Engineering 33 (2024) 263–275

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Climate Change—Article

Construction and Application of a Regional Kilometer-Scale Carbon Source and Sink Assimilation Inversion System (CCMVS-R)

Lifeng Guo [a], Xiaoye Zhang [b,a,*], Junting Zhong [a,*], Deying Wang [a], Changhong Miao [b], Licheng Zhao [c], Zijiang Zhou [c], Jie Liao [c], Bo Hu [d], Lingyun Zhu [e], Yan Chen [f]

[a] *Monitoring and Assessment Center for GHGs and Carbon Neutrality & Key Laboratory of Atmospheric Chemistry of CMA, Chinese Academy of Meteorological Sciences, Beijing 100081, China*
[b] *Laboratory of Climate Change Mitigation and Carbon Neutrality, Henan University, Zhengzhou 450001, China*
[c] *National Meteorological Information Center, Beijing 100081, China*
[d] *Shanxi Meteorological Administration, Taiyuan 030002, China*
[e] *Shanxi Provincial Institute of Meteorological Sciences, Taiyuan 030002, China*
[f] *Jiangsu Provincial Climate Center, Nanjing 210000, China*

ARTICLE INFO

Article history:
Received 31 August 2022
Revised 1 February 2023
Accepted 27 February 2023
Available online 14 June 2023

Keywords:
CCMVS-R
Regional carbon assimilation system
Anthropogenic carbon emissions
CO_2
POD 4DVar

ABSTRACT

CO_2 is one of the most important greenhouse gases (GHGs) in the earth's atmosphere. Since the industrial era, anthropogenic activities have emitted excessive quantities of GHGs into the atmosphere, resulting in climate warming since the 1950s and leading to an increased frequency of extreme weather and climate events. In 2020, China committed to striving for carbon neutrality by 2060. This commitment and China's consequent actions will result in significant changes in global and regional anthropogenic carbon emissions and therefore require timely, comprehensive, and objective monitoring and verification support (MVS) systems. The MVS approach relies on the top-down assimilation and inversion of atmospheric CO_2 concentrations, as recommended by the Intergovernmental Panel on Climate Change (IPCC) Inventory Guidelines in 2019. However, the regional high-resolution assimilation and inversion method is still in its initial stage of development. Here, we have constructed an inverse system for carbon sources and sinks at the kilometer level by coupling proper orthogonal decomposition (POD) with four-dimensional variational (4DVar) data assimilation based on the weather research and forecasting–greenhouse gas (WRF-GHG) model. Our China Carbon Monitoring and Verification Support at the Regional level (CCMVS-R) system can continuously assimilate information on atmospheric CO_2 and other related information and realize the inversion of regional and local anthropogenic carbon emissions and natural terrestrial ecosystem carbon exchange. Atmospheric CO_2 data were collected from six ground-based monitoring sites in Shanxi Province, China to verify the inversion effect of regional anthropogenic carbon emissions by setting ideal and real experiments using a two-layer nesting method (at 27 and 9 km). The uncertainty of the simulated atmospheric CO_2 decreased significantly, with a root-mean-square error of CO_2 concentration values between the ideal value and the simulated after assimilation was close to 0. The total anthropogenic carbon emissions in Shanxi Province in 2019 from the assimilated inversions were approximately 28.6% (17%–38%) higher than the mean of five emission inventories using the bottom-up method, showing that the top-down CCMVS-R system can obtain more comprehensive information on anthropogenic carbon emissions.

1. Introduction

Climate change, which is currently being discussed on an international scale, refers mainly to climate warming due to increased greenhouse gases (GHGs). Climate change, global warming, and the anthropogenic greenhouse effect are of great concern at home and abroad. The report by Working Group I of the Intergovernmental Panel on Climate Change (IPCC) known as the *Fifth Assessment Report* clearly states that the increase in GHGs (including carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O)) and changes in land use will cause a series of effects, including an

* Corresponding authors.
E-mail addresses: xiaoye@cma.gov.cn (X. Zhang), zhongjt@cma.gov.cn (J. Zhong).

https://doi.org/10.1016/j.eng.2023.02.017

increase in average temperature, a change in extreme temperature, increased frequency of catastrophic weather, an increase in precipitation at the latitude of 30°N, and a decrease in rainfall in the tropics, which have been occurring since the 1970s [1]. In addition, the melting of glaciers due to global warming has led to a rise in sea level, with a global average sea level rise of 0.19 m between 1901 and 2010 and an average growth rate of 1.7 mm·a^{-1}. CO_2 is the most important anthropogenic GHG in the earth's atmosphere. Due to the impact of human activities (e.g., fossil fuel consumption and cement production), the concentration of atmospheric CO_2 has increased from (289.9 ± 3.3) parts per million (ppm) in 1900 to (398.8 ± 7.3) ppm in 2010–2019—an increase of 38%, with rapid growth occurring in the last 50 years [2,3]. According to the Paris Agreement, in order to limit global warming to below 1.5 °C by 2100 [4], the IPCC's *Sixth Assessment Report (AR6)* recommends substantial and sustained reductions in anthropogenic CO_2 emissions. To limit global warming to 1.5, 1.7, or 2.0 °C, the IPCC's 6th Assessment Panel Working Group I estimates remaining carbon budgets of 140, 230, and 370 petagrams of carbon (PgC) from 1 January 2020, respectively. However, CO_2 emissions from human activities are still increasing rapidly, with emissions from human activities such as fossil fuel use, cement production, and land-use change reaching 10.30 PgC·a^{-1} from 2006 to 2015 [5,6]; moreover, anthropogenic emissions of CO_2 were (10.90 ± 0.90) PgC·a^{-1} from 2010 to 2019. The remaining carbon budget will be consumed within a few decades based on current estimates of anthropogenic emissions of approximately 10.00 PgC·a^{-1} [7]. These anthropogenic carbon emissions are distributed among the three components of the earth system, with 46% being trapped in the atmosphere ((5.10 ± 0.02) PgC·a^{-1}), 23% absorbed by the oceans ((2.50 ± 0.60) PgC·a^{-1}), and 31% absorbed by vegetation in terrestrial ecosystems ((3.40 ± 0.90) PgC·a^{-1}) [2].

In September 2020, the Chinese government announced that "China will increase its autonomous national contribution, adopt more vigorous policies and measures, strive to peak CO_2 emissions by 2030, and achieve carbon neutrality by 2060" [8–10]. This solemn commitment shows the world that China is determined to follow a green, low-carbon, high-quality development path. China's commitment to achieving carbon neutrality by 2060 is consistent with the Paris Agreement's 2 °C targets for achieving zero emissions [11]. The core of the IPCC's emissions inventory guidelines is the governance mechanism of monitoring, reporting, and verification (MRV)—an essential element of establishing an evaluation system for implementing policies and programs. One of these critical elements is verifying each country's carbon emissions every five years, starting in 2023. The two main methods for estimating anthropogenic carbon emissions are known as the "bottom-up" and "top-down" methods. For example, global and national-scale CO_2 emission data are estimated by the Carbon Dioxide Information Analysis Center (CDIAC) under the US Department of Energy (1751–2006). Global CO_2 emission data are also collected by the International Energy Agency (IEA), whose database covers global CO_2 emissions from fossil fuels in more than 140 countries and regions (1971–2005), disaggregated by industry and fuel type.

However, the inventory survey method can generally be used to estimate only relatively static carbon emissions (usually on an annual scale). The statistical data on which this method is based are often incomplete or even erroneous, and there are significant differences and errors in the emission factors. Due to these limitations and errors, the emission fluxes estimated via the inventory survey method are subject to large uncertainties [2,6,7,12–16]. The availability of multisource CO_2 concentration observations—including high-precision ground-based observations and observations from the GHGs observing satellites (GOSAT [17] and GOSAT-2 [18]), the orbiting carbon observatories (OCO-2 [19,20] and OCO-3 [21]), and TanSat [22,23]—provides data support for studying global/regional carbon assimilation systems [24]. Therefore, top-down anthropogenic carbon flux estimation methods are increasingly favored by researchers [25–30]. CO_2 satellite observation assimilation will play an important role in global and regional carbon cycle studies and anthropogenic CO_2 accounting [31]. The European Space Agency (ESA) plans to use the same approach to estimate anthropogenic CO_2 emissions and to support a global inventory study in 2028 [32].

China has proposed reaching its carbon peak by 2030 and becoming carbon neutral by 2060 [8–10]. However, there is an urgent need to accurately calculate anthropogenic emissions from different regions in order to implement emission-reduction measures and assess whether the actions taken are appropriate and effective. To respond to the national carbon-neutral strategy, this paper develops a regional high-precision carbon assimilation system called the China Carbon Monitoring and Verification Support at the Regional level (CCMVS-R). In future research, we will introduce additional high-precision ground-based CO_2 concentration observation network data and multisource satellite observation data, which are expected to invert anthropogenic carbon emissions at the provincial/urban scale with high spatial and temporal resolutions. The structure of this paper is as follows: Section 1 provides an introduction; Section 2 introduces the materials and methods; Section 3 gives an analysis of the CCMVS-R ideal experiment and Shanxi inversion results; and Section 4 is the conclusion.

2. Materials and methods

2.1. Data information

2.1.1. Meteorological inputs

The boundary field weather-driven data used by the CCMVS-R system include National Centers for Environmental Prediction (NCEP) Final Operational Global (FNL) analysis data,[†] with a resolution of 1° × 1° and containing information on 26 standard isobars (10–1000 hPa) in the troposphere and the surface layer. Global data analysis is performed every 6 h at 00:00, 06:00, 12:00, and 18:00 Coordinated Universal Time (UTC).

2.1.2. Anthropogenic carbon flux data

Anthropogenic carbon flux data mainly include carbon emissions due to fossil fuel combustion and fire carbon flux, where fossil fuel combustion emissions are a direct human influence on the global carbon cycle. The anthropogenic carbon emissions selected for this paper are shown in Table 1 [6,7,12,33–36].

2.1.3. VPRM model-driven data

The Vegetation photosynthesis respiration model (VPRM) is a carbon flux estimation model based on remote-sensing data. The spatial information and meteorological field information (2 m temperature and downward shortwave radiation data) within the target area required for driving the model are provided by the mesoscale atmospheric transport model (weather research and forecasting (WRF)) [37]; the surface cover data required for model input are provided by high-precision, vegetation-type data (SYNMAP with 1 km resolution); and the enhanced vegetation index (EVI) and land surface water index (LSWI) required for model input are extracted from high-resolution remote-sensing data (MOD09A1)[‡]. The EVI, LSWI, and other parameter data correspond to the target area grid through the application of spatial interpolation methods.

[†] https://rda.ucar.edu/datasets/ds083.2/.

