.29km	16-daily
CAMS	XCO_2	0.75°×0.75°	3-hourly
GLASS	LAI	0.05°×0.05°	8-daily
	FVC		
ODIAC	CO_2 emissions	1km×1km	Monthly
LandScan	Population	1km×1km	Yearly
SRTM15	DEM	500m×500m	\
Era5	Temperature	0.25°×0.25°	Hourly
	Relative humidity		
	Surface pressure		
	Total precipitation		
	Boundary layer height		
TCCON	XCO_2	\	2min

2.1.1 Satellite XCO2 retrievals

Satellite XCO_2 retrievals were sourced from OCO-2. The OCO mission made the global, space-based measurements of atmospheric CO_2 with the precision, resolution, and coverage needed to characterize CO_2 sources and sinks on regional scales (Crisp et al., 2004). The OCO-2 satellite was launched in July 2014 to measure reflected solar radiation in specific narrow spectral bands in the NIR and SWIR regions with high spectral resolution. Subsequently, the measured spectra are converted into estimated XCO_2 values through various physics-based algorithms and sources of prior information (Taylor et al., 2023). This study selected OCO-2 L2 XCO_2 retrieval from 2015 to 2020 as the data source for the target dataset of the model, which can be accessed at https://disc.gsfc.nasa.gov. The data has a spatial resolution of 2.25km × 1.29 km and a revisiting interval of 16 days. Fig. 1 shows the coverage of OCO-2 retrieval in one revisiting interval.

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

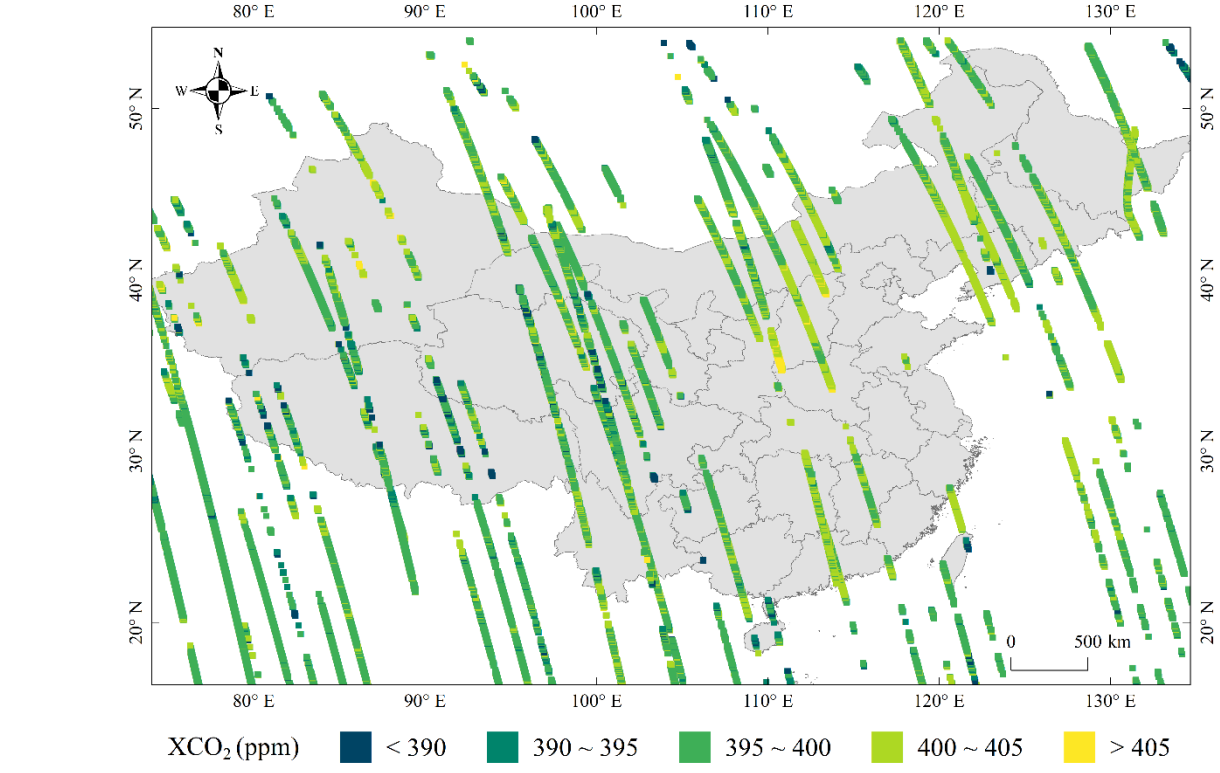


Fig. 1. Spatial distribution of OCO-2 XCO_2 retrievals from January 1 to January 16, 2015 (a revisiting interval) in China.

2.1.2 Reanalyzed XCO2 data

Reanalyzed XCO_2 data was sourced from Copernicus Atmospheric Monitoring Service (CAMS) global greenhouse gas reanalysis (EGG4). European Centre for Medium-Range Weather Forecasts (ECMWF) combines model data with various observations (Such as GOSAT XCO_2 products) using 4D-Var data assimilation method to generate a global CAMS XCO_2 reanalysis dataset (Engelen et al., 2009; Massart et al., 2016). Reanalysis does not require timely release of data, so there is more time to improve data quality. In this study, the CAMS reanalyzed XCO_2 data were collected as the main input data for model with a spatial resolution of 0.75°×0.75° and a temporal resolution of 3 hours. The data can be accessed at https://ads.atmosphere.copernicus.eu/.

2.1.3 Vegetation parameters

Terrestrial vegetation growth is the dominant driver of the seasonality of the concentration of atmospheric CO_2 (Tiwari et al., 2013; Yuan et al., 2018). There is a certain correlation between vegetation growth and seasonal cycle of CO_2 concentration. This study selected LAI and FVC as vegetation parameters to be involved in modeling.

LAI and FVC was obtained from The Global Land Surface Satellite (GLASS) products. GLASS LAI product use

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

generalized regression neural networks (GRNNs) to generate real LAI. This generalized regression neural network combines satellite LAI data and MODIS surface reflectance data (Liang et al., 2021). Considering the low computational efficiency of GRNNs, the multivariate adaptive regression splines (MARS) model was proposed for generating GLASS FVC products from MODIS data (Yang et al., 2016). In this study, the LAI and FVC products has a spatial resolution of 0.05º×0.05º and a temporal resolution of 8 days. The products can be accessed at http://www.glass.umd.edu/.

2.1.4 Anthropogenic factors

There is a close relationship between atmospheric CO_2 concentration and human activities (Fang et al., 2011). To characterize the impact of human activities on atmospheric CO_2 concentration, carbon emissions and population density were added as anthropogenic factors to the input variables of the model.

The Open-source Data Inventory for Anthropogenic CO_2 (ODIAC) is a global high spatial resolution gridded emissions data product that distributes CO_2 emissions from fossil fuel combustion (Oda et al., 2018). ODIAC carbon emissions data product is generated from national emission estimates through spatiotemporal disaggregation. Carbon emission data product was obtained with a spatial resolution of 1km×1km and a temporal resolution of one month from the ODIAC fossil fuel emission dataset (https://db.cger.nies.go.jp/dataset/ODIAC/DL_odiac2022.html).

Landscan data is a global population distribution dataset developed by Oak Ridge National Laboratory (ORNL). It is the only global and annually updated population dataset since 2000 which has been widely used in various studies (Calka and Bielecka, 2019). In this study, the population dataset from Landscan had a spatial resolution of 1km×1km and a temporal resolution of one year, which can be accessed at https://landscan.ornl.gov/.

2.1.5 DEM and meteorological parameters

Due to the higher density of CO_2 compared to air, the concentration of CO_2 is influenced by elevation. And population distribution is closely related to terrain as well. Therefore, in this study, DEM with a spatial resolution of 500m×500m obtained from SRTM15 was introduced in the model input as a geographic covariate (Tozer et al., 2019).

Besides, it has been proven that meteorological parameters such as precipitation and temperature have a correlation with atmospheric CO_2 concentration (Golkar et al., 2020; Perez et al., 2018; Royer, 2006). The meteorological parameters used in this study were derived from the ERA5 reanalysis dataset, which used the same assimilation scheme as CAMS (Hersbach et al., 2020). These meteorological parameter data include relative humidity, temperature, boundary layer height, surface pressure, and total precipitation, all of which had a spatial resolution of 0.25º×0.25º and a temporal resolution of one hour. Considering that OCO-2 satellite measures XCO_2 at a local overpass time of approximately 13:30, only CAMS and ERA5 reanalysis data with time of 6:00 UTC was used in this study. The meteorological parameters data can be accessed at https://cds.climate.copernicus.eu/.

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

2.1.6 In situ measurement

TCCON is a global network of ground-based Fourier transform spectrometers with 34 globally dispersed sites, including 27 operational sites, 3 future sites and 4 previous sites. Column abundances of CO_2, CO, CH_4, N_2O and other molecules that absorb in the near-infrared are remotely measured by the network, which can be directly comparable with the near-infrared total column measurements from space-based instruments (Wunch et al., 2011). In this study, the measurements from Hefei site and Xianghe site was used for model accuracy verification. Table B1 shows the basic information of the sites. And the measurements can be accessed at https://tccondata.org/.

2.1.7 Data preprocessing

To ensure the reliability of the data, high-quality XCO_2 retrievals were acquired through quality control (xco2_quality_flag=0). With the grid center as the interpolated point, IDW was used for XCO_2 retrievals in each grid (0.1°×0.1°) to obtain the CO_2 concentration representing the grid. As for the input data of the model, XCO_2 reanalysis, LAI, FVC, DEM and five meteorological parameter data from the ERA5 dataset were resampled using bilinear interpolation to a spatial resolution of 0.1°×0.1°. Population data and CO_2 emission data are also aggregated into a spatial resolution of 0.1°×0.1°. We assume that LAI and FVC remain unchanged for 8 days, population remains unchanged for a year, while the monthly CO_2 emission data is evenly distributed to each day in month. In addition, due to OCO-2 satellite measuring XCO_2 at a local overpass time of approximately 13:30, XCO_2 measurements between 12:00 and 15:00 daily were screened from TCCON in situ measurement for model validation.

2.2 Method

2.2.1 The framework for generating XCO2 dataset

Fig. 2 presents the workflow for generation of atmospheric CO_2 concentration dataset in this study. In order to generate full-coverage and high-resolution daily XCO_2 in China, CAMS XCO_2 reanalysis, vegetation parameter data, anthropogenic factors data, DEM data, meteorological parameters, and OCO-2 XCO_2 data were collected for DSC-DF-LGB model training, while TCCON data was collected for validation. Then, the input data (i.e. data other than OCO-2 XCO_2 and TCCON) was reprojected and the spatial resolution was adjusted to 0.1°×0.1° using bilinear resampling or aggregation. After that, the OCO-2 XCO_2 data from 2015 to 2020 was extracted and quality control was performed through xco2_quality_flag=0. For OCO-2 XCO_2 located in the same grid (0.1°×0.1°), calculated the XCO_2 value of that grid using IDW. Subsequently, the processed OCO-2 XCO_2 data (target XCO_2) was spatiotemporally matched with the processed input data to construct a dataset for model training. In addition, XCO_2 measurements from 12:00 to 15:00 at the Hefei and Xianghe TCCON sites were extracted for model accuracy validation. When training the DSC-DF-LGB model, the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were used as evaluation indicators for the model. Firstly, the model estimation performance was assessed based on 10-fold cross

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

validation and then used the TCCON sites to further verify the accuracy of model estimation. Finally, the trained DSC-DF-LGB model was used to generate full-coverage and high-resolution daily XCO_2 concentration in China from 2015 to 2020.

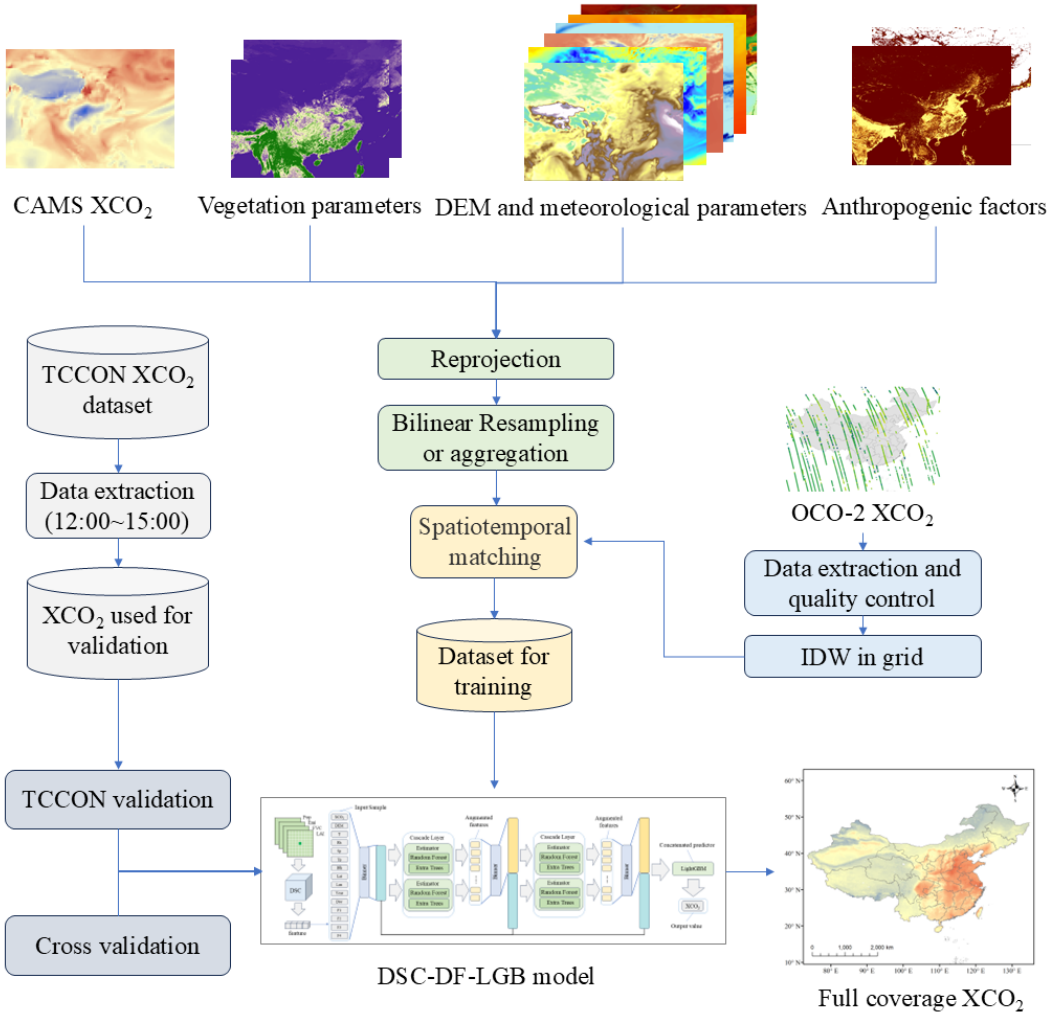


Fig. 2. The workflow for generation of atmospheric CO_2 concentration dataset

2.2.2 Model description

In this study, DSC-DF-LGB was used to generate atmospheric CO_2 datasets. The model structure is shown in Fig. 3. This model consisted of a Deep Separable Convolutional Neural Network and a deep forest concatenated with LightGBM. Given that each variable is independent, DSC (Fig. A1) is used to capture the local features of each variable instead of CNN. Deep Forest (Zhou and Feng, 2019) had 2 cascaded layers, each with 2 estimators (each estimator includes 1 RF and 1 ERT). The cascaded layers were mainly used to generate augmented features. These augmented features and input variables were fed

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

Open Access Earth System Science Data Discussions

together into LightGBM for calculating and outputting estimated XCO_2. The general form of the estimation model is shown in Eq. (1).

$$estimated\ XCO_{2_j} = f_j\left(XCO_{2_j}, DEM_j, T_j, RH_j, Sp_j, Tp_j, Blh_j, Lat_j, Lon_j, Year_j, Doy_j, F1_j, F2_j, F3_j, F4_j\right) \quad (1)$$

Where j is a specific grid, estimated XCO_2 represents the value calculated through DSC-DF-LGB model, f_j is the function of the DF-LGB model at the grid position, and XCO_{2j}, DEM_j, etc. are the various input features described in the second section. Besides, longitude (Lon), latitude (Lat), year, and day of year (Doy) are added to the input data to represent spatiotemporal information. F1, F2, F3 and F4 represent the feature values extracted by DSC.

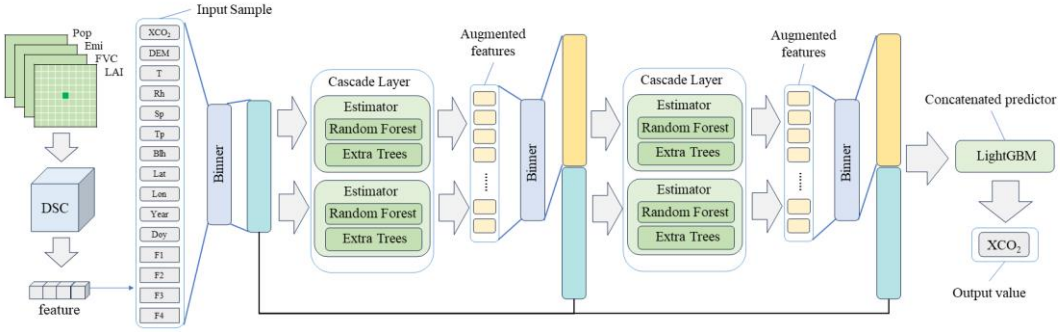


Fig. 3. The structure of the DSC-DF-LGB model.

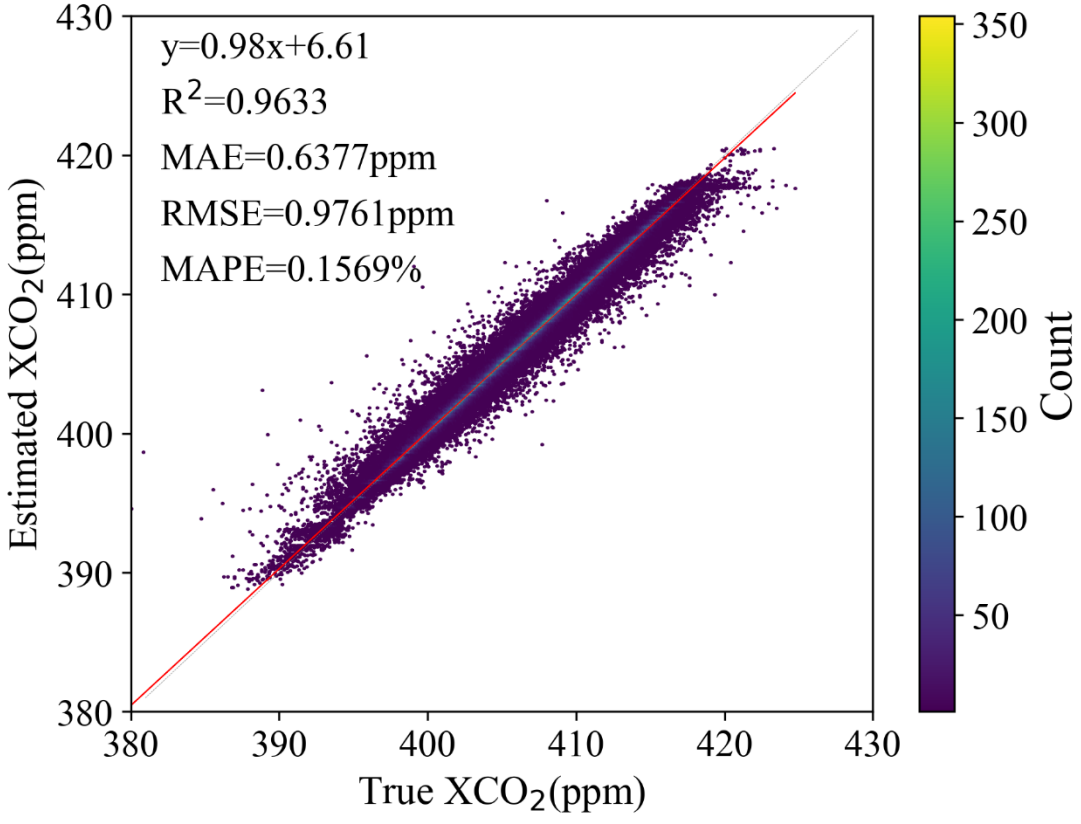
3 Results and Discussions

3.1 Model performance evaluation

Firstly, the overall performance of the DSC-DF-LGB model was evaluated based on 10-fold CV. Fig. 4 reveals the estimation performance of the model through comparison between XCO_2 estimated by model and true XCO_2. The CV result showed that R^2 reached 0.9633, while MAE, RMSE, and MAPE were as low as 0.6377, 0.9761, and 0.1569%, indicating that DSC-DF-LGB has excellent estimation performance. However, when estimating XCO_2, there is a tendency to overestimate low values and underestimate high values. This is due to the robustness of the model itself, which can identify and suppress numerical anomalies. Next, TCCON measurements were used for independent data validation of the model. Fig. 5 illustrates the validation result of XCO_2 estimated by DSC-DF-LGB, and compares the result with CAMS XCO_2 and OCO-2 XCO_2. The estimated XCO_2 have high consistency with TCCON, and the R^2, MAE, RMSE, and MAPE are 0.8786, 1.1956 ppm, 1.5452 ppm, and 0.2910%, respectively. Compared with the validation results of CAMS XCO_2 with R^2 and RMSE of 0.7005, 2.4267 ppm, the estimated XCO_2 accuracy has been significantly improved. Surprisingly, the estimated XCO_2 accuracy is also higher than OCO-2 XCO_2. due to the small amount of OCO-2 XCO2 used for validation and most of the data being located in Xianghe. Table 2 presents the validation results for estimating XCO_2, CAMS XCO_2, and OCO-2 XCO_2 at each site. At both sites, the validation results for estimating XCO_2 were superior to CAMS XCO_2. At the Xianghe site, the validation results of CAMS XCO_2 and OCO-2 XCO_2 were not satisfactory, with R^2 less than 0.5. On the contrary,

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

the estimated XCO_2 at Xianghe site showed quite good validation results, with R^2 close to 0.8. This excellent validation result demonstrates the strong spatial generalization ability of the model and further proved the feasibility of generating full-coverage and high spatiotemporal resolution XCO_2 data through the DSC-DF-LGB model.



225 **Fig. 4.** Comparison between XCO_2 estimated by DSC-DF-LGB model and true XCO_2

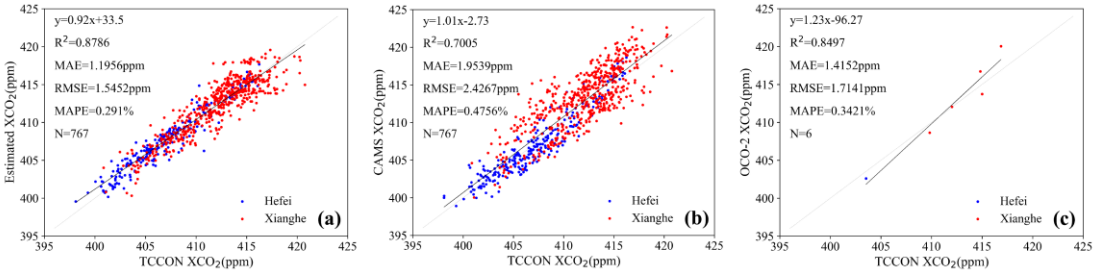


Fig. 5. Comparison of TCCON sites verification results. (a)-(c) represent the validation results of the estimated XCO_2, CAMS XCO_2, and OCO-2 XCO_2

230

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

Open Access Earth System Science Data Discussions

Table 2 Verification accuracy for each TCCON site

Data source	Hefei				Xianghe			
	R^2	MAE (ppm)	RMSE (ppm)	MAPE (%)	R^2	MAE (ppm)	RMSE (ppm)	MAPE (%)
Estimated XCO_2	0.9092	0.9881	1.2137	0.2429	0.7949	1.2941	1.6799	0.3138
CAMS XCO_2	0.8793	1.0833	1.3993	0.2659	0.4364	2.3674	2.7849	0.5752
OCO-2 XCO_2	\	0.9257	0.9257	0. 2294	0.4573	1.5131	1.8315	0.3647

3.2 The ability of the model to generate earlier and later datasets

Due to the limitation of OCO-2 XCO_2 retrievals, DSC-DF-LGB model can only train data after 2015. In order to understand the model's ability to estimate past XCO_2, only data from 2016 to 2020 was used for model training, and 10-fold CV was used to evaluate model performance. Then the generalization of the model was tested using dataset in 2015 as an independent test set. As shown in Table 3, the CV result was good, with R^2, MAE, RMSE, and MAPE of 0.9478, 0.6395 ppm, 0.9798 ppm, and 0.1568%, respectively. However, during independent dataset testing, there is no linear relationship between input and output data, with an R^2 of -0.1767. It indicates that the model has weak temporal generalization ability and cannot estimate past XCO_2. This is mainly due to the inherent shortcomings of decision tree based models and their inability to capture the trend of XCO_2 in 2015 without training.

