npj | climate and atmospheric science

Article

Published in partnership with CECCR at King Abdulaziz University

https://doi.org/10.1038/s41612-024-00824-w

A new global carbon flux estimation methodology by assimilation of both in situ and satellite CO_2 observations

Wu Su[1], Binghao Wang[1], Hanyue Chen[2], Lin Zhu[3], Xiaogu Zheng[4,5] & Song Xi Chen[2,6,7]

Accurate estimation of carbon removal by terrestrial ecosystems and oceans is crucial to the success of global carbon mitigation initiatives. The emergence of multi-source CO_2 observations offers prospects for an improved assessment of carbon fluxes. However, the utility of these diverse observations has been limited by their heterogeneity, leading to much variation in estimated carbon fluxes. To harvest the diverse data types, this paper develops a multi-observation carbon assimilation system (MCAS), which simultaneously integrates both satellite and ground-based observations. MCAS modifies the ensemble Kalman filter to apply different inflation factors to different types of observation errors, addressing the heterogeneity between satellite and in situ data. In commonly used independent validation datasets, the carbon flux derived from MCAS outperformed those obtained from a single source, demonstrating a 20% reduction in error compared to existing carbon flux products. We use MCAS to conduct ecosystem and ocean carbon flux inversion for the period of 2016–2020, which reveals that the 5-year average global net terrestrial and ocean sink was 1.84 ± 0.60 and 2.74 ± 0.49 petagrams, absorbing approximately 47% of human-caused CO_2 emissions together, which were consistent with the global carbon project estimates of 1.82 and 2.66 petagrams. All these facts suggest MCAS is a better methodology than those for assimilating single-source observation only.

With the increasing availability of observational techniques, the top-down atmospheric inversion method is gaining traction for its ability to estimate carbon flux using a range of CO_2 measurements, including in situ and satellite observations[1–3]. The existing carbon flux inversion methods consist of two categories according to the kind of observational data used. The first one involves using the in situ observations, with CarbonTracker being the representative along with other studies[4–7]. The other one employs satellite observations. As shown in Fig. S1 in Supplementary Information (SI), satellite observations provide broader spatial and temporal coverage and, as a result, more research has been conducted in recent years on satellite data for carbon flux inversion[8–11]. However, the quality of satellite data is considered to be lower than that from the in situ observations[12–14]. Moreover, the satellite data provide aggregated air column concentration of CO_2 (XCO_2) only, while the in situ data provide information at different altitudes.

Although projects such as the orbiting carbon observatory-2 model intercomparison project (OCO-2 MIP, see Table S1) have begun exploring the different results obtained by assimilating observations from various sources, they have not shown that assimilating multiple sources of data would provide better results. The essence of this issue may lie in the heterogeneity between satellite and in situ observations[15–17]. Most assimilation systems do not account for this heterogeneity, leading to assimilation results that can be even less accurate than those obtained by assimilating single-source observations[18]. Some studies[19,20] have begun to explore the advantages of assimilating multi-source observational data, while they usually assume that satellite data are less accurate and therefore require additional processing rather than truly addressing the heterogeneity issues among different sources of observational data.

In this paper, we proposed a carbon inversion system that assimilates multi-source CO_2 data, including in situ and satellite observations. To account for the heterogeneity among the multiple data sources, a modified ensemble Kalman filter method with multiple inflation of the forecast and observation error covariance matrix[21] is employed to establish a multi-

[1]Center for Big Data Research, Peking University, Beijing, China. [2]Center for Statistical Science, Peking University, Beijing, China. [3]Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites, National Satellite Meteorological Center (National Center for Space Weather), Beijing, China. [4]Shanghai Zhangjiang Mathematics Institute, Shanghai, China. [5]International Global Change Institute, Hamilton, New Zealand. [6]Institute of Carbon Neutrality, Peking University, Beijing, China. [7]Department of Statistics and Data Science, Tsinghua University, Beijing, China. e-mail: songxichen@pku.edu.cn

observation carbon assimilation system (MCAS). The multiple inflation method automatically balances the heterogeneity between multi-source observations to achieve better data assimilation results. MCAS is used to simultaneously assimilate in situ CO_2 measurements and satellite XCO_2 retrievals, yielding the global terrestrial ecosystem and ocean carbon fluxes with 3-hourly and 1° × 1° resolution from 2016 to 2020. The carbon fluxes estimated by the MCAS have demonstrated significant advantages over control experiments that assimilate only in situ or satellite observations and other similar products, achieving the best performance in two independent validation datasets. Our approach may provide a pathway to fully exploit the diversity of CO_2 observation data, allowing better emissions planning.