[‡] https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD09A1.

Table 1
Information on the CO_2 emission inventories.

No.	Emission inventory	Refs.
1	Emissions database for global atmospheric research (EDGAR)	[12,33]
2	Open-source data inventory for anthropogenic CO_2 (ODIAC)	[6,34]
3	Global carbon project (GCP)	[7]
4	Global infrastructure emission database (GID)	[35]
5	China high-resolution emission gridded data (CHRED)	[36]

2.1.4. Initial and boundary CO_2 concentrations

The initial and boundary conditions for each type of CO_2 emission and concentration field (i.e., anthropogenic sources and background fields) are obtained from the global output product of the CarbonTracker system [38,39]. The CarbonTracker system data product has a spatial resolution of $1° \times 1°$ in China, a 34-layer vertical division, and a temporal resolution of 3 h. The CO_2 concentration fields included in this product include CO_2 total concentrations, biogenic combustion sources, and fossil fuel source concentrations.

2.1.5. CO_2 concentration observation data

The observation data in this paper comprise pseudo-observation data and actual observation data. Among these, the pseudo-observations are the CO_2 concentrations simulated by the WRF-GHG model [40], based on the ideal anthropogenic carbon emissions and then extracted as the observed values to participate in the assimilation. The actual observation data are from six high-precision ground stations in Shanxi Province (Table S1 in Appendix A), which are used for the assimilation inversion and analysis of anthropogenic carbon emissions in Shanxi. Strict quality control of CO_2 observations is required before they enter the assimilation system. The current trace-gas analyzer is based on Picarro, which can theoretically achieve an observation accuracy of 0.1 ppm using the calibration of multiple known concentrations of a standard gas; however, there are various problems in actual observations, such as missing standard gas concentration offset, abnormal sample gas offset, and inappropriate water vapor removal, which affect the observation quality. Thus, strict and uniform data quality control is needed to ensure reliable and accurate observation data. This study uses an integrated quality control, screening, and fitting system to process CO_2 observations from six ground-based stations in Shanxi. Finally, we produced a global ground-based high-precision CO_2 concentration dataset with complete quality control and uniform standards.

2.2. Introduction to the WRF-GHG model

The CCMVS-R was developed based on the WRF-GHG model (V3.9.1), which is based on the mesoscale weather forecasting model WRF and mainly simulates the GHGs of CO_2, carbon monoxide (CO), and CH_4. The WRF model was developed by the National Center for Atmospheric Research (NCAR) in the 1990s. The WRF-GHG model can simulate the exchange of GHGs between the atmosphere and terrestrial ecosystems, taking into account the effects of atmospheric dispersion and transport on GHGs, and can simulate and predict the distribution of GHGs in time and space. WRF-GHG was originally developed to simulate atmospheric GHGs (CO_2, CH_4, and CO) in the Amazon Basin in the Balanço Atmosférico Regional de Carbono na Amazônia (BARCA) project. For CO_2, it is a direct dynamic coupling of the WRF model and the VPRM model, while the VPRM is mainly used to simulate the high-resolution spatial and temporal distribution of natural carbon flux. Structurally, for the CO_2 module, WRF-GHG was developed based on the WRF model with chemistry (WRF-Chem) by adding the corresponding input and output variables (ebio_gee, ebio_res, e_co2, co2_ant, co2_bio, co2_oce, co2_bck, etc.) to the registry.ghg registry and adding module_ghg_fluxes.F and other GHG modules.

The various source concentration variables CO_2_ANT, CO_2_BIO, CO_2_OCE, and CO_2_BCK are shown in Table 2 and represent CO_2 concentration variations from anthropogenic sources, CO_2 concentration variations from biological activities, CO_2 concentration variations from ocean fluxes, and atmospheric background concentration values, respectively. The CO_2 flux source in question can not only be obtained from an external file—that is, called offline (ext.)—but can also be calculated online in mode (int.). The time step can be set in half-hourly, daily, or WRF time steps, all of which can be controlled via the namelist.input file.

Due to space limitations, the CCMVS-R regional high-precision carbon assimilation system is optimized for natural carbon flux simulation by means of the VPRM ecological diagnostic model, which will be described in detail in other research results. The natural carbon flux is the result of the optimization; therefore, in this study, the natural carbon flux is considered to have no error, and only the anthropogenic carbon emissions are assimilated and inverted.

2.3. The CCMVS-R building process

The CCMVS-R was mainly developed using Fortran and Python in order to realize multicore parallel computing. Many researchers favor the Fortran language as a scientific computing language, while Python is an object-oriented interpreted computer programming language with powerful and rich computational libraries. The CCMVS-R building process consists of four main parts.

(1) Monte Carlo methods are run to generate ensemble perturbations, and then proper orthogonal decomposition (POD) and four-dimensional variational (4DVar) partitioning are applied to compute the eigenvalues and eigenvectors of the ensemble.

(2) The obtained basis vector (i.e., anthropogenic carbon emissions) and the mean state vector replace the initial anthropogenic carbon emissions of the WRF-GHG model, and the parallel pooling task is submitted to the Linux system.

(3) A scripting program to dynamically listen to the ensemble tasks is run. Once all tasks are completed, the simulation results for each ensemble and its mean value are resolved, and the POD 4DVar method is applied to construct a cost function in conjunction with the CO_2 concentration observations.

(4) The *a posteriori* anthropogenic carbon emission data obtained by the cost function solution are used to update the anthropogenic carbon emission input of the WRF-GHG model, and the model reruns this assimilation window. The resulting CO_2 concentration product is the simulated value of the CO_2 concentration after anthropogenic carbon emission inventory optimization.

The regional carbon assimilation system runs in a cycle according to the above operation until all assimilation is finished, which is schematically shown in Fig. 1.

Table 2
CO_2 concentration variables defined in registry.ghg.

Concentration variable (ppm)	Variable description
CO_2_ANT	Changes in CO_2 concentration from anthropogenic emissions
CO_2_BIO	Changes in CO_2 concentration from biogenic activity
CO_2_OCE	Changes in CO_2 concentration from ocean fluxes
CO_2_BCK	Atmospheric CO_2 background concentration

2.4. Introduction of efficient assimilation methods for POD 4DVar

In this study, we applied POD to the 4DVar algorithm [41,42]. The CCMVS-R was constructed based on the WRF-GHG model, which largely reduces the computer resources required for assimilation inversion and improves the computational efficiency and estimation accuracy with a spatial resolution up to the kilometer level. POD 4DVar is a straightforward method that uses the Monte Carlo method to generate a 4D sample ensemble similar to the ensemble Kalman filtering method (EnKF), where the distribution is Gaussian with a mean of 0 and 30% uncertainty. Since the POD orthogonal basis has optimality in the least-squares sense, the POD method is applied to extract the orthogonal basis from the forecast ensemble in four dimensions, which can capture more information of the forecast ensemble and better characterize the spatial structure and temporal evolution of the 4D variables. This method does not require an integral adjoint model and is straightforward in operation and maintenance.

Since the traditional 4DVar data assimilation method requires tangent and adjoint linear operators, which are very difficult to obtain for nonlinear model operators, POD 4DVar can make the implicit optimization problem explicit by capturing the spatiotemporal evolutionary characteristics of the data with a set of basis vectors. This approach not only simplifies the data assimilation process but also retains two significant advantages of traditional 4DVar data assimilation: ① The physical model provides a strong dynamic constraint, and ② observations from multiple periods can be introduced simultaneously in a single assimilation window. The assimilation effect is pronounced when the POD 4DVar method is used for the data assimilation of models with significant initial field errors and forecasted model uncertainties. For conventional 4DVar data assimilation, the cost function (Eq. (1)) is constructed from the difference between the simulated values of the model and the observed values, and the *a posteriori* anthropogenic carbon emissions are obtained by deriving the cost function ($J\left(\vec{x_0}\right)$).

$$J\left(\vec{x_0}\right) = \left(\vec{x_0} - \vec{x_b}\right)^{\mathrm{T}} B^{-1}\left(\vec{x_0} - \vec{x_b}\right) + \sum_{j=1}^{S}\left[\vec{y_j} - H_j(\vec{x_j})\right]^{\mathrm{T}} R_j^{-1}\left[\vec{y_j} - H_j(\vec{x_j})\right] \tag{1}$$

$$\vec{x_j} = M_{t_0 \to t_j}(\vec{x_0}) \tag{2}$$

where the superscript T represents the matrix transpose, $\vec{x_b}$ is a background value, j represents the CO_2 concentration observation time series, H_j represents the observation operator, the $\vec{y_j}$ represents the observation at a series of times t_j, $j = 1, \ldots, S$, $M_{t_0 \to t_j}$ represents the forecast model, and the matrices $\boldsymbol{B}$ and $\boldsymbol{R}$ represent the flux background error and concentration observation error covariance, respectively. $\vec{x_0}$ represents the control variables at the initial moment of the assimilation window. In the cost function, the control variable $\vec{x_0}$ is connected with $\vec{x_j}$ through the forward integration of Eq. (2). It is clear that the cost function on the gradient value of $\vec{x_0}$ is difficult to obtain. However, the POD 4DVar method greatly simplifies its calculation. In the POD 4DVar assimilation algorithm, assuming that there are S time series in the assimilation window, N random perturbation fields are generated using the Monte Carlo method, and each perturbation field is added to the initial background field at $t = t_0$ to produce N initial fields $\vec{x_n}(t_0)$, n = 1, 2, …, N. The state variables of all sets within the used assimilation window ($\vec{X_n}$) are obtained by means of Eq. (3).

$$\vec{X_n} = \left(\vec{x_n}(t_0), \vec{x_n}(t_1), \ldots, \vec{x_n}(t_S)\right), n = 1, 2, \ldots, N \tag{3}$$

The next point is how to obtain the appropriate basis vectors. The average of the N set state vectors is represented as follows:

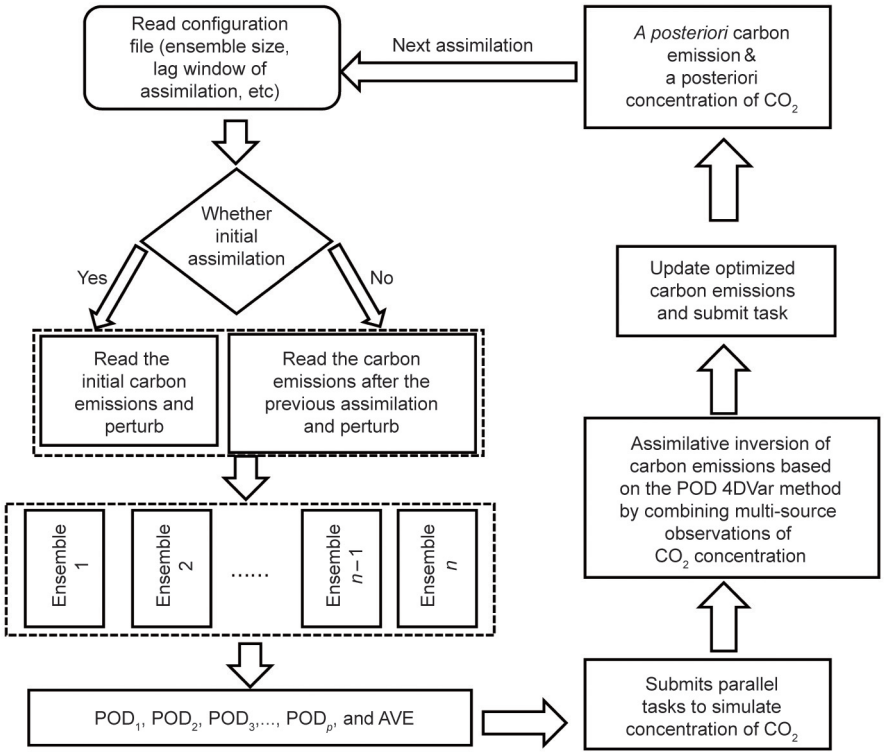

Fig. 1. Schematic diagram of the CCMVS-R construction. AVE represents the average value of the ensemble; p: the number of POD modes.

$$\bar{X} = \frac{1}{N}\sum_{n=1}^{N}\overrightarrow{X_n} \tag{4}$$

$$\delta X_n = \vec{X}_n - \bar{X},\ n = 1,\ \ldots,\ N \tag{5}$$

In the construction of the matrix $\boldsymbol{A}(M \times N) = (\delta X_1,\ \ldots,\ \delta X_n)$, $M = M_g \times M_v \times (S+1)$, where M_g and M_v denote the number of model space lattice points and the number of model variables, respectively (here, only anthropogenic carbon emissions are represented). To calculate the POD modes, it is necessary to solve the eigenvalue problem.

$$(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})_{M\times M}V = \lambda V \tag{6}$$

In practice, $M >> N$; therefore, to efficiently and conveniently solve the problem, the problem is transformed into an $N \times N$ eigenvalue problem, as follows:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}V = \lambda V$$

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}V = \boldsymbol{A}\lambda V$$

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}V) = \lambda(\boldsymbol{A}V) \tag{7}$$

This translates into the problem of finding the eigenvalues of an $N \times N$ matrix, and the computational efficiency is greatly improved, which can meet the need for efficient computation of regional assimilation systems with large dimensional characteristics and high spatial and temporal resolution.

$$T^{*}V_n = \lambda_n V_n,\ n = 1,2,\ \ldots,\ N \tag{8}$$

where $T^{*} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})_{N\times N}$, V_n is the nth column of eigenvector V, and λ_n is the eigenvalue λ of the nth row. The nonzero eigenvalues λ_n $(1 \le n \le N)$ are selected for the orthonormal, and the POD pattern is given by $\phi_n = \frac{AV_n}{\sqrt{\lambda_n}}(1 \le n \le N)$. In the 4D space, analysis variable $\vec{X}_a$ can be rewritten in POD, which is represented as follows:

$$\vec{X}_a = \bar{X} + \sum_{n=1}^{p}\beta_n\phi_n \tag{9}$$

where p is the number of POD modes, the size of which is determined according to the following equation:

$$p = \min\left\{p, I(P) = \frac{\sum_{n=1}^{p}\lambda_n}{\sum_{n=1}^{N}\lambda_n} : I(P) \ge \gamma\right\} \tag{10}$$

where $0 < \gamma < 1$ denotes the percentage of the total information captured by the reduced dimensional space $D^p = \mathrm{span}\{\phi_1, \ldots, \phi_p\}$. To capture most information of the POD group, γ is allowed to take values of approximately 1. Reconstructing the cost function Eq. (1) according to Eq. (9), the control variable $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ replaces the state variable at the initial moment $\overrightarrow{X_0}$ so that the control variables are treated explicitly in the cost function, and the tangent operator and the adjoint mode are not needed to determine the minimum cost function. The problem of finding the minimum of the cost function is transformed into a general optimization algorithm to find β. However, the problem still requires an iterative process, and it is still computationally intensive to find β. This problem is solved according to the following flow. First, the POD model matrix is constructed as follows:

$$\boldsymbol{\Phi} = (\phi_1, \phi_2, \ldots, \phi_p) \tag{11}$$

where $\phi_n = (\phi_n(t_0), \phi_n(t_1), \ldots, \phi_n(t_S))^{\mathrm{T}}, n = 1,2,\ldots,p$. The transformation in Eq. (11) is given by

$$\boldsymbol{\Phi} = (\Phi_0, \Phi_1, \ldots, \Phi_S)^{\mathrm{T}} \tag{12}$$

where $\boldsymbol{\Phi}_j = (\phi_1(t_j), \phi_2(t_j), \ldots, \phi_p(t_j))$. Eq. (9) is rewritten as follows:

$$\vec{X}_a = \bar{X} + \Phi\beta \tag{13}$$

Rewriting the cost function using Eq. (13),

$$J(\beta) = (\bar{x}(t_0) + \Phi_0\beta - \vec{X}_b)^{\mathrm{T}}\boldsymbol{B}^{-1}(\bar{x}(t_0) + \Phi_0\beta - \vec{X}_b)$$

$$+\sum_{j=1}^{S}[\vec{y_j} - H_j\bar{x}(t_j) - H_j\Phi_j\beta]^{\mathrm{T}}\boldsymbol{R}_j^{-1}[\vec{y_j} - H_j\bar{x}(t_j) - H_j\Phi_j\beta] \tag{14}$$

where $\vec{y_j} = (y_{j1}, \ldots, y_{jm_j})$ is the CO_2 concentration observation vector, and m_j is the size of $\vec{y_j}$. The N sets of observations can be generated by perturbing the observations using a Gaussian distribution perturbation with an uncertainty based on the site observation accuracy, which is represented as follows:

$$Y_{i,j} = \vec{y}_j + \varepsilon_i,\ i = 1, \ldots, N \tag{15}$$

where $\varepsilon_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, \ldots, \varepsilon_{i,m_j})^{\mathrm{T}}$; the mean value of perturbation of the N sets is 0, which is represented by the matrix as $\boldsymbol{E}_j = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)$; and the observation error covariance matrix can be evaluated as follows:

$$\boldsymbol{R}_j = \frac{\boldsymbol{E}_j\boldsymbol{E}_j^{\mathrm{T}}}{N-1},\ j = 1, \ldots, S \tag{16}$$

The background error covariance $\boldsymbol{B}$ is constructed in a similar way to the observation error covariance $\boldsymbol{R}$.

Since $\boldsymbol{R}_j^{-1}$ is a symmetric matrix, the gradient of the cost function $(\nabla J(\beta))$ is easily obtained by a simple calculation:

$$\nabla J(\beta) = (\Phi_0)^{\mathrm{T}}\boldsymbol{B}^{-1}(\bar{x}(t_0) - \vec{x}_b + \Phi_0\beta) + \sum_{j=1}^{S} -[H_j\Phi_j]^{\mathrm{T}}\boldsymbol{R}_j^{-1}$$
$$\times [\vec{y_j} - H_j\bar{x}(t_j) - H_j\Phi_j\beta] \tag{17}$$

To find the minimum value of the cost function, it is necessary to calculate $\nabla J(\beta) = 0$; that is,

$$((\Phi_0)^{\mathrm{T}}\boldsymbol{B}^{-1}\Phi_0 + \sum_{j=1}^{S} -[H_j\Phi_j]^{\mathrm{T}}\boldsymbol{R}_j^{-1}[H_j\Phi_j])\beta$$

$$= \sum_{j=1}^{S}[H_j\Phi_j]^{\mathrm{T}}\boldsymbol{R}_j^{-1}\left[\vec{y_j} - H_j\bar{x}(t_j)\right] - (\Phi_0)^{\mathrm{T}}\boldsymbol{B}^{-1}(\bar{x}(t_0) - \vec{x_b}) \tag{18}$$

After the above series of transformations, Eq. (18) can be obtained directly without an iterative process, which greatly simplifies the computation and improves the efficiency of the assimilation inversion.

2.5. CCMVS-R experimental program settings

In this paper, the simulation effect of the system is verified by two sets of experimental schemes. Scheme 1 verifies the system's validity and convergence; therefore, the observation data are also assumed to be pseudo-observations. Scheme 2 is based on the high-precision ground-based observation data from Shanxi Province, and an assimilation inversion of the anthropogenic carbon emissions in Shanxi in 2019 is performed. The two experimental schemes maintain the same settings except for the different observation data and use two layers of regional nested simulations with resolutions of 27 and 9 km, in which the numbers of grid points in the x–y directions for the two domains are 64 × 56 and 64 × 91, respectively. The core nested area is selected from Shanxi, a province with large carbon emissions (Fig. 2), where the green circles represent six CO_2 concentration observation stations. The assimila-

tion inversion is only performed for the 9 km nested region, with the first layer providing the boundary conditions for this region.

In this experiment, we set the assimilation window size to eight days, the simulation time to nine days, and the spin-up to one day to make the meteorological field simulation results as stable as possible (Fig. 3). As a key parameter in the assimilation system, the size of the assimilation window must be set according to the actual conditions of the ground-based and satellite observations. The primary physical process parameterization schemes used in the WRF-GHG simulations include rapid radiative transfer model (RRTM) longwave radiation [43], Dudhia shortwave radiation [44], Yonsei University (YSU) boundary layer parameterization [45], the Noah land surface model (LSM) [46], and Kain–Fritsch cumulus parameterization [47]. These physical parameterization schemes have been verified by numerous experiments and have good simulation effects.

In this paper, 80 ensemble members are used for assimilation inversion. In theory, a sufficiently large number of samples can cover more information on state variables, and the assimilation effect should be better. However, considering the computational and time costs, the ensemble size should be reasonably allocated according to the computational resources. In addition, ten nodes with approximately 300 cores are involved in the assimilation inversion computation. The details of the Scheme 1 are outlined below.

(1) Assuming that the raw data of the Emissions Database for Global Atmospheric Research (EDGAR) are ideal carbon emissions, the CO_2 concentrations simulated based on this inventory are used as pseudo-observations to participate in the assimilation of the ideal experiment.

(2) An uncertainty of 500% is added to the original EDGAR inventory, FFpri = FF + FF×rand, as the initial emission data to participate in the assimilation (where the resulting negative values are taken as absolute values), where FFpri is the *a priori* anthropogenic carbon emissions, FF is the assumed ideal anthropogenic carbon emissions, and rand is the random disturbance factor.

(3) The CO_2 concentration and boundary field data are from the CarbonTracker system (1° × 1°).

(4) The VPRM mode input data are from MOD09A1 data (500 m, 8 d).

(5) The WRF meteorological boundary field data are from FNL data (2019, 1° × 1°).

(6) There are six pseudo-observation stations whose locations are at the six CO_2 concentration monitoring stations in Shanxi (Fig. 2). The red symbol marks the location of a hypothetical ideal validation station not involved in assimilation (Fig. 2).

The Scheme 2 is consistent with the ideal scenario setup in terms of meteorological driving data, boundary conditions, and initial conditions, but the observation data are from six high-precision ground stations in Shanxi Province, and the 2018 EDGAR emission inventory is used to provide the *a priori* anthropogenic carbon emissions to participate in the assimilation.

3. Results and discussion

3.1. Evaluating assimilation effects based on ideal experiments

The ideal experiments designed in this paper are all conducted in the model to examine whether the constructed CCMVS-R carbon

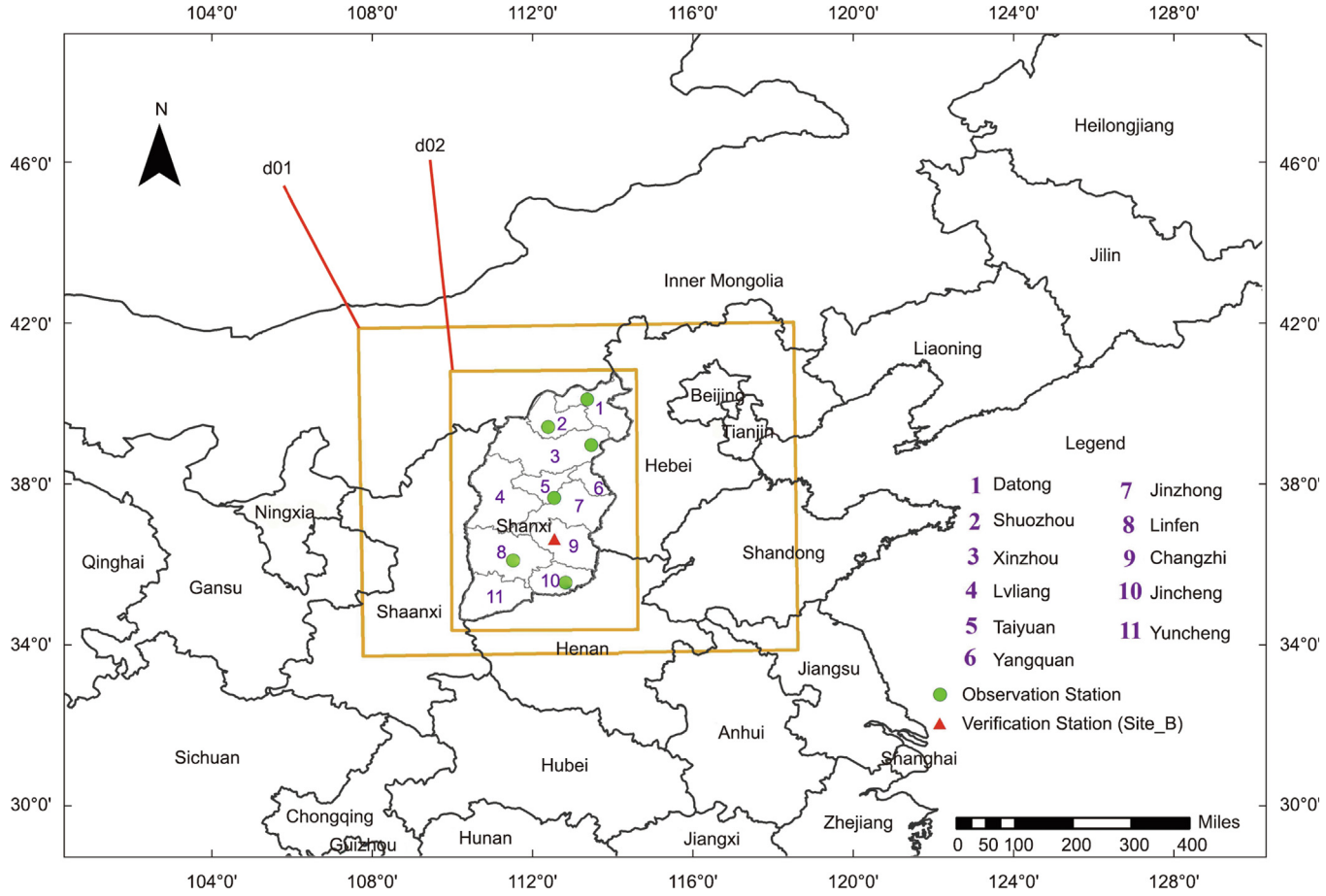

Fig. 2. Schematic diagram of the nested simulation area. The d01 represents the first layer, and the d02 represents the inner nested layer.

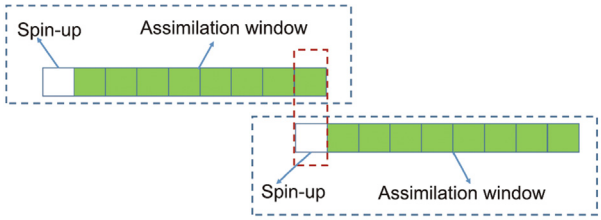

Fig. 3. Schematic diagram of the assimilation system operation cycle.

assimilation system can assimilate well to invert anthropogenic carbon emissions. Fig. 4(a) shows the assumed ideal anthropogenic carbon emissions using the 2018 EDGAR emission inventory. It is evident from the distribution that some regions, such as Taiyuan, Linfen, Changzhi, and Datong, have higher emissions, while other areas have lower emissions, which is closely related to the distribution of industrial and residential areas in Shanxi. Fig. 4(b) shows the distribution obtained by randomly assigning a 500% uncertainty perturbation to each grid based on ideal anthropogenic carbon emissions. From this, it can be seen that the whole distribution is significantly higher than the ideal anthropogenic carbon emissions, which we take as *a priori* anthropogenic carbon emissions for the analysis of the CCMVS-R assimilation effect. Fig. 4(c) shows the distribution of anthropogenic carbon emissions after assimilation. The spatial distribution tends to be relatively similar to the assumed ideal anthropogenic carbon emissions, and some high-value areas of anthropogenic carbon emissions before the assimilation are corrected. For the statistical analysis of the corresponding grid of ideal anthropogenic carbon emissions, *a priori* anthropogenic carbon emissions, and *a posteriori* anthropogenic carbon emissions, the correlation between *a priori* anthropogenic carbon emissions and ideal anthropogenic carbon emissions is 0.89 with a root-mean-square error (RMSE) of 4.89×10^4 $mol \cdot km^2 \cdot h^{-1}$. The correlation between the assimilated and ideal anthropogenic carbon emissions is improved to 0.94 with an RMSE of 5.18×10^3 $mol \cdot km^2 \cdot h^{-1}$. Although the correlation does not improve much, with an increase of 5.5% relative to the *a priori* anthropogenic carbon emissions, the RMSE decreases significantly, with a decrease of 89.4% relative to the *a priori* anthropogenic carbon emissions and a noticeable assimilation effect.

Fig. 5(a) shows a cumulative histogram of anthropogenic carbon emissions in Shanxi Province during one assimilation process; the ideal carbon flux is 5.24×10^6 $mol \cdot h^{-1}$, the *a priori* carbon flux after the perturbation is 1.82×10^7 $mol \cdot h^{-1}$, and the *a posteriori* anthropogenic carbon emissions after the assimilation are 5.97×10^6 $mol \cdot h^{-1}$. The *a priori* anthropogenic carbon emissions are 3.47 times the ideal anthropogenic carbon emissions, and the *a posteriori* anthropogenic carbon emissions are 1.14 times the ideal anthropogenic carbon emissions. Fig. 5(b) shows the difference between the anthropogenic carbon emissions and the corresponding grid points of the ideal anthropogenic carbon emissions before and after assimilation. The black line shows the difference between the *a priori* and ideal anthropogenic carbon emissions, while the red line is the difference between the *a posteriori* anthropogenic carbon emissions and ideal anthropogenic carbon emissions. It can be seen from the figure that the difference between the grid points after assimilation is mainly concentrated near the 0 axis line. The anthropogenic carbon emissions after assimilation have very obviously improved.

Figs. 4 and 5 display the assimilation effect in terms of the spatial distribution and emission statistics. In contrast, Fig. 6 analyzes the assimilation effect in terms of the difference between the simulated and pseudo-observed CO_2 concentrations before and after assimilation. The black line represents the pseudo-observed CO_2 concentrations, the red line represents the simulated CO_2 concentrations before assimilation, and the blue line represents the simulated CO_2 concentrations after assimilation, where Figs. 6(a) and (b) show the hourly average CO_2 concentrations at the validation station that did not participate in assimilation and at the Taiyuan station. The simulated values of CO_2 concentrations before and after assimilation at validation station Site B—which did not participate in the assimilation—agree with the pseudo-observation data, despite some deviations. Figs. 6(c) and (d) show the simulated hourly CO_2 concentrations in one assimilation process at the Taiyuan station and at the validation station without assimilation, which aligns with Figs. 6(a) and (b). The simulation effect is improved after assimilation, the correlation between the CO_2 concentration after assimilation and the ideal result at the two monitoring stations is close to 1, and the RMSE is close to 0. These results indicate that the CCMVS-R can capture the daily variation information of the stations well in the ideal experiment and is expected to assimilate the inverse regional anthropogenic carbon emissions with higher accuracy based on high-precision ground-based observation data.

3.2. Inverse anthropogenic CO_2 emissions at high resolution in Shanxi

Fig. 7 shows a comparison of 9 km *a posteriori* anthropogenic carbon emissions and *a priori* anthropogenic carbon emissions in Shanxi Province in 2019 by means of CCMVS-R inversion. This comparison indicates that the bottom-up statistical method may underestimate the emissions in this region. The overall distribution of anthropogenic carbon emissions in Shanxi is stronger in the central part and weaker in the eastern and western parts. In the central region, anthropogenic carbon emissions are linearly clustered in a north–south direction. Three emission centers with high intensity are located at the junction of Shuozhou–Datong, the south side of Taiyuan, and the junction of Linfen–Yuncheng, respectively.

