A Review of Anthropogenic Ground-Level Carbon Emissions Based on Satellite Data

Kai Hu, Qi Zhang, Shen Gong, Fuying Zhang, Liguo Weng, Shanshan Jiang, and Min Xia, *Member, IEEE*

***Abstract*—The severity of the global warming issue emphasizes the critical importance of utilizing carbon satellite data to estimate ground-level carbon dioxide emissions. However, existing reviews have not kept pace with the latest research developments. Therefore, this article provides an overview of relevant work in the global carbon emissions field to address this knowledge gap. Through visual analysis using Citespace software, the article outlines two methods for quantifying carbon dioxide: 1) ground-level observations; and 2) satellite remote sensing. Despite the unique advantages of ground-level observations, satellite remote sensing is crucial for its extensive spatial coverage and long-term continuity in understanding carbon cycling, drawing significant attention. In addition, the article integrates the application of machine learning in the carbon emissions field, dividing it into two parts: Direct estimation based on ground emission inventory data and estimation of ground-level carbon emissions based on carbon satellite data. This innovative approach combines satellite observational data with ground data to accurately estimate the current ground-level carbon emissions with robust spatial distribution characteristics.**

***Index Terms*—Carbon satellite, machine learning, retrieval algorithm, XCO_2.**

I. Introduction

CARBON dioxide (CO_2) is a major greenhouse gas, contributing to 70% of the greenhouse effect. Human emissions have led to an increase in atmospheric CO_2 concentrations, resulting in a global temperature rise of about 1° C [1]. The IPCC urges significant emission reduction measures to limit global warming and achieve carbon neutrality.

The study of atmospheric CO_2 by the IPCC is categorized into upper and lower layers, with boundary height being a critical factor. CO_2 in the upper layer is a result of historical emissions over thousands of years, representing the residual effects that carbon sinks have not had the opportunity to absorb [2]. Meanwhile, CO_2 in the lower layer primarily originates from current human activities, influencing future global temperatures [3], [4].

Manuscript received 20 July 2023; revised 5 December 2023 and 12 January 2024; accepted 14 January 2024. Date of publication 18 January 2024; date of current version 19 April 2024. This work was supported by National Natural Science Foundation of China under Grant 42275156. *(Corresponding authors: Qi Zhang; Shen Gong; Fuying Zhang; Shanshan Jiang.)*

Kai Hu, Liguo Weng, and Min Xia are with the Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing 210044, China.

Qi Zhang, Shen Gong, Fuying Zhang, and Shanshan Jiang are with the Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: 20211249164@nuist.edu.cn; 20211249032@nuist.edu.cn; 001624@nuist.edu.cn; jss@nuist.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3355549

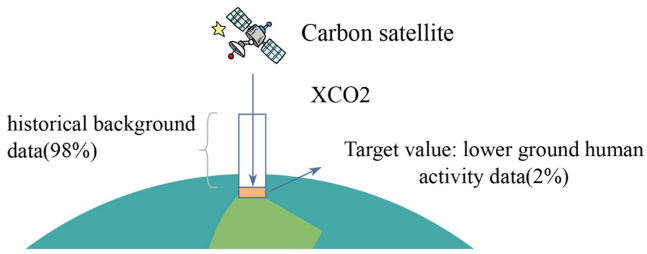

Fig. 1. Carbon satellite observation of carbon flux column.

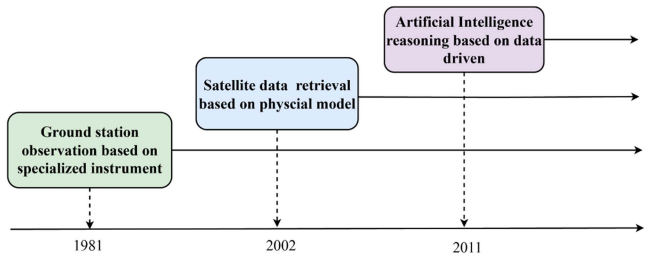

Fig. 2. Development timeline of carbon emission observation technology.

The extraction of anthropogenic emission data from total CO_2 concentration data presents several challenges, stemming from the following factors.

1) As depicted in Fig. 1, XCO_2 concentrations observed by carbon satellites predominantly capture background data in the upper troposphere, with only a 2% representation of the current anthropogenic CO_2 flux from the lower ground.
2) CO_2 is a long-lived gas with historical implications, persisting in the environment for centuries.
3) The total amount of CO_2 on the upper surface is significantly influenced by seasonal changes and atmospheric motion.

CO_2 monitoring encompasses ground observation, satellite data retrieval, and data-driven methodologies. The current emphasis is on data-driven estimation utilizing satellite data, but progress is impeded by a scarcity of pertinent datasets. Fig. 2 illustrates the stages of carbon emission monitoring, while Fig. 3 highlights the primary research areas. In addition, Fig. 4 provides an overview of the article's structure.

The first method focuses on ground-level CO_2 monitoring using various instruments (Y1). These instruments reflect human activities such as fossil fuel combustion and electricity

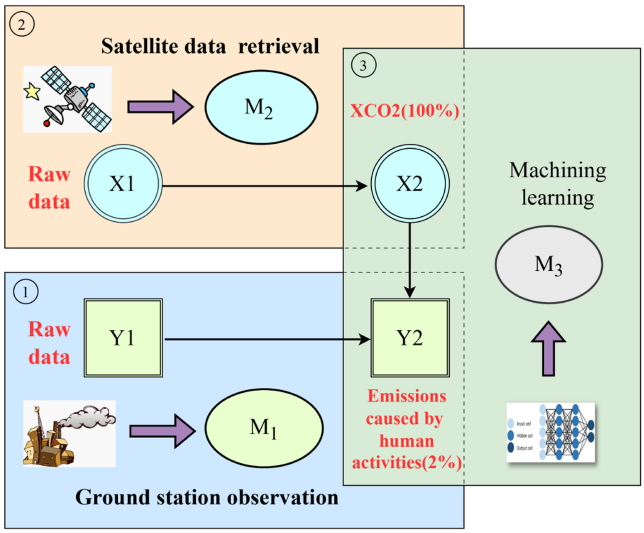

Fig. 3. Technology road map.

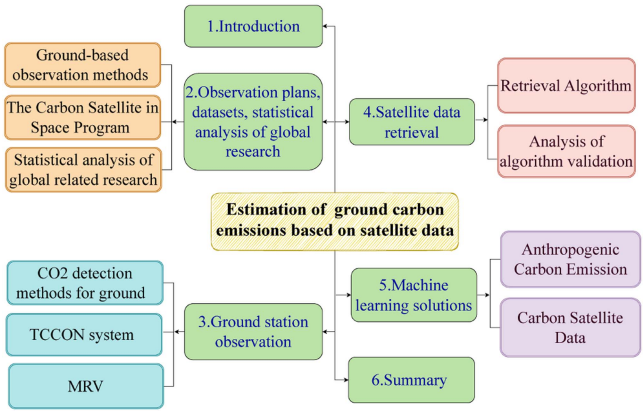

Fig. 4. Structure diagram of this article.

usage. The calculated CO_2 emissions (Y2) based on Y1 data and relevant standards are determined by the model M1. Machine learning (ML) techniques can be employed for predictive analysis. In CO_2 ground observation, Dong et al. [5] summarized macroscopic estimation methods applicable at national, provincial, and city levels, addressing challenges in regional energy variations in 2014. Ground-based observations of fixed sources were recommended for enhanced reliability. Zhang et al. [6] found on-site measurements more accurate than emission factor approaches in monitoring CO_2 emissions from fixed sources. Physical optical methods, particularly nondispersive infrared spectroscopy, were recognized internationally. Chen et al. [7], using CiteSpace software, analyzed the past decade's progress in domestic and international carbon emissions. They noted a global trend in employing mathematical models for quantitative analysis. Debone et al. [8] observed frequent use of neural networks and regression models in predictive analyses of carbon emissions, energy consumption, and economic growth.

The second method focuses on satellite data retrieval based on physical model. X_1 represents the original data of carbon satellite detection of atmospheric carbon dioxide concentration, while X_2 is XCO_2 data, which is derived after processing X_1 using retrieval algorithms. M_2 is physical model. Section III provides a detailed description of the first program. David et al. [9] comprehensively reviewed the use of satellite data for emission estimation, particularly questioning retrieval techniques for quantifying eight major atmospheric pollutants and greenhouse gases. Yue et al. [10] compared carbon dioxide concentration measurements from satellites and ground-based observations, introducing carbon dioxide detection satellites and validating results with ground truth proxy observations. Lees et al. [11] detailed the use of remote sensing to estimate ecosystem carbon flux, explaining the scope of satellite data for estimation and assessing their capabilities in ecosystem carbon flux. These studies mainly compared atmospheric remote sensing with traditional ground observations for measuring carbon dioxide. Using satellite-based monitoring with XCO_2 data allows for global-scale coverage, enabling the study of spatial distribution and trends in CO_2 flux worldwide.

The third method employs data-driven approaches to address human-induced carbon emissions in XCO_2 columns via satellite remote sensing, overcoming limitations in ground observations. Carbon assimilation techniques, integrating observational data with numerical models, have improved accuracy. For instance, Kaminski et al. [12] estimated urban fossil fuel CO_2 emissions using the CCFFDAS assimilation system. Miyazaki et al. enhanced accuracy and predictability by estimating CO_2/NOx ratios and ground-level fossil fuel CO_2 emissions with a Kalman filter [13].

While these techniques enhance spatial distribution, they rely on numerical models and prior knowledge. The advent of artificial intelligence, specifically machine learning, allows direct pattern learning from data, reducing reliance on prior knowledge and proving more suitable for carbon observation tasks. Although machine learning is applied for carbon dioxide estimation, there's limited exploration of spatial distribution. Hakkarainen et al. [2] differentiated regions with anthropogenic CO_2 emissions using clustering algorithms, deriving XCO_2 anomaly maps. However, the study did not account for the long-term nature and seasonal variations of CO_2 emissions. Yang et al. [14] used artificial neural networks to estimate ground carbon emissions in XCO_2 for China during 2010–2014, revealing potential spatial patterns. For satellite XCO_2 data estimation, auxiliary data is crucial. Combining meteorological auxiliary data, as seen in Wang et al.'s [15] approach for nighttime PM2.5, enhances model accuracy and reliability. Mustafa et al. [16] improved prediction models by incorporating NPP auxiliary data into XCO_2 data, achieving high consistency between anthropogenic emissions and satellite estimates. In recent research, Zhang et al. [17] proposed an unsupervised deep model that successfully mapped carbon satellite data to emission data, allowing timely monitoring of regional and corporate carbon emissions. Wang et al. [18] adopted a zone-based approach, combining Open-source Data Inventory for Anthropogenic

CO_2 (ODIAC) anthropogenic emission inventory and potential temperature data for remote sensing monitoring and analysis of spatiotemporal changes in China's anthropogenic carbon emissions.

This article systematically reviews research pertaining to ground-level carbon emissions resulting from human activities, utilizing the Web of Science as the primary information source. We scrutinize the efforts of previous scholars in developing algorithms for estimating lower ground carbon, encompassing both satellite and ground observations [19]. Furthermore, we delve into the interplay between machine learning and carbon emissions in both spatial and terrestrial domains. The significant contributions of this article are as follows.

1) We utilized Citespace software to analyze domestic and international research on carbon emissions, aiming to identify contributions and forefront issues. Key areas explored include carbon trading, emission source localization, and carbon emission estimation at various scales.
2) We outlined the developmental history of remote sensing carbon satellites and ground observation instruments, comparing their advantages and disadvantages. This analysis informs future instrument optimization and algorithm improvement, empowering researchers to address challenges in atmospheric CO_2 concentration monitoring.
3) We summarized XCO_2 datasets from GOSAT and OCO-2 at different time intervals, alongside selected ground-based anthropogenic emission datasets. This information serves as a foundational data source for exploring trends in atmospheric carbon dioxide concentration and the impact of human emissions.
4) This review pioneers a comprehensive exploration of ground carbon emission estimation from a machine learning perspective. Beyond enhancing accuracy and introducing direct ground estimation, our focus on data-driven differentiation provides a novel research perspective. This not only contributes to the field, but also outlines key focal points for future carbon emission estimation research.

II. Observation Schemes, Datasets, Statistical Analysis of Global Related Research

Initially, CO_2 observation programs predominantly revolve around two methodologies: 1) ground-based observation; and 2) carbon satellite observation. This article sequentially presents the ground-based observation scheme, followed by the carbon satellite observation scheme, delineating their chronological development over time.

A. Ground-Based Observation Methods and Datasets

Traditionally, ground-based sampling has been a reliable method for detecting challenging-to-measure CO_2 emissions. This approach uses sensors on vents, vehicles, and ships to measure CO_2 concentrations at ground level, investigating generation and consumption mechanisms for spatial and temporal distribution understanding [21].

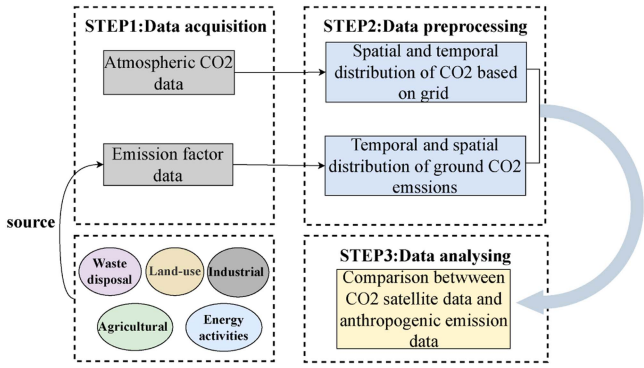

Fig. 5. Structure diagram of this article.

Various methods exist for detecting and verifying terrestrial CO_2 emissions [20]. Verification methods estimate emissions using energy activity and conversion coefficients, which, while simple, can be data-uncertain. Achieving global uniform real-time CO_2 monitoring is challenging due to limitations in ground-based observation schemes [22]. Developed countries have established costly Measuring, Reporting, and Verification (MRV) policy systems, which are demanding for developing nations. In addition, bottom-up carbon emission data may face adjustments and falsifications by emission sources and local governments [23].

1) Anthropogenic Emissions Dataset: Anthropogenic emission data is used to estimate the concentration of CO_2. It primarily includes CO_2 emissions from various sources such as agricultural activities, industrial production, land use changes, and waste disposal. Fig. 5 illustrates the factors involved in anthropogenic emission data. Six datasets related to anthropogenic CO_2 emissions were introduced. Table I is a summary table of the anthropogenic emission inventory dataset [24].

In 1944, the World Development Indicators (WDI) was established, offering comprehensive global development data in csv format spanning 1996–2020 [25]. BP Amoco's acquisition in 2000 and subsequent rebranding to BP marked a period of increased global energy consumption, rising from 342.23 EJ in 1990 to over 576 EJ in 2020, as reported in BP's statistical review of world energy [26].

The ODIAC project, launched in 2008 by JAXA, provides high-resolution (1 km $\times$ 1 km) emissions data for fossil fuel CO_2 emissions in nc file format, covering 2000–2019 [27]. In 2012, Tsinghua University developed the Multiresolution Emission Inventory for China (MEIC), offering high-resolution emission inventories for various sectors in nc files [28], [29].

Dabo Guan's team created the China Emission Accounts and Datasets System (CEADS) database in 2016, using a combined accounting approach for Scope 1 and Scope 2 emissions [30]. Fig. 6 illustrates the line graph of CO_2 emissions from fossil fuels in China from 2010 to 2019 based on the CEADS inventory.

B. Carbon Satellite in Space Program and Datasets

With technological advancements, traditional ground-based methods for observing atmospheric CO_2 concentrations have

TABLE I
ANTHROPOGENIC EMISSION INVENTORY DATASET

Name	WDI	BP	CARMA	ODIAC	MEIC	CEADS
Data formats	CSV	points	points	1° × 1°	grid/points / Surface	points
Year	1944	2000	2007	2008	2012	2016
Unit	Million tons	Million tons	ton	ton	Million tons	million tons/ 10 billion kJ
Statistics Department	Financial	Oil, gas		Point sources nonpoint sources	Power	Industrial
	Public Sector		–	Cement production	Industrial	Agriculture
	Healthcare			Gas flflaring	Civilian	Mining
	Mining Sector			International aviation	Transportation and Agriculture	Service
Producer	WDI	BP	Global Development Center United States	Center for global environment research, national institute for environment studies	MECI team	Natural Science Foundation of China, Ministry of Science and Technology
Resource	https://data cata log. worldbank. org/	https://www. bp.com	http://carma. or g/plant	http://db. cger.nies. go.jp/data set/ODIAC/	http://meic model.org /?page_id =127	https://www. ceads.net. cn/data/ province/

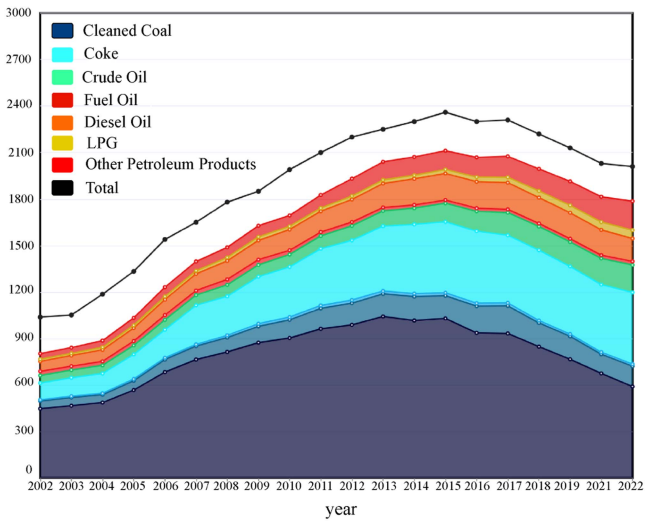

Fig. 6. Emission inventory chart of different energy sources in China [22].

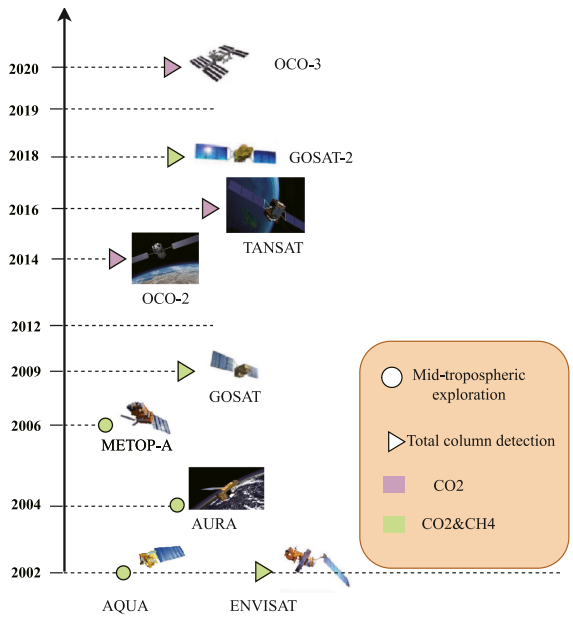

Fig. 7. Launching timeline of satellites for CO_2 detection including scheduled satellite plans.

expanded. Carbon satellites, equipped with high spatial and temporal resolution through satellite remote sensing, have become crucial for obtaining global and regional CO_2 concentration data [31]. Over nearly two decades, these satellites have revealed spatial and temporal characteristics using a top-down observation approach, offering data less susceptible to manipulation or errors by local authorities [32], [33]. Fig. 7 depicts the timeline of human-launched CO_2 detection satellites, while Table II details instruments and performance of major satellites [34].

In 2002, NASA launched the AQUA satellite, while ESA launched the ENVISAT satellite. These two satellites are equipped with the Atmospheric Infrared Sounder (AIRS) and Scanning Imaging Absorption Spectrometer (SCIAMACHY), respectively, making them among the first instruments capable of monitoring CO_2 concentrations from space. AIRS is primarily used to support medium-term numerical weather prediction [35], while SCIAMACHY is the first instrument capable of accurately measuring both CO_2 and CH4 concentrations [29].

In 2004, NASA launched the AURA satellite, which carries the Tropospheric Emission Spectrometer (TES) [36]. The TES instrument measures the infrared radiation emitted by the atmosphere, providing a global map of tropospheric ozone and its photochemical precursors with a spatial resolution of 5.3 km × 8.5 km, compared to the previous AIRS and SCIAMACHY

TABLE II
MAJOR EXISTING OR PREVIOUS SATELLITES FOR CO_2 DETECTION AND THEIR MEASUREMENT PERFORMANCE [11]

Satellite	AQUA	ENVISAT	AURA	METOP-A	GOSAT	OCO-2	TANSATA	GOSAT-2	OCO-3
Sponsoring agency	NASA	ESA	NASA	ESA/EUM ETSA	MOE/JAXA/ NIES	NASA	CAS/MOS T/CMA	MOE/JAX A/NIES	NASA
Country/ region	United States	Europe	United States	Europe	Japan	United States	China	Japan	United States
Instrument	AIRS	SCIAMAC HY	TES	IASI	TANSO-FTS/ CAI	Grating Spectrom eter	Carbon Spec	TANSO/ CAI-2	Grating Spectrom eter
Operational period	2002–	2002–2012	2004–	2006–	2009	2014–	2016–	2018–	2019–
Repeat Cycle	1 day	6 days	2 days	0.5 days	3 days	16 days	16 days	6 days	–
Spatial Resolution (km)	50×50	30×60	5.3×8.5	50×50	~ [b]10×10	1.29× 2.25	~2×2	10×10	1.6×2.2
Spectral Range (μm)	3.7~15.4	0.24~2.4	3.2~15.4	3.6~15.5	0.76~14.3	0.76~ 2.08	0.76~2.08	0.76~14.3	0.76~ 2.08
CO_2 sensitivity	Mid-troposphere	Total column including near surface	Mid-troposphere	Mid-troposphere	Total column including near surface	Total column including near surface	Total column including near surface	Total column including near surface	Total column including near surface
CO_2 uncertainty (ppm)[a]	1.5	14	–	2	4	<1	4	0.5	<1
Cost (USD)	–	300 million	–	–	373 million	465 million	–	404 million	110 million
Reference	Aumann et al.,(2003); National Research Council, (2010)	(Bovensma nn et al., (1999); National Research Council, (2010)	Beer et al. (2001)	Clerbaux et al.,(2009); National Research Council, (2010)	Kuze et al., (2009); National Research Council, (2010)	Connor et al.,(2016); Crisp et al.,(2017); National Research Council, (2010)	Liu et al., 2018b; Yang et al., (2018)	(Glumb et al.,(2014); JAXA, (2018)	Eldering et al. (2019)

[a]:CO_2 uncertainty (ppm) stands for the estimate of random errors and additional systematic errors unaccounted for or otherwise eliminated from the total error.[b]:The~ in Spatial resolution indicates approximately.

instruments, which had lower spatial resolutions. The TES instrument also provides better results for CO_2 observations.

In 2006, the European Meteorological Office launched the METOP-A satellite carrying the Infrared Atmospheric Sounding Interferometer. This instrument is capable of detecting atmospheric concentrations of CO_2, CH_4, SO_2, and NO_2, but at a coarse spatial resolution of 50 km×50 km [37].

In 2009, the Ministry of the Environment of Japan launched the GOSAT satellite, which is the world's first greenhouse gas detection satellite. It uses the Near Infrared Sensor-Fourier Transform Spectrometer (TANSO-FTS) and the Cloud and Aerosol Imager (CAI) to observe global CO_2 and CH4 concentration distribution information and quantitatively analyze greenhouse gas sources and spatial and temporal variation characteristics. The satellite has a spatial resolution of 10 km×10 km, and it can observe the same site every three days [38].

In 2009, NASA initiated the Orbiting Carbon Observatory (OCO) mission, but unfortunately, the launch failed. However, in 2014, the OCO-2 satellite was successfully launched. OCO-2 carries three high-resolution grating spectrometers that use the sunlight reflected from the Earth's surface to make accurate measurements of CO_2. The sunlight entering the spectrometer passively passes through the atmosphere twice, once from the sun to the Earth and once from the Earth's surface to the spectrometer. OCO-2 has a spatial resolution of 1.29 km×2.25 km and an observation period of 16 days [10].

In 2016, China launched TANSAT, a scientific experimental satellite designed for detecting and monitoring carbon dioxide. It carries a carbon dioxide detector as well as cloud and aerosol detectors that provide data on the seasonal concentration of carbon dioxide as climate changes [39]. The TANSAT satellite has a spatial resolution of 2 km×2 km and an observation period of 16 days.

In 2018, the Ministry of the Environment and the Aerospace Research Agency of Japan launched the GOSAT-2 satellite. This satellite is equipped with a more advanced pair of TANSO-2 and a CAI compared to its predecessor, the GOSAT satellite. The satellite collects atmospheric CO_2 concentration data with an unprecedentedly low uncertainty of only 0.5, making it an important tool for studying global carbon cycle and climate change [40].

In 2019, NASA's OCO-3 was installed on the International Space Station, which provides more intensive observations at sampling sites, especially in high-latitude regions [41]. The

planned lifetime task is three years, measuring total column CO_2 and solar-induced chlorophyll fluorescence [42].

1) CO_2 Satellite Database: XCO_2 is defined as the total number of CO_2 molecules in the column divided by the total number of dry air molecules in the column [43]

$$X_{CO_2} = \frac{\int_0^\infty N_{CO_2}(z)dz}{\int_0^\infty N_{\text{air}}(z)dz} \tag{1}$$

where N_{CO_2} is the number density (number of carbon dioxide molecules per cubic meter) associated with altitude (z), and $N_{\text{air}(z)}$ is the number of dry air densities associated with altitude.

The GOSAT and OCO-2 satellites are essential for XCO_2 measurement and widely used in research. GOSAT TANSO-FTS data is crucial for testing the ACOS retrieval algorithm, aiding XCO_2 retrieval from raw space data and calibrating OCO-2 missions. A GOSAT and TCCON XCO_2 retrieval comparison helps identify ACOS GOSAT XCO_2 product biases.

Since OCO-2's 2014 launch, researchers estimate global and regional CO_2 fluxes through top-down atmospheric retrieval methods [44]. Processed at ESMO and JPL's SDOS, OCO-2 mission data includes Level 1 A, Level 1B, and Level 2 products [45]. The more compact "lite file" product includes XCO_2 estimates and global bias correction. OCO-2 data, delivered in both "forward" and "backward" processing, offers higher scientific application quality in backward data.

Periodically updated, GOSAT and OCO-2 data are described in series [1] and [2], respectively. [1]. GOSAT Data Series

The ACOS algorithm, initially developed for OCO, was adapted for GOSAT data in 2009, processed using the Level-2 "full physical" (L2FP) retrieval algorithm. The dataset includes location, XCO_2, evaluation parameters, and recording angles in HDF5/NetCDF formats. ACOS_L2FP_V2.8 processed all GOSAT NIR soundings between 2009 and 2011 [46].

In subsequent versions, such as ACOS_L2FP_V3.3 (2010) [47], solar-induced chlorophyll fluorescence (SIF) was incorporated, improving surface pressure accuracy [48]. ACOS_L2FP_V3.5 (2014) introduced changes like replacing aerosol types and using TCCON version GGG2014 for XCO_2 bias correction [49], [50]. ACOS_L2FP_v7.3 (2016) and ACOS_L2FP_v9 (2019) extended the time record and improved algorithm features, with v9 covering April 2009 to June 2020 [51]. GOSAT v9, utilizing Bayesian optimal estimation, proves effective for studying carbon cycle phenomena over a decade or more [25], [52].

[2]. OCO-2 data series

The OCO-2 dataset, available in NetCDF formats, features a 1.75 km spatial resolution and 16-day temporal resolution, covering the entire world. It provides information on location, time, XCO_2, XCO_2 evaluation parameters, and recording angle.

In 2015, Eldering et al. [54] used OCO-2 V7 data to estimate CO_2 fluxes [53]. In 2019, OCO-2 released V9, addressing sub-footprint pointing offset and coding errors, enhancing accuracy.

The 2021 OCO-2 v10r product, using ACOS_L2FP, improves XCO_2 data quality and quantity. It reduces biases, achieving a single detection accuracy of approximately 0.8 ppm over land and 0.5 ppm over water [55].

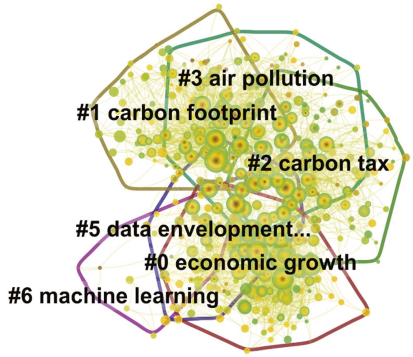

Fig. 8. Keyword clustering map.

Table III summarizes XCO_2 data products obtained from the mentioned satellites. To achieve a comprehensive CO_2 flux, retrieval algorithms are crucial for extracting the total column number of atmospheric trace gases from NIR nadir spectra. Section IV reviews various methods proposed for this purpose.

C. Statistical Analysis of Global Related Research

Citespace is a widely used citation visualization and analysis tool known for its user-friendly interface and clear visualizations. It plays a crucial role in analyzing source regions, identifying key scholars, understanding research hotspots, and tracking the evolution of literature in related fields. This tool facilitates the creation of network knowledge maps [56], [57]. In this section, we leverage Citespace software to visualize and analyze global carbon emission-related literature. Our analysis includes comparing keyword mapping and the publication volume across different countries in the field of carbon emissions. In addition, we utilize the pyecharts module in Python to extract the number of paper contributions from the Web of Science over the last two years and create a world map.

In this study, a Web of Science advanced search was conducted using the keywords "carbon emission" and "XCO_2," with a time frame of 2005 to 2022 and a time slice of 1. A total of 1353 relevant documents were selected and analyzed for their keywords, national institutions, and topics using visualization tools.

Fig. 8 reveals key clusters in carbon emissions research using data from the Web of Science. Among them, #0 delves into the connection between economic growth and increased carbon emissions due to industrial energy consumption. #1 examines the overall carbon footprint from human activities, while #2 scrutinizes the impact of carbon tax on energy use. #6 employs machine learning for regional CO_2 emission estimation. The recent emphasis on large-scale remote sensing, aided by machine learning, enhances data accuracy, supporting green initiatives, and carbon neutrality goals.

Table IV summarizes research themes identified through clustering. The earliest and enduring theme is the greenhouse effect, with its significant impact on various aspects of the environment and global affairs. This concern has garnered global attention. In recent years, carbon emissions have gained

TABLE III
XCO_2 DATASET

Dataset	Year	Satellite	Format	Features	Resource
ACOS_L2FP_ V2.8 [46]	2009	GOSAT	HDF5	v2.8 retrieval accuracy approaches the 1–2 ppm requirement for accurate monitoring of carbon fluxes on a regional scale	https://disc.gsfc.nasa.gov/data collection /ACOS_L2S_2.8.html
ACOS_L2FP_V3.3 [48]	2010	GOSAT	HDF5	Updated spectroscopy-ABSCO coefficients V4.1.1; reduced a priori value of aerosol optical depth to 0.05	https://disc.gsfc.nasa.gov/data collection /ACOS_L2S_3.3.html
ACOS_L2FP_ V3.5 [50]	2014	GOSAT	HDF5	In addition to the empirical screening of data quality, a "warning level" was implemented	https://disc.gsfc.nasa.gov/data collection /ACOS_L2S_3.5.html
ACOS_L2FP_7.3 [49]	2016	GOSAT	HDF5	V7.3 products are distributed with many atmospheric retrieval studies	https://disc.gsfc.nasa.gov/data collection /ACOS_L2S_7.3.html
OCO-2_V7 [54]	2017	OCO-2	NetCDF	The V8 data generally reduce regional scale bias	https://disc.gsfc.nasa.gov/data sets /OCO_2_L2_Lite_FP_7r
ACOS_L2FP_9r [52]	2019	GOSAT	HDF5	The measured global CO_2 estimates cover data -sparse areas where ground-based measurements are difficult to make.	https://daac.gsfc.nasa.gov/datasets /ACOS_L2_Lite_FP_9r
OCO-2_V9r [54]	2019	OCO-2	NetCDF	V9 product data addresses several bias corrections to improve accuracy and coverage.	https://daac.gsfc.nasa.gov/datasets /OCO_2_L2_Lite_FP_9r
OCO-2_V10r [55]	2021	OCO-2	NetCDF	OCO-2 level 2 bias correction XCO_2 and other fields selected from the full physical retrieval.	https://daac.gsfc.nasa.gov/datasets/OCO_2 _L2_Lite_FP_10r

TABLE IV
ANTHROPOGENIC EMISSION INVENTORY DATASET

Number	Category	Concentration Year	Representative Literature
1	#0 GHG	2005–2022	Temperature-associated increases in the global soil respiration record [61].
2	#2Remote sensing satellite data	2015–2022	An Application of Remote Sensing Data in Mapping Landscape-Level Forest Biomass for Monitoring the Effectiveness of Forest Policies in Northeastern China [62].
3	#5 High Resolution	2016–2021	The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes [63].
4	#4CarbonEmissions	2017–2022	Satellite-derived PM2.5 concentration trends over Eastern China from 1998 to 2016: Relationships to emissions and meteorological parameters [64].

TABLE V
COMPARISON OF CO_2 DETECTION METHODS FOR STATIONARY EMISSION SOURCES

Method name	Instrumentation	Costs	Operation and maintenance	Detection method	Responsiveness	Accuracy	Stability
Chemical method	Chemical absorption method	Moderate	Complex maintenance	Laboratory	Slower	relatively low	Poor
	Gas	Chromatography	Moderate	Easy	Laboratory Moderate	High	Good
Electrochemical methods		Low	Easy	Laboratory	Moderate	High	General
Physical method	Non-spectroscopic infrared spectroscopy	High	Regular maintenance	On site	Quick	High	General
	Fourier transform infrared spectroscopy	High	Regular maintenance	Laboratory	Quick	High	Good

TABLE VI
METHODS OF GROUND-BASED OBSERVATIONS

Fields	Year	Methods	Advantages
Carbon dioxide instrumentation testing	1981	Gas separation in the laboratory	Simple, convenient, low cost, and better test results
TCCON system	2004	Inverse performance of CO_2 from near-infrared (NIR) solar absorption spectra by spectroscopy	Accurate retrieval of CO_2 concentration with high precision
Carbon verification	2011	Measurement, and report verification by third-party agencies	Authoritative, fair and able to develop better measures to deal with carbon emissions

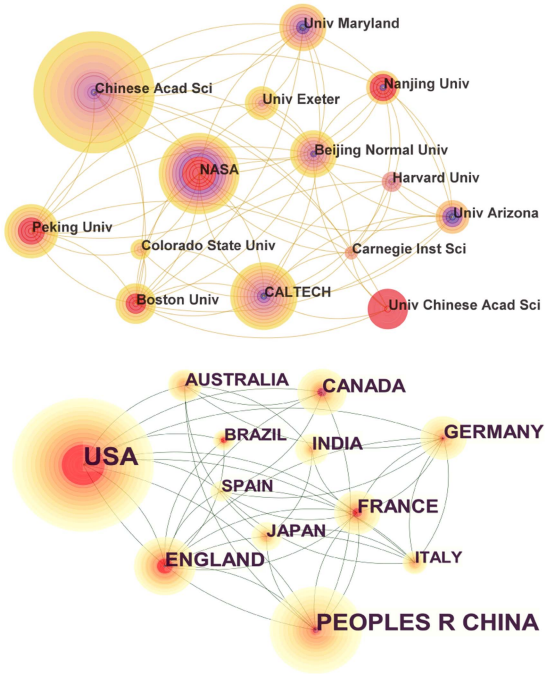

Fig. 9. Mapping of country, institutional cooperation networks.

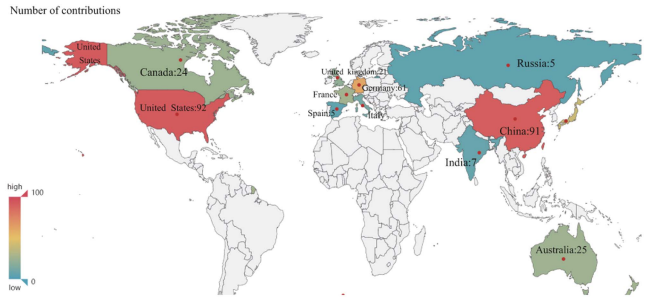

Fig. 10. Contribution of different countries in the field of carbon emissions for the two years 2020–2023.

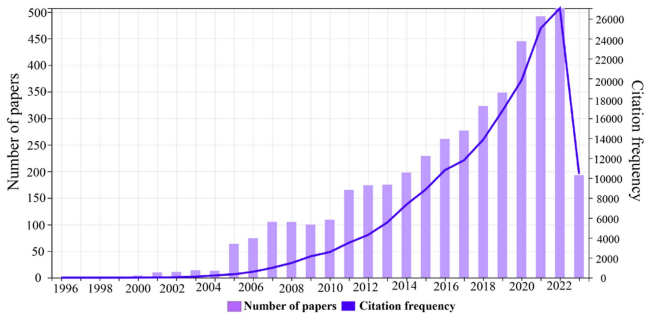

Fig. 11. Frequency of citations to articles on carbon emissions and the number of published papers.

prominence, notably after the Paris Climate Agreement took effect in November 2016, prompting countries to propose emission reduction targets aligned with international objectives [62].

Fig. 9 shows the graph of the number of articles published by relevant countries and institutions about remote sensing carbon emissions. The larger the circle, the more frequently it appears and the more articles are published; the circle layer from inside to outside shows the time from past to present. The red color is highlighting nodes where relevant literature has attracted attention in the field. Connecting lines indicate the existence of continuity between different national institutions. From Fig. 9, it can be seen that the United States has the largest node among countries, the largest number of publications, and the earliest time, indicating that the United States has laid the foundation of research in the field of remote sensing carbon emissions, and has played an important role in theoretical and research innovation with far-reaching influence. China ranked second in the number of publications, and although it started late, it is developing rapidly, thanks to the Chinese government's emphasis on energy conservation and emission reduction, but there is still a gap compared to the United States.

The Chinese Academy of Sciences leads in carbon emissions publications among the top ten institutions, including NASA, Caltech, and others. Despite China's significant overall research output, its emphasis on remote sensing technology in carbon emissions research lags behind Western counterparts.

As shown in Fig. 10, China and the United States have contributed the most papers on the topic of remote sensing carbon emissions. This is likely due to the fact that both countries are major players in the global economy and energy industry, and have therefore invested heavily in research on remote sensing technology for carbon emission monitoring. As a result, the number of paper contributions from developed regions is generally higher than that from developing regions, as the latter may have less resources and funding available for such research.

Fig. 11 presents the publication count and citation frequency of literature on carbon emissions from 1996 to the first half of 2023, obtained from the Web of Science database. The publication count and citation frequency of carbon emissions literature serve as indicators of the level of academic research activity in this field. As depicted in the figure, the discipline of remote sensing applied to carbon emissions has undergone three distinct stages: 1) initial stage; 2) slow growth stage; and 3) rapid growth stage. This indicates the continuous attention of both domestic and international scholars towards research in remote sensing for carbon emissions.

In summary, Citespace aims to assist researchers in gaining a better understanding of the knowledge structure related to carbon emissions, identifying research hotspots and trends, and providing valuable insights and information for academic research, disciplinary development, and climate decision-making.

D. Section Summary

This section summarizes several key elements related to carbon emission observation, encompassing a carbon emission observation program comprising space carbon satellites and ground observation methods. The primary remote sensing satellites utilized are GOSAT and OCO-2, while ground observation

methods include the gas detector method, ground base station verification method, and MRV. In addition, this subsection introduces relevant datasets, incorporating those from carbon satellites and anthropogenic emissions. Finally, the section employs the Citespace visualization tool to showcase recent papers on carbon emission research in terms of regions, keywords, and topic clustering. Section III spotlights the TCOON system. In addition to the retrieval algorithm, ground-based CO_2 detection methods are explored, considering that the column concentration of atmospheric CO_2 in satellite observations encompasses CO_2 emissions resulting from human activities.

Fig. 12. Total Carbon Column Observing Network (TCCON) base stations worldwide [72].

III. Ground Station Observation

In this section, we highlight the essential role of ground observations in exploring carbon cycle patterns. We discuss various ground observation systems, including simple instruments and the widely used TCCON system, crucial for validating satellite-derived data. The principles for analyzing ground carbon monitoring are detailed, comparing their strengths and weaknesses. In addition, we introduce the globally adopted MRV monitoring system, contributing to an improved understanding of carbon cycle processes.

A. CO_2 Detection Methods for Ground-Based Stationary Emission Sources

In the past, ground-based sampling and detection instruments were the only reliable means of detecting changes in global atmospheric CO_2 concentrations. CO_2 gas detection from stationary emission sources involved detecting and quantitatively analyzing CO_2 gas using detectors [63]. Table V summarizes the advantages and disadvantages of these methods [64].

Chemical methods involve manual determination of CO_2 in gases, either by direct laboratory testing or calibration of CO_2 concentrations from stationary emission sources [65]. These methods do not require standard gas calibration and are easy to operate. However, they are not suitable for continuous on-site detection of stationary emission sources due to poor operational timeliness. Gas chromatography is a more precise method for offline determination of CO_2 with high sensitivity, but it requires sampling and preprocessing, which limits its timeliness [66].

The electrochemical method is a widely used and cost-effective instrument for detecting CO_2 due to its lightweight instrumentation and ease of operation. However, its selectivity for gas separation is limited and the sensor cannot operate for long periods, which hinders its further development [67].

Physical-optical methods primarily use nondispersive infrared absorption spectroscopy to detect CO_2, but they are susceptible to external interference, requiring regular instrument calibration for reliability [68]. Despite this vulnerability, these methods are easy to maintain and offer high precision across a broad measurement range, making them applicable in various fields. Another method, Fourier Transform Infrared Spectroscopy (FTIR), has evolved from laboratory techniques to real-time monitoring of industrial emissions, such as those from power plants and factories. FTIR features rapid scanning, simultaneous detection of over 50 compounds, minimal cross-interference, high resolution, and sensitivity, without the need for frequent calibration. However, it comes with drawbacks like high costs, lack of portability, the need for regular maintenance, and higher software workload. FTIR finds widespread application in ground observation platforms, including the National Oceanic and Atmospheric Administration's Earth System Research Laboratory (NOAA ESRL), Total Carbon Column Observing Network (TCCON), Global Greenhouse Gas Reference Network, Comprehensive Observation Network for Trace Gases by Airborne Instruments (CONTRAIL), Intercontinental Chemical Transport Experiment–North America, In-service Aircraft for a Global Observing System, and the Carbon in Arctic Reservoirs Vulnerability Experiment. Among these, the TCCON system stands out as one of the most extensively deployed global observation systems.

B. TCCON System

The Total Carbon Column Observing Network (TCCON) was established in 2004 as a ground-based FTS network to measure the precision and validate the accuracy of CO_2. These instruments record solar absorption spectra, providing high-precision greenhouse gas time series data. TCCON aims to enhance carbon cycle understanding, validate satellite data, and serve as a standard for satellite-ground comparisons [69].

TCCON sites worldwide, strategically placed for impact on CO_2 concentrations, utilize Bruker 120HR or 125HR spectrometers equipped with sun trackers. Gold-plated mirrors reduce photon noise, achieving a 2 cm spectral resolution [70]. While TCCON offers high-precision validation data, its limited spatial coverage is a drawback. Few stations exist in desert areas, and there is no representation in the Pacific or Central Asia [71]. TCCON's location in low-emission areas may not capture changes in CO_2 concentrations in regions with significant emissions, presenting a need for real-time detection methods with broader coverage. Here is Fig. 12 illustrating the global distribution of TCCON sites.[1]

[1]https://tccon-wiki.caltech.edu/

C. Measuring, Reporting, and Verification

The Bali Roadmap, established during the 13th Conference of the Parties to the UNFCCC in December 2007, urges countries to undertake Nationally Appropriate Mitigation Action to combat climate change. This entails devising strategies and implementing measures tailored to each country's circumstances for reducing greenhouse gas emissions. Ensuring transparency and accountability, the Bali Roadmap highlights the significance of MRV for tracking and reporting progress toward emission reduction targets.

Carbon verification is crucial for effectively monitoring greenhouse gas emissions, ensuring accurate data for carbon emissions supervision, management, and trading. The MRV mechanism, mandated by the United Nations, guides this process. In the measurement stage, standardized guidelines and accounting methodologies are employed for accurate and scientifically valid greenhouse gas emissions data. The reporting stage involves making the reporting process public, with specified enterprises or facilities participating based on set thresholds. Third-party verification ensures periodic checks, maximizing data accuracy and credibility [73], [74]. Developed countries have established robust greenhouse gas measurement policies. China, in 2007, introduced the "China National Program for Addressing Climate Change" and adopted standards like the "IPCC Guidelines," "ISUCA," and "Enterprise Standard" for greenhouse gas accounting. However, these estimates need validation, and their frequency falls short for real-time carbon emissions estimation [75], [76].

D. Section Summary

In the third section, we delved into the detection principles of ground-based fixed emission sources, comparing the advantages and disadvantages of different detection principles. Currently, ground monitoring systems primarily employ the Fourier-transform infrared spectroscopy principle, with a particular focus on analyzing the widely used TCCON system globally. TCCON has garnered attention due to its high precision and global distribution. Through a comparative analysis of its superior detection principles, we gained a deeper understanding of its role in monitoring ground carbon emissions. Furthermore, Table VI provides a summary of these three methods.

IV. Satellite Data Retrieval

The technique of using atmospheric retrieval modeling to infer greenhouse gas carbon sources and sinks originated in the 1980s [77], while the retrieval of carbon concentrations using satellite remote sensing began in the early 21st century.This section introduces atmospheric retrieval algorithms and the g-b FTS retrieval algorithm.

A. Atmospheric Retrieval Algorithm

The physically-based retrieval algorithms proposed in this section are as follows:

1) Weighted Function Modified Differential Optical Absorption Spectroscopy (WFM-DOAS);
2) Bayesian Error Subsequent Diffusion (BESD);
3) National Institute for Environmental Studies (NIES);
4) Remote Sensing Technology Company (RemoTeC);
5) Atmospheric CO_2 Observations from Space (ACOS).

We have introduced these algorithms, compared their advantages and disadvantages, and indicated the satellites they are suitable for, as detailed in Table VIII.

1) WFM-DOAS Algorithm: The WFM-DOAS retrieval algorithm was developed in 2006 to calculate the column concentration of target gas using an unconstrained linear least squares method based on scaling preselected vertical sections. It was initially developed for the SCIAMACHY on the European Space Agency's Envisat-1 satellite. The reference spectrum used for linear fitting includes trace gas total column weighting function, temperature section displacement weighting function, and low-order polynomial.The retrieval formula for XCO_2 is as follows [78]:

$$XCO_2 = \frac{\sum(W(j) \cdot \Delta XCO_2(j))}{\sum W(j)}. \tag{2}$$

XCO_2 represents the retrieved column-averaged CO_2 mixing ratio. $W(j)$ represents the weight factor at wavelength point j. $\Delta XCO_2(j)$ represents the change in XCO_2 at wavelength point j.

In 2012, Heyman et al. [79] discovered a bias in the early WFMDv2.1 version, as it overlooked the impact of intractable scattering in thin cloud layers. To address this, version v2.2 was developed, specifically correcting biases in thin cloud layers, particularly in the presence of strong water vapor absorption and the O2-A band. This improved treatment of clouds enhances XCO_2 dataset accuracy, considering height-correlated effects of temperature and pressure. It also corrects variations in absorption cross-section, ensuring higher measurement precision. This algorithm is designed for the SCIAMACHY satellite [80].

2) RemoTeC Algorithm: The RemoTec algorithm was developed in 2009 by the Netherlands Institute for Space Research in collaboration with KIT. It retrieves the total column of CO_2 by parameterizing the number of particles, their height distribution, and microphysical properties [81]. The retrieval formula for XCO_2 is as follows:

$$XCO_2 = f \cdot \alpha \cdot \beta. \tag{3}$$

f is a correction factor used to adjust for measurement biases. α represents atmospheric profile parameters, including temperature profiles, humidity profiles, and atmospheric composition profiles. β represents the sensitivity function of XCO_2, describing the response of measurement wavelengths to XCO_2.

RemoTec algorithm has two versions: 1) Simplified Retrieval Proxy (SRPR); and 2) Full Physics Retrieval (SRFP). SRPR is a simplified and efficient version suitable for scenarios with lower precision requirements. SRFP is more comprehensive, considering atmospheric aerosol parameters and particle height distribution, making it suitable for research and applications requiring detailed information [82]. Both versions use TANSO-CAI/GOSAT cloud screening and are compatible with GOSAT and OCO-2 satellites [83].

TABLE VII
NIR SPECTRAL BANDS USED IN THE ACOS SEARCH

Band number	Name	Spectral range (cm^{-1})	Number of GOSAT channels
1	O_2A	12950~13190	1203
2	Weak CO_2	6166~6286	601
3	Strong CO_2	4810~4897	436

3) NIES Algorithm: In 2011, NIES in Japan developed an algorithm to extract column abundances of greenhouse gases efficiently from short infrared spectra obtained by TANSO FTS's thermal infrared and near-infrared sensors [84].

NIES algorithm continually enhances retrieval accuracy by updating data sources, improving physical models, and refining processing methods [85]. It can simultaneously retrieve multiple atmospheric parameters such as CO_2 concentration, CH_4 concentration, and aerosol optical thickness. This versatility positions NIES algorithm favorably for comprehensive atmospheric component research and monitoring. The algorithm incorporates metrics like DFS, MSR, and AOD to assess its performance [84], [86]. However, it has some drawbacks, including data dependency and sensitivity to missing or noisy observational data, and its complexity is rooted in sophisticated physical models and retrieval methods. Overall, the NIES algorithm demonstrates improved atmospheric retrieval accuracy compared to the preceding algorithms [87]. It is compatible with GOSAT, OCO-2, and SCIAMACHY satellites.

4) BESD Algorithm: In 2010, Buchwitz et al. [88] the BESD algorithm. The BESD algorithm is an optimal estimation algorithm developed for evaluating SCIAMACHY measurements, it is used to retrieve XCO_2 from the SCAMACHY nadir measurements.

BESD algorithm employs the principle of optimal estimation to minimize differences between simulated and observed measurements. It can retrieve independent parameters across multiple bands. However, the algorithm has some drawbacks, such as sensitivity to data gaps leading to inaccurate results. It relies on prior information to improve retrieval accuracy and is specifically developed for the SCIAMACHY satellite.

5) ACOS Retrieval Algorithm: In 2012, Crisp et al. [74] introduced the ACOS algorithm which utilizes the best estimation method to optimize the input parameters of the forward model to match the simulated spectra with the observed spectra, while constrained by a priori information. The matched spectra are obtained from three OCO NIR bands: band 1 near 0.76 μm, band 2 μm near 1.6 μm, and band 3 μm near 2.1 μm. The wavebands and their spectral ranges can be found in Table VII. The state vector x comprises the forward model parameters that are to be optimized.

All channels in the three bands are aggregated into an observation vector y. Mathematically, the observed value Y of the simulated state vector x takes the following form:

$$Y = F(x, b) + \epsilon. \tag{4}$$

In the ACOS algorithm, the forward model of the retrieval is denoted as F in formula (4). The fixed set of input parameters is represented by b, and ϵ contains estimates of the instrument noise and forward model error. The flow of the ACOS 2-stage retrieval algorithm is illustrated in Fig. 13 [51]. The first input is the "Level-1B Product," which contains calibration, spectrally resolved emissivity, and geometric information for each of the three spectral bands. In the prescreening step, poor-quality data and cloudy scenes are removed. The filtered data is then passed to the core of the algorithm, which is the XCO_2 retrieval step. Here, a priori states are constructed based on meteorological inputs and observed spectra.

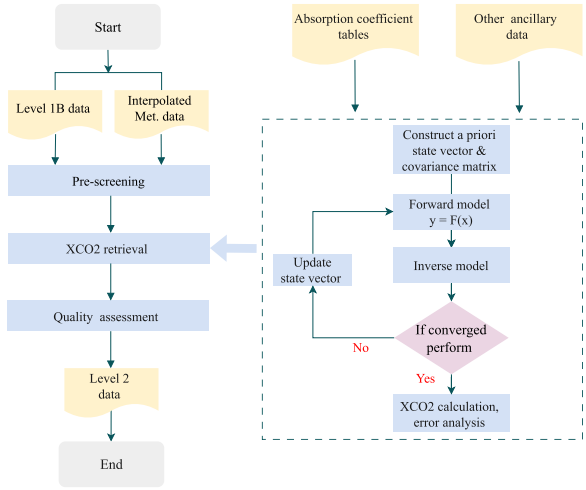

Fig. 13. Flow of the basic steps of the ACOS Level 2 algorithm.

The ACOS algorithm is a commonly used method for retrieving atmospheric CO_2, offering global-scale CO_2 distribution information. Utilizing spectral data from multiple near-infrared bands, it provides richer information for estimating CO_2 concentrations. The algorithm streamlines the process, avoiding the need for complex models and external data, and relies relatively less on prior information. ACOS enables nonintrusive observation of atmospheric CO_2 concentrations without perturbing the atmospheric system or local adoption. However, the algorithm is influenced by observational conditions such as cloud cover and atmospheric optics. It is suitable for GOSAT and OCO-2 satellites.

B. g-b FTS-Based Retrieval Algorithm

The Ground-Based Fourier Transform Spectrometer (g-b FTS) has been operational at the Amiens site since 2014, as part of the TCCON network. Using the Fourier transform principle, g-b FTS decomposes spectral signals of different wavelengths into frequency signals, obtaining spectral information on various chemical substances in the atmosphere. It provides detailed vertical profiles of the atmosphere near or at the ground level. With a time resolution of approximately 2 min, g-b FTS offers observational data over short periods, facilitating the monitoring of short-term variations in atmospheric components. Compared to satellite greenhouse gas remote sensing measurements, g-b FTS is less affected by atmospheric factors. However, ground-based observations are constrained by geographical location, potentially limiting the comprehensive monitoring of global atmospheric changes. In addition, ground observations are susceptible to weather conditions such as cloud cover and precipitation, leading to potential data gaps and inaccuracies [73].

TABLE VIII
SUMMARY OF RETRIEVAL ALGORITHMS

Algorithm	Year	Method	Available data	Advantages
WFM-DOAS	2006	Linear least squares	SCIAMACHY	Calculation of target gas column concentration
RemoTec	2009	Retrieval	GOSAT and OCO-2	Total CO_2 column can be retrieved
BESD	2010	Based on optimal estimation	SCIAMACHY OCO-2 and GOSAT	Retrieval of the difference in XCO_2-related scattering
NIES	2011	Unbiased cloud detection method	GOSAT	Retrieval of column abundance by short infrared spectroscopy
ACOS	2012	Based on optimal estimation	GOSAT and OCO-2	Estimated XCO_2 and L2 product data values

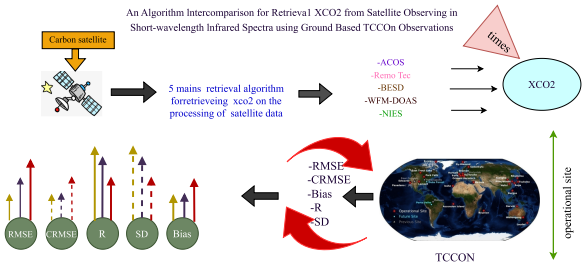

Fig. 14. Validation of the retrieval of XCO_2 from satellite observations based on ground-based TCCON observations (adapted from [69].

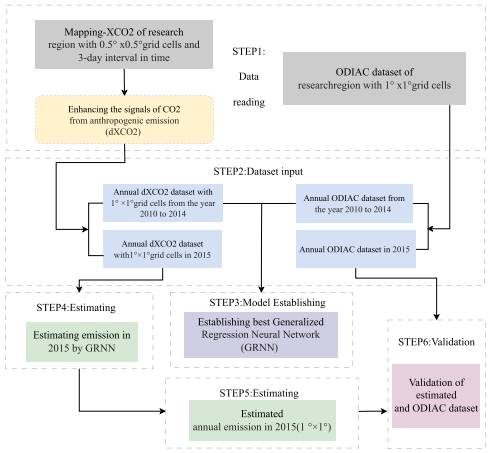

Fig. 15. Flowchart for estimating anthropogenic emissions.

The retrieval algorithms represented by this instrument include the GFIT retrieval algorithm developed by the Jet Propulsion Laboratory [89], the PROFFIT retrieval algorithm from the Karlsruhe Institute of Technology [90], and the SFIT retrieval algorithm developed by RINSLANDC and others [91]. The PROFFIT and SFIT retrieval algorithms, based on the optimal estimation approach, are widely used and continuously updated. PROFFIT is suitable for retrieving gas concentrations in the near-infrared spectral range, while SFIT is applicable for retrieving gas concentrations in the midinfrared spectral range. To assess the accuracy of algorithm products, various widely used evaluation metrics were employed, with different benchmark TCCON measurement points selected as reference data. Accuracy assessment metrics include Root Mean Square Error (RMSE), Central Root Mean Square Error (CRMSE), Bias (BIAS), Correlation Coefficient (R), and Standard Deviation (SD).

C. Section Summary

This section discusses the pros and cons of different atmospheric retrieval algorithms and their use cases, as outlined in Table VIII. It also analyzes the distinctions between the g-b FTS retrieval algorithm and other atmospheric retrieval methods. In conclusion, each algorithm has its strengths and limitations. For detailed insights into CO_2 concentrations at a specific regional site, the g-b FTS retrieval algorithm may be suitable. For global-scale climate monitoring and research, atmospheric retrieval algorithms provide more comprehensive data. The choice of the appropriate algorithm depends on factors like research scale and data quality. Section V will delve further into Method 3 from Fig. 3.

Fig. 14 shows how carbon satellites use retrieval algorithms to generate XCO_2 products by identifying absorption spectra. These products are then validated against high-resolution XCO_2 data from ground-based Fourier transform spectrometers, using metrics like RMSE, CRMSE, SD, and Bias to assess accuracy. The validated XCO_2 results are sent back to the satellite for storage.

V. APPLICATION OF MACHINE LEARNING IN CARBON EMISSION ESTIMATION

In recent years, there has been increasing attention on estimating CO_2 emissions at national, regional, and global levels, gradually becoming a crucial component of public policies and strategies to address climate change.

Currently, many studies employ machine learning techniques for estimating CO_2 emissions. While these studies provide overall estimates, they primarily focus on assimilation techniques, emission inventories, or other auxiliary data. However, there are reliability issues in assessing ground emission data, prompting the need to establish an independent, objective, and comprehensive system for measuring CO_2 emissions.

To overcome these limitations, researchers propose a method that uses machine learning to directly infer ground carbon emissions from satellite monitoring of CO_2 concentration, validated against ground emission inventories. However, compared to ground inventory data, the application of machine learning in monitoring CO_2 concentration from satellite observations is relatively limited, leading to a less comprehensive understanding of emission distributions in different geographical locations or spatial dimensions. Future developments include using satellite

TABLE IX
ACRONYMS

ACOS	Atmospheric CO_2 observations from space	JAXA	Japan Aerospace Exploration Agency
AD	Absolute Deviation	JPL	Jet Propulsion Laboratory
AIRS	Atmospheric InfraRed Sounder	LSTM	Long Short-Term Memory
ANN	Artificial Neural Network	MAE	Mean Absolute Error
ARIMA	Autoregressive Integrated Moving Average	MAPE	Average absolute percentage error
ARMA	Autoregressive Moving Average	MEIC	Multi-resolution Emission Inventory for China
ASE	Automatic air Sampling Equipment	MLP	MultiLayer Perceptron
BA	Bat Algorithm	MOE	Ministry of Environment
NOAA	National Oceanic and Atmospheric Administration	MSE	Mean Square Error
Bees	Bees Algorithm	MSR	Mean Square Residual
BESD	Bremen optimal Estimation DOAS	NASA	National Aeronautics and Space Administration
BP	British Petroleum	NDIR	Non Dispersive InfraRed
BPNN	Back-propagation neural network	NGBM	Nonlinear Grey Bernoulli Model
CAI	Cloud and Aerosol Imager	NIR	Near InfraRed
CDIAC	Carbon Dioxide Information Analysis Center	NIES	National Institute for Environmental Studies
CARMA	CARbon Monitoring for Action	NOAA	National Oceanic and Atmospheric Administratin
CEADs	Carbon Emission Accounts & Datasets	NRBF	Normalized Radial Basis Function
Chi2	Chi-Square Statistic	ODIAC	Open-source Data Inventory for Anthropogenic CO_2
CRMSE	Center Root Mean Square Error	OCO	Orbiting Carbon Observatory
CME	Continuous Measuring Equipment	OMI	Ozone Monitoring Instrument
DOAS	Differential Optical Absorption Spectroscopy	PRCC	Partial Rank Correlation Coefficient
DFS	Discrete Fourier Series	RBNN	Radial Basis Neural Network
EC	Eddy Covariance	RMSE	Root Mean Square Error
ENVISAT	European eNVIronment SATellite	PPDF	Path Probability Distribution Function
ESA	European Space Agency	RNN	Recurrent Neural Network
ESRL	Earth System Research Laboratory	SCIAMACHY	Scanning Imaging Atmospheric Absorption Spectrometer
FTS	Fourier Transform Spectrometer	SIF	Solar-Induced Fluorescenc
g-b	Ground-based high resolution	SWIR	Short Wave InfraRed
GHG	Green House Gas	TANSO	Thermal And Near-infrared Sensor

XCO_2 data to finely depict the spatial distribution of ground carbon emissions, thereby enhancing monitoring precision and comprehensiveness.

A. Estimation of Anthropogenic Carbon Emission Inventory Data Based on Machine Learning

1) Carbon Emission Prediction Based on Classical Statistics: Over the past few decades, researchers have used classical statistical and econometric methods to model or predict the growth of carbon emissions [92], [93]. Regression-based analysis has been an effective method to study carbon emissions, incorporating other independent variables such as economic development, energy consumption, technology and innovation, and population [94]. However, given that the independent variables of carbon emissions are often nonlinear and nonstationary [95], some commonly used time series models, such as Grey Model (GM), Nonlinear Grey Bernoulli Model (NGBM), and AutoRegressive Integrated Moving Average (ARIMA) model [93], aim to enhance the accuracy and reliability of predictions. By combining these time series models with regression analysis

TABLE X
ACRONYMS

b	b is a fixed set of input parameters in the forward model of formula 2.
$e^{-\frac{1}{2\sigma^2}x^2}$	This is the expression form of the Gaussian function.
f()	Activation functions for all hidden layer elements of formula 8.
F(x,b)	F(x,b) is the inverse forward model, where x is the simulated state vector and b is a fixed set of input parameters.
F_i	F is the average observed CO_2 data in ppm. F_i:Observed CO_2 data in ppm, i∈1∼n
F_i^*	F is the average observed CO_2 data in ppm. F_i^*: CO_2 data in ppm for each algorithmic product, i∈1∼n
S_a	S_a is a priori covariance matrix.
S_ϵ	$S_\epsilon$is the observation error covariance matrix., ∈ contains the estimation of instrument noise and forward model error
$u^{(0)}(k)$	$u^{(0)}(k)$ represents the initial value of the time series of formula 4, k = 1, 2, ... , m
x_i	Network input for neuron i in the input layer
$Z^{(1)}(k)$	$Z(1)(k)$ represents the observed values of the first order differences of formula 5 . k is the variable representing the index of each observation in a time series. k =1 , 2, ..., m
ϵ	It contains the estimation of instrument noise and forward model error of formula 2
υ_0	υ_0 is a bias term in formula 8, which adjusts the intercept of the model.
ω	It is the weight of the RBF module.

methods, a more comprehensive understanding and prediction of the growth trends in carbon emissions can be achieved [96].

For instance, Pao et al. [99] utilized the GM to predict the dynamic changes in three variables in Brazil from 2008 to 2013, including pollutant emissions, energy consumption, and output. The model's input data comprised pollutant emissions, energy consumption, and output, while the output data was the carbon dioxide emissions. This model exhibited advantages in prediction over the ARIMA model [98]. In another study, Pao et al. in 2012 used the NGBM to forecast carbon emissions in China from 2004 to 2009. They proposed a numerical method to optimize the parameters of NGBM and compared it with the ARIMA and GM models. The NGBM model showed smaller RMSE and stronger predictive ability compared to the ARIMA model [100]. Furthermore, Zheng et al. [97] analyzed the spatiotemporal distribution of near-surface and midtropospheric carbon dioxide concentrations globally using GOSAT L2 dataset and ground observations from atmospheric background stations. They employed the ARIMA model to predict the future changes in carbon dioxide concentrations in near-surface regions of China. The results indicated a slight advantage of the ARIMA model in fitting and predicting GOSAT remote sensing dataset.

These three classical statistical methods do not have an absolute ranking in terms of superiority or inferiority; instead, they each have scenarios in which they are more suitable. The ARIMA model is suitable for linear and stationary time series data, demonstrating good fitting capability for trends and seasonality but may perform less effectively for nonlinear and nonstationary data. The GM model is appropriate for situations with limited data and challenges in establishing precise models. However, it may not perform as well as other complex models in cases of long-term prediction and large datasets [101]. The NGBM model is mainly used for long-term prediction, particularly well-suited for large-scale data compared to the traditional GM model [102]. It exhibits better generalization capabilities but requires parameter optimization, and its fitting effectiveness is limited for certain types of data. Nevertheless, the challenge for CO_2 emissions lies in dependence on historical data and the difficulty of accurately capturing complex nonlinear relationships. In situations influenced by various factors, statistical-based methods may not comprehensively consider the impact of these factors. In 2015, Falat et al. [103] conducted multiple comparative experiments and found that artificial neural networks outperformed statistical regression analysis, simple time series models, and other methods. Artificial neural networks were considered a more suitable choice for estimating the carbon dioxide emissions of each country.

2) Neural Network-Based Carbon Emission Prediction: Artificial Neural Networks (ANNs) are computational models inspired by the human brain's structure. Comprising interconnected neurons arranged in layers, including input, hidden, and output layers, ANNs utilize adjustable weights and activation functions for information processing [104]. Common types include backpropagation neural networks (BPNN), feedforward neural networks (MLP), radial basis function networks, and long short-term memory (LSTM) neural networks. ANNs find applications in diverse tasks such as pattern recognition, classification, and prediction. They have shown promising results in predicting carbon dioxide emissions for different countries [105], [106]

Jiang et al. [107] utilized a backpropagation (BP) neural network to analyze environmental factors for CO_2 flux in urban forestland. Input variables included wind speed, air temperature, water vapor concentration, atmospheric pressure, and water vapor pressure. The model predicted CO_2 flux based on these inputs. Alex et al. [74] expanded on World Development Indicators (WDI) data, introducing variables like population, financial development, and industrialization. They used a BP neural network to study carbon emissions in countries like Australia, Brazil, China, India, and the United States.

Shabani et al. [108] enhanced prediction accuracy for Iranian agricultural CO_2 emissions using a multimodel neural network approach, integrating multiple models for superior results.

In 2021, Rezaei et al. [109] achieved an R2 value of 0.98 for Southeast Asian countries' CO_2 production using neural radial basis function (NRBF) and tansig function neural networks. Assad forecasted carbon dioxide levels in Middle Eastern countries with a Multilayer Perceptron (MLP) network, finding logsig activation provided the highest accuracy. Jassim et al. [110] used artificial neural networks to predict CO_2 emissions from highway trucks. Other methods, like Support Vector Machines, have been applied to model CO_2 emissions [111]. In 2022, Hu et al. [112] employed LSTM neural networks to predict China's total carbon emissions. Comparing with an ARIMA-BP model, they found LSTM neural networks had superior predictive performance.

These examples illustrate the gradually maturing application of neural networks in the field of ground carbon emissions. This development is reflected not only in technological updates and optimizations, but also in the more comprehensive and accurate modeling and prediction of CO_2 emissions for different countries, regions, and specific industries. Researchers are transitioning from traditional BP models to LSTM models and experimenting with different activation functions in the models, continually exploring and optimizing deeper neural network architectures. Such efforts aim to better adapt to complex environmental factors and variable data, thereby achieving significantly improved predictive outcomes for the models.

B. Machine Learning for Ground Carbon Emission Analysis With Carbon Satellite Data

When conducting this research, several technical and data-related challenges were encountered:

First Challenge. Limitations in Measurement Orbits: Due to the restricted measurement orbits of carbon satellites, data collection faced challenges, limiting the possibility of multiple measurements of ground carbon sources. Matching with fixed monitoring stations for multiple comparisons may present considerable challenges.

Second Challenge. Discrete Imaging in Strip-Like Patterns: Carbon satellites typically image in a strip-like manner, resulting in discontinuous data. Carbon concentration is recorded at discrete positions, posing a challenge for machine learning methods that typically handle continuous data. Researchers often need to consider how to fill or estimate XCO_2 data for unobserved areas during data processing and modeling.

Third Challenge. Difficulty in Obtaining Wind Field Information: Obtaining accurate wind field information is difficult, and wind conditions can impact the quantification of ground carbon emissions. However, obtaining precise wind field information is extremely challenging, and usually, only average wind speed and direction data over larger spatial and longer temporal scales are available. These challenges collectively increase the complexity of studying ground carbon emissions.

These challenges collectively increase the complexity of studying ground carbon emissions. The following section introduces innovative approaches currently adopted by researchers to address these technical and data-related challenges.

1) Ground Carbon Emission Estimation Based on Clustering Methods: K-means clustering relies on the average Euclidean distance between data points, assigning them to clusters based on proximity to the mean. EM clustering, based on a Gaussian distribution mixture, involves iterative assignment and updating steps. In this study, K-means clustering, driven by NO_2 levels, initially divides data into four groups based on increasing NO_2 tropospheric column values. The results are then used for EM clustering, offering finer refinement, and a better understanding of the correlation between XCO_2 anomalies and NO_2 tropospheric columns [113].

Unlike directly estimating ground carbon emission data, estimating ground carbon emissions by combining satellite monitoring data poses significant challenges due to the complexity introduced by trends, seasonality, long lifetimes, and the large-scale atmospheric background of CO_2 compared to short-term air pollutants like NO_2. The analysis of anthropogenic CO_2 emissions becomes intricate as CO_2 plumes generated by ground emission sources may cause local enhancements in near-source observational data from carbon satellites. Diffusion and flow of gases lead to the separation of these local enhancements from the background XCO_2. Such local enhancements are considered reflections of ground carbon emission data.

In 2016, Hakkarainen et al. [2] first proposed that to highlight regions relevant to ground carbon dioxide emissions, it is necessary to calculate the anomaly (ground carbon emissions) of XCO_2 by subtracting the median $\overline{XCO_2}$ from daily observations at a particular moment. The formula is as follows:

$$XCO_2(\text{anomaly}) = XCO_2(\text{specific moment}) - \overline{XCO_2}. \tag{5}$$

First, ground carbon emission data were computed by subtracting daily background concentrations from each satellite observation. This process helps eliminate seasonal variations and mitigates the spatial distribution impact on data points. The daily median is determined when there are at least 100 measurements available on the same day within the selected study area.

Second, the average of all XCO_2 anomalies within defined grid cells is calculated. This average is computed only when there are at least ten available measurement values in each grid cell. The study utilized $1^\circ \times 1^\circ$ grid cells for interpolation analysis, using OCO-2 XCO_2 version 7r data. As an auxiliary dataset, NO_2 tropospheric column measurement data from the Ozone Monitoring Instrument (OMI) were employed. ODIAC data served as a basis for comparing with XCO_2 anomalies. Two unsupervised machine learning methods: 1) k-means clustering; and 2) expectation-maximization clustering, were applied in the analysis.

These methods reveal that the spatial features of detected CO_2 emission regions in the XCO_2 anomaly map align closely with the ODIAC emission inventory. The study establishes a

positive correlation between average XCO_2 anomaly values and CO_2 emission values, marking the first research to directly differentiate ground carbon emissions from OCO-2 observations. However, the study notes the oversight of seasonal variations' impact on ground carbon emissions, suggesting the need for seasonal considerations for a more comprehensive analysis.

2) Ground Carbon Emission Estimation Based on the GRNN Method: GRNN is a neural network employing radial basis functions and kernel regression. Unlike backpropagation, GRNN doesn't need iteration and can estimate functions between input and output datasets directly from training data. It consists of the following four layers:

1) input;
2) pattern;
3) summation;
4) output.

The number of input variables depends on observed and collected datasets, and this information is sent to the pattern layer. To follow neural network computations, it is essential to standardize all training data initially, as per the formula [114].

$$d(x_0 - x_i) = \sum_{j=1}^{p} \left(\frac{x_{0j} - x_{ij}}{\sigma} \right)^2 \tag{6}$$

p represents the dimension of the variable x_i, and σ is the diffusion parameter. Its optimal value is determined by minimizing the RMSE between the training data and the predicted values of the dependent variable. The predicted target variable is the emission of ODIAC CO_2, defined by the following equation:

$$\hat{y}(x_0) = \frac{\sum_{i=1}^{n} y_i e^{-d(x_0, x_i)}}{\sum_{i=1}^{n} e^{-d(x_0, x_i)}}. \tag{7}$$

x_i is the input of a training sample, y_i corresponds to the target output for the respective training sample. The weights are determined by $e^{-d(x_0, x_i)}$. Hakkarainen et al. did not consider seasonal variations causing periodic CO_2 fluctuations. Its structural flowchart as shown in Fig. 15. They calculated three-day interval anomaly values and derived annual averages, eliminating regional seasonal changes and minimizing atmospheric transport effects. Yang generated monthly and annual average $dXCO_2$ datasets, validated against ODIAC emissions. Notably, Yang introduced GRNN model to illustrate the nonlinear relationship between $dXCO_2$ and ODIAC, a novel approach in this study [14].

Yang used $dXCO_2$ and ODIAC emission data from 2010 to 2014 for training, comprising 5415 samples. The model predicted 2015 ODIAC emissions using $dXCO_2$ from that year, validated against preprocessed ODIAC data. The correlation between GOSAT XCO_2 anomalies and Chinese ODIAC CO_2 emissions showed an R2 of 0.82, supporting ground carbon emission validation and regional estimation based on XCO_2 variation, especially in high-emission regions. However, the model had higher uncertainty in low/no emission and point source areas. Annual $dXCO_2$ estimation was influenced by biosphere CO_2 absorption and flux, requiring auxiliary data like primary productivity and nighttime light. Yang's method offers rapid ground carbon emission updates and a new approach. In 2021, Mustafa et al. [16] improved on Yang's dataset, focusing on East and West Asia, using ACOS V10r and adding net primary productivity (NPP) data. Results showed a smoother CO_2 emission distribution with a 13% increase in R2. Detection in deserts remains challenging. The study introduced NPP data, enriching the model's information on ecosystem CO_2 dynamics.

Incorporating meteorological data may enhance the model, especially in special regions. Considering factors like wind speed and temperature improves atmospheric CO_2 simulation accuracy. Including meteorological data enhances ground carbon emission estimation, particularly in complex surface feature regions.

3) Ground Carbon Emission Estimation Based on Deep Learning: The Transformer model, based on self-attention mechanisms in deep learning, has shown effectiveness in processing sequential data, making it a valuable tool for handling satellite remote sensing data. Unlabeled data were used to mask each input sequence with a masking ratio ρ, inputting the masked data into the Transformer-based neural network architecture to output representations of the masked data [115].

Studies utilizing machine learning for satellite-based estimation of ground carbon emissions often overlook crucial factors like wind speed, humidity, and temperature. In 2022, Zhao et al. addressed this gap by integrating multiple data sources, developing a data processing algorithm, and proposing a novel deep learning model for estimating carbon dioxide emissions from individual ground sources. The methodology involved three main steps: 1) retrieval based on carbon sources; 2) handling anomalous data; and 3) linear prediction. The study effectively leveraged carbon satellite data characteristics, employing a Transformer model for mask pretraining and linear regression for supervised modeling, establishing a mapping from satellite data to carbon emissions [17].

C. Section Summary

This section covers two main aspects of machine learning in carbon emissions research. The first part focuses on modeling the relationship between ground variables (such as fossil fuel consumption, population, and GDP) and carbon emissions, using classical statistics, time series models, and neural networks. The second part explores the application of machine learning in analyzing carbon satellite data, addressing challenges like limited measurement orbits and discrete strip imaging. Researchers employ innovative methods, including clustering and GRNN, to estimate ground carbon emissions directly from satellite observations. The introduction of deep learning, specifically Transformer models, presents new possibilities for data processing. However, challenges such as satellite data gaps and improving model adaptability remain areas for future development.

VI. Summary and Perspective

This review focuses on three main aspects of machine learning in the current field of carbon dioxide monitoring: 1) ground observation; 2) satellite observation; and 3) the integrated application of satellite and ground observations. We comprehensively discuss widely used ground carbon emission inventories and

the development stages of XCO_2 datasets derived from carbon satellite retrievals. Utilizing the Citespace software, we provide a comprehensive review of global research in the carbon emission domain. The subsequent sections offer detailed analyses of different ground observation stations, their applicable scopes, adopted algorithms, and an introduction to the widely used MRV system. We delve into atmospheric retrieval algorithms, clarify the strengths and weaknesses of various algorithms, and guide the selection of algorithms for different carbon satellites. The final focus is on the application of machine learning in carbon emission estimation, covering direct estimation methods for ground carbon emissions, such as using ARIMA, GM, and ANN to establish relationships between various variables and carbon emissions. Despite the effectiveness of these models, their reliance on self-reported information makes them subjective.

The estimation of ground carbon emissions based on satellite carbon columns provides an impartial measurement system. Researchers have successfully retrieved ground carbon emission data using carbon columns, distinguishing background concentrations and seasonal variations in emissions through clustering and GRNN models. In addition, using Transformer models with multisource data to estimate emissions from individual ground sources shows promise. While these methods have achieved some success, challenges related to satellite limitations, discrete imaging, atmospheric dynamics, and satellite noise necessitate further research in this nascent field.

Future CO_2 monitoring research faces both challenges and opportunities. To advance, researchers may consider focusing on the following.

1) *Enhanced Remote Sensing:* Improve data techniques, refining algorithms to reduce aerosol and cloud interference. Fill data gaps using neural networks for better spatiotemporal resolution.
2) *Integrated Ground Emission Modeling:* Combine diverse datasets for comprehensive emission models, considering semantic, elevation, and meteorological data for a nuanced understanding of spatial distribution.
3) *Continual Machine Learning Improvement:* Evolve machine learning for better precision. Explore deeper levels of algorithms to model intricate carbon emission relationships effectively.
4) *Satellite Technology Innovation:* Innovate satellites for higher spatiotemporal resolution, catering to diverse monitoring needs. These advancements will bolster the accuracy and real-time capabilities of carbon emission estimation methods, supporting in-depth climate studies and policy formulation.
5) To achieve the goal of carbon neutrality in the future, it is not only necessary to monitor carbon emission sources, but also to conduct in-depth research on the complex interactions between carbon sinks and carbon sources.

In summary, future research will focus on improving remote sensing precision, integrating diverse emission model data, advancing machine learning technologies, innovating satellite technologies, and conducting comprehensive studies on the relationship between carbon sinks and sources in the context of future carbon neutrality goals.These comprehensive research methods will provide a more scientific and practical foundation for achieving carbon neutrality, addressing climate change, and promoting sustainable development.

REFERENCES

[1] O. Schneising et al., "Atmospheric greenhouse gases retrieved from SCIAMACHY: Comparison to ground-based FTS measurements and model results," *Atmospheric Chem. Phys.*, vol. 12, no. 3, pp. 1527–1540, 2012.

[2] J. Hakkarainen, I. Ialongo, and J. Tamminen, "Direct space-based observations of anthropogenic CO_2 emission areas from OCO-2," *Geophysical Res. Lett.*, vol. 43, no. 21, pp. 11400–11406, 2016, doi: 10.1002/2016GL070885.

[3] L. Liang, Q. Wang, S. Qiu, D. Geng, and S. Wang, "NEP estimation of terrestrial ecosystems in China using an improved CASA model and soil respiration model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10203–10215, 2023, doi: 10.1109/JSTARS2023.3325774.

[4] L. Liang et al., "Remote sensing estimation and spatiotemporal pattern analysis of terrestrial net ecosystem productivity in China [J]," *Remote Sens.*, vol. 14, 2022, Art. no. 1902, doi: 10.3390/rs14081902.

[5] G. Dong, T. Ruan, and Z. Chen, "A review of carbon emission estimation methods in the carbon trading process," *Syst. Simul. Technol. Appl.*, vol. 15, pp. 108–112, 2015.

[6] N. Zhang, Z. Zhang, and L. Kang, "A review of CO_2 emission detection methods for stationary emission sources," *J. China Acad. Environ. Manage. Cadres*, vol. 26, 2016, Art. no. 60.

[7] J. Chen and Y. Ren, "Recent advances in research on land use carbon emissions over the past decade," in *Proc. Ind. Archit. Academic Exchange Conf.*, 2020, pp. 17–21.

[8] D. Debone, V. P. Leite, and S. G. E. K. Miraglia, "Modelling approach for carbon emissions, energy consumption and economic growth: A systematic review," *Urban Climate*, vol. 37, 2021, Art. no. 100849.

[9] D. G. Streets et al., "Emissions estimation from satellite retrievals: A review of current capability," *Atmospheric Environ.*, vol. 77, pp. 1011–1042, 2013.

[10] T. Yue, L. Zhang, M.-W. Zhao, Y. Wang, and J. Wilson, "Space-and ground-based CO_2 measurements: A review," *Sci. China Earth Sci.*, 59, pp. 2089–2097, 2016.

[11] K. Lees, T. Quaife, R. Artz, M. Khomik, and J. M. Clark, "Potential for using remote sensing to estimate carbon fluxes across Northern Peatlands — A review," *Sci. Total Environ.*, vol. 615, pp. 857–874, 2017.

[12] T. Kaminski et al., "Assessing the impact of atmospheric CO_2 and NO2 measurements from space on estimating city-scale fossil fuel CO_2 emissions in a data assimilation system," *Front. Remote Sens.*, vol. 3, 2022, Art. no. 887456, doi: 10.3389/frsen.2022.887456.

[13] K. Miyazaki and K. Bowman, "Predictability of fossil fuel CO_2 from air quality emissions," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 1604, doi: 10.1038/s41467-023-37264-8.

[14] S. Yang, L. Lei, Z. Zeng, Z. He, and H. Zhong, "An assessment of anthropogenic CO_2 emissions by satellite-based observations in China," *Sensors*, vol. 19, 2019, Art. no. 44.

[15] Y. Wang, M. Wang, B. Huang, S. Li, and Y. Lin, "Estimation and analysis of the nighttime PM2.5 concentration based on LJ1-01 images: A case study in the Pearl River Delta urban agglomeration of China," *Remote Sens.*, vol. 1317, 2021, Art. no. 3405.

[16] F. Mustafa et al., "Neural network based estimation of regional scale anthropogenic CO_2 emissions using OCO-2 dataset over East and West Asia," *Atmospheric Meas. Techn. Discuss.*, pp. 1–17, 2021.

[17] Z. Zhang, J. Gu, J. Zhao, J. Huang, and H. Wu, "Near real-time CO_2 emissions based on carbon satellite and artificial intelligence," *Atmospheric Ocean. Phys.*, vol. 29, pp. 4376–4389, 2022.

[18] Y. Wang, M. Wang, F. Teng, and Y. Ji, "Remote sensing monitoring and analysis of spatiotemporal changes in China's anthropogenic carbon emissions based on XCO_2 data," *Remote Sens.*, vol. 15, no. 12, 2023, Art. no. 3207.

[19] K. Hu, D. Zhang, M. Xia, M. Qian, and B. Chen, "LCDNet: Lightweighted cloud detection network for high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4809–4823, 2022, doi: 10.1109/JSTARS.2022.3181303.

[20] K. Paustian, N. H. Ravindranath, and A. R. V. Amstel, *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Hayama, Japan: Institute for Global Environmental Strategies, 2006.

[21] H. Bovensmann et al., "A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications," *Atmospheric Meas. Techn.*, vol. 3, no. 4, pp. 781–811, 2010, doi: 10.5194/amt-3-781-2010.

[22] R. M. Duren and C. C. Miller, "Measuring the carbon emissions of megacities," *Nature Climate Change*, vol. 2, pp. 560–562, 2012.

[23] K. Hu, M. Li, M. Xia, and H. Lin, "Multi-scale feature aggregation network for water area segmentation," *Remote Sens.*, vol. 14, no. 1, 2022, Art. no. 206, doi: 10.3390/rs14010206.

[24] T. Lei et al., "Adaptive CO_2 emissions mitigation strategies of global oil refineries in all age groups," *One Earth*, vol. 4, pp. 1114–1126, 2021.

[25] M. Kiel et al., "How bias correction goes wrong: Measurement of XCO_2 affected by erroneous surface pressure estimates," *Atmospheric Meas. Techn.*, vol. 12, no. 4, pp. 2241–2259, 2019, doi: 10.5194/amt-12-2241-2019.

[26] Concepts R., "BP statistical review of world energy 2001 in review," British Petroleum, 2001.

[27] C. D. Elvidge et al., "A fifteen year record of global natural gas flaring derived from satellite data," *Energies*, vol. 2, pp. 595–622, 2009.

[28] M. Li, H. Liu, and G. Geng, "Anthropogenic emission inventories in China: A review," *Atmospheric Chem. Phys. Discuss.*, vol. 4, no. 6, pp. 834–866, 2018.

[29] G. E. Bodeker, H. Shiona, and H. Eskes, "Indicators of Antarctic ozone depletion," *Atmospheric Chem. Phys.*, vol. 5, pp. 2603–2615, 2005.

[30] Y. Shan, Q. Huang, D. Guan, and K. Hubacek, "China CO_2 emission accounts 2016–2017," *Sci. Data*, vol. 7, 2020, Art. no. 54.

[31] D. Munasinghe, S. Cohen, and K. Gadiraju, "A review of satellite remote sensing techniques of river delta morphology change," *Remote Sens. Earth Syst. Sci.*, vol. 4, no. 1, pp. 44–75, 2021, doi: 10.1007/s41976-021-00044-3.

[32] Y. Sun, "A study of carbon verification methods in China," vol. 80, pp. 19–89, BUCT, 2018.

[33] Y. LiU, L. Yao, and J. Wang, "The current status of application of carbon satellite data in China," *Satell. Appl.*, pp. 46–50, 2022.

[34] H. Zhong, "Study on the correlation analysis between the change of atmospheric CO_2 concentration by satellite remote sensing and anthropogenic emission," *CAS*, 2021, Art. no. 79.

[35] H. H. Aumann et al., "AIRS/AMSU/HSB on the aqua mission: Design, science objectives, data products, and processing systems," *IEEE Trans Geosci Remote Sens*, vol. 41, no. 2, pp. 253–264, Feb. 2003.

[36] P. E. Dennison et al., "High spatial resolution mapping of elevated atmospheric carbon dioxide using airborne imaging spectroscopy: Radiative transfer modeling and power plant plume detection," *Remote Sens. Environ.*, 139, pp. 116–129, 2013, doi: 10.1016/j.rse.2013.08.001.

[37] H. Suto et al., "Thermal and near-infrared sensor for carbon observation fourier transform spectrometer-2 (TANSO-FTS-2) on the greenhouse gases observing SATellite-2 (GOSAT-2) during its first year in orbit," *Atmospheric Meas. Techn.*, vol. 14, no. 3, pp. 2013–2039, 2021, doi: 10.5194/amt-14-2013-2021.

[38] R. Frankenber et al., "The Orbiting Carbon Observatory (OCO-2): Spectrometer performance evaluation using pre-launch direct sun measurements," *Atmospheric Meas. Techn.*, vol. 8, pp. 301–313, 2015.

[39] W. Jiahui, L. Liang, L. Han, C. Chunyang, and G. Di, "Interpretation of the report on temporal dynamics and spatial distribution of global carbon source and sink," in *Proc. 8th Int. Conf. Agro-Geoinformatics*, 2019, pp. 1–4.

[40] Y. Oishi, H. Ishida, and Y. Nakajima, "The impact of different support vectors on GOSAT-2 CAI-2 L2 cloud discrimination," *Remote Sens.*, vol. 9, 2017, Art. no. 12362017.

[41] W. V. Kim et al., "Retrieving XCO_2 from GOSAT FTS over East Asia using simultaneous aerosol information from CAI," *Remote Sens.*, vol. 8, 2016, Art. no. 994.

[42] T. E. Taylor et al., "OCO-3 early mission operations and initial (vEarly) XCO_2 and SIF retrievals," *Remote Sens. Environ.: Interdiscipl. J.*, vol. 251, 2020, Art. no. 112032.

[43] C. W. O'Dell et al., "Improved retrievals of carbon dioxide from orbiting carbon observatory-2 with the version 8 ACOS algorithm," *Atmospheric Meas. Techn.*, vol. 11, 2018, Art. no. 6539.

[44] J. Hakkarainen, I. Ialongo, S. Maksyutov, and D. Crisp, "Analysis of four years of global XCO_2 anomalies as seen by orbiting carbon observatory-2," *Remote Sens.*, vol. 11, 2019, Art. no. 850.

[45] T. E. Taylor, C. W. O'Dell, C. Frankenberg, P. T. Partain, and M. Gunson, "Orbiting carbon Observatory-2 (OCO-2) cloud screening algorithms; Validation against collocated MODIS and CALIOP data," *Atmospheric Meas. Techn.*, vol. 9, no. 12, pp. 12663–12707, 2015.

[46] C. W. O'Dell et al., "The ACOS CO_2 retrieval algorithm – Part 1: Description and validation against synthetic observations," *Atmospheric Meas. Techn.*, vol. 5, pp. 99–121, 2012, doi: 10.5194/amt-5-99-2012.

[47] C. Frankenberg, C. W. O'Dell, L. Guanter, and J. McDuffie, "Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: Implications for its retrieval and interferences with atmospheric CO_2 retrievals," *Atmospheric Meas. Techn.*, vol. 5, pp. 2081–2094, 2012.

[48] F.-M. Bréon, L. David, P. Chatelanaz, and F. Chevallier, "On the potential of a neural network-based approach for estimating XCO_2 from OCO-2 measurements," *Atmospheric Meas. Techn.*, vol. 15, pp. 5219–5234, 2021.

[49] H. Lindqvist et al., "Does GOSAT capture the true seasonal cycle of carbon dioxide?," *Atmospheric Chem. Phys.*, vol. 15, no. 22, pp. 13023–13040, 2015.

[50] A. Chatterjee et al., "Influence of El Niño on atmospheric CO_2 over the tropical pacific ocean: Findings from NASA's OCO-2 mission," *Science*, vol. 358, 2017, Art. no. 466.

[51] P. I. Palmer, L. Feng, D. F. Baker, F. Chevallier, H. Bösch, and P. Somkuti, "Net carbon emissions from African biosphere dominate pan-tropical atmospheric CO_2 signal," *Nature Commun.*, vol. 10, 2019, Art. no. 3344.

[52] T. E. Taylor et al., "An eleven year record of XCO_2 estimates derived from GOSAT measurements using the NASA ACOS version 9 retrieval algorithm," *Earth Syst. Sci. Data Discuss.*, vol. 2021, pp. 1–53, 2021.

[53] S. Crowell et al., "The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network," *Atmospheric Chem. Phys.*, vol. 19, pp. 9797–9831, 2019.

[54] H. Peiro et al., "Four years of global carbon cycle observed from the orbiting carbon observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7," *Atmospheric Chem. Phys.*, vol. 22, pp. 1097–1130, 2022.

[55] F. Chevallier, G. Broquet, B. Zheng, P. Ciais, and A. Eldering, "Large CO_2 emitters as seen from satellite: Comparison to a gridded global emission inventory," *Geophysical Res. Lett.*, vol. 49, 2022, Art. no. e2021GL097540.

[56] C. Chen and L. J. Leydesdorff, "Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis," *Annu. Rev. Inf. Sci. Technol.*, vol. 65, no. 2, pp. 334–351, 2013.

[57] J. Hou, X. Yang, and C. Chen, "Emerging trends and new developments in information science: A document co-citation analysis (2009–2016)," *Scientometrics*, vol. 115, pp. 869–892, 2018.

[58] X. Wang et al., "An application of remote sensing data in mapping landscape-level forest biomass for monitoring the effectiveness of forest policies in Northeastern China," *Environ. Manage.*, vol. 52, no. 3, pp. 612–620, 2013.

[59] A. Eldering et al., "The orbiting carbon Observatory-2 early science investigations of regional carbon dioxide fluxes," *Science*, vol. 358, 2017, Art. no. 5745.

[60] K. Gui et al., "Satellite-derived PM 2.5 concentration trends over eastern China from 1998 to 2016: Relationships to emissions and meteorological parameters," *Environ. Pollut.*, vol. 247, pp. 1125–1133, 2019.

[61] L. Lai, "The carbon emission effect of land use in China," *NJU*, vol. 157, pp. 1–157, 2010.

[62] Y. Yang et al., "Ultrasound-boosted selectivity of CO in CO_2 electrochemical reduction," *Ultrasonics Sonochemistry*, vol. 76, no. 17, 2021, Art. no. 105623, doi: 10.1016/j.ultsonch.2021.105623.

[63] I. A. Digdaya et al., "A direct coupled electrochemical system for capture and conversion of CO_2 from oceanwater," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 4412.

[64] P. Liu, J. Guo, H. G. Im, and W. Roberts, "The effects of CO_2/CH_4 ratio on soot formation for autothermal reforming of methane at elevated pressure," *Combustion Flame*, vol. 258, 2022, Art. no. 112379, doi: 10.1016/j.combustflame.2022.112379.

[65] P. O. Wennberg et al., "The total carbon column observing network (TCCON)," in *Proc. AGU Fall Meeting Abstr.*, 2005, pp. 2087–2122.

[66] W. Brault, "New approach to high-precision fourier transform spectrometer design," *Appl. Opt.*, vol. 35, no. 16, pp. 2891–2896, 1996, doi: 10.1364/AO.35.002891.

[67] R. A. Washenfelder et al., "Carbon dioxide column abundances at the Wisconsin tall tower site," *J. Geophysical Res.*, vol. 111, 2006, Art. no. D22305.

[68] M. M. D. S. Pupo and C. Kortlever, "Electrolyte effects on the electrochemical reduction of CO_2," *ChemPhysChem*, vol. 20, pp. 2926–2935, 2019, doi: 10.1002/cphc.201900680.

[69] J. Hedelius et al., "Intercomparability of XCO_2 and XCH4 from the United States TCCON sites," *Atmospheric Meas. Techn.*, vol. 10, pp. 1481–1493, 2017, doi: 10.5194/amt-10-1481-2017.

[70] C. Liu, A. Pan, and D. Xie, "Policy suggestions for improving enterprise carbon emission accounting system in China," *Economy*, vol. 11, pp. 42–45, 2014.

[71] S. Zheng, X. Zhang, and H. Liu, "A few thoughts on the construction of MRV management mechanism in China's carbon market," *China Econ. Trade J.*, vol. 2, pp. 9–10, 2016.

[72] L. Duan, W. Wei, and F. Tang, "Analysis of greenhouse gas emissions of electrolytic alumni enterprises," *Energy Efficiency*, vol. 38, no. 2, pp. 171–172, 2019.

[73] Y. He et al., "The construction of the basic framework of China's carbon emission statistics and accounting system," *Statist. Inf. Forum*, vol. 30, pp. 30–36, 2015.

[74] A. O. Acheampong and P. Boateng, "Modelling carbon emission intensity: Application of artificial neural network," *J. Cleaner Prod.*, vol. 10, pp. 833–856, 2019.

[75] N. Ahmad, L. Du, J. Lu, J. Wang, H.-Z. Li, and M. Hashmi, "Modelling the CO_2 emissions and economic growth in Croatia: Is there any environmental Kuznets curve?," *Energy*, vol. 123, pp. 164–172, 2017.

[76] X. Tan, Y. Zou, Z. Guo, K. Zhou, and Q. Yuan, "Deep contrastive self-supervised hashing for remote sensing image retrieval," *Remote Sens.*, vol. 14, 2022, Art. no. 3643, doi: 10.3390/rs14153643, 2022.

[77] J. Heymann et al., "SCIAMACHY WFM-DOAS XCO_2: Comparison with CarbonTracker XCO_2 focusing on aerosols and thin clouds," *Atmospheric Meas. Techn.*, vol. 5, no. 8, pp. 1935–1952, 2012, doi: 10.5194/amt-5-1935-2012.

[78] W. Peters et al., "An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker," *Proc. Nat. Acad. Sci.*, vol. 104, pp. 18925–18930, 2007.

[79] T. Krings et al., "MAMAP – A new spectrometer system for column-averaged methane and carbon dioxide observations from aircraft: Retrieval algorithm and first retrievals for point source emission rates," *Atmospheric Meas. Techn.*, vol. 4, no. 9, pp. 1735–1758, 2011, doi: 10.5194/amt-4-1735-2011.

[80] A. Butz, O. P. Hasekamp, C. Frankenberg, and I. Aben, "Retrievals of atmospheric CO_2 from simulated space-borne measurements of backscattered near-infrared sunlight: Accounting for aerosol effects," *Appl. Opt.*, vol. 48, no. 18, pp. 3322–3336, 2009.

[81] A. Butz et al., "Toward accurate CO_2 and CH_4 observations from GOSAT," *Geophysical Res. Lett.*, vol. 38, 2011, Art. no. L14812.

[82] A. Butz, O. Hasekamp, C. Frankenberg, J. Vidot, and I. Aben, "CH4 retrievals from space-based solar backscatter measurements: Performance evaluation against simulated aerosol and cirrus loaded scenes," *J. Geophys. Res.*, vol. 115, 2010, Art. no. D24302, doi: 10.1029/2010JD014514.

[83] Y. Yoshida et al., "Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the greenhouse gases observing satellite," *Atmospheric Meas. Techn. Discuss.*, vol. 4, no. 6, pp. 4791–4833, 2011.

[84] Y. Yoshida et al., "Improvement of the retrieval algorithm for GOSAT SWIR XCO_2 and XCH4 and their validation using TCCON data," *Atmospheric Meas. Techn.*, vol. 6, pp. 1533–1547, 2013, doi: 10.5194/amt-6-1533-2013.

[85] M. Reuter, M. Buchwitz, O. Schneising, J. Heymann, H. Bovensmann, and J. P. Burrows, "A method for improved SCIAMACHY CO_2 retrieval in the presence of optically thin clouds," *Atmospheric Meas. Techn.*, vol. 3, pp. 209–232, 2010.

[86] B. Dils et al., "The greenhouse gas climate change initiative (GHG-CCI): Comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH4 retrieval algorithm products with measurements from the TCCON," *Atmospheric Meas. Techn.*, vol. 7, pp. 1723–1744, 2014, doi: 10.5194/amt-7-1723-2014.

[87] S. Karbasi, H. Malakooti, M. Rahnama, and M. Azadi, "Study of mid-latitude retrieval XCO_2 greenhouse gas: Validation of satellite-based shortwave infrared spectroscopy with ground-based TCCON observations," *Sci. Total Environ.*, vol. 836, 2022, Art. no. 155513, doi: 10.1016/j.scitotenv.2022.155513.

[88] M. Zhang, Z. Wang, and C. Liu, "A few thoughts on fixed source flue gas comparison monitoring," *Chin. J. Environ. Manage.*, vol. 24, pp. 56–58, 2014.

[89] B. J. Connor, V. Sherlock, G. Toon, D. Wunch, and P. O. Wennberg, "GFIT2: An experimental algorithm for vertical profile retrieval from near-IR spectra," *Atmospheric Meas. Techn.*, vol. 9, no. 8, pp. 3513–3525, 2016, doi: 10.5194/amt-9-3513-2016.

[90] D. Liu, Y. Huang, Z. Cao, X. Lu, and X. Liu, "The influence of instrumental line shape degradation on gas retrievals and observation of greenhouse gases in Maoming, China," *Atmosphere*, vol. 12, 2021, Art. no. 863.

[91] F. Hase et al., "Application of portable FTIR spectrometers for detecting greenhouse gas emissions of the major city Berlin," *Atmospheric Meas. Techn.*, vol. 8, no. 7, pp. 3059–3068, 2015, doi: 10.5194/amt-8-3059-2015.

[92] C. Gallo, F. Contó, and M. Fiore, "A neural network model for forecasting CO_2 emission," *AGRIS On-Line Papers Econ. Informat.*, vol. 6, pp. 31–36, 2014.

[93] H. Pao, H.-C. Fu, and E. Tseng, "Forecasting of CO_2 emissions, energy consumption and economic growth in China using an improved Grey model," *Energy*, vol. 40, pp. 400–409, 2012.

[94] N. Ahmad, L. Du, J. Lu, J. Wang, H.-Z. Li, and E. Hashmi, "Modelling the CO_2 emissions and economic growth in Croatia: Is there any environmental Kuznets curve?," *Energy*, vol. 123, pp. 164–172, 2017.

[95] C. Gallo, F. Contó, and M. Fiore, "A neural network model for forecasting CO_2 emission," *Agris On-Line Papers Econ. Informat.*, vol. 6, pp. 31–36, 2014.

[96] J. Shah, D. Vaidya, and A. Shah, "A comprehensive review on multiple hybrid deep learning approaches for stock prediction," *Intell. Syst. Appl.*, vol. 65, 2022, Art. no. 200111.

[97] J. Zheng, "Research on the time-series data processing and prediction model of atmospheric CO_2," Anhui University of Technology, Ma'anshan, China, 2020.

[98] J. Zhou, R. Fang, Y. Li, Y. Zhang, and B. Peng, "Parameter optimization of nonlinear grey bernoulli model using particle swarm optimization," *Appl. Math. Comput.*, vol. 207, pp. 292–299, 2009.

[99] D. Y. Zhu, S. B. Ding, and X. Y. Xu, "Grey dynamic model of predicting the long-term effects of drugs and its application to pirenzepine," *Zhongguo Yao Li Xue Bao*, vol. 11, no. 6, pp. 481–484, 1990.

[100] S. Liu and J. Forrest, "Grey models for decision making," in *Grey Systems*. Berlin, Germany: Springer, 2010, pp. 197–223, doi: 10.1007/978-3-642-16158-2_7.

[101] H. Pao, C.-M. Tsai, and E. Abstracts, "Modeling and forecasting the CO_2 emissions, energy consumption, and economic growth in Brazil," *Fuel Energy Abstr.*, vol. 36, 2011, Art. no. 5.

[102] C.-I Chen, H. L. Chen, S. Chen, and N. Simulation, "Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear grey Bernoulli model NGBM(1, 1)," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 13, pp. 1194–1204, 2008.

[103] L. Falat and L. Pancikova, "Quantitative modelling in economics with advanced artificial neural networks," *Procedia Econ. Finance*, vol. 34, pp. 194–201, 2015, doi: 10.1016/S2212-5671(15)01619-6.

[104] S. H. Aljahdali, F. Sheta, and C. Debnath, "Estimating software effort and function point using regression, support vector machine and artificial neural networks models," in *Proc. IEEE/ACS 12th Int. Conf. Comput. Syst. Appl.*, 2015, pp. 1–8.

[105] V. S. Dave and R. Dutta, "Neural network based models for software effort estimation: A review," *Artif. Intell. Rev.*, vol. 42, pp. 295–307, 2012.

[106] K. Hu, C. Weng, C. Shen, T. Wang, L. Weng, and M. Xia, "A multi-stage underwater image aesthetic enhancement algorithm based on a generative adversarial network," *Eng. Appl. Artif. Intell.*, vol. 123, 2023, Art. no. 106196, doi: 10.1016/j.engappai.2023.106196.

[107] L. Jiang, M. Luo, and T. Sun, "The environmental impact factors of CO_2 fluxes in urban woodlands using neural network analysis," *Guangxi Phys.*, vol. 33, pp. 19–21, 2012.

[108] E. Shabani, B. Hayati, E. Pishbahar, M. A. Ghorbani, and P. Ghahremanzadeh, "A novel approach to predict CO_2 emission in the agriculture sector of Iran based on inclusive multiple model," *J. Cleaner Prod.*, vol. 279, 2021, Art. no. 123708.

[109] H. S. H. Jassim, W. Lu, and P. Olofsson, "Assessing energy consumption and carbon dioxide emissions of off-highway trucks in earthwork operations: An artificial neural network model," *J. Cleaner Prod.*, vol. 198, pp. 364–380, 2018.

[110] Z. Wang, G. S. Ma, D. Y. Gong, J. Sun, and L. Zhang, "Application of mind evolutionary algorithm and artificial neural networks for prediction of profile and flatness in hot strip rolling process," *Neural Process. Lett.*, vol. 50, pp. 2455–2479, 2019.

[111] N. Ali, M. E. H. Assad, H. F. Fard, B. Jourdehi, I. Mahariq, and M. A. Al-Shabi, "CO_2 emission modeling of countries in Southeast of Europe by using artificial neural network," *Proc. SPIE*, vol. 12120, 2022, Art. no. 121200F.

[112] J. Hu, Z. Luo, and F. Li, "Prediction of China's carbon emission intensity under "carbon peaking" target: Analysis based on LSTM and ARIMA-BP models," *Financial Sci.*, vol. 2, pp. 89–101, 2022.

[113] C. C. Aggarwal, *Data Mining: The Textbook*. New York, NY, USA: Springer, 2015.

[114] H. K. Cigizoglu and M. Alp, "Generalized regression neural network in modelling river sediment yield," *Adv. Eng. Softw.*, vol. 37, pp. 63–68, 2006.

[115] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.