Results

Assimilation of CO_2 observations from in situ and satellite

The proposed MCAS was used to estimate carbon fluxes for terrestrial ecosystems and ocean-spanning over the years from 2016 to 2020. To highlight the efficacy of leveraging multi-source observations, we conducted three distinct inversion procedures with the MCAS: (1) MCAS(in situ): inversion using only in situ observations; (2) MCAS(Sat): inversion using only satellite observations; and (3) MCAS(in situ&Sat): inversion using both in situ and satellite observations. The details of the MCAS procedures are available in "Methods".

Table 1 shows the average carbon budget for the period of 2016–2020. For comparison, the results of the global carbon budget from the CarbonTracker version 2022 (CT2022,[22]) and the global carbon project (GCP,[23]), are also presented. The annual average terrestrial ecosystem carbon sink estimated by MCAS(in situ&Sat) falls between that of MCAS(in situ) and MCAS(Sat), being 0.5 PgC $year^{-1}$ higher than MCAS(in situ) and 0.3 PgC $year^{-1}$ lower than MCAS(Sat). On the contrary, the ranking for the size of the ocean carbon sink is reversed, with MCAS(in situ) having the highest ocean carbon sink estimate, MCAS(Sat) the lowest, and MCAS(in situ&Sat) positioned between the two. As the global net carbon flux will result in changes in atmospheric CO_2 concentration, we also calculated the global carbon budget imbalance, a measure of the mismatch between net emissions (total emissions − total sinks) and the NOAA's 5-year average CO_2 growth rate (5.20 PgC $year^{-1}$), which is estimated directly using in situ observations[23–25]. The budget imbalance of MCAS(in situ), MCAS(Sat), and MCAS(in situ&Sat) were 0.09 PgC $year^{-1}$, 0.10 PgC $year^{-1}$, and 0.02 PgC $year^{-1}$. Hence, as shown in Table 1, the proposed MCAS(in situ&Sat) that assimilates both in situ and satellite data had a smaller global carbon budget imbalance than the other two MCAS results. It is also worth noting that the inversion results are highly sensitive to the selection of prior carbon fluxes. We listed the inversion results for two sets of different priors in Table S3, showing significant differences between them.

Figure 1a–c presents the annual averages of global terrestrial ecosystem and ocean fluxes for 2016–2020 estimated by the three MCAS procedures. While ocean carbon fluxes show relatively less spatial variations, significant spatial differences appear in the terrestrial carbon fluxes among them. It is noted that results inferred using satellite observations display smoother spatial estimates than MCAS(in situ). Figure 1d illustrates the spatial differences between the inversion results of MCAS(in situ&Sat) and MCAS(in situ). Overall, by incorporating satellite data, MCAS(in situ&Sat) decreases the global average ocean carbon sink over the years 2016 to 2020 from 3.17 ± 0.41 of MCAS(in situ) to 2.74 ± 0.49 PgC $year^{-1}$. For terrestrial flux, it shows significant spatial differences between MCAS(in situ&Sat) and MCAS(in situ) in regions with fewer in situ observations, such as South America and Africa. MCAS(in situ&Sat) shows larger terrestrial ecosystem carbon sinks in these regions, while ecosystem carbon sinks are lower than those from MCAS(in situ) in the Northern Hemisphere, especially in the Eastern United States and Canada, Eurasia except northern Siberian. Figure 1e illustrates the spatial differences between MCAS(in situ&Sat) and MCAS(Sat). Compared to MCAS(Sat), MCAS(in situ&Sat) shows a much lower terrestrial ecosystem sink in the Northern Hemisphere, but larger in the Southern Hemisphere, while the situation is reversed for the oceanic sink. Figure 1f shows the difference between MCAS(Sat) and MCAS(in situ). MCAS(Sat) significantly reduces the carbon sink in the European and the ocean carbon sink in the Northern Hemisphere while increasing the carbon sink in the western part of North America and North Siberian compared to MCAS(in situ). Table S2 and Fig. S2 present the annual average carbon fluxes across the 11 TransCom (Atmospheric Tracer Transport Model Intercomparison Project) regions for the period 2016-2020 and the *p*-values for the significant difference between carbon flux estimates based on 64 ensemble runs by the MCAS(in situ) and MCAS(sat) procedures, which suggests MCAS(in situ) and MCAS(Sat) are significantly different at 8 out of the 11 TranCom regions. These differences might be attributed to the varying distribution of observation data across regions. For example, in Boreal North America and Tropical South America, there are significant differences between MCAS(in situ) and MCAS(Sat). In the former region, MCAS(in situ&Sat) aligns more closely with MCAS(in situ) because of the abundance of in situ data, while satellite observations might be limited by high latitude restriction. In the Tropical South America region, where in situ data is scarce, MCAS(in situ&Sat) is closer to MCAS(Sat). See Section S1 for further discussion on the regional carbon fluxes in conjunction with climatic conditions.

We used two independent datasets, the ObsPack validation set (5% of the ObsPack data, non-assimilated) and the TCCON dataset as described in the section "Validation data". The performance of five carbon flux products which include the prior, CT2022, and the proposed MCAS(in situ), MCAS(Sat), and MCAS(in situ&Sat) were evaluated. We ran the GEOS-

Table 1 | Mean global carbon budgets and their standard deviations during the 2016–2020 period estimated in this study, as well as those of the CarbonTracker version 2022 (CT2022) and the global carbon project (GCP) (PgC $year^{-1}$)

		MCAS(in situ)	MCAS(Sat)	MCAS(in situ&Sat)	CT2022	GCP
Fossil CO_2 emissions		9.80	9.80	9.80	9.80	9.76
Terrestrial fluxes	NEE[a]	−3.49 ± 0.84	−4.29 ± 0.45	−3.99 ± 0.60	−3.03	/
	Wildfire	2.15	2.15	2.15	2.15	/
	Net land flux[b]	−1.34 ± 0.84	−2.14 ± 0.45	−1.84 ± 0.60	−0.88	−1.82
Oceanic fluxes		−3.17 ± 0.41	−2.36 ± 0.59	−2.74 ± 0.49	−3.62	−2.66
Global net CO_2 fluxes[c]		5.29 ± 0.95	5.30 ± 0.66	5.22 ± 0.73	5.30	5.28
Atmospheric CO_2 growth[d]		5.20	5.20	5.20	5.20	5.20
Budget Imbalance[e]		0.09	0.10	0.02	0.10	0.08

[a]NEE represents the net ecosystem exchange.
[b]Net land flux = NEE + Wildfire.
[c]Global net CO_2 flux = Fossil + NEE + Wildfire + Oceanic.
[d]Atmospheric CO_2 growth was obtained from https://gml.noaa.gov/ccgg/trends/gl_gr.html (last access: 1 January 2024) and was estimated with a factor of 2.124 PgC ppm^{-1} [23].
[e]The difference between net CO_2 flux and atmospheric CO_2 growth.

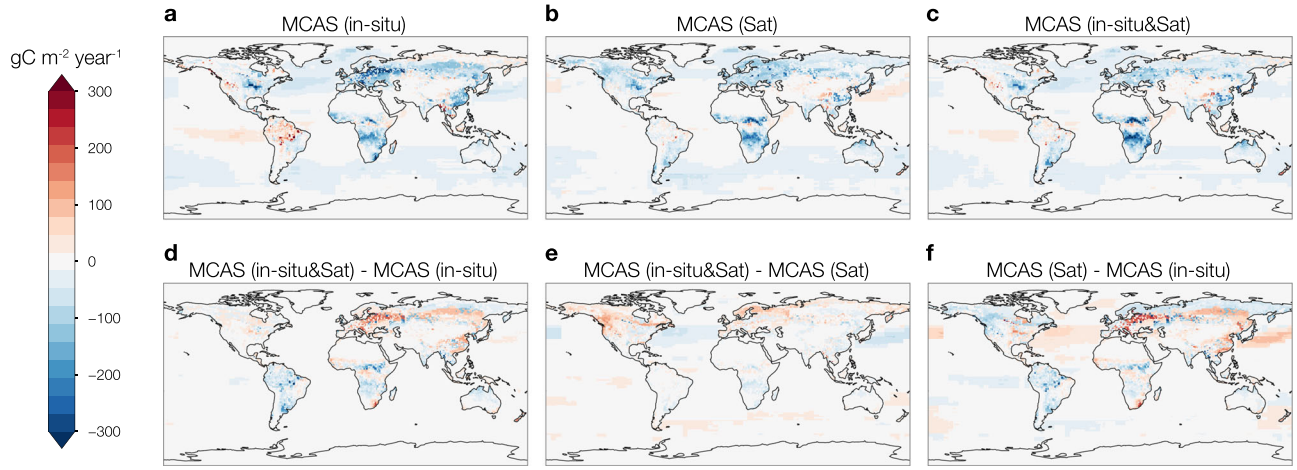

Fig. 1 | Average global terrestrial ecosystem and ocean carbon flux for the period 2016-2020 (PgC year^{-1}). **a** MCAS(in situ), **b** MCAS(Sat), **c** MCAS(in situ& Sat), where blue represents carbon sinks and red represents carbon sources. The spatial pattern of the difference in carbon flux inversion between: **d** MCAS(in situ& Sat) and MCAS(in situ), **e** MCAS(in situ& Sat) and MCAS(Sat), **f** MCAS(Sat) and MCAS(in situ).

Table 2 | Evaluation results for the Prior, CT2022, MCAS(in situ), MCAS(Sat), and MCAS(in situ&Sat) carbon flux estimate on the ObsPack validation set and TCCON dataset over 2016–2020, which include the root mean square error (in ppm), the correlation coefficient (CORR), and the average bias (in ppm) between the forecasts and the observations

Products	ObsPack validation set			TCCON		
	RMSE	CORR	BIAS	RMSE	CORR	BIAS
Prior	2.827	0.911	0.327	1.752	0.942	0.892
CT2022	2.656	0.916	0.100	1.530	0.947	0.497
MCAS(in situ)	2.498	0.924	0.202	1.549	0.946	0.571
MCAS(Sat)	2.702	0.909	−0.494	1.408	0.949	0.190
MCAS(in situ&Sat)	2.132	0.942	−0.004	1.385	0.950	0.111

Chem model with different products as the carbon flux modules in GEOS-Chem to obtain atmospheric CO_2 concentration forecasts spanning 2016–2020. Subsequently, these forecasts were compared with the observed CO_2 values in the ObsPack validation set and the TCCON dataset. The evaluation results, presented in Table 2, include the root mean square error (RMSE), the correlation coefficient (CORR), and the average difference between the predicted and observed values (BIAS).

Table 2 reveals that the carbon fluxes obtained by data assimilation consistently outperformed the prior carbon flux. MCAS(in situ&Sat), incorporating both in situ and satellite observations, exhibits the best performance in both validation datasets. On the in situ validation set from ObsPack, the results of assimilating in situ data outperformed MCAS(Sat), which only assimilated satellite observations. On the XCO_2 observations from TCCON, the results of assimilating satellite observations outperformed MCAS(in situ), which only assimilates in situ observations. The proposed MCAS(in situ&Sat), which assimilates both in situ and satellite data, consistently outperformed the other two on both datasets. Compared to MCAS(in situ), MCAS(in situ&Sat) reduces the RMSE by approximately 15% on the ObsPack validation set and 4% on TCCON. Compared to MCAS(Sat), MCAS(in situ&Sat) reduced the RMSE by approximately 21% on the ObsPack validation set and only 2% on TCCON. Additionally, Table S4 and Fig. S3 specifically use the 5% reserved aircraft data from ObsPack to evaluate the inversion results, along with the performance above and below 3000m, respectively. The results show that MCAS(in situ&Sat) consistently performs best, followed by MCAS(in situ), with MCAS(Sat) performing slightly worse. Below 3000 m, the advantage of MCAS(in situ&Sat) becomes more pronounced, while the limitations of MCAS(Sat) are more evident. This highlights the limitation of satellite observations in the lower atmosphere, as they only provide column concentrations and are less sensitive to near-surface CO_2 emissions and absorption. However, above 3000 m, while MCAS(in situ&Sat) remains the best, MCAS(Sat) shows improved performance, indicating that satellite observations effectively constrain CO_2 concentrations in the upper atmosphere. These findings underscore the importance of combining in situ and satellite observations for more accurate carbon flux estimates.

These results demonstrate that MCAS(in situ) and MCAS(Sat) have their respective advantages and disadvantages in different aspects, and the proposed MCAS(in situ&Sat) can handle heterogeneity between multi-source data, making the results of assimilating both types of data superior to assimilating either type alone.

Comparison between MCAS and other carbon flux products

We discussed the results of MCAS, including (in situ) and (in situ&Sat), with CT2022 (CarbonTracker 2022). MCAS(in situ&Sat) utilized the same in situ observation data and prior carbon fluxes as the others, and OCO-2 Level 2 data were used for satellite observations additionally. Table S5 details the similarities and differences between MCAS and CT2022. Figure 2a, b presents the annual averages of global terrestrial ecosystem and ocean fluxes for 2016–2020 estimated by the Prior and CT2022. It can be seen from Table 1 that the total carbon sinks (terrestrial ecosystems + oceans) estimated by MCAS(in situ), MCAS(in situ&Sat), and CT2022 over the 5 years are quite close, at 6.66 PgC year^{-1}, 6.73 PgC year^{-1}, and 6.65 PgC year^{-1}, respectively. However, the terrestrial ecosystem carbon sink estimated by MCAS(in situ&Sat) is higher by 0.5 PgC year^{-1} and 0.96 PgC year^{-1} compared to MCAS(in situ) and CT2022, respectively, while the ocean sink correspondingly decreases. MCAS(in situ&Sat) estimates the ratio of net land carbon sink to ocean carbon sink to be 2:3, which is consistent with the result reported by GCP. However, the two in situ inversions, MCAS(in situ) and CT2022, provide a larger estimate for the ocean carbon sink. These differences can be seen in detail in Figs. 1d and 2d. Compared to CT2022, MCAS(in situ&Sat) shows a smaller carbon sink in North America and a larger carbon sink in Europe.

From Table 2, MACS(in situ&Sat) reduces the RMