https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

Atmospheric Measurement Techniques Discussions

Fast retrieval of XCO_2 over East Asia based on the OCO-2 spectral measurements

Fengxin Xie[1,2], Tao Ren[1], Changying Zhao[1], Yuan Wen[3], Yilei Gu[3], Minqiang Zhou[4], Pucai Wang[4], Kei Shiomi[5], and Isamu Morino[6]

[1]China-UK Low Carbon College, Shanghai Jiao Tong University, Shanghai, China
[2]Atmosphere and Ocean Research Institute, the University of Tokyo, Chiba, Japan
[3]Shanghai Institute of Satellite Engineering, Shanghai, China
[4]Institute of Atmospheric Physics, Chinese Academy of Science, Beijing, China
[5]Earth Observation Research Center, Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan
[6]Satellite Remote Sensing Section and Satellite Observation Center, Earth system Division, National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki 305-8506, Japan

Correspondence: Tao Ren (tao.ren@sjtu.edu.cn)

Abstract. The increase in greenhouse gas concentrations, particularly CO_2, has significant implications for global climate patterns and various aspects of human life. Spaceborne remote sensing satellites play a crucial role in high-resolution monitoring of atmospheric CO_2. However, the next generation of greenhouse gas monitoring satellites is expected to face challenges such as low retrieval efficiency and insufficient retrieval accuracy. To address these challenges, this study focuses on enhancing the retrieval of column-averaged dry air mole fraction of carbon dioxide (XCO_2) using spectral data from the OCO-2 satellite. A novel approach based on neural network (NN) models is proposed to tackle the nonlinear inversion problems associated with XCO_2 retrieval. The study employs a data-driven supervised learning method and explores two distinct training strategies. Firstly, training is conducted using experimental data obtained from the inversion of traditional optimization models, which are released as the OCO-2 satellite products. Secondly, training is performed using a simulated dataset generated by an accurate forward calculation model. The inversion and prediction performance of the machine learning model for XCO_2 is compared, analyzed, and discussed for the observed region. The results demonstrate that the model trained on simulated data accurately predicts XCO_2 in the target area. Furthermore, when compared to OCO-2 satellite product data, the developed XCO_2 retrieval model achieves rapid predictions (<1 ms) with high precision (2 ppm or approximately 0.5%). The accuracy of the machine learning model's retrieval results is validated against reliable data from TCCON sites, demonstrating its capability to capture CO_2 seasonal variations and annual growth trends effectively.

1 Introduction

Since the industrial revolution, human activities have released large amounts of greenhouse gases, primarily carbon dioxide, into the atmosphere. This continual increase in emissions has led to global warming and disrupted human societies and ecosystems (Zehr, 2015). Accurately estimating atmospheric carbon fluxes is critical for implementing effective emission reduction strategies at national and regional levels. However, precise carbon flux estimates require assimilating carbon dioxide concentration data across regions, using measurements of atmospheric column-averaged dry air mole fraction of carbon dioxide (XCO_2) (Jin et al., 2021). Direct measurement methods like balloons or aircraft have challenges obtaining global-scale data. Currently, the main monitoring approach uses spectrometers to record spectra in CO_2 absorption bands, followed by inversion algorithms to derive XCO_2. The two primary monitoring methods are ground-based monitoring stations and satellite remote sensing.

The Total Carbon Column Observing Network (TCCON) provides ground-based monitoring of atmospheric carbon dioxide through a global network of high-precision Fourier transform spectrometers (Wunch et al., 2011, 2015). However, TCCON sites are sparsely distributed and cannot be deployed in regions with unfavorable geography or harsh climate. Consequently, the network lacks the extensive spatial coverage required for comprehensive global carbon monitoring and carbon cycle analysis. Nevertheless, the ultra-high spectral resolution of TCCON spectrometers enables highly accurate retrievals of XCO_2. Under clear sky conditions, TCCON precision can reach 0.1% (<0.4 ppm). Under relatively clear conditions with minimal clouds and aerosols, precision remains within 0.25% (<1 ppm) (Messerschmidt et al., 2011). Due to such high precision and accuracy, TCCON data are invaluable for validating satellite-based XCO_2 products (Cogan et al., 2012; Wunch et al., 2017; Liang et al., 2017) and comparing them to carbon cycle models. However, the spatial limitations of the network underscore the need for satellite remote sensing to provide regular global measurements of atmospheric carbon dioxide.

High-spectral-resolution greenhouse gas monitoring satellites employ spectrometers on orbit to measure solar radiation spectra after interaction with the Earth's atmosphere and ground surface (Meng et al., 2022). Unlike ground monitoring, satellite remote sensing is not limited spatially or temporally, offering potential for high-resolution dynamic global and regional concentration monitoring. Consequently, satellite remote sensing has become vital for future greenhouse gas monitoring worldwide. Notable ongoing passive CO_2 observation missions include China's TanSat (Liu et al., 2018), Japan's GOSAT (2009) and GOSAT-2 (2018) (Hamazaki et al., 2005; Kuze et al., 2009; Imasu et al., 2023), and the United States' OCO-2 (2014) and OCO-3 (2018) (Crisp et al., 2017; Eldering et al., 2019). Upcoming missions are France's MicroCarb by CNES (Cansot et al., 2023), ESA's CO_2M (Sierk et al., 2021) and GOSAT-GW (Matsunaga and Tanimoto, 2022). The next-generation greenhouse gases monitoring satellites mainly address the challenge of improving the spatial and temporal resolutions of observations. However, single satellites still have resolution, coverage, and meteorological limitations for regional emission monitoring. Enhancing satellite sensor performance alone cannot produce datasets sufficient for monitoring carbon sources and sinks. Improving the accuracy and efficiency of satellite data inversion is also crucial. Integrating data from multiple satellites into a coordinated system is necessary to fully capture dynamic changes in regional carbon sources and sinks. Developing new high-precision, high-throughput inversion methods to efficiently derive accurate greenhouse gas concentration distributions from satellite data is a key challenge needing attention.

The mainstream inversion algorithms (O'Dell et al., 2012; Crisp et al., 2012; Yoshida et al., 2013) for retrieving greenhouse gas concentrations from high-spectral-resolution satellite measurements are based on nonlinear Bayesian optimization theory (Rodgers, 2000) and full physical models. In essence, these algorithms operate by iteratively adjusting estimated gas concentration profiles and other atmospheric-surface parameters in a radiative forward model to minimize the mismatch between simulated and observed spectra. More specifically, the inversion process starts with an initial atmospheric state guess, including trace gas concentration profiles as functions of pressure/altitude. Radiative transfer equations are then solved to simulate the top-of-atmosphere radiance spectrum observed by the satellite for this atmospheric state. The simulated spectrum is compared to the actual observed spectrum, calculating the difference, covariance and “cost function". The input gas profiles and model parameters are iteratively adjusted to reduce the cost function over multiple rounds of radiative transfer simulations. Once simulated spectra closely match observations, the model state is output as the retrieved concentration profile. However, executing these complex optimizations requires computationally expensive interpolation of high-spectral-resolution gas absorption reference data in each iteration. Running the radiative forward model repeatedly for every adjusted atmospheric state also leads to slow overall inversion. Consequently, optimization-based retrievals struggle to match increasing satellite observation volumes and throughput needs. This inherent inefficiency has become a major obstacle to operational greenhouse gas monitoring using current and planned high-resolution spectrometers. While rigorous, standard nonlinear optimization retrievals lack the speed and scalability required for high-precision satellite-based greenhouse gas mapping. Overcoming this bottleneck necessitates new inversion approaches that can ingest high-resolution spectral data and retrieve concentrations with both accuracy and computational efficiency.

In recent years, machine learning has demonstrated exceptional performance across various research fields, with the discovery of potential nonlinear relationships between data being one of its fundamental and crucial applications. Regarding the important applications of carbon dioxide (CO_2) retrieving, Carvalho et al. (2010) attempted to retrieve the vertical CO_2 profiles using spectral data from SCIAMACHY’s 6 channels (1000-1700 nm). The overall precision and bias of the retrieved results were estimated to be approximately 1.0% and less than 3.0%, respectively. Gribanov et al. (2010) developed a two-hidden-layer multilayer perceptron (MLP) model to retrieve CO_2 vertical concentrations for the GOSAT instrument mode, achieving an inversion accuracy better than 1 ppm for CO_2 column-averaged values and better than 4 ppm for surface CO_2 concentrations for the test samples. Zhao et al. (2022) proposed a two-step machine learning model, utilizing simulated training samples, to predict XCO_2 from GOSAT observations in Australia from 2010 to 2016. David et al. (2021) and Bréon et al. (2022) attempted to establish correlations between XCO_2 in the European Centre for Medium-Range Weather Forecasts’ CAMS (Copernicus Atmosphere Monitoring Service) database and OCO-2 satellite monitoring spectra using multilayer perceptron artificial neural network models. However, their recent research (Bacour et al., 2023) indicates that when the test dataset extends beyond the time range covered by the training dataset, the predicted results show a slight bias, approximately 2.5 ppm per year. Practical deployment of machine learning techniques for remote sensing demands additional research into the generalization performance of models on new observational data distributions beyond those encountered during training.

In the present paper, a proof-of-concept study demonstrates a novel machine learning strategy to accurately and efficiently retrieve atmospheric XCO_2 value from OCO-2 satellite spectral measurements. The model rapidly retrieves XCO_2 directly from OCO-2 spectral data, eliminating the need for repetitive radiative transfer simulations required by traditional nonlinear optimization retrieval algorithms. Additionally, the model enables prediction of future XCO_2 values. The method was validated by comparing the retrieved XCO_2 against OCO-2 satellite version-10r products and ground-based TCCON measurements, confirming the accuracy of our efficient spectral inversion approach. This provides an effective solution for rapid inversion of large-scale, high-spectral-resolution remote sensing data from multiple sources in the future.

https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

2 The machine learning based XCO_2 retrieval model

2.1 Targeted area and data screening

This proof-of-concept study aims to develop and validate an accurate and efficient machine learning-based XCO_2 retrieval model applied to the long OCO-2 time series for the East Asian region. Currently, similar global XCO_2 retrieval models rely on computationally intensive physical models. Our goal is to demonstrate a more efficient data-driven approach using MLP neural networks.

Before developing the machine learning based fast retrieval model, we implemented several preprocessing steps on the OCO-2 observational dataset (OCO-2 Science Team et al., 2020a) for the target East Asian area spanning between 20°N-45°N and 110°E-145°E, as shown in Fig. 1. Specifically, we filtered the data both spatially and temporally to focus only on observations within this geographic region and time period of interest (2016-2021). Additionally, we filtered the data to only include “Nadir" mode observations marked as “Good" based on the quality flag indicator (“xco2_quality_flag” = 0 in OCO-2 Lite v10r files (OCO-2 Science Team et al., 2020b)), as these represent the highest quality OCO-2 measurements.

Several TCCON ground stations located in this region (e.g. Hefei, Saga, Tsukuba, Xianghe, Anmyeondo and Rikubetsu), as shown in Fig. 1, provide valuable ground-truth XCO_2 data for validating the MLP model predictions. If the model can accurately reproduce the TCCON observations from corresponding OCO-2 measurements, it suggests the model has learned meaningful relationships between the satellite data and underlying CO_2 concentrations.

Furthermore, successful demonstration of accurate XCO_2 retrieval over East Asia is a first step toward expanding this approach globally. The model could be retrained or supplemented with additional regional data to extend coverage. By combining reliable regional MLP models, global XCO_2 maps could be retrieved. This “jigsaw puzzle" strategy would further validate the feasibility of global-scale machine learning-based XCO_2 retrievals from satellite observations.

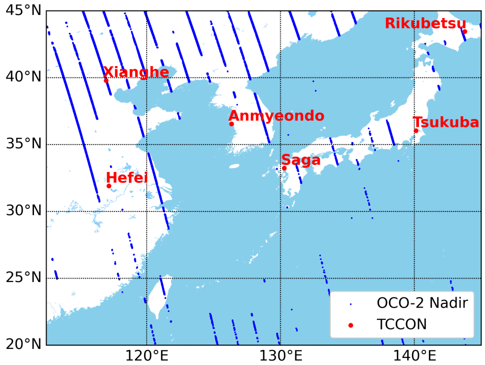

Figure 1. The target area for the East Asia region, distribution of observation points (from OCO-2 L2std v10r files) of OCO-2 Nadir mode in January 2016, and the distribution of TCCON sites in this area. The map was plotted by Python-Basemap 1.3.4 version.

2.2 The artificial neural network architecture

This study introduces a multilayer perceptron (MLP) neural network model for estimating XCO_2 from OCO-2 satellite observations. Inspired by David et al. (2021) and Bréon et al. (2022), the "MLP-XCO_2" model input layer is designed based on the measurement principles of OCO-2 and atmospheric radiative transfer effects on the observed spectra, the artificial neural networks architect is shown in Fig. 2. Specifically, the MLP model input layer consists of spectral information, prior atmospheric data, and geographical observation information as summarized in Table 1 and explained below.

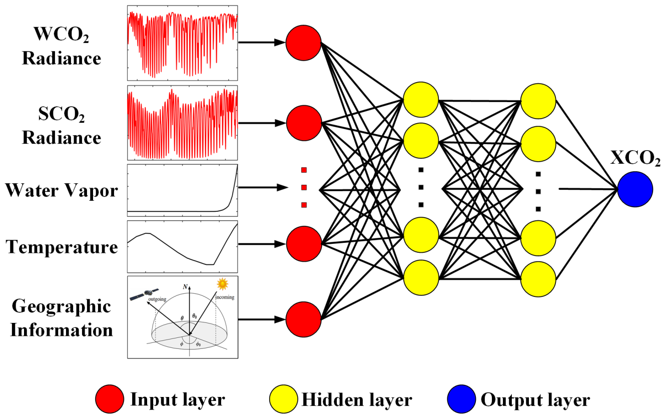

Figure 2. Schematic diagram of the MLP model.

Table 1. Detailed lists of the input layer of the MLP-XCO_2 model

Input elements	Variables	Number
Spectral information	WCO_2	525
	SCO_2	755
Prior data	Water vapor profile	20
	Temperature profile	20
	Surface pressure	1
Geographical information	Solar zenith	1
	Solar azimuth	1
	Satellite zenith	1
	Satellite azimuth	1
	Sun-Earth distance	1
	Relative velocity	1
Total		1327

Table 2. Wavelength spacing of the input spectra

Band	Spectral range [μm]	Spectral points [μm]
WCO_2	1.5990-1.6151	λ_1 = 1.5990, $\lambda_{i+1} = \lambda_i + 10^{-4}(6.10 - 3.60\lambda_i)$, i = 1-524
SCO_2	2.0478-2.0779	λ_1 = 2.0478, $\lambda_{i+1} = \lambda_i + 10^{-4}(7.58 - 3.48\lambda_i)$, i = 1-754

Spectral Information: The OCO-2 satellite instrument measures high-resolution spectra in three spectral bands centered around 0.76, 1.6, and 2.0 μm, referred to as the O_2-A, weak CO_2 (WCO_2), and strong CO_2 (SCO_2) bands, respectively (OCO-2 Science Team et al., 2019a). However, only the WCO_2 and SCO_2 bands are used as inputs for current XCO_2 retrievals. The O_2-A band is excluded as it lacks significant information needed to directly estimate XCO_2, based on radiative transfer principles. Instead, the O_2-A band is primarily used in OCO-2's operational full-physics algorithm for rapid cloud and aerosol screening prior to CO_2 retrieval (O'Dell et al., 2012), saving substantial computational costs. Each OCO-2 spectral band is sampled by 1024 detector pixels. However, over time some detectors have degraded or become unstable in the space environment, resulting in pixels being flagged as “bad samples" in quality filters (Marchetti et al., 2019). To maximize high-quality training data, additional preprocessing is performed on the WCO_2 and SCO_2 bands. Specifically, the beginning and ending spectral ranges corresponding to the most degraded detectors are removed. The remaining good quality spectra are re-sampled into 525 and 755 wavelength points for the WCO_2 and SCO_2 bands, respectively (spectral points in wavelength are detailed in Table 2). To enhance the CO_2 absorption line information, each input spectrum is normalized by dividing the mean radiance within a nearby spectrally transparent window lacking absorption features (1.6056-1.6059 μm using 10 points for WCO_2; 2.0602-2.0607 μm using 15 points for SCO_2).

Prior Data: The traditional inversion algorithm utilized in the OCO-2 satellite retrieves atmospheric temperature and water vapor profiles by optimizing a single parameter for each. However, accurate retrievals of XCO_2 require complete water vapor profile information to compute the layer weighting functions. Furthermore, variations in temperature between atmospheric layers directly impact the spectral absorption of greenhouse gases like carbon dioxide and water vapor, altering the propagation of solar radiation signals through the atmosphere. This ultimately affects the intensity of radiation signals received by the satellite sensor. Therefore, this study incorporates prior temperature and water vapor data (from OCO-2 MET v10r files (OCO-2 Science Team et al., 2019b)) for all defined atmospheric layers as inputs into the deep neural network model to improve representation of the true atmospheric state. Additionally, surface pressure is another input. Although the key retrieval information for surface pressure comes from the O_2-A band, machine learning models based on simulated data essentially predict XCO_2 by fitting the “correct solutions." As long as any input surface pressure is provided, the forward calculation model can simulate the corresponding correct spectrum. Whether the O_2-A band is required to provide the necessary surface pressure information is irrelevant. Therefore, in the input data of our model, O_2-A information was not added to increase the complexity and training difficulty of the neural network.

Geographical Information: The model is designed to accept four key observation geometry angles that are determined by the relative positions of the Sun, satellite, and ground observation point. These include the satellite zenith angle, satellite

https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

azimuth angle, solar zenith angle, and solar azimuth angle. The solar zenith angle features prominently as a cosine term in the radiative transfer equation that defines the atmospheric radiative processes. The other angles are provided in radians. Additionally, time-dependent satellite measurements including the Earth-Sun distance and the velocity of the satellite relative to the Earth's surface are input into the model. The Earth-Sun distance has a direct scaling effect on the upper limit of the solar spectral intensity distribution. The relative velocity impacts the spectral mapping between the OCO-2 spectrometer grating points and wavelengths. Both factors directly influence the intensity distribution of the measured high-resolution radiance spectra.

3 Satellite product data based machine learning model

In this section, we first developed the MLP-XCO_2 model using the OCO-2 v10r product dataset. The primary goal was to optimize the hyperparameters of the MLP-XCO_2 network. On one hand, we aimed to confirm whether the “slow bias", as shown in Bacour et al. (2023) is a universal issue across machine learning models with similar architectures. On the other hand, by fixing the hyperparameters of the MLP-XCO_2 network structure, we sought to develop a comparable model using simulated data in later sections. In theory, MLP models using identical hyperparameters should possess the same fitting and generalization abilities. By first presenting results from a model trained solely on satellite product data, we can demonstrate the limitations of these satellite data-based models. This then motivates the development of new machine learning strategies to overcome these limitations, as discussed in later sections.

Following the target areas and data screening methods discussed previously, we collected observational data from the OCO-2 v10r L1B database and XCO_2 results from the L2Lite database. Specifically, we obtained data from March, June, September and December of 2016-2018 to serve as the training set. Data from the same corresponding months in 2019-2020 were reserved as the test set to evaluate the predictive capability of the MLP-XCO_2 model on future cases. Robust future prediction is essential for the model to satisfy requirements for large-scale, high-precision greenhouse gas distribution retrieval. In total, the training set contained 86613 samples from 2016-2018, the 2019 test set has 24204 samples, and the 2020 test set contains 32292 samples.

To balance model complexity and performance, the MLP-XCO_2 architecture (Fig. 2) comprises five hidden layers, with 1000, 500, 300, 100 and 20 nodes, respectively. All hidden layers use ReLU activation functions. The output layer contains a single node to predict XCO_2 values, with a linear activation function.

Figure 3 presents results for the MLP-XCO_2 model on two different test sets. The first subplot shows predicted XCO_2 values on the 2016-2018 training data. The model achieves high accuracy on this in-sample data, with Root Mean Square Error (RMSE) below 1 ppm. This demonstrates its strong interpolation capability within the training time range. The second subplot depicts out-of-sample test results on 5% of the training data that was excluded from model fitting. Performance remains highly accurate on these held-out points, further validating the model's robustness for XCO_2 prediction within the 2016-2018 period.

Figure 4 evaluates model generalization to the 2019 and 2020 test sets, which are outside the training time range. Here, we observed a noticeable positive bias in the predictions. Furthermore, this offset increases from 2019 to 2020. The 2.5-3 ppm

https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

growth in bias aligns with the observed rise in global average XCO_2 of approximately 2.5-3 ppm/year over this period. This 180 discrepancy indicates the MLP-XCO_2 model fails to fully capture the underlying upward trend in atmospheric CO_2. While prediction is excellent within the training period, the model does not extrapolate well to future years experiencing CO_2 growth. This highlights limitations of models trained solely on historical satellite data, motivating the development of new techniques to incorporate external information about temporal CO_2 dynamics.

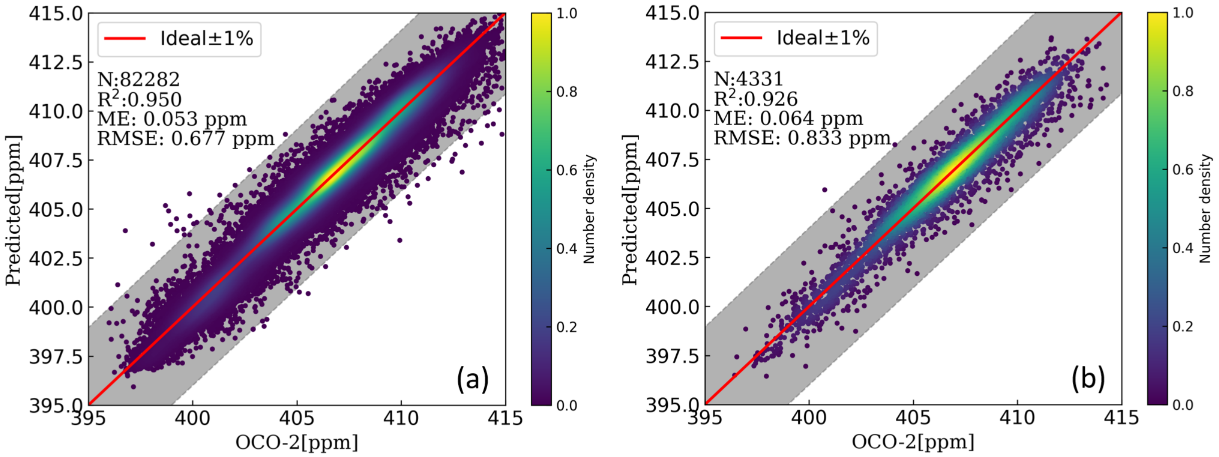

Figure 3. Comparison of XCO_2 results predicted by the MLP-XCO_2 model versus results retrieved by OCO-2 v10r product from 2016-2018. Panel (a) is for the 95% training data, while (b) is for the 5% test data (not involved in training). The solid red lines in the figure correspond to perfect agreement, where shadow areas around the solid red lines represent ±1% of XCO_2 deviations.

4 Simulation data based machine learning model

185 In the previous section, the MLP-XCO_2 model showed excellent interpolation within the training data range but exhibited bias when predicting outside this period. To eliminate this bias, we propose using an accurate forward model to simulate representative training data that covers future atmospheric conditions. If we can pre-generate atmospheric profiles that capture possible future states, and simulate their corresponding spectral radiance using a accurate forward model, the MLP-XCO_2 model can pre-learn future satellite observations. This could prevent the incremental annual bias and enable accurate XCO_2 190 prediction. The effectiveness of this approach depends on the forward model accuracy and representativeness of the simulated atmospheres (Zhao et al., 2022).

It is therefore critical to select an appropriate radiative transfer forward model with proven reliability in simulating spectral radiance under varying atmospheric conditions. The model must precisely capture the relationship between trace gas concentrations, meteorological states, and resulting spectral signatures. With accurate simulations, the machine learning model can 195 generalize robustly to future atmospheric scenarios. The representative training data should span the expected range of atmospheric variability in XCO_2 and interfering species like water vapor. Broad sampling of the state space is key for the model

https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

Atmospheric Measurement Techniques Discussions

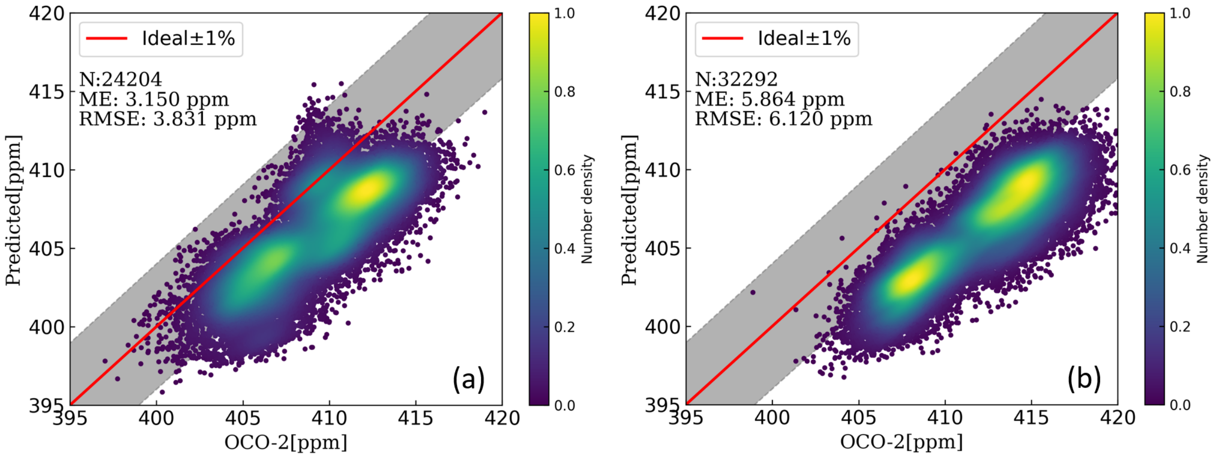

Figure 4. Comparison of XCO_2 results predicted by the MLP-XCO_2 model versus results retrieved by OCO-2 v10r product from 2019-2020. Panel (a) is for 2019, while (b) is for 2020. The solid red lines in the figure correspond to perfect agreement, where shadow areas around the solid red lines represent ±1% of XCO_2 deviations.

to learn a robust mapping to XCO_2 across multiple atmospheric regimes. The following sections describe our approach for accurate spectral radiative transfer simulations and possible (realistic) atmospheric profiles generations.

4.1 Forward model

In this study, we developed a forward radiative transfer calculation model using the ReFRACtor (Reusable Framework for Retrieval of Atmospheric Composition) software (McDuffie et al., 2018). ReFRACtor is an extensible framework for multi-instrument atmospheric radiative transfer and retrieval, originally derived from the operational OCO-2 retrieval program. Although ReFRACtor contains both radiative transfer and retrieval capabilities, we only utilized the radiative transfer component. Specifically, we configured ReFRACtor to simulate top-of-atmosphere radiance spectra that would be observed by OCO-2. These simulated observations were then used to generate a large training dataset for our machine learning model, MLP-XCO_2.

The OCO-2 satellite primarily observes the radiative spectra in the short-wave infrared (SWIR) band. Over the range of SWIR, the impact of thermal emission can be ignored when simulating the spectra (Crisp et al., 2021). These spectra are detected by the OCO-2 satellite detectors after downwelling absorption, surface reflection, and upwelling re-absorption in the atmosphere. To simulate OCO-2's observed spectra in the weak CO_2 band (WCO_2) around 1.6 µm and strong CO_2 band (SCO_2) around 2.06 µm, the ReFRACtor model numerically solves the Eq. (1) of the radiative transfer equation (RTE) (Modest and Mazumder, 2021):

$$\mu \frac{dI(\tau,\mu,\phi)}{d\tau} = -I(\tau,\mu,\phi) + J(\tau,\mu,\phi) \tag{1}$$

where I_η is the observed spectra, μ is the cosine of the observation zenith angle (e.g., $\mu = \cos\theta$), τ is the vertical optical depth which can be column-integrated from the molecular absorption coefficients and optical path, ϕ is the azimuthal angle relative to the observation point for the satellite and the sun, and J represents the scattering components and inhomogeneous source term, describing both single scattering and multiple scattering contributions. The term J in RTE can be expressed as Eq. (2):

$$J(\tau,\mu,\phi) = \frac{\omega}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} P(\tau,\mu,\phi;\mu',\phi')\,I(\tau,\mu',\phi')\,d\mu' d\phi' + \frac{\omega}{4\pi}P(\tau,\mu,\phi;\mu',\phi')\,I_0 \exp(-\tau/\mu_0) \quad (2)$$

where ω is the single scattering albedo, P is the scattering phase function, μ' and ϕ' are the cosine and azimuth angle of the incident direction angle in each direction, μ_0 is the cosine of the solar zenith, and I_0 is the solar intensity in the top of atmosphere.

The ReFRACtor model uses a hybrid model to solve RTE. Specifically, the radiative transfer software LIDORT (Spurr, 2008) is applied for the scalar and Jacobian computation. Concurrently, the two-order scattering model (Natraj and Spurr, 2007) is utilized for the additional radiance correction. Within this framework, the ReFRACtor model comprehensively considers five types of scatter particles for each sounding: two types of clouds, two types of tropospheric aerosols, and one type of stratospheric aerosol. The single scattering optical properties for each cloud and aerosol particle, including cross-section, single scattering albedo, and scattering phase matrix, have been pre-computed and tabulated for the forward calculations. Furthermore, the model determines surface reflectance as a quadratic spectral albedo for each band which is derived from the bidirectional reflectance distribution function (BRDF).

An essential step for developing the forward calculation model is referencing the pre-computed look-up table of H_2O and CO_2 to obtain the required spectral absorption coefficients. In this study, the ABSCO v5.1 database (Absorption Coefficient Table (Payne et al., 2020)) was applied for this purpose. Additionally, we identified and corrected an overestimation of the solar continuum in ReFRACtor compared to the OCO-2 Level 2 algorithm (Crisp et al., 2021). Without this correction, there would have been approximately 3 % overestimation in the 1.6 μm band and 6.5 % in the 2.06 μm band. By reducing the solar continuum, our forward model aligned better with the OCO-2 spectral measurements. These configurations of the absorption coefficients and solar continuum were essential to accurately simulate OCO-2 spectra for generating training data across a variety of observing conditions.

To assess the performance of the forward model, we selected four distinct global locations in the year of 2017. The goal was to replicate the OCO-2 observed spectra for both the WCO_2 1.6 μm absorption band and the SCO_2 2.06 μm absorption band at the four locations. By accessing the OCO-2 L2std database, we acquired atmospheric conditions and pertinent geographical data (including spectral albedo, surface pressure, and observation angles) specific to these chosen locations. The outcomes of our simulations for these four locations are visually depicted in Fig. 5 and Fig. 6, respectively for the two bands, with accompanying residual plots displayed in the lower panels. It is worth noting that the simulated results exhibit a high level of agreement with the observed OCO-2 spectra. Impressively, the relative error remains under 1%, underlining the robustness of the established forward model. The remarkable agreements between the observed and simulated spectra indicates the excellent performance of the forward radiative transfer model. This performance is particularly evident in accurately replicating the satellite observations from OCO-2. As a result, this forward model serves as a reliable tool for the development of machine learning models trained using simulated spectral data.

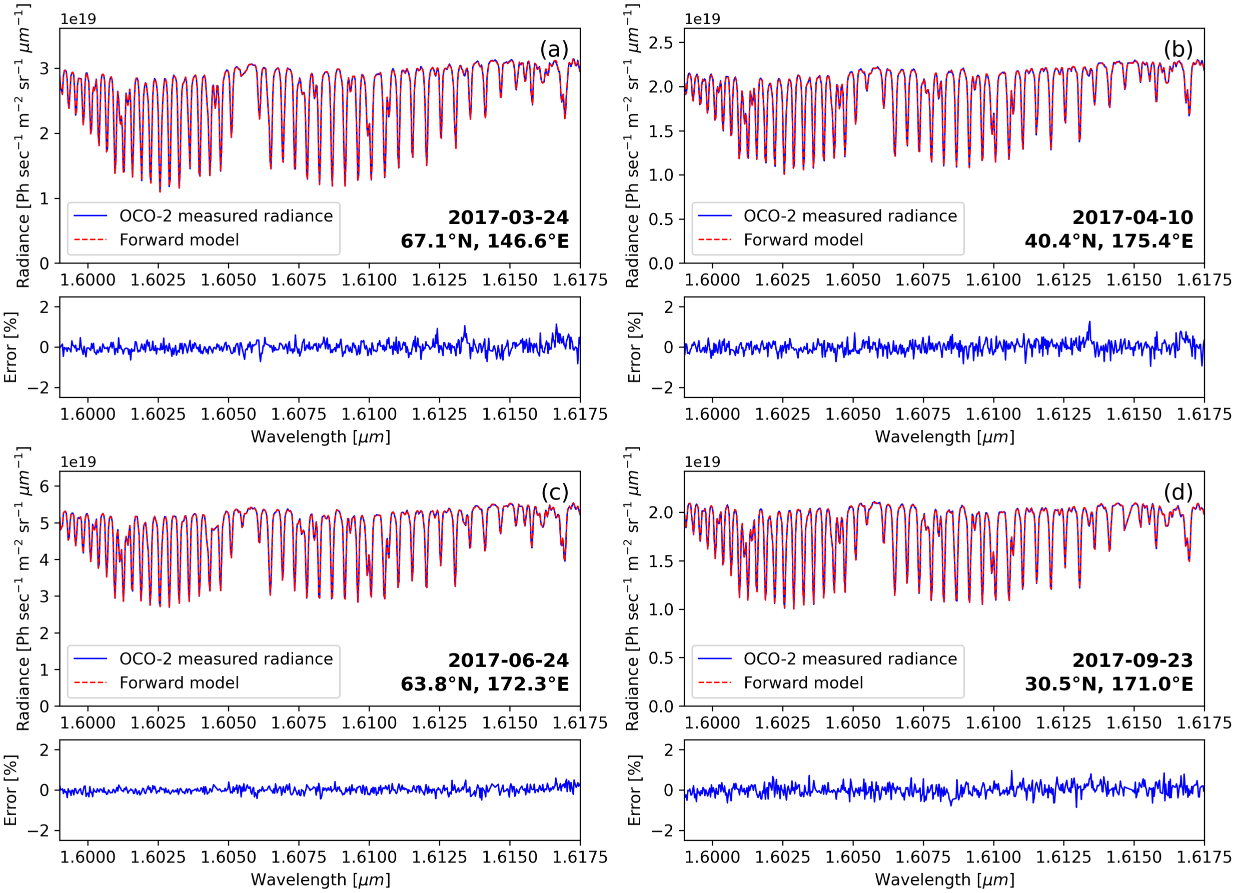

Figure 5. Comparisons of the OCO-2 observed spectra with the simulated ones from the proposed forward calculation model in WCO_2 band. The lower panel shows the relative error between the spectrum observed by the OCO-2 satellite and that simulated by the forward calculation model. Subplots (a)-(d) correspond to test samples from four different regions.

4.2 Training data generation

To optimize the training of the MLP-XCO_2 model, it is essential that the input training vectors cover a wide range of realistic variations. Although the idea of randomizing all input parameters to enhance diversity might appear attractive, simulating satellite spectra involves managing a multitude of interdependent variables. In addition to the CO_2 vertical profile, factors such as surface pressure, temperature profile, water vapor, aerosols, and observation geometry must be accurately represented. Randomizing all of these parameters would require an impractical amount of data and could result in combinations that have no real-world relevance. For example, the four viewing angles determined by the sun, observation point, and the OCO-2 satellite have fixed combinations during the satellite's regular operation. Therefore, randomly selecting angle combinations lacks practical significance. To ensure that the training data covers valid variations, we conducted an analysis of historical

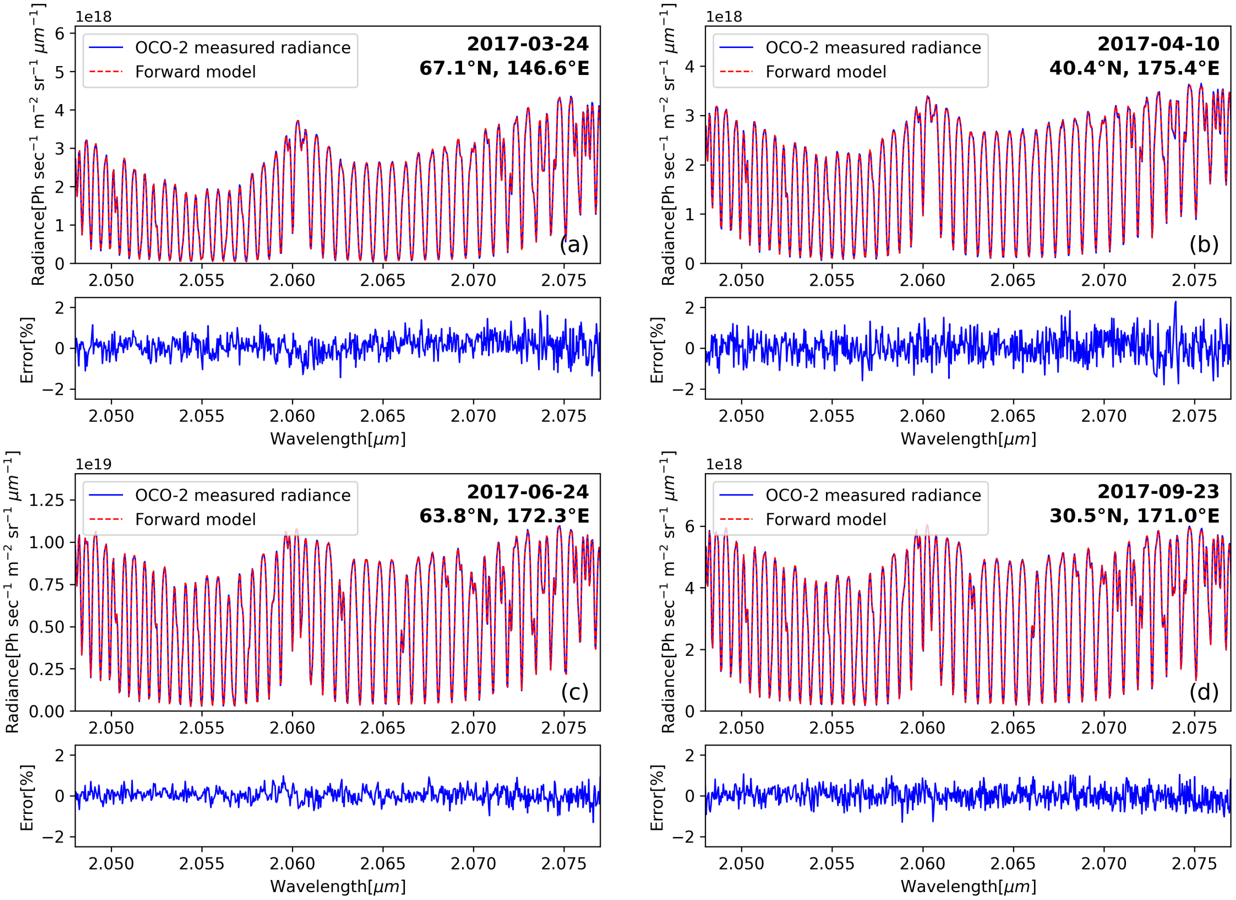

Figure 6. Comparisons of the OCO-2 observed spectra with the simulated ones from the proposed forward calculation model in SCO_2 band. The lower panel shows the relative error between the spectrum observed by the OCO-2 satellite and that simulated by the forward calculation model. Subplots (a)-(d) correspond to test samples from four different regions.

OCO-2 retrievals. This analysis revealed consistent seasonal patterns and year-to-year trends in most parameters. This supports the idea of selecting representative samples from statistical distributions rather than relying on complete randomization. By carefully considering the relationships between parameters and the realities of satellite observations, we can create a reasonably sized training dataset that effectively captures the range of expected predictions.

Among all the input parameters, the generation of the vertical CO_2 profile holds special significance. This dataset essentially defines the MLP-XCO_2 model's range of applicability. In the context of the ReFRACtor model, which serves as the basis for our forward model, the atmospheric CO_2 profile is divided into 20 sub-layers based on pressure. To gain insights into the atmospheric CO_2 concentration in each of these sub-layers, we conducted a statistical analysis using OCO-2 retrieved CO_2 profiles in the East Asia region from 2016 to 2018. The resulting box plots, as depicted in Fig. 7, reveal a gradual increase

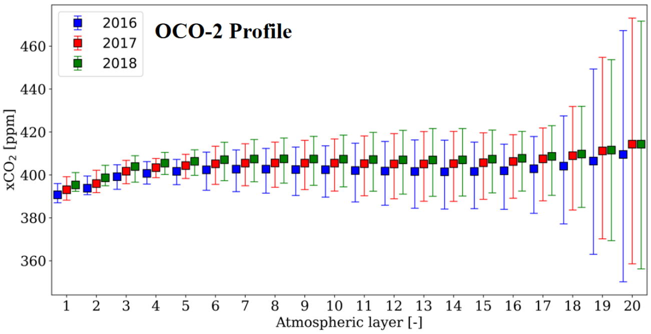

Figure 7. Box plots of the vertical distribution of CO_2 profiles (from OCO-2 L2std files) retrieved by the OCO-2 satellite over East Asia in Nadir mode from 2016 to 2018. The horizontal-axis represents the atmospheric layers from layer 1 (top of atmosphere) to layer 20 (near-surface). The upper and lower bounds of each box show the maximum and minimum CO_2 concentrations recorded within that layer for each year.

in CO_2 concentration uncertainty as we move from the upper atmosphere to the surface. This presents a particular challenge when dealing with standardized atmospheric CO_2 profiles,

The generation of the vertical CO_2 profile is especially critical among all input parameters. This dataset theoretically determines the generalization domain of the MLP-XCO_2 model. In the forward model based on the ReFRACtor model, atmospheric CO_2 profile is segmented into 20 sub-layers by pressure. By statistically analyzing the OCO-2 retrieved CO_2 profiles in target East Asia area from 2016-2018, the box plots for atmospheric CO_2 concentration in each sub-layer are shown in Fig. 7. From the upper atmosphere down to the ground surface, the uncertainty of CO_2 concentrations gradually increases. This challenges the ability for the standardization of atmospheric CO_2 profiles, particularly closer to the Earth's surface. Fortunately, a consistent year-on-year rise in CO_2 concentrations in each sub-layer has been observed over time. Consequently, in our research, we have proposed a method for generating subsequent CO_2 atmospheric profiles. We incrementally increase the CO_2 concentration by 2.5 ppm annually, starting from the 2016 OCO-2 retrieved CO_2 vertical profile. This approach ensures that we encompass a range of plausible atmospheric CO_2 distributions with realistic shapes, enabling the generation of simulated spectra for the designated training years.

In addition to the CO_2 profile, Fig. 8 illustrates the year-to-year trends of various observed parameters essential for the forward calculation model in the East Asian region. These parameters, although they display seasonal variations, consistently exhibit annually cyclic patterns. Given that the OCO-2 satellite conducts global observations in cycles of approximately half a month (15-16 days), this study employed observation parameters and priori data for atmospheric profiles, except for CO_2, from the year 2016 as a reference. These reference data were repetitively utilized for generating simulations in subsequent years. Regarding the quadratic spectral albedo, the constant term in the training data samples is uniformly set to 1 (to be normalized before being processed by the neural network). The slope and the quadratic coefficient are stochastically sampled within the range of values corresponding to the retrieval results based on the OCO-2 L2 products.

https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

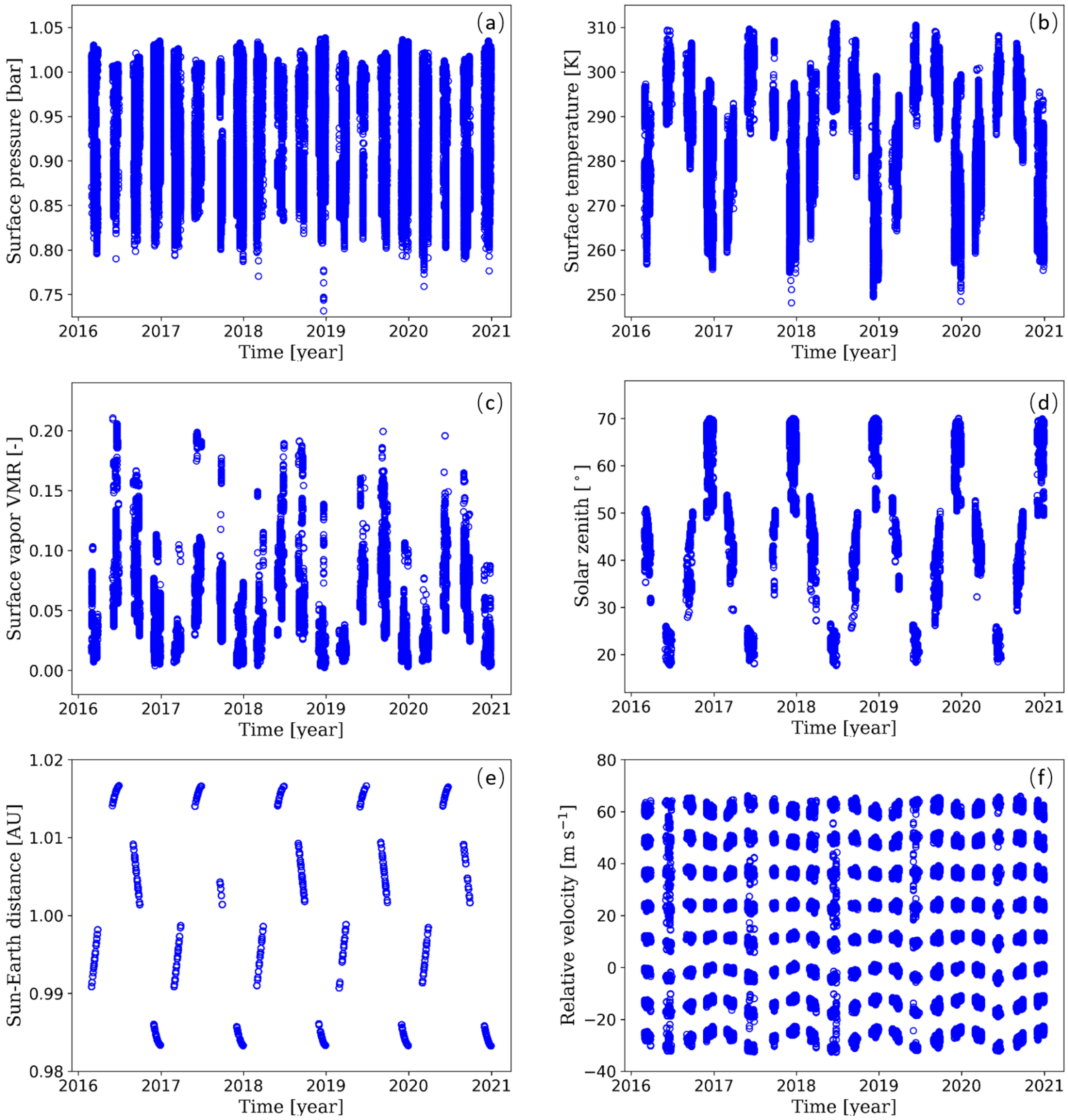

Figure 8. Scatter plots of atmospheric parameters required for forward calculation models (excluding CO_2 profiles) from 2016 to 2020, sourced from the OCO-2 L2 product. Panel (a) is the surface pressure, (b) is the surface temperature, (c) is the near-surface water vapor concentration, (d) is the solar zenith, (e) is the Sun-Earth distance, and (f) is the Earth-satellite relative velocity.

In summary, based on 10000 uniformly sampled observation data from the OCO-2 satellite throughout the year 2016, we created a total of 50000 sets of new data and the forward model was used to generate the corresponding simulated spectra for each set. These simulated samples serve as the foundational dataset for training the new MLP-XCO_2 machine learning model. It's important to note that this new model relies solely on the data recorded by the OCO-2 satellite in 2016 as its reference. Through data augmentation techniques, we have enabled the model to accurately and efficiently perform XCO_2 retrieval for the "future" years from 2017 to 2021.

5 Results and discussions

5.1 Comparison with the OCO-2 satellite product data

To evaluate the retrieval capability of the MLP-XCO_2 model trained on simulated data, the neural network architecture and hyperparameters were intentionally kept identical to the previous model trained on actual OCO-2 satellite product data. Keeping these factors constant isolates the training data source as the only major difference between the models. This enables a direct, apples-to-apples comparison of how the training data affects model performance.

Figure 9 (a) shows the retrieval results on 5% of the training data that was excluded from model training. Setting aside this test subset is a standard technique for evaluating model performance on new examples. The accurate predictions of the MLP-XCO_2 model on the test data suggest the model has learned generalizable patterns not overfit to the training data. Figure 9 (b) shows the comparison of the retrieval results of the MLP-XCO_2 model on real OCO-2 satellite spectral observations in 2016. Figure 10 displays XCO_2 predictions from 2017 to 2020 using test data consistent with Fig. 3 and Fig. 4. As the simulated training data was generated based on 2016 OCO-2 measurements, testing on 2017-2020 data evaluates the model's ability to make predictions beyond the time frame of the training data.

The scatter plots demonstrate the MLP-XCO_2 model trained on simulated data can accurately and stably predict the annual XCO_2 growth trend, maintaining an Mean Error (ME) within 0.1 ppm and RMSE around 2 ppm (0.5%), respectively. Compared to models trained relying solely on historical satellite product data, the key advantage is the ability to make reasonable forecasts of future atmospheric XCO_2 levels. By generating possible realistic future prior information for the atmospheric conditions and using an accurate forward model to simulate the corresponding spectra, the approach avoids inherent biases when extrapolating beyond the distribution of the training data. Rather than simply extending trends, the model is constrained by fundamental physical relationships to interpolate within realistic bounds. This transforms the prediction task into a well-posed interpolation problem versus an unconstrained extrapolation. The simulated data provides a physical regularization that makes the model's outputs to be scientifically sound. By training on synthetic data spanning potential future scenarios, the model learns robust representations not tightly coupled to specifics of the training data time period. This enables high-fidelity inversion and prediction of XCO_2 even for future time periods beyond available measurements.

5.2 Comparison with the TCCON data

A comparison of the retrieved results from the OCO-2 satellite showed that the RMSE of our developed MLP-XCO_2 model was around 2 ppm. In other words, our results could be worse or better than OCO-2 satellite, requiring further comparison with

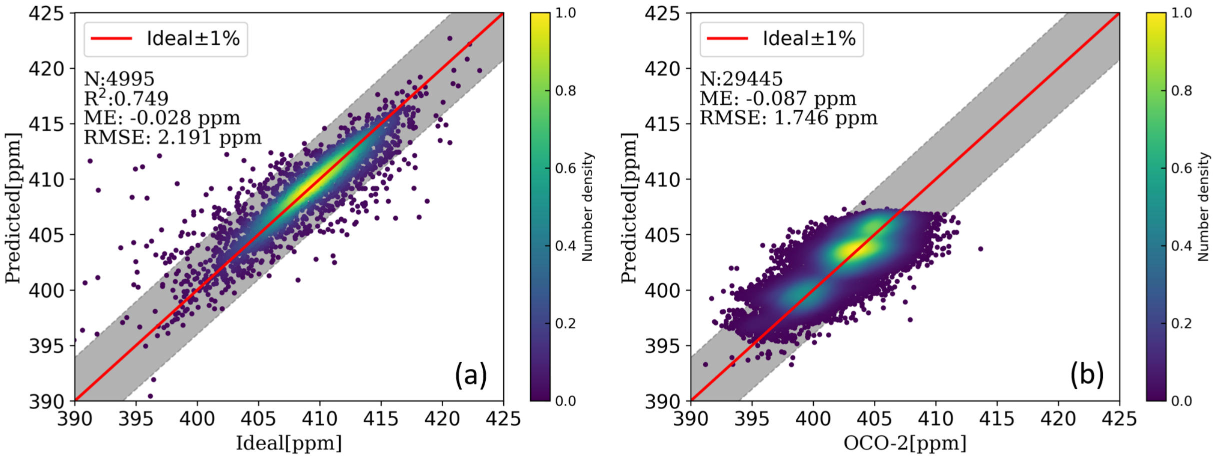

Figure 9. Panel (a) is the comparison of XCO_2 results predicted by the MLP-XCO_2 model from 5% test data (not involved in training). Panel (b) is the comparison of XCO_2 results predicted by the MLP-XCO_2 model versus results retrieved by OCO-2 v10r product from 2016 test data.

ground-based measurements. To further validate the accuracy of the MLP-XCO_2 model, we compared the XCO_2 retrievals from the OCO-2 v10r Nadir mode products, the MLP-XCO_2 model outputs, and ground-based measurements from 5 TCCON sites within the study region (Fig. 1). As summarized in Table 3, spatiotemporal screening was applied to the TCCON and OCO-2 data to obtain comparable observations. The 5 TCCON sites included were: Tsukuba (Morino et al., 2022b), Saga (Shiomi et al., 2022), Hefei (Liu et al., 2022), Xianghe (Zhou et al., 2022) and Rikubetsu (Morino et al., 2022a). The Anmyeondo site was excluded from this analysis as the XCO_2 data was not updated in the TCCON GGG2020 database, and was only available until early 2018 in the GGG2014 database.

Figure 11-1 presents time series comparisons of XCO_2 retrievals from the different TCCON sites, MLP-XCO_2 model, and OCO-2 Nadir observations. Figure 11-2 displays the box plots of the differences between the MLP-XCO_2 model results, OCO-2 products, and TCCON site data. The plots at each of the five TCCON sites demonstrate the simulated data-trained MLP-XCO_2 model accurately predicts XCO_2 from the OCO-2 spectra. The model successfully captures seasonal variations and the long-term XCO_2 growth trend over the 4-year study period. The reliable performance over time and across multiple TCCON sites further validates the model has learned generalizable representations of carbon cycle processes rather than overfitting to specifics of the simulated training data. By using realistic future simulations for training, the model provides robust and unbiased XCO_2 retrievals across a range of atmospheric conditions.

5.3 Retrieval efficiency

In this study, the ReFRACtor forward model required 12.16 seconds per simulation case (two absorption bands) using an AMD Ryzen-7 5800X computer. The OCO-2 retrieval based on Bayesian optimization typically needs over three iterations to con-

https://doi.org/10.5194/amt-2023-224
Preprint. Discussion started: 20 November 2023

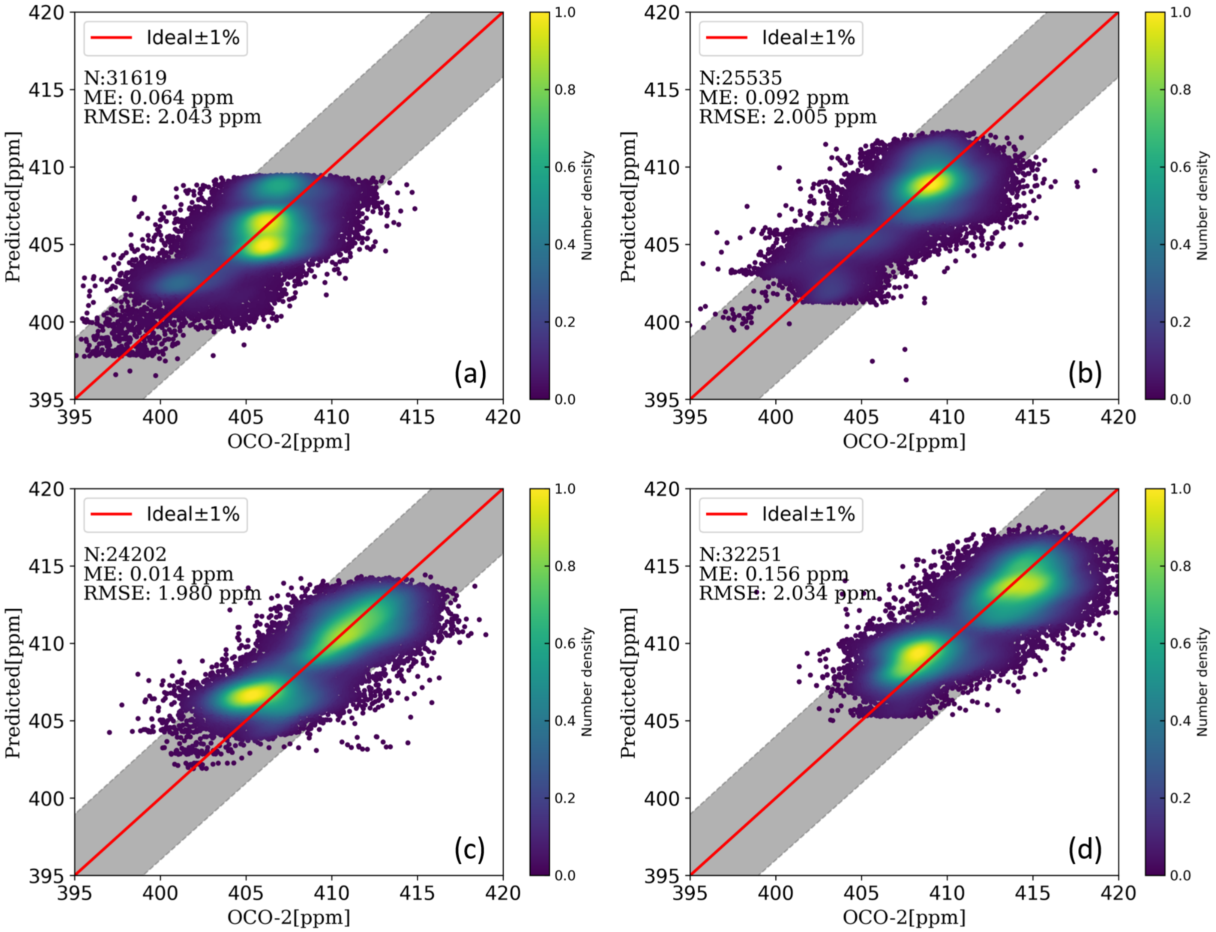

Figure 10. Comparison of XCO_2 results predicted by the MLP-XCO_2 model versus results retrieved by OCO-2 v10r product from 2017-2020. Panel (a), (b), (c), and (d) display the predictions of the MLP-XCO_2 model from 2017 to 2020, respectively.

Table 3. Spatio-temporal screening conditions for TCCON sites and OCO-2 satellite Nadir mode observations

TCCON site	Local time	Observed location	Test sample	Reference
Tsukuba	$12:48-12:58$	$36.05°N \pm 0.2°, 140.12°E \pm 0.2°$	520	Morino et al. (2022b)
Saga	$13:30-13:40$	$33.24°N \pm 0.2°, 130.29°E \pm 0.2°$	60	Shiomi et al. (2022)
Hefei	$13:20-13:30$	$31.90°N \pm 0.2°, 117.17°E \pm 0.2°$	763	Liu et al. (2022)
Xianghe	$13:15-13:25$	$39.80°N \pm 0.2°, 116.96°E \pm 0.2°$	1082	Zhou et al. (2022)
Rikubetsu	$13:20-13:30$	$43.46°N \pm 0.2°, 143.77°E \pm 0.2°$	254	Morino et al. (2022a)

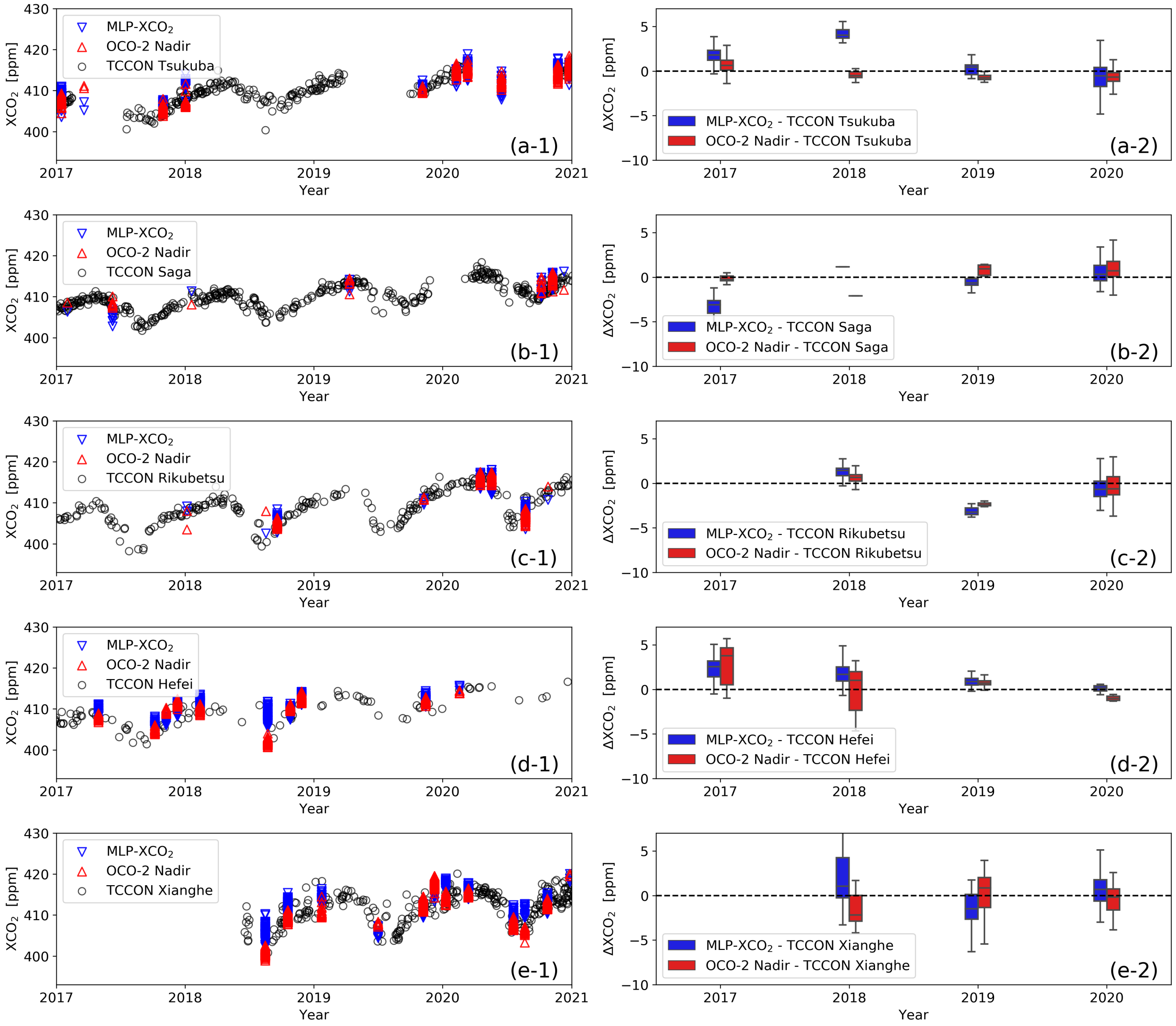

Figure 11. Comparisons of XCO_2 results from 2017 to 2020 across five TCCON sites. Panel (a-1)-(e-1) show the time series comparisons of XCO_2 retrievals from the different TCCON sites, MLP-XCO_2 model, and OCO-2 L2Lite Nadir observations for the Tsukuba, Saga, Hefei, Xianghe and Rikubetsu site, respectively, with data screening conditions as defined in Table 3. Panel (a-2)-(e-2) present the boxplots depicting the differences (ΔXCO_2) between the MLP-XCO_2 model and OCO-2 products in comparison to the TCCON results for each year. The boxes showing the middle half of the data, from the 25% to the 75% percentiles. The median (50%) is represented by the line within each box. The whiskers encompass the central 90% of the data, extending from the 5% to the 95% percentiles.

verge, indicating at least 36.48 seconds per retrieval. In contrast, the MLP-XCO_2 model demonstrated remarkable efficiency on the same hardware. It required just 0.71 seconds total to retrieve XCO_2 from 1846 OCO-2 test spectra across first three TCCON sites (Tsukuba, Saga and Hefei site), averaging 0.38 milliseconds per sample. This rapid inversion drastically reduces processing times compared to traditional methods. While machine learning models need significant upfront time for training data generation and hyperparameter tuning, the prediction is extremely fast once deployed. This enables near real-time processing ideal for operational satellite data streams. Furthermore, the precision and efficiency of neural networks makes them well-suited to meet future demands of high-resolution global greenhouse gas monitoring, enabling millisecond-scale XCO_2 retrievals suitable for large-scale satellite analysis.

6 Conclusions

This proof-of-concept study aims to use the efficient regression inversion capability of machine learning method to develop machine learning models based on simulated atmospheric radiative transfer data for efficient inversion of satellite observed spectra to retrieve XCO_2. This helps overcome the low efficiency in traditional optimization-based iterative algorithms for XCO_2 retrievals. In the presented study, XCO_2 inversion models using both satellite product based and simulation based data were developed, trained and tested. Long time series inversion and prediction of OCO-2 observations over East Asia were also performed using the developed models. The results were compared with OCO-2 and TCCON retrievals, showing the simulation data based machine learning models can effectively eliminate lagging biases while achieving millisecond-level (<1 ms) inversion efficiency, high accuracy (around 2 ppm), and long-term prediction stability.

Code availability. The ReFRACtor model and its OCO retrieval implementation can be accessed from the Github ReFRACtor repository (https://github.com/ReFRACtor, last accessed in August 2023). To obtain the training data generator and MLP-XCO_2 model, please send requests to Tao Ren (tao.ren@sjtu.edu.cn).

Data availability. The OCO-2 products (including OCO-2 L1B, Met, L2std and L2Lite files) are available from Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/datasets/, last access: March 2023). The TCCON site products are available from TCCON DATA ACHIEVE (https://tccondata.org/, last access: March 2023).

Financial support. This research has been supported by the National Natural Science Foundation of China (Grants No. 52276077)

Author contributions. FX and TR designed the study. FX developed the forward model and the machine learning code, carried out the tests and result analysis under the supervision of TR. FX and TR prepared the manuscript. All authors reviewed the manuscript.

Competing interests. The corresponding author has declared that none of the authors has any competing interests.

365 *Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We gratefully acknowledge the TCCON Data Archive hosted by CaltechDATA at https://tccondata.org for providing the TCCON data for our study. Our thanks also go to the OCO-2 Science Team for the OCO-2 project. We appreciate our colleagues for their feedback and our affiliated institutions for their unwavering support.

References

Bacour, C., Bréon, F.-M., and Chevallier, F.: On the challenge posed by the estimation of XCO_2 from OCO-2 observations in near-real time based on artificial neural network, IWGGMS-19, 2023.

Bréon, F.-M., David, L., Chatelanaz, P., and Chevallier, F.: On the potential of a neural-network-based approach for estimating XCO_2 from OCO-2 measurements, Atmospheric Measurement Techniques, 15, 5219–5234, https://doi.org/10.5194/amt-15-5219-2022, 2022.

Cansot, E., Pistre, L., Castelnau, M., Landiech, P., Georges, L., Gaeremynck, Y., and Bernard, P.: MicroCarb instrument, overview and first results, in: International Conference on Space Optics — ICSO 2022, edited by Minoglou, K., Karafolas, N., and Cugny, B., vol. 12777, p. 1277734, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2690330, 2023.

Carvalho, A. R., Ramos, F. M., and Carvalho, J. C.: Retrieval of carbon dioxide vertical concentration profiles from satellite data using artificial neural networks, Trends in Computational and Applied Mathematics, 11, 205–216, https://doi.org/10.5540/tema.2010.011.03.0205, 2010.

Cogan, A., Boesch, H., Parker, R., Feng, L., Palmer, P., Blavier, J.-F., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., et al.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): comparison with ground-based TCCON observations and GEOS-Chem model calculations, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD018087, 2012.

Crisp, D., Fisher, B., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L., Castano, R., Connor, B., Deutscher, N., et al.: The ACOS CO_2 retrieval algorithm–part II: global XCO_2 data characterization, Atmospheric Measurement Techniques, 5, 687–707, https://doi.org/10.5194/amt-5-687-2012, 2012.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., et al.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmospheric Measurement Techniques, 10, 59–81, https://doi.org/10.5194/amt-10-59-2017, 2017.

Crisp, D., O'Dell, C., Eldering, A., Fisher, B., et al.: Orbiting carbon observatory (OCO) - 2 level 2 full physics algorithm theoretical basis document Version 3.0 – Rev 1, https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO_L2_ATBD.pdf, 2021.

David, L., Bréon, F.-M., and Chevallier, F.: XCO_2 estimates from the OCO-2 measurements using a neural network approach, Atmospheric Measurement Techniques, 14, 117–132, https://doi.org/10.5194/amt-14-117-2021, 2021.

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data, Atmospheric Measurement Techniques, 12, 2341–2370, https://doi.org/10.5194/amt-12-2341-2019, 2019.

Gribanov, K., Imasu, R., and Zakharov, V.: Neural networks for CO_2 profile retrieval from the data of GOSAT/TANSO-FTS, Atmospheric and Oceanic Optics, 23, 42–47, https://doi.org/10.1134/S1024856010010094, 2010.

Hamazaki, T., Kaneko, Y., Kuze, A., and Kondo, K.: Fourier transform spectrometer for greenhouse gases observing satellite (GOSAT), in: Enabling sensor and platform technologies for spaceborne remote sensing, vol. 5659, pp. 73–80, SPIE, https://doi.org/10.1117/12.581198, 2005.

Imasu, R., Matsunaga, T., Nakajima, M., Yoshida, Y., Shiomi, K., Morino, I., Saitoh, N., Niwa, Y., Someya, Y., Oishi, Y., et al.: Greenhouse gases Observing SATellite 2 (GOSAT-2): mission overview, Progress in Earth and Planetary Science, 10, 33, https://doi.org/10.1186/s40645-023-00562-2, 2023.

Jin, Z., Tian, X., Han, R., Fu, Y., Li, X., Mao, H., Chen, C., and GAO, J.: Tan-Tracker global daily NEE and ocean carbon fluxes for 2015-2019 (TT2021 dataset), https://doi.org/10.11888/Meteoro.tpdc.271317, 2021.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Applied optics, 48, 6716–6733, https://doi.org/10.1364/AO.48.006716, 2009.

Liang, A., Gong, W., Han, G., and Xiang, C.: Comparison of satellite-observed XCO$_2$ from GOSAT, OCO-2, and ground-based TCCON, Remote Sensing, 9, 1033, https://doi.org/10.3390/rs9101033, 2017.

Liu, C., Wang, W., Sun, Y., and Shan, C.: TCCON data from Hefei, China, Release GGG2020R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A., https://doi.org/10.14291/tccon.ggg2020.hefei01.R0, 2022.

Liu, Y., Wang, J., Yao, L., Chen, X., Cai, Z., Yang, D., Yin, Z., Gu, S., Tian, L., Lu, N., et al.: The TanSat mission: preliminary global observations, Science Bulletin, 63, 1200–1207, https://doi.org/10.1016/j.scib.2018.08.004, 2018.

Marchetti, Y., Rosenberg, R., and Crisp, D.: Classification of anomalous pixels in the focal plane arrays of Orbiting Carbon Observatory-2 and-3 via machine learning, Remote Sensing, 11, 2901, https://doi.org/10.3390/rs11242901, 2019.

Matsunaga, T. and Tanimoto, H.: Greenhouse gas observation by TANSO-3 onboard GOSAT-GW, in: Sensors, Systems, and Next-Generation Satellites XXVI, vol. 12264, pp. 86–90, SPIE, https://doi.org/10.1117/12.2639221, 2022.

McDuffie, J., Bowman, K. W., Hobbs, J., Natraj, V., Val, S., Sarkissian, E., and Thill, M. D.: ReFRACtor: Reusable Software Framework for Retrieval of Satellite Atmospheric Composition, in: AGU Fall Meeting Abstracts, vol. 2018, pp. A11F–2282, 2018.

Meng, G., Wen, Y., Zhang, M., Gu, Y., Xiong, W., Wang, Z., and Niu, S.: The status and development proposal of carbon sources and sinks monitoring satellite system, Carbon Neutrality, 1, 32, 2022.

Messerschmidt, J., Geibel, M., Blumenstock, T., Chen, H., Deutscher, N., Engel, A., Feist, D. G., Gerbig, C., Gisi, M., Hase, F., et al.: Calibration of TCCON column-averaged CO$_2$: the first aircraft campaign over European TCCON sites, Atmospheric Chemistry and Physics, 11, 10 765–10 777, https://doi.org/10.5194/acp-11-10765-2011, 2011.

Modest, M. F. and Mazumder, S.: Radiative heat transfer, Academic press, 2021.

Morino, I., Ohyama, H., Hori, A., and Ikegami, H.: TCCON data from Rikubetsu, Hokkaido, Japan, Release GGG2020R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A., https://doi.org/10.14291/tccon.ggg2020.rikubetsu01.R0, 2022a.

Morino, I., Ohyama, H., Hori, A., and Ikegami, H.: TCCON data from Tsukuba, Ibaraki, Japan, 125HR, Release GGG2020R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A., https://doi.org/10.14291/tccon.ggg2020.tsukuba02.R0, 2022b.

Natraj, V. and Spurr, R. J.: A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering–absorbing media, Journal of Quantitative Spectroscopy and Radiative Transfer, 107, 263–293, https://doi.org/10.1016/j.jqsrt.2007.02.011, 2007.

OCO-2 Science Team, Gunson, M., and Eldering, A.: OCO-2 Level 1B calibrated, geolocated science spectra, Retrospective Processing V10r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/6O3GEUK7U2JG, accessed: [August 2023], 2019a.

OCO-2 Science Team, Gunson, M., and Eldering, A.: OCO-2 Level 2 meteorological parameters interpolated from global assimilation model for each sounding, Retrospective Processing V10r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/OJZZW0LIGSDH, accessed: [August 2023], 2019b.

OCO-2 Science Team, Gunson, M., and Eldering, A.: OCO-2 Level 2 bias-corrected XCO_2 and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing V10r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/6SBROTA57TFH, accessed: [August 2023], 2020a.

OCO-2 Science Team, Gunson, M., and Eldering, A.: OCO-2 Level 2 geolocated XCO_2 retrievals results, physical model, Retrospective Processing V10r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/E4E140XDMPO2, accessed: [August 2023], 2020b.

O'Dell, C., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., et al.: The ACOS CO_2 retrieval algorithm–Part 1: Description and validation against synthetic observations, Atmospheric Measurement Techniques, 5, 99–121, https://doi.org/10.5194/amt-5-99-2012, 2012.

Payne, V. H., Drouin, B. J., Oyafuso, F., Kuai, L., Fisher, B. M., Sung, K., Nemchick, D., Crawford, T. J., Smyth, M., Crisp, D., et al.: Absorption coefficient (ABSCO) tables for the Orbiting Carbon Observatories: version 5.1, Journal of Quantitative Spectroscopy and Radiative Transfer, 255, 107 217, https://doi.org/10.1016/j.jqsrt.2020.107217, 2020.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World scientific, 2000.

Shiomi, K., Kawakami, S., Ohyama, H., Arai, K., Okumura, H., Ikegami, H., and Usami, M.: TCCON data from Saga, Japan, Release GGG2020R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A., https://doi.org/10.14291/tccon.ggg2020.saga01.R0, 2022.

Sierk, B., Fernandez, V., Bézy, J.-L., Meijer, Y., Durand, Y., Courrèges-Lacoste, G. B., Pachot, C., Löscher, A., Nett, H., Minoglou, K., et al.: The Copernicus CO2M mission for monitoring anthropogenic carbon dioxide emissions from space, in: International Conference on Space Optics—ICSO 2020, vol. 11852, pp. 1563–1580, SPIE, https://doi.org/10.1117/12.2599613, 2021.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, Light scattering reviews 3: Light scattering and reflection, pp. 229–275, https://doi.org/10.1007/978-3-540-48546-9_7, 2008.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 369(1943), 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., and Wennberg, P. O.: The Total Carbon Column Observing Network's GGG2014 Data Version, Pasadena, California, https://doi.org/10.14291/TCCON.GGG2014.DOCUMENTATION.R0/1221662, 2015.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., et al.: Comparisons of the orbiting carbon observatory-2 (OCO-2) XCO_2 measurements with TCCON, Atmospheric Measurement Techniques, 10, 2209–2238, https://doi.org/10.5194/amt-10-2209-2017, 2017.

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G., Wunch, D., et al.: Improvement of the retrieval algorithm for GOSAT SWIR XCO_2 and XCH_4 and their validation using TCCON data, Atmospheric Measurement Techniques, 6, 1533–1547, https://doi.org/10.5194/amt-6-1533-2013, 2013.

Zehr, S.: The sociology of global climate change, Wiley Interdisciplinary Reviews: Climate Change, 6, 129–150, https://doi.org/10.1002/wcc.328, 2015.

Zhao, Z., Xie, F., Ren, T., and Zhao, C.: Atmospheric CO_2 retrieval from satellite spectral measurements by a two-step machine learning approach, Journal of Quantitative Spectroscopy and Radiative Transfer, 278, 108 006, https://doi.org/10.1016/j.jqsrt.2021.108006, 2022.

Zhou, M., Wang, P., Nan, W., Yang, Y., Kumps, N., Hermans, C., and De Mazière, M.: TCCON data from Xianghe, https://doi.org/10.14291/tccon.ggg2020.xianghe01.R0, 2022.