Table 3 Model validation results for estimating past XCO_2

Verification method	R^2	MAE (ppm)	RMSE (ppm)	MAPE (%)
CV	0.9478	0.6395	0.9798	0.1568
Independent dataset testing	-0.1767	2.5324	2.8644	0.6352

Actually, most variables have data from 2020 onwards, but the time range of the generated XCO_2 dataset is limited by CAMS to 2020. Therefore, we attempted to remove CAMS variables and use the remained ones to train the DSC-DF-LGB model. Similar to section 3.1, CV and independent site validation are used to evaluate model performance. As shown in Fig. 6, the model performed well in cross validation, with R^2, MAE, RMSE, and MAPE values of 0.9621, 0.6542 ppm, 0.9925 ppm, and 0.1602%, respectively. However, the model demonstrated poor performance in independent site validation, with R^2, MAE, RMSE, and MAPE of 0.8101, 1.4736 ppm, 1.9322 ppm, and 0.3585%, respectively. Compared with the model trained on datasets with CAMS, there is only a slight loss of accuracy in CV, while estimation accuracy decreased in independent site validation, with RMSE increasing from 1.55 ppm to 1.93 ppm. In addition, if data from 2020 onwards is added for training model, it may improve estimation accuracy to an extent. Therefore, it is feasible to generate later XCO_2 datasets by DSC-DF-LGB model trained on datasets without CAMS, but the estimation accuracy will inevitably decrease.

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

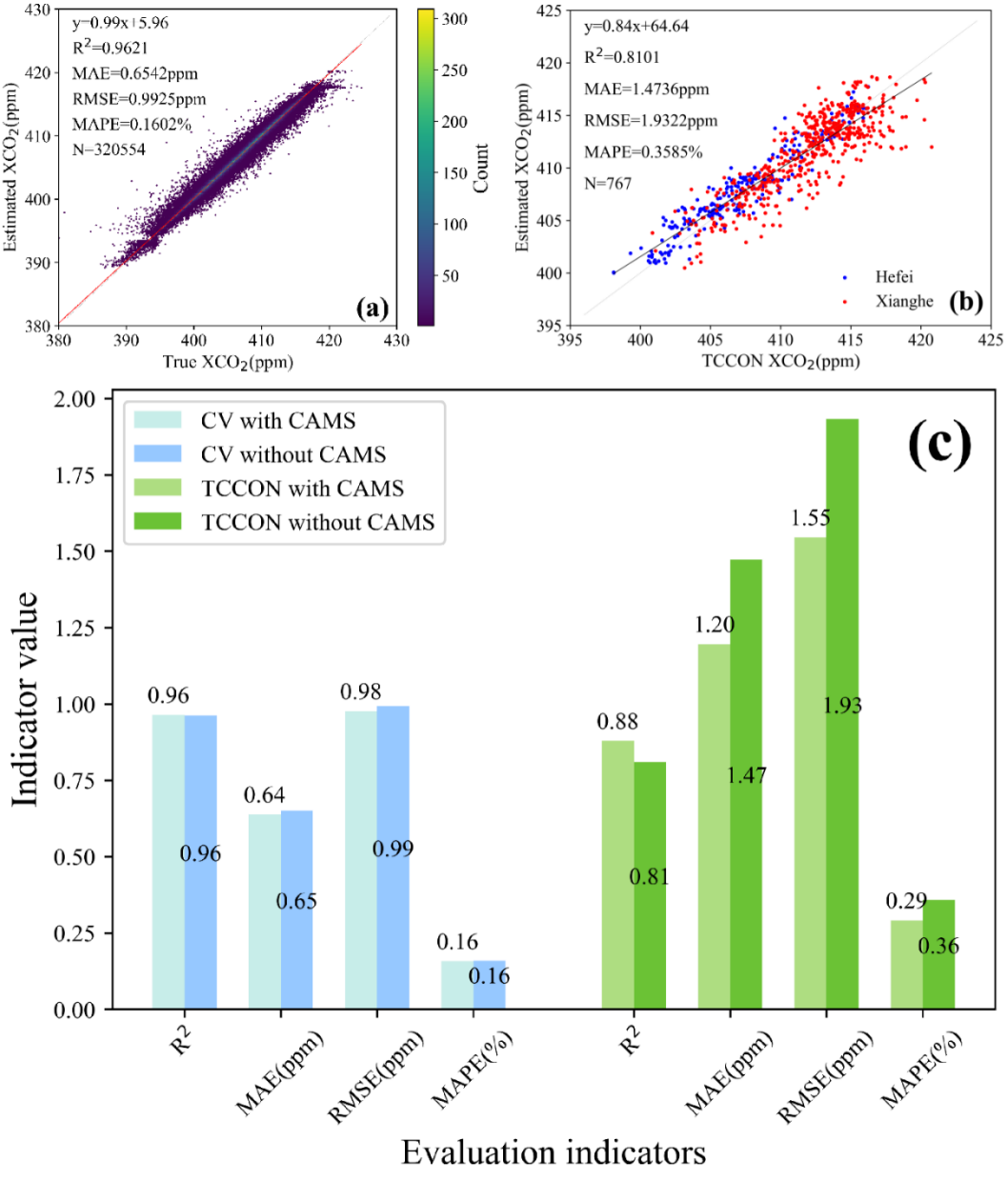


Fig. 6. Performance evaluation of DSC-DF-LGB models trained on datasets without CAMS. (a) shows the CV validation result of the model, (b) represents the TCCON independent site validation result, and (c) demonstrates the comparison results with the model trained on datasets with CAMS.

3.3 The spatiotemporal distribution of XCO_2 in China

Fig.7 shows the spatial distribution of monthly average XCO_2 in 2018. The seasonal variation of atmospheric CO_2

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

concentration in China is significant. The atmospheric CO_2 concentration is higher in spring (from March to May), with the highest CO_2 concentration occurring in April. Subsequently, the atmospheric CO_2 concentration began to decrease and reached its lowest point in August. And from August to April of the following year, the atmospheric CO_2 concentration continued to rise again. The seasonal variation of atmospheric CO_2 is mainly caused by the growth and death of vegetation. Vegetation undergoes vigorous photosynthesis in summer, absorbing a large amount of CO_2 from the atmosphere. In the autumn and winter seasons, when vegetation dies and decomposes, CO_2 is released into the air (Tiwari et al., 2013). In addition, the atmospheric CO_2 concentration in Northeast China was lower in July and August, which is related to the strong carbon sequestration ability of black soil (Liu et al., 2019).

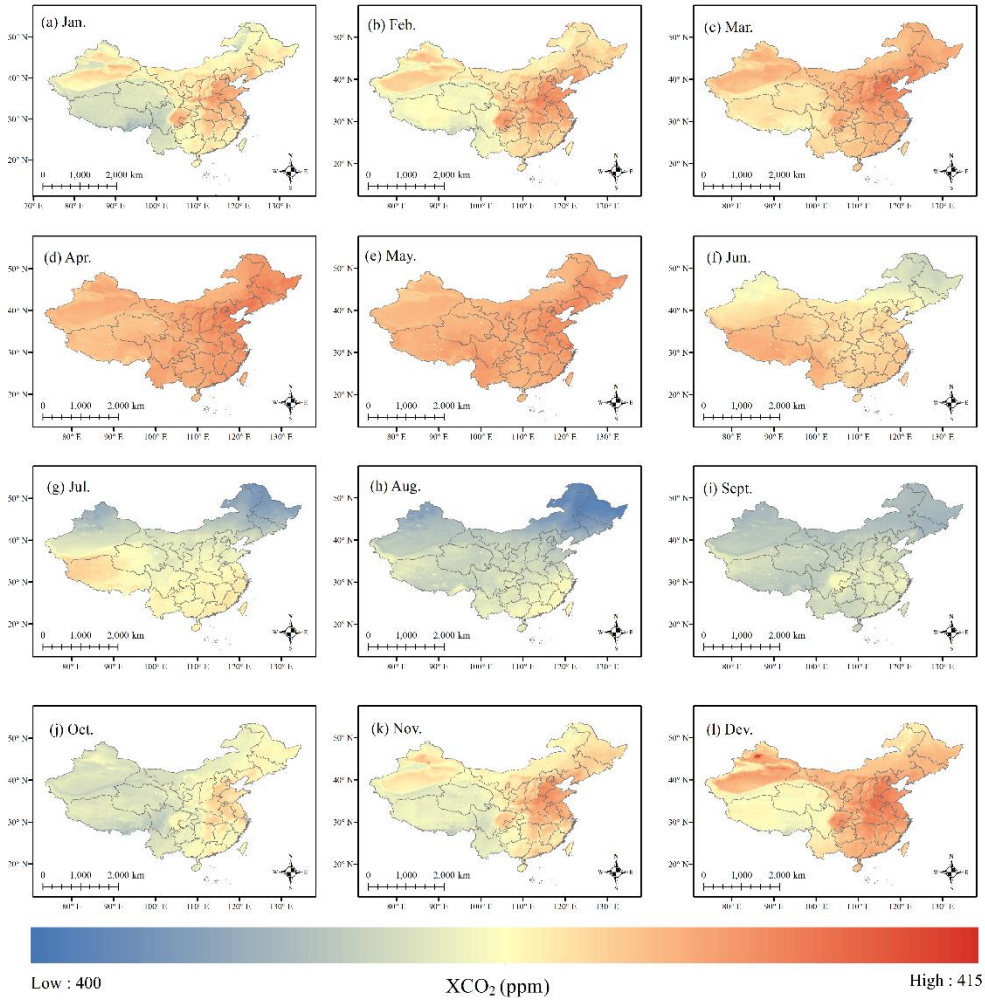


Fig. 7. Monthly average XCO_2 in China in 2018. (a)-(l) show the spatial distribution of XCO_2 from January to December.

Fig. 8 shows the spatial distribution of annual average XCO_2 in 2018, which is compared to carbon emissions. The atmospheric CO_2 concentration exhibits significant spatial heterogeneity. The CO_2 concentration in the eastern coastal areas is significantly higher than that in the western regions. This distribution pattern is basically consistent with that of carbon

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

emissions, indicating an inseparable relationship between excessive carbon emissions and the increase in atmospheric CO_2 concentration. Excessive carbon emissions have led to higher atmospheric CO_2 concentrations, especially in the Beijing Tianjin Hebei (BTH), Yangtze River Delta (YRD), and Pearl River Delta (PRD) regions where industry is developed. Besides, the significant carbon emissions from the Chengdu Chongqing Economic Circle (CCEC) have also led to higher atmospheric CO_2 concentrations in the region compared to surrounding cities. This indicates that the concentration of atmospheric CO_2 is significantly influenced by human activities.

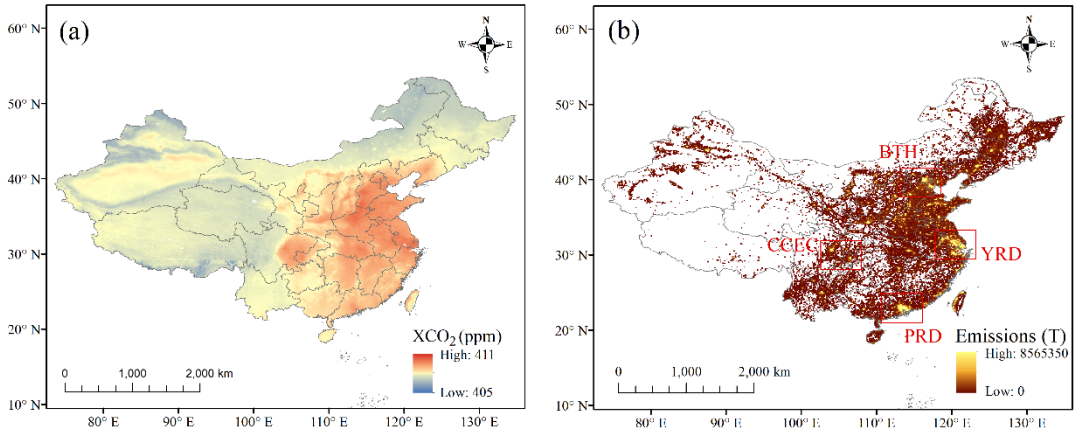


Fig. 8. Comparison of Annual Average XCO_2 and Total Carbon Emissions in 2018. (a) represents the spatial distribution of average XCO_2 in 2018, (b) represents the spatial distribution of carbon emissions in 2018.

Subsequently, the annual average XCO_2 in China from 2015 to 2020 was calculated. As shown in Fig. 9, the annual average XCO_2 in China has increased from 400.11 ppm in 2015 to 413.04 ppm in 2020. The annual average growth is 2.59 ppm/yr, manifesting an overall downward trend. Except for the annual average XCO_2 growth of 3.25 ppm in 2016, the growth in all other years is below 3 ppm. This indicates the tremendous efforts made by the Chinese government to limit global climate change after signing the Paris Agreement.

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

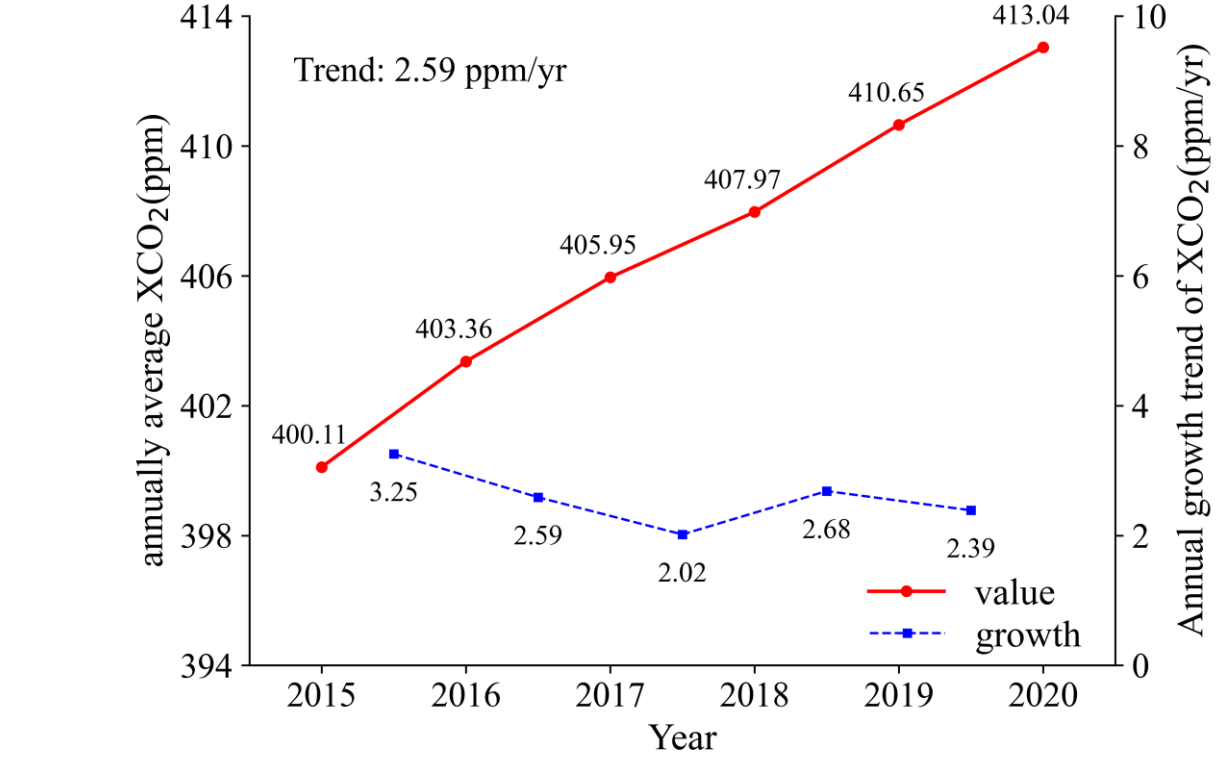


Fig. 9. Annual average XCO_2 and annual growth in China from 2015 to 2020.

4 Code and Data availability

The full-coverage 0.1° daily XCO_2 dataset and the code are available at https://zenodo.org/doi/10.5281/zenodo.12696674 (Huang, 2024).

5 Conclusion

This study generated a daily full-coverage XCO_2 dataset based on DSC-DF-LGB. By calculating, monthly and annual average XCO_2 datasets can be also generated. Compared to CAMS reanalysis, the generated XCO_2 dataset has higher spatial resolution and data accuracy. However, although the model has strong spatial generalization, its temporal generalization is poor, making it unable to generate earlier or later XCO_2 datasets. Subsequently, an in-depth analysis was conducted on the spatiotemporal distribution of XCO_2 in China. The results show that the XCO_2 concentration is highest in April and lowest in August, demonstrating a clear seasonal cycle, mainly due to the growth and death of vegetation; The spatial distribution of XCO_2 is consistent with that of carbon emissions, with higher CO_2 concentrations in the BTH, YRD, PRD and CCEC in China, which indicates that the spatial distribution of XCO_2 is significantly influenced by human activities; The XCO_2 is

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

300 increasing year by year, but its growth is gradually slowing down. This dataset generated in this study can be further used for related research on carbon sources and sinks.

Appendix A

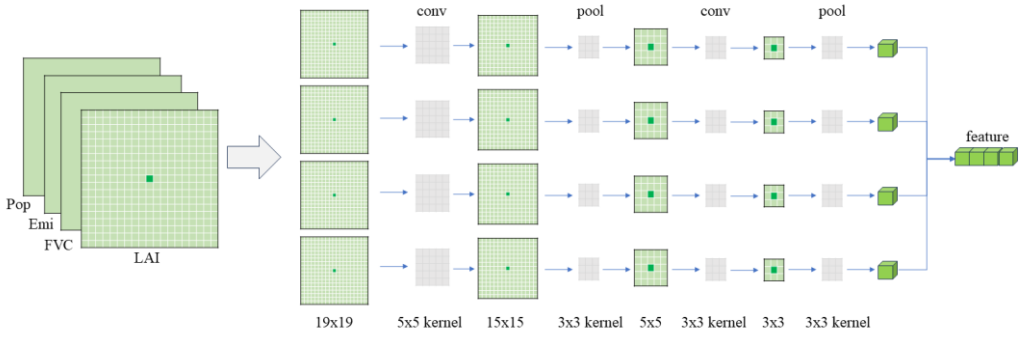


Fig. A1. The structure of DSC. The window size is 19x19. And features are extracted through two convolutional layers and average
305 pooling layer.

Appendix B

Table B1 TCCON site information

ID	Name	Longitude	Latitude	Data range
1	Hefei	117.17	31.90	2015.11~2022.12
2	Xianghe	116.96	39.80	2018.06~2023.02

Author contributions

Conceptualization: HY and XH; methodology: XH and HF; validation: XH, QL and LC; Investigation: YQ, YY and XH;
310 writing – original draft preparation: XH; writing – review and editing: HY, GF and XH; funding acquisition: HY.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgment

This research was supported by the Third Xinjiang Scientific Expedition Program (Grant No.2022xjkk1006), the Xinjiang

https://doi.org/10.5194/essd-2024-371
Preprint. Discussion started: 28 October 2024

315 Uygur Autonomous Region Key Research and Development Program (2022B01012-1), the National Natural Science Foundation of China (41971335), the Science and Technology Innovation Project of Jiangsu Provincial Department of Natural Resources (2023018,2022004), the Open Research Project of The Hubei Key Laboratory of Intelligent Geo-Information Processing (KLIGIP-2023-A04).

The OCO-2 XCO_2 retrievals used in this study were produced by the OCO-2 project at the Jet Propulsion Laboratory, 320 California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center.

References

Calka, B., Bielecka, E.: Reliability analysis of landscan gridded population data. The case study of Poland, Isprs Int J Geoinf, 8 (5), 222, https://doi.org/10.3390/ijgi8050222, 2019.

325 Chen, X., Wang, Z., Shangguan, Y., Yu, J., Hu, B., Shen, Q., Xue, J., Zhang, X., Shi, Z.: Estimating monthly surface ozone using multi-source satellite products in China based on deep forest model, Atmos Environ, 307, 119819, https://doi.org/10.1016/j.atmosenv.2023.119819, 2023.

Crisp, D., Atlas, R.M., Breon, F.M., Brown, L.R., Burrows, J.P., Ciais, P., Connor, B.J., Doney, S.C., Fung, I.Y., Jacob, D.J., Miller, C.E., O'Brien, D., Pawson, S., Randerson, J.T., Rayner, P., Salawitch, R.J., Sander, S.P., Sen, B., Stephens, 330 G.L., Tans, P.P., Toon, G.C., Wennberg, P.O., Wofsy, S.C., Yung, Y.L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., Schroll, S.: The orbiting carbon observatory (OCO) mission, Adv Space Res, 34 (4), 700-709, https://doi.org/10.1016/j.asr.2003.08.062, 2002.

Cui, L., Yang, H., Qiao, Y., Huang, X., Feng, G., Lv, Q., Fan, H.:Estimating high spatio-temporal resolution XCO_2 using spatial features deep fusion model, Atmos Res, 308, 107542, https://doi.org/10.1016/j.atmosres.2024.107542, 2024.

335 Engelen, R.J., Serrar, S., Chevallier, F.: Four-dimensional data assimilation of atmospheric CO_2 using airs observations, Journal of Geophysical Research: Atmospheres, 114 (D3), https://doi.org/10.1029/2008JD010739, 2009.

Fang, J.Y., Zhu, J.L., Wang, S.P., Yue, C., Shen, H.H.: Global warming, human-induced carbon emissions, and their uncertainties, Sci China Earth Sci, 54 (10), 1458-1468, https://doi.org/10.1007/s11430-011-4292-0, 2011.

Fu, Y., Liao, H., Tian, X., Gao, H., Cai, Z., Han, R.: Sensitivity of the simulated CO_2 concentration to inter-annual variations 340 of its sources and sinks over east Asia, Adv Clim Chang Res, 10 (4), 250-263, https://doi.org/10.1016/j.accre.2020.03.001, 2019.

Golkar, F., Al-Wardy, M., Saffari, S.F., Al-Aufi, K., Al-Rawas, G.: Using OCO-2 satellite data for investigating the variability of atmospheric CO_2 concentration in relationship with precipitation, relative humidity, and vegetation over Oman, Water (Basel), 12 (1), 101. https://doi.org/10.3390/w12010101, 2020.

345 He, C., Ji, M., Grieneisen, M.L., Zhan, Y.: A review of datasets and methods for deriving spatiotemporal distributions of atmospheric CO_2, J Environ Manage, 322, 116101, https://doi.org/10.1016/j.jenvman.2022.116101, 2022a.

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 XCO_2 across China based on OCO-2 satellite retrievals and Carbontracker output, Geophys Res Lett, 49, e2022GL098435, https://doi.org/10.1029/2022GL098435, 2022b.

He, Q., Ye, T., Chen, X., Dong, H., Wang, W., Liang, Y., Li, Y.: Full-coverage mapping high-resolution atmospheric CO_2 concentrations in China from 2015 to 2020: spatiotemporal variations and coupled trends with particulate pollution, J Clean Prod, 428, 139290, https://doi.org/10.1016/j.jclepro.2023.139290, 2023.

He, S., Yuan, Y., Wang, Z., Luo, L., Zhang, Z., Dong, H., Zhang, C.: Machine learning model-based estimation of XCO_2 with high spatio-temporal resolution in China, Atmosphere (Basel), 14 (3), 13, https://doi.org/10.3390/atmos14030436, 2023.

He, Z., Lei, L., Zhang, Y., Sheng, M., Wu, C., Li, L., Zeng, Z., Welp, L.R.: Spatio-temporal mapping of multi-satellite observed column atmospheric CO_2 using precision-weighted kriging method, Remote Sens (Basel), 12 (3), 576, https://doi.org/10.3390/rs12030576, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N.: The era5 global reanalysis, Q J R Meteorol Soc, 146 (730), 1999-2049, https://doi.org/10.1002/qj.3803, 2020.

Huang, S., Zhang, X., Chen, N., Ma, H., Zeng, J., Fu, P., Nam, W., Niyogi, D.: Generating high-accuracy and cloud-free surface soil moisture at 1km resolution by point-surface data fusion over the southwestern U.S, Agric for Meteorol, 321, 17, https://doi.org/10.1016/j.agrformet.2022.108985, 2022.

Huang, X.: Full-coverage daily 0.1° XCO2 in China, Zenodo [data set], https://zenodo.org/doi/10.5281/zenodo.12696674, 2024.

IPCC: Summary for policymakers. In: global warming of 1.5°c. An IPCC special report on the impacts of global warming of 1.5°c above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, Cambridge university press, Cambridge, UK and New York, NY, USA, pp. 3-24, https://doi.org/10.1017/9781009157940.001, 2018.

Kump, L.R.: What drives climate? Nature, 408 (6813), 2, https://doi.org/10.1038/35047188, 2000.

Lan, X., Tans, P., W., K.: Thoning: trends in globally-averaged CO_2 determined from NOAA global monitoring laboratory measurements, Version 2024-06, https://doi.org/10.15138/9n0h-zh07, 2024.

Li, J., Jia, K., Wei, X., Xia, M., Chen, Z., Yao, Y., Zhang, X., Jiang, H., Yuan, B., Tao, G., Zhao, L.: High-spatiotemporal resolution mapping of spatiotemporally continuous atmospheric CO_2 concentrations over the global continent, Int J Appl Earth Obs Geoinf, 108, 102743, https://doi.org/10.1016/j.jag.2022.102743, 2022.

Li, K., Bai, K., Jiao, P., Chen, H., He, H., Shao, L., Sun, Y., Zheng, Z., Li, R., Chang, N.: Developing unbiased estimation of atmospheric methane via machine learning and multiobjective programming based on tropomi and gosat data, Remote Sens Environ, 304, 114039, https://doi.org/10.1016/j.rse.2024.114039, 2024.

Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L.: Estimating ground-level pm2.5 by fusing satellite and station observations: a geo-intelligent deep learning approach, Geophys Res Lett, 44 (23), 11, 911-985, 993, https://doi.org/10.1002/2017GL075710, 2017.

Li, T., Wu, J., Wang, T.: Generating daily high-resolution and full-coverage XCO_2 across China from 2015 to 2020 based on OCO-2 and cams data, Sci Total Environ, 893, 164921, https://doi.org/10.1016/j.scitotenv.2023.164921, 2023.

Li, Z., Yuan, Q., Su, X.: High-spatial-resolution surface soil moisture retrieval using the deep forest model in the cloud environment over the Tibetan plateau, Geo Spat Inf Sci, https://doi.org/10.1080/10095020.2024.2307931, 2024.

Liang, A., Gong, W., Han, G., Xiang, C.: Comparison of satellite-observed XCO_2 from GOSAT, OCO-2, and ground-based TCCON, Remote Sensing (Basel, Switzerland), 9 (10), 1033. https://doi.org/10.3390/rs9101033, 2017.

Liang, A., Pang, R., Chen, C., Xiang, C.: XCO_2 fusion algorithm based on multi-source greenhouse gas satellites and Carbontracker, Atmosphere (Basel), 14 (9), 1335, https://doi.org/10.3390/atmos14091335, 2023.

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., Zhao, X., Zhou, J.: The global land surface satellite (glass) product suite, Bull Am Meteorol Soc, 102 (2), E323-E337, https://doi.org/10.1175/BAMS-D-18-0341.1, 2021.

Liu, S., Fan, R., Yang, X., Zhang, Z., Zhang, X., Liang, A.: Decomposition of maize stover varies with maize type and stover management strategies: a microcosm study on a black soil (Mollisol) in northeast China, J Environ Manage, 234, 226-236, https://doi.org/10.1016/j.jenvman.2019.01.008, 2019.

Liu, Y., Wang, J., Yao, L., Chen, X., Cai, Z., Yang, D., Yin, Z., Gu, S., Tian, L., Lu, N., Lyu, D.: The Tansat mission: preliminary global observations, Sci Bull (Beijing), 63 (18), 1200-1207, https://doi.org/10.1016/j.scib.2018.08.004, 2018.

Massart, S., Agustí-Panareda, A., Heymann, J., Buchwitz, M., Chevallier, F., Reuter, M., Hilker, M., Burrows, J.P., Deutscher, N.M., Feist, D.G., Hase, F., Sussmann, R., Desmet, F., Dubey, M.K., Griffith, D.W.T., Kivi, R., Petri, C., Schneider, M., Velazco, V.A.: Ability of the 4-d-var analysis of the GOSAT BESD XCO_2 retrievals to characterize atmospheric CO_2 at large and synoptic scales, Atmos Chem Phys, 16 (3), 1653-1671, https://doi.org/10.5194/acp-16-1653-2016, 2016.

Miller, J.B., Tans, P.P., Gloor, M.: Steps for success of oco-2, Nat Geosci, 7 (10), 691, https://doi.org/10.1038/ngeo2255, 2014.

Mustafa, F., Bu, L., Wang, Q., Yao, N., Shahzaman, M., Bilal, M., Rana, W.A., Iqbal, R.: Neural-network-based estimation of regional-scale anthropogenic CO_2 emissions using an orbiting carbon observatory-2 (OCO-2) dataset over east and west Asia, Atmos Meas Tech, 14 (11), 7277-7290, https://doi.org/10.5194/amt-14-7277-2021, 2021.

Oda, T., Maksyutov, S., Andres, R.J.: The open-source data inventory for anthropogenic CO_2, version 2016 (odiac2016): a

global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth Syst Sci Data, 10 (1), 87-107, https://doi.org/10.5194/essd-10-87-2018, 2018.

Perez, I.A., Sanchez, M.L., Garcia, M.A., Pardo, N., Fernandez-Duque, B.: The influence of meteorological variables on CO_2 and CH_4 trends recorded at a semi-natural station, J Environ Manage, 209, 37-45, https://doi.org/10.1016/j.jenvman.2017.12.028, 2018.

Petit, J.R., Raynaud, D.: Forty years of ice-core records of CO_2, Nature, 579 (7800), 505-506, https://doi.org/10.1038/d41586-020-00809-8, 2020.

Ran, Y., Li, X.: Tansat: a new star in global carbon monitoring from China. Sci Bull (Beijing), 64 (5), 2, https://doi.org/10.1016/j.scib.2019.01.019, 2019.

Royer, D.L.: CO_2-forced climate thresholds during the phanerozoic, Geochim Cosmochim Acta, 70 (23), 5665-5675, https://doi.org/10.1016/j.gca.2005.11.031, 2006.

Siabi, Z., Falahatkar, S., Alavi, S.J.: Spatial distribution of XCO2 using oco-2 data in growing seasons. J Environ Manage, 244, 110-118. https://doi.org/10.1016/j.jenvman.2019.05.049, 2019.

Song, Z., Chen, B., Huang, J.: Combining himawari-8 AOD and deep forest model to obtain city-level distribution of $PM_{2.5}$ in China, Environ Pollut, 297, 118826, https://doi.org/10.1016/j.envpol.2022.118826, 2022a.

Song, Z., Chen, B., Zhang, P., Guan, X., Wang, X., Ge, J., Hu, X., Zhang, X., Wang, Y.: High temporal and spatial resolution $PM_{2.5}$ dataset acquisition and pollution assessment based on FY-4A TOAR data and deep forest model in China, Atmos Res, 274, 14, https://doi.org/10.1016/j.atmosres.2022.106199, 2022b.

Tang, Y., Deng, R., Liang, Y., Zhang, R., Cao, B., Liu, Y., Hua, Z., Yu, J.: Estimating high-spatial-resolution daily $PM_{2.5}$ mass concentration from satellite top-of-atmosphere reflectance based on an improved random forest model, Atmos Environ, 302, 119724, https://doi.org/10.1016/j.atmosenv.2023.119724, 2023.

Taylor, T.E., O'Dell, C.W., Baker, D., Bruegge, C., Chang, A., Chapsky, L., Chatterjee, A., Cheng, C., Chevallier, F., Crisp, D., Dang, L., Drouin, B., Eldering, A., Liang, F., Fisher, B., Fu, D., Gunson, M., Haemmerle, V., Keller, G.R., Kiel, M., Kuai, L., Kurosu, T., Lambert, A., Laughner, J., Lee, R., Liu, J., Mandrake, L., Marchetti, Y., Mcgarragh, G., Merrelli, A., Nelson, R.R., Osterman, G., Oyafuso, F., Palmer, P.I., Payne, V.H., Rosenberg, R., Somkuti, P., Spiers, G., To, C., Weir, B., Wennberg, P.O., Yu, S., Zong, J.: Evaluating the consistency between OCO-2 and OCO-3 XCO_2 estimates derived from the NASA ACOS version 10 retrieval algorithm, Atmos Meas Tech, 16 (12), 3173-3209, https://doi.org/10.5194/amt-16-3173-2023, 2023.

Tiwari, Y.K., Revadekar, J.V., Ravi Kumar, K.: Variations in atmospheric carbon dioxide and its association with rainfall and vegetation over India, Atmos Environ, 68, 45-51, https://doi.org/10.1016/j.atmosenv.2012.11.040, 2013.

Tozer, B., Sandwell, D.T., Smith, W.H.F., Olson, C., Beale, J.R., Wessel, P.: Global bathymetry and topography at 15 arc sec: srtm15+, Distributed by Opentopography, https://doi.org/10.5069/g92r3pt9, 2019.

Voigt, C.: The compliance and implementation mechanism of the paris agreement, Review of European, Comparative & International Environmental Law, 25 (2), 161-173, https://doi.org/10.1111/reel.12155, 2016.

Wang, Q., Mustafa, F., Bu, L., Zhu, S., Liu, J., Chen, W.: Atmospheric carbon dioxide measurement from aircraft and comparison with OCO-2 and Carbontracker model data, Atmos Meas Tech, 14 (10), 6601-6617, https://doi.org/10.5194/amt-14-6601-2021, 2021.

Wang, W., He, J., Feng, H., Jin, Z.: High-coverage reconstruction of XCO_2 using multisource satellite remote sensing data in Beijing–Tianjin–Hebei region, International Journal of Environmental Research and Public Health, 19 (17), 10853, https://doi.org/10.3390/ijerph191710853, 2022.

Wang, Y., Yuan, Q., Li, T., Yang, Y., Zhou, S., Zhang, L.: Seamless mapping of long-term (2010-2020) daily global XCO_2 and XCH_4 from the greenhouse gases observing satellite (GOSAT), orbiting carbon observatory 2 (oco-2), and cams global greenhouse gas reanalysis (cams-egg4) with a spatiotemporally self-supervised fusion method, Earth Syst Sci Data, 15 (8), 3597-3622, https://doi.org/10.5194/essd-15-3597-2023, 2023.

Wang, Y., Yuan, Q., Li, T., Zhu, L.: Global spatiotemporal estimation of daily high-resolution surface carbon monoxide concentrations using deep forest, J Clean Prod, 350, 131500, https://doi.org/10.1016/j.jclepro.2022.131500, 2022.

Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A., Liu, L., Wu, H., Song, Y.: Improved 1 km resolution $PM_{2.5}$ estimates across China using enhanced space–time extremely randomized trees, Atmos Chem Phys, 20 (6), 3273-3289, https://doi.org/10.5194/acp-20-3273-2020, 2020.

Wunch, D., Toon, G.C., Blavier, J.L., Washenfelder, R.A., Notholt, J., Connor, B.J., Griffith, D.W.T., Sherlock, V., Wennberg, P.O.: The total carbon column observing network, Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences, 369 (1943), 2087-2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Yang, H., Feng, G., Xiang, R., Xu, Y., Qin, Y., Li, S.: Spatio-temporal validation of airs CO_2 observations using GAW, HIPPO and TCCON. Remote Sens (Basel), 12 (21), 3583, https://doi.org/10.3390/rs12213583, 2020.

Yang, L., Jia, K., Liang, S., Liu, J., Wang, X.: Comparison of four machine learning methods for generating the glass fractional vegetation cover product from MODIS data, Remote Sens (Basel), 8 (8), 682, https://doi.org/10.3390/rs8080682, 2016.

Yang, Q., Yuan, Q., Li, T.: Ultrahigh-resolution $PM_{2.5}$ estimation from top-of-atmosphere reflectance with machine learning: theories, methods, and applications, Environ Pollut, 306, 119347, https://doi.org/10.1016/j.envpol.2022.119347, 2022.

Yoshida, Y., Ota, Y., Eguchi, N., Kikuchi, N., Nobuta, K., Tran, H., Morino, I., Yokota, T.: Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the greenhouse gases observing satellite, Atmos Meas Tech, 4 (4), 717-734, https://doi.org/10.5194/amt-4-717-2011, 2011.

Yuan, Q., Li, S., Yue, L., Li, T., Shen, H., Zhang, L.: Monitoring the variation of vegetation water content with machine learning methods: point–surface fusion of MODIS products and GNSS-IR observations, Remote Sensing (Basel, Switzerland), 11 (12), 1440, https://doi.org/10.3390/rs11121440, 2019.

Yuan, W., Piao, S., Qin, D., Dong, W., Xia, J., Lin, H., Chen, M.: Influence of vegetation growth on the enhanced seasonality of atmospheric CO_2, Global Biogeochem Cycles, 32 (1), 32-41, https://doi.org/10.1002/2017GB005802,

2018.

Yue, T., Zhang, L., Zhao, M., Wang, Y., Wilson, J.: Space- and ground-based CO_2 measurements: a review. Science China Earth Sciences, 59 (11), 2089-2097, https://doi.org/10.1007/s11430-015-0239-7, 2016.

Zhang, H., Luo, M., Zhan, W., Zhao, Y., Yang, Y., Ge, E., Ning, G., Cong, J.: Himic-monthly: a 1 km high-resolution atmospheric moisture index collection over China, 2003-2020, Sci Data, 11 (1), 425, https://doi.org/10.1038/s41597-024-03230-2, 2024.

Zhang, L., Li, T., Wu, J.: Deriving gapless CO_2 concentrations using a geographically weighted neural network: China, 2014–2020, Int J Appl Earth Obs Geoinf, 114, 103063, https://doi.org/10.1016/j.jag.2022.103063, 2022.

Zhang, L.L., Yue, T.X., Wilson, J.P., Zhao, N., Zhao, Y.P., Du, Z.P., Liu, Y.: A comparison of satellite observations with the XCO_2 surface obtained by fusing TCCON measurements and geos-chem model outputs, Sci Total Environ, 601, 16, https://doi.org/10.1016/j.scitotenv.2017.06.018, 2017.

Zhang, M., Liu, G.: Mapping contiguous XCO_2 by machine learning and analyzing the spatio-temporal variation in China from 2003 to 2019, Sci Total Environ, 858, 159588, https://doi.org/10.1016/j.scitotenv.2022.159588, 2023.

Zhao, C., Liu, Z., Wang, Q., Ban, J., Chen, N.X., Li, T.: High-resolution daily aod estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region, Atmos Environ, 203, 70-78, https://doi.org/10.1016/j.atmosenv.2019.01.045, 2019.

Zhou, Z., Feng, J.: Deep forest, Natl Sci Rev, 6 (1), 13, https://doi.org/10.1093/nsr/nwy108, 2019.