The inversion results were compared with five sets of anthropogenic carbon emission inventory products—namely, global infrastructure emission database (GID), open-source data inventory for anthropogenic CO_2 (ODIAC), EDGAR, China high-resolution emission gridded data (CHRED), and global carbon project (GCP). The comparison results (Table 3) show that the statistical values of the *a posteriori* inventory in each city of Shanxi Province are closer to those of CHRED. However, in general, the *a posteriori* anthropogenic carbon emissions from the CCMVS-R assimilation inversion are approximately 28.6% (17%–38%) higher than those of the five inventory products, especially the emissions of the ODIAC anthropogenic carbon emission inventory product, which may be considerably underestimated.

Fig. 8 portrays a statistical histogram of the *a posteriori* anthropogenic carbon emissions and five sets of emission inventory products for prefecture-level cities in Shanxi, which shows that the region with the highest *a posteriori* anthropogenic carbon emissions is the city of Linfen. Taking Linfen and Datong—the two major carbon-emitting cities in Shanxi Province—as an example, the statistical analysis is based on the *a posteriori* emissions and the three inventory products GCP, GID, and ODIAC in 2019. The results show that the statistical value of the anthropogenic carbon emissions of the GID data product for Linfen is greater than that for Datong, which is consistent with the statistical results of the *a posteriori* emissions; however, for the other two sets of inventories (GCP and ODIAC), Datong has higher carbon emissions.

To verify the reasonableness of the *a posterior* anthropogenic carbon emissions and analyze the pollution concentration of CO_2's associated gas CO in the atmosphere, CO pollution concentration observations were obtained from the China National Environmental Monitoring Center[†], and the collected CO concentration observations from 6 + 6 monitoring stations within Datong and

[†] https://www.cnemc.cn/.

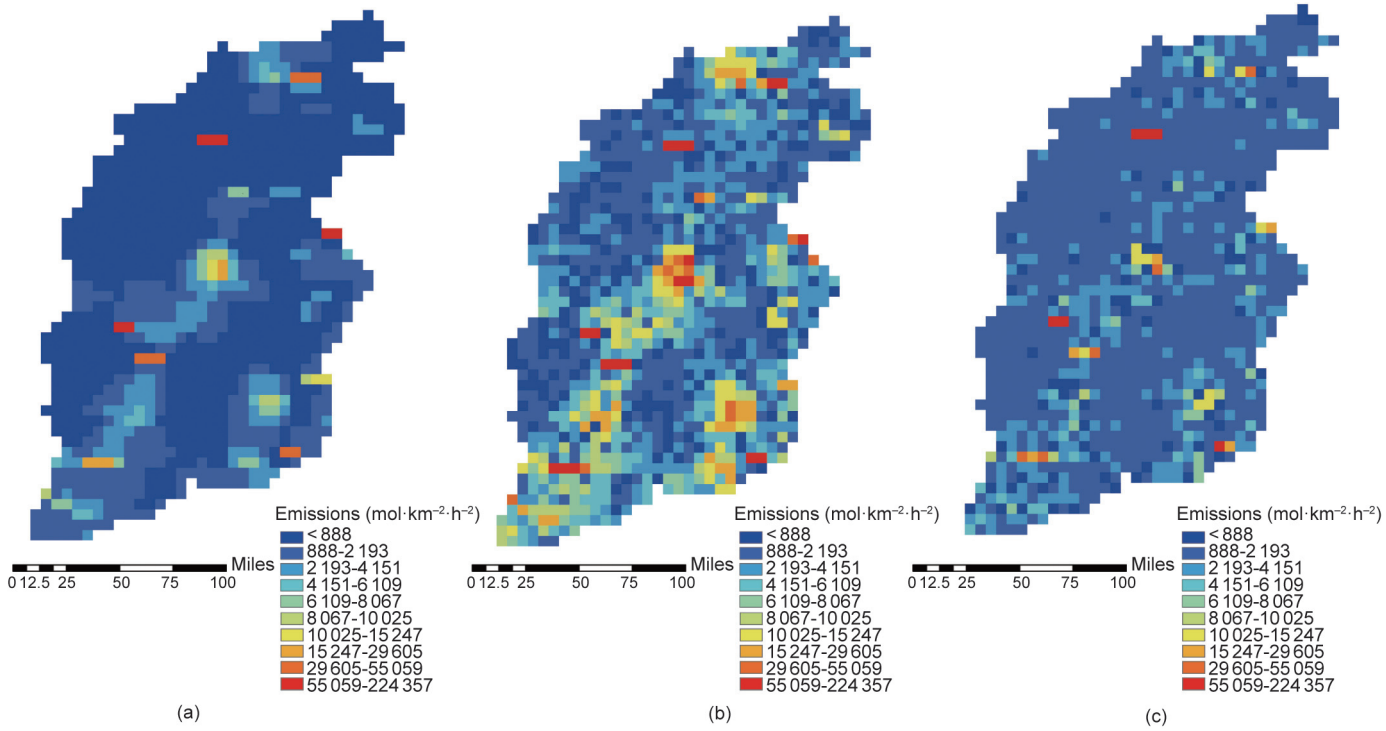

Fig. 4. Distribution of anthropogenic carbon emissions and ideal anthropogenic carbon emissions before and after assimilation. (a) Ideal anthropogenic carbon emissions; (b) *a priori* anthropogenic carbon emissions; (c) *a posteriori* anthropogenic carbon emissions.

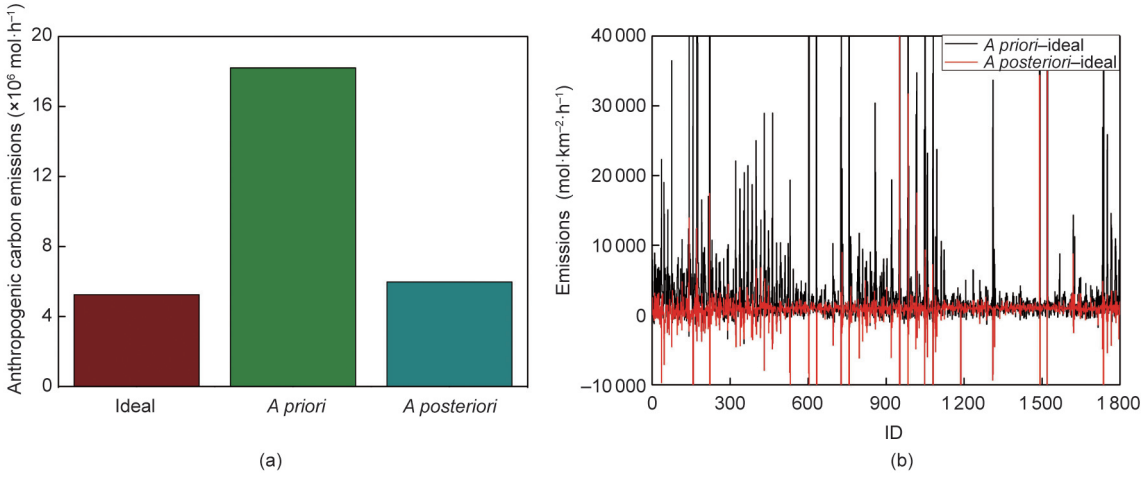

Fig. 5. Comparison of anthropogenic carbon emissions and ideal anthropogenic carbon emissions before and after assimilation in Shanxi Province. (a) Statistical bar graph; (b) difference line graph. ID: the number assigned to a grid point for anthropogenic emissions.

Linfen in 2019 were statistically analyzed by city where the stations are located, by month. The analysis was used to represent the average observed concentrations in Datong and Linfen. Fig. 9(a) shows the monthly average CO concentration observations in Datong (black line) and Linfen (red line). Except for March, May, and July, when those in Linfen are lower than those in Datong, the CO concentration observations in Linfen are similar to or significantly higher than those in Datong, especially in winter. Fig. 9(b) shows the annual average observation of CO concentrations in Datong and Linfen. Those in Linfen (red bar) are approximately 0.16 $mg \cdot m^{-3}$ higher than those in Datong (black bar), and the area of Linfen is approximately 1.43 times larger than that of Datong. These findings verify that the *a posteriori* anthropogenic carbon emissions in Linfen are higher than those in Datong, which is reasonable. ODIAC and GCP are inventory products with a time lag that use bottom-up statistical methods to determine the emission factors—methods that vary greatly and have large errors [48,49]. Thus, the obtained results may have a significant bias.

Fig. 10 shows scatter plots of the simulated CO_2 concentration of the CCMVS-R system before and after the assimilation of anthropogenic carbon emissions from 1 February to 30 April 2019 at the Datong and Taiyuan monitoring stations in Shanxi, compared with the daily average (24 h) of the observed values. The horizontal axis represents the observed values, while the vertical axis represents

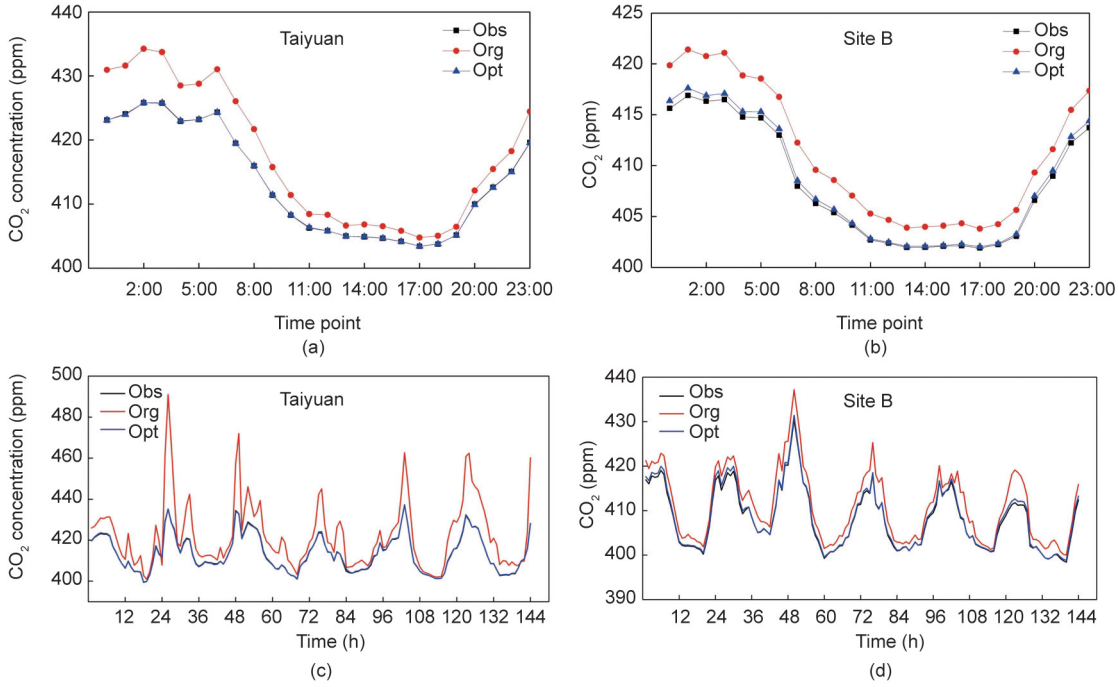

Fig. 6. Comparison analysis of simulated and pseudo-observed CO_2 concentrations before and after assimilation: (a, b) show the hourly average CO_2 concentrations of (a) Taiyuan and (b) Site B, while (c) and (d) show the simulated hourly CO_2 concentrations in one assimilation process of (c) Taiyuan and (d) Site B. Obs represents the pseudo-observed CO_2 concentrations, Org represents the simulated CO_2 concentrations before assimilation, and Opt represents the simulated CO_2 concentrations after assimilation.

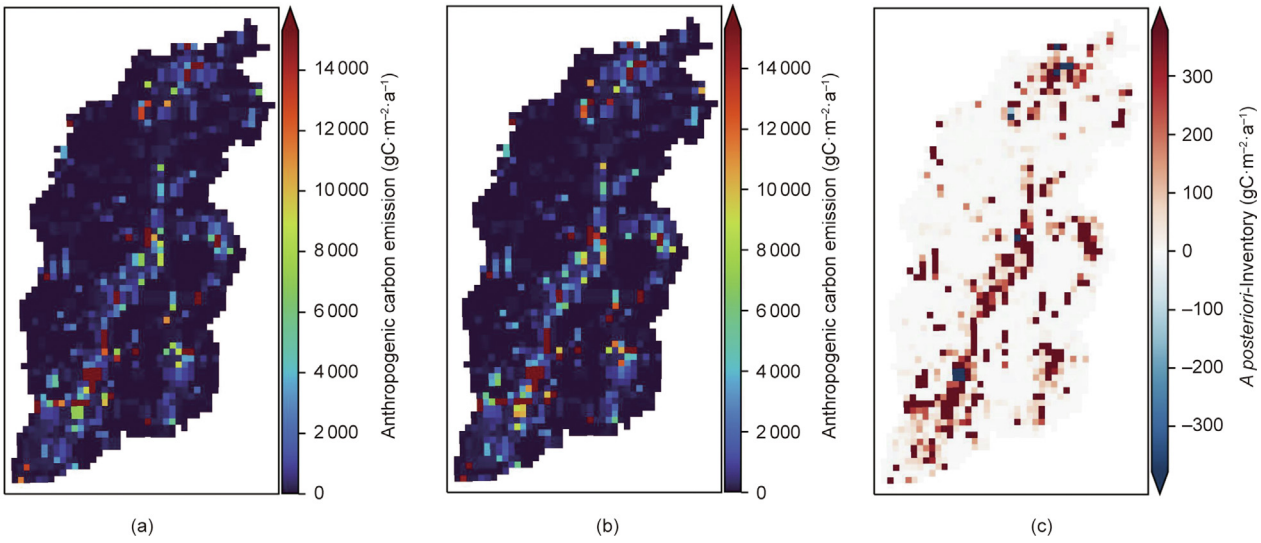

Fig. 7. Comparative analysis of anthropogenic carbon emissions (9 km) in Shanxi Province before and after assimilation. (a) *A priori* anthropogenic carbon emissions; (b) *a posteriori* anthropogenic emissions; and (c) differences between *a priori* and *a posteriori* anthropogenic carbon emissions.

the simulated values before and after the assimilation. The correlation coefficients between the simulated and observed CO_2 concentrations after assimilation by the CCMVS-R system at the Datong and Taiyuan monitoring stations are as high as 0.80 and 0.83, respectively, with mean deviations of −0.25 and 1.01 ppm. Compared with the correlation coefficients of 0.60 and 0.76 and mean deviations of −5.92 and −11.1 ppm at the Datong and Taiyuan monitoring stations before assimilation, there is a significant improvement. The simulation effect before the assimilation is similar to that reported in other studies, such as in the work of Diao et al. [50]. In their study using the WRF-GHG model without the assimilation of atmospheric CO_2 data, Diao et al. obtained a correlation coefficient of approximately 0.64 and an average deviation of approximately 6.87 ppm compared with the observed values. Those results showed that the emissions were closer to the actual emissions after the optimization of anthropogenic carbon emissions. The results provided in Fig. 10 show that the deviation between the system-simulated values and the observed values is

Table 3
Comparison analysis of the after-assimilation inventory and five inventory products.

City/Province	GCP	ODIAC	EDGAR	CHRED	GID
Datong	0.90	1.07	0.84	0.86	0.68
Taiyuan	0.87	0.77	0.70	1.04	0.78
Xinzhou	1.20	1.14	1.14	1.05	1.11
Jinzhong	0.69	0.45	0.65	0.63	0.65
Jincheng	1.72	1.07	1.43	0.80	1.58
Shuozhou	0.40	1.08	0.79	0.95	0.85
Yuncheng	0.60	0.39	0.56	0.79	0.63
Changji	1.48	0.80	1.10	0.75	0.89
Yangquan	0.58	0.71	0.86	0.82	0.58
Linfen	0.24	0.19	0.25	0.80	0.46
Lvliang	0.82	1.30	1.15	0.75	2.20
Shanxi	0.70	0.62	0.67	0.83	0.75
Mean	0.38	0.31	0.29	0.18	0.39

Value: inventory products divided by the post-assimilation inventory. Mean: AVERAGE(ABS(Value-1)).

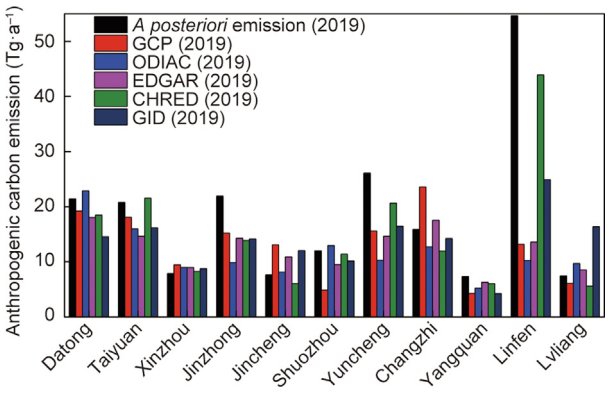

Fig. 8. Comparison analysis of *a posteriori* anthropogenic carbon emissions and inventory product statistics of prefecture-level cities in Shanxi Province.

larger for some dates (especially within the purple rectangle), which is strongly related to the system deviation, the observation error, and the difficulty of capturing small-scale anthropogenic carbon emission events. However, in general, the system-simulated values after the assimilation can capture well the daily averaged variation in CO_2 concentration information.

Fig. 11 shows the daily average time series of the CCMVS-R system's simulated values and station observations from 1 February to 30 April 2019 for the Datong and Taiyuan monitoring stations, with the horizontal axis representing the date and the vertical axis representing the CO_2 concentration values (ppm) of the assimilated CCMVS-R system's simulated values (red line) versus the observed values (black line). The line graph in Fig. 11 shows that the CCMVS-R system's simulated values can well capture the information on the average daily changes in CO_2 concentration. However, the simulation effect of the Taiyuan monitoring station is inferior to that of the Datong monitoring station—especially the deviation of the system-simulated values from the observed values in the dates marked by the purple rectangle boxes in Fig. 11. This deviation may be due to the following reasons: ① The Taiyuan monitoring station is located in an area with intensive emissions (Fig. 7), which has a significant impact on the observations; in addition, the observed data are influenced by anthropogenic emissions, which can cause the observation results to fluctuate greatly within a day. After averaging, there are significant differences between the model's simulated values and observations and those at the Datong monitoring station. ② The surface is strongly mixed by turbulence after sunrise due to solar radiation, and vegetation photosynthesis is simultaneously enhanced; this results in the concentration of CO_2 in the near-surface atmosphere decreasing continuously and reaching the lowest value at approximately 15:00. With the subsequent continuous weakening of photosynthesis and atmospheric turbulence, the concentration of CO_2 in the near-surface atmosphere increases gradually. At night, the atmospheric boundary layer tends to stabilize; CO_2 accumulates in the atmospheric boundary layer, and its concentration increases continuously and reaches the daily maximum before sunrise. These meteorological factors directly affect the atmospheric CO_2 concentration. Therefore, the simulation accuracy of WRF-GHG on meteorological factors (including temperature and downward shortwave radiation, etc.) affects the simulation

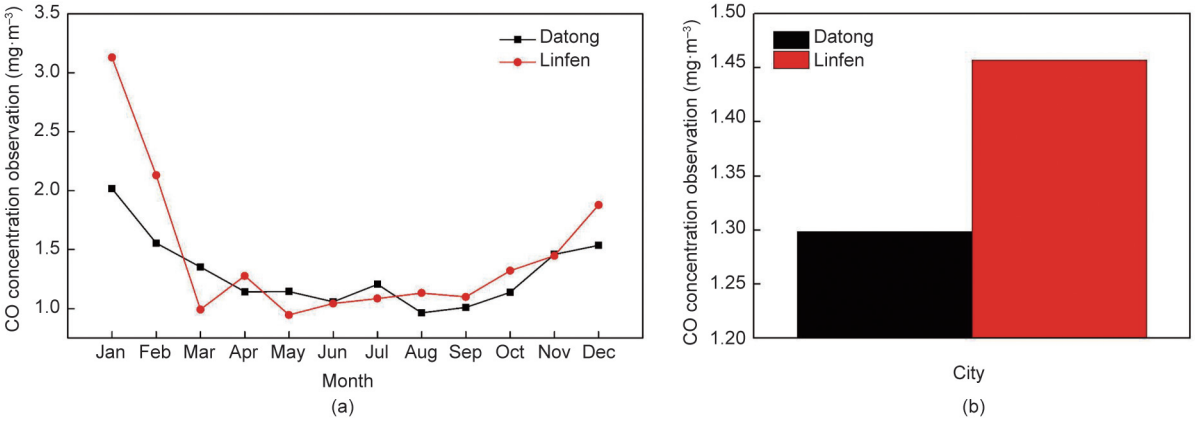

Fig. 9. Comparative analysis of CO concentration observations at the Datong and Linfen air quality monitoring stations in Shanxi. (a) Monthly average of CO concentration observations in 2019; (b) annual average of CO concentration observations.

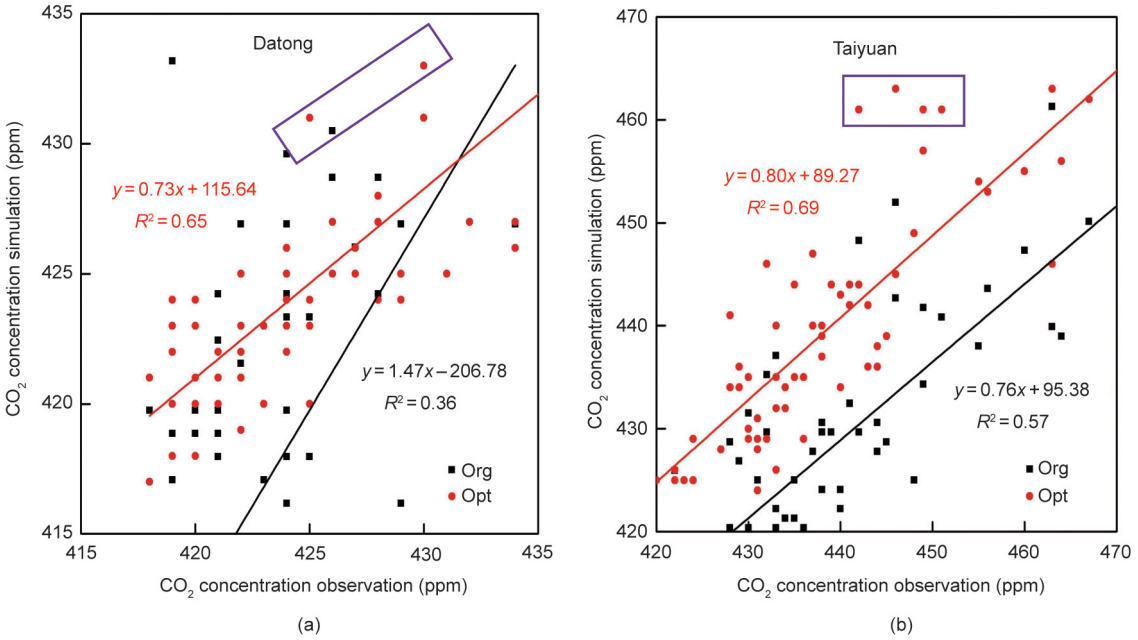

Fig. 10. Comparison analysis of simulated and observed CO_2 concentrations in the CCMVS-R system. (a) Datong monitoring station; (b) Taiyuan monitoring station. Black markers represent the simulated values before assimilation, and red markers represent the simulated values after assimilation.

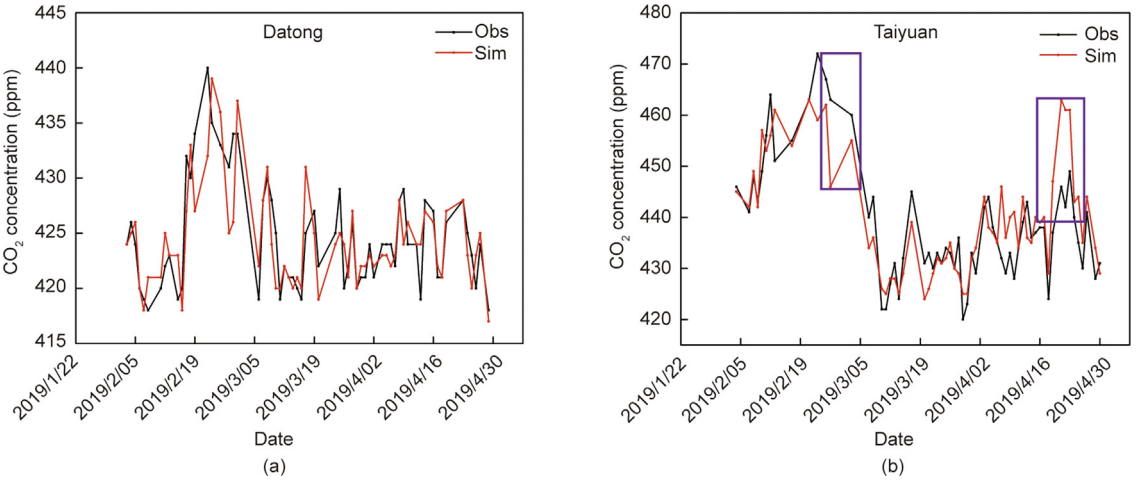

Fig. 11. Analysis of the daily variation in the simulated (Sim) and observed (Obs) values of the CCMVS-R system (black line represents observed values and red line represents simulated values). (a) Datong monitoring station; (b) Taiyuan monitoring station.

effect of atmospheric CO_2 concentration to a certain extent, which is the reason for the large deviation between the simulated and observed values during some periods.

3.3. Advantages of our CCMVS-R system

The good performance of the CCMVS-R system stems from its two main advantages: ① The CCMVS-R was developed based on the WRF model, and the inverse anthropogenic carbon emissions and CO_2 concentration data are more realistic with the combination of atmospherically driven data and a model parameterization scheme. This is a significant improvement over the anthropogenic carbon emissions obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) and GOSAT data using model recursion by Guo et al. [51]. Guo et al.'s method lacked consideration of atmospheric transmission models, resulting in a lack of further limitations from meteorological factors, and hence had considerable uncertainty. ② The CCMVS-R is based on the 4DVar assimilation inversion method of POD to build the assimilation system, which essentially reduces the computer resources required for assimilation inversion while improving the computational efficiency and estimation accuracy. Our approach solves the critical technical issue to a certain level of high spatial and temporal resolution and huge dimensionality of the regional carbon assimilation system, operational cost, and computational efficiency limitation [49,52,53].

4. Conclusions

Based on the WRF-GHG regional model, we used the POD 4DVar method to construct a regional high-precision carbon assimilation system, CCMVS-R. The system realizes the assimilation inversion of

anthropogenic carbon emissions at the kilometer level. It provides an effective means of accurately estimating anthropogenic carbon emissions at the regional and local scales. Two nested layers were designed with the resolutions of the outer and inner regions set at 27 and 9 km, respectively, to verify the effectiveness of the CCMVS-R system. The core nested area was within Shanxi Province, which is associated with large carbon emissions.

Two schemes were also designed for verification using six ground-based sites with high-precision CO_2 monitoring from Shanxi. Scheme 1 used simulated values from the WRF-GHG model extracted from six ground-observation stations as pseudo-observation data for assimilation inversion. In comparison, Scheme 2 used the information from six high-precision ground-observation stations for assimilation inversion, while the other settings remained the same. The simulation results of the ideal experiment showed that ① the anthropogenic carbon emissions after assimilation converged well to the assumed ideal carbon flux, with significant correction of some high-value regions of the *a priori* carbon flux. The correlation between anthropogenic carbon emissions and the ideal carbon flux before and after CCMVS-R assimilation improved from 0.89 before assimilation to 0.94 after assimilation, and the RMSE decreased from 4.89×10^4 $mol \cdot km^2 \cdot h^{-1}$ before assimilation to 5.18×10^3 $mol \cdot km^2 \cdot h^{-1}$ after assimilation—a decrease of 89.4%. This experiment illustrates the effective convergence of the CCMVS-R system. ② The results of the anthropogenic carbon emissions after assimilation were used as input to the CCMVS-R system. The simulation ability of CCMVS-R on CO_2 concentration was greatly improved. In this experiment, the pseudo-observations simulated by the Taiyuan monitoring station, which participates in the assimilation, and Site B, a hypothetical observation that does not participate in the assimilation, were selected to verify the simulation ability of the CCMVS-R on CO_2 concentration before and after the assimilation. The experimental results showed that the simulation effect was significantly improved after assimilation. The correlation between the CO_2 concentration after assimilation and the ideal value was close to 1, and the RMSE between them was close to 0.

Scheme 2 assimilated six high-precision ground station observations in Shanxi Province to obtain inverse anthropogenic carbon emissions in 2019. The spatial distribution of the *a posteriori* anthropogenic carbon emissions was reasonable. Anthropogenic carbon emissions increased in most cities in Shanxi, particularly in Linfen, showing that the bottom-up emission inventory for this city may be underestimated. The results of the *a posteriori* anthropogenic carbon emission statistics were compared with the mean of five anthropogenic carbon emission inventory products (GID, ODIAC, EDGAR, CHRED, and GCP), showing that the *a posteriori* anthropogenic carbon emissions from CCMVS-R assimilation inversion were approximately 28.6% (17%–38%) higher than those of the five inventory products, especially for the ODIAC inventory products, which may be more underestimated than the other inventories due to the time lag and the error of the bottom-up statistical method.

In addition, the relationship between the simulated and observed CO_2 concentrations after assimilation at the Datong and Taiyuan monitoring stations from 1 February to 30 April 2019 was compared and analyzed. The results showed that the CCMVS-R has sound simulation effects, with the correlation coefficients between the simulated and observed CO_2 concentrations after assimilation being as high as 0.80 and 0.83, with mean deviations of −0.25 and 1.01 ppm, respectively. Their correlation coefficients and mean deviations were significantly improved compared with those before the assimilation at the Datong and Taiyuan monitoring stations. The simulated values of the system after assimilation were able to capture the information on the average daily variation in CO_2 concentration well. The inversion results show that the CCMVS-R regional high-precision carbon assimilation system has good application prospects.

The quantity and quality of observational data limited the development of the regional carbon assimilation system. In future studies, multisource CO_2 concentration observational data (i.e., ground-based observations, OCO-2/OCO-3, GOSAT/GOSAT-2, TanSat, etc.) will be introduced to the CCMVS-R, and the companion gas CO of CO_2 will be introduced into the regional atmospheric CO_2 assimilation inversion system to compensate for the low inversion accuracy in regions without sufficient observational data. This will potentially improve the inversion accuracy of the regional carbon fluxes further.

Acknowledgments

This research was supported by the General Project of Top-Design of Multi-Scale Nature-Social Models, Data Support and Decision Support System for NSFC Carbon Neutrality Major Project (42341202), and the Basic Scientific Research Fund of the Chinese Academy of Meteorological Sciences (2021Z014). The authors are grateful to all the organizations and groups that provided indispensable datasets that we used in this study. We acknowledge the CO_2 boundary conditions provided by Dr. Huifang Zhang.

Compliance with ethics guidelines

Lifeng Guo, Xiaoye Zhang, Junting Zhong, Deying Wang, Changhong Miao, Licheng Zhao, Zijiang Zhou, Jie Liao, Bo Hu, Lingyun Zhu, and Yan Chen declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2023.02.017.

References

[1] Pachauri R, Meyer L, Plattner G, Stocker T. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. Geneva: Intergovernmental Panel on Climate Change; 2014.
[2] Canadell JG, Monteiro PM, Costa MH, Da Cunha LC, Cox PM, Alexey V, et al. Final government distribution—chapter 5: global carbon and other biogeochemical cycles and feedbacks. Report. Geneva: Intergovernmental Panel on Climate Change; 2021 Sep. Report No.: AR6 WGI.
[3] Chandra N, Patra PK, Niwa Y, Ito A, Iida Y, Goto D, et al. Estimated regional CO_2 flux and uncertainty based on an ensemble of atmospheric CO_2 inversions. Atmos Chem Phys 2022;22(14):9215–43.
[4] Ma J, Li L, Wang H, Du Yi, Ma J, Zhang X, et al. Carbon capture and storage: history and the road ahead. Engineering 2022;14:33–43.
[5] Le Quéré C, Andrew RM, Canadell JG, Sitch S, Korsbakken JI, Peters GP, et al. Global carbon budget 2016. Earth Syst Sci Data 2016;8(2):605–49.
[6] Oda T, Maksyutov S, Andres RJ. The Open-source Data Inventory for Anthropogenic Carbon dioxide (CO_2), version 2016 (ODIAC2016): a global, monthly fossil-fuel CO_2 gridded emission data product for tracer transport simulations and surface flux inversions. Earth Syst Sci Data 2018;10(1):87–107.
[7] Jones MW, Andrew RM, Peters GP, Janssens-Maenhout G, De-Gol AJ, Ciais P, et al. Gridded fossil CO_2 emissions and related O_2 combustion consistent with national inventories 1959–2018. Sci Data 2021;8(1):2.
[8] The Central People's Government of the People's Republic of China. China's achievements, new goals and new measures for nationally determined contributions. New York City: The United Nations Framework Convention on Climate Change; 2020. Chinese.
[9] Liu Z, Sun T, Yu Y, Ke P, Deng Z, Lu C, et al. Real-time carbon emission accounting technology toward carbon neutrality. Engineering 2022;14:44–51.
[10] Wei YM, Chen K, Kang JN, Chen W, Wang XY, Zhang X. Policy and management of carbon peaking and carbon neutrality: a literature review. Engineering 2022;14:52–63.
[11] Zhang HN, Shen RR, Zhang XP, Kang JJ, Yuan JH. Implications and pathways of China's carbon neutrality: a review. Clim Chang Res 2022;18(2):240–52. Chinese.

[12] Crippa M, Solazzo E, Huang G, Guizzardi D, Koffi E, Muntean M, et al. High resolution temporal profiles in the emissions database for global atmospheric research. Sci Data 2020;7(1):121.
[13] Gregg JS, Andres RJ, Marland G. China: Emissions pattern of the world leader in CO_2 emissions from fossil fuel consumption and cement production. Geophys Res Lett 2008;35(8):L08806.
[14] Liu Z, Guan D, Wei W, Davis SJ, Ciais P, Bai J, et al. Reduced carbon emission estimates from fossil fuel combustion and cement production in China. Nature 2015;524(7565):335–8.
[15] Rayner P, Raupach M, Paget M, Peylin P, Koffi E. A new global gridded data set of CO_2 emissions from fossil fuel combustion: methodology and evaluation. J Geophys Res D Atmospheres 2010;115:D19306.
[16] Wang Y, Munger J, Xu S, McElroy MB, Hao J, Nielsen C, et al. CO_2 and its correlation with CO at a rural site near Beijing: implications for combustion efficiency in China. Atmos Chem Phys 2010;10(18):8881–97.
[17] Kuze A, Suto H, Nakajima M, Hamazaki T. Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. Appl Opt 2009;48(35):6716–33.
[18] Glumb R, Davis G, Lietzke C. The TANSO-FTS-2 instrument for the GOSAT-2 greenhouse gas monitoring mission. In: Proceedings of 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS 2014); 2014 Jul 13–18; Quebec City, QC, Canada. New York City: IEEE; 2014. p. 1238–40.
[19] Eldering A, O'Dell CW, Wennberg PO, Crisp D, Gunson MR, Viatte C, et al. The Orbiting Carbon Observatory-2: first 18 months of science data products. Atmos Meas Tech 2017;10(2):549–63.
[20] Eldering A, Wennberg PO, Crisp D, Schimel DS, Gunson MR, Chatterjee A, et al. The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. Science 2017;358:6360.
[21] Eldering A, Taylor TE, O'Dell CW, Pavlick R. The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data. Atmos Meas Tech 2019;12(4):2341–70.
[22] Liu Yi, Wang J, Yao Lu, Chen Xi, Cai Z, Yang D, et al. The TanSat mission: preliminary global observations. Sci Bull 2018;63(18):1200–7.
[23] Yang D, Liu Yi, Cai Z, Chen Xi, Yao Lu, Lu D. First global carbon dioxide maps produced from TanSat measurements. Adv Atmos Sci 2018;35(6):621–3.
[24] Zhang Q, Li M, Wei C, Mizzi AP, Huang Y, Gu Q. Assimilation of OCO-2 retrievals with WRF-Chem/DART: a case study for the Midwestern United States. Atmos Environ 2021;246:118106.
[25] Bréon F, Broquet G, Puygrenier V, Chevallier F, Xueref-Remy I, Ramonet M, et al. An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements. Atmos Chem Phys 2015;15(4):1707–24.
[26] Kort EA, Frankenberg C, Miller CE, Oda T. Space-based observations of megacity carbon dioxide. Geophys Res Lett 2012;39:L17806.
[27] Pillai D, Buchwitz M, Gerbig C, Koch T, Reuter M, Bovensmann H, et al. Tracking city CO_2 emissions from space using a high-resolution inverse modelling approach: a case study for Berlin. Germany Atmos Chem Phys 2016;16(15):9591–610.
[28] Schneising O, Heymann J, Buchwitz M, Reuter M, Bovensmann H, Burrows J. Anthropogenic carbon dioxide source areas observed from space: assessment of regional enhancements and trends. Atmos Chem Phys 2013;13(5):2445–54.
[29] Wong K, Fu D, Pongetti T, Newman S, Kort E, Duren R, et al. Mapping CH_4: CO_2 ratios in Los Angeles with CLARS-FTS from mount Wilson. California Atmos Chem Phys 2015;15(1):241–52.
[30] Ye X, Lauvaux T, Kort EA, Oda T, Feng S, Lin JC, et al. Constraining fossil fuel CO_2 emissions from urban area using OCO-2 observations of total column CO_2. J Geophys Res Atmos 2020;125(8):e2019JD030528.
[31] Zhongming Z, Linong L, Xiaona Y, Wangqiang Z, Wei L. 2019 refinement to the 2006 IPCC guidelines for National Greenhouse Gas Inventories. Geneva: Intergovernmental Panel on Climate Change; 2019.
[32] Janssens-Maenhout G, Pinty B, Dowell M, Zunker H, Andersson E, Balsamo G, et al. Toward an operational anthropogenic CO_2 emissions monitoring and verification support capacity. Bull Am Meteorol Soc 2020;101(8):E1439–51.
[33] Crippa M, Oreggioni G, Guizzardi D, Muntean M, Schaaf E, Lo Vullo E, et al. Fossil CO_2 and GHG emissions of all world countries. Luxemburg: Publication Office of the European Union; 2019.
[34] Oda T, Maksyutov S. A very high-resolution (1 km× 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights. Atmos Chem Phys 2011;11(2):543–56.
[35] Liu J, Tong D, Zheng Y, Cheng J, Qin X, Shi Q, et al. Carbon and air pollutant emissions from China's cement industry 1990–2015: trends, evolution of technologies, and drivers. Atmos Chem Phys 2021;21(3):1627–47.
[36] Cai B, Liang S, Zhou J, Wang J, Cao L, Qu S, et al. China high resolution emission database (CHRED) with point emission sources, gridded emission data, and supplementary socioeconomic data. Resour Conserv Recycling 2018;129:232–9.
[37] Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost G, Skamarock WC, et al. Fully coupled "online" chemistry within the WRF model. Atmos Environ 2005;39(37):6957–75.
[38] Peters W, Jacobson AR, Sweeney C, Andrews AE, Conway TJ, Masarie K, et al. An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. Proc Natl Acad Sci USA 2007;104(48):18925–30.
[39] Zhang HF, Chen BZ, van der Laan-Luijkx IT, Chen J, Xu G, Yan JW, et al. Net terrestrial CO_2 exchange over China during 2001–2010 estimated with an ensemble data assimilation system for atmospheric CO_2. J Geophys Res D Atmospheres 2014;119(6):3500–15.
[40] Beck V, Koch T, Kretschmer R, Marshall J, Ahmadov R, Gerbig C, et al. The WRF Greenhouse Gas Model (WRF-GHG). Jena: Max Planck Institute for Biogeochemistry; 2010.
[41] Tian X, Xie Z. Effects of sample density on the assimilation performance of an explicit four-dimensional variational data assimilation method. Sci China Ser D Earth Sci 2009;52(11):1849–56.
[42] Tian X, Xie Z, Sun Q. A POD-based ensemble four-dimensional variational assimilation method. Tellus A Dyn Meteorol Oceanogr 2011;63(4):805–16.
[43] Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD. Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. J Geophys Res D Atmospheres 2008;113(D13):D13103.
[44] Dudhia J. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J Atmos Sci 1989;46(20):3077–107.
[45] Hong SY, Noh Y, Dudhia J. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev 2006;134(9):2318–41.
[46] Chen F, Dudhia J. Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon Weather Rev 2001;129(4):569–85.
[47] Kain JS. The Kain-Fritsch convective parameterization: an update. J Appl Meteorol 2004;43(1):170–81.
[48] Feng T, Zhou W, Wu S, Niu Z, Cheng P, Xiong X, et al. High-resolution simulation of wintertime fossil fuel CO_2 in Beijing, China: characteristics, sources, and regional transport. Atmos Environ 2019;198:226–35.
[49] Martin CR, Zeng N, Karion A, Mueller K, Ghosh S, Lopez-Coto I, et al. Investigating sources of variability and error in simulations of carbon dioxide in an urban region. Atmos Environ 2019;199:55–69.
[50] Diao Y, Huang J, Liu C, Cui J, Liu S. A modeling study of CO_2 flux and concentrations over the Yangtze River Delta using the WRF-GHG model. Chin J Atmos Sci 2015;39:849–60. Chinese.
[51] Guo M, Xu J, Wang X, He H, Li J, Wu L. Estimating CO_2 concentration during the growing season from MODIS and GOSAT in East Asia. Int J Remote Sens 2015;36(17):4363–83.
[52] Fischer ML, Parazoo N, Brophy K, Cui X, Jeong S, Liu J, et al. Simulating estimation of California fossil fuel and biosphere carbon dioxide exchanges combining in situ tower and satellite column observations. J Geophys Res D Atmospheres 2017;122(6):3653–71.
[53] Zheng T, French NH, Baxter M. Development of the WRF-CO_2 4D-Var assimilation system v1.0. Geosci Model Dev 2018;11(5):1725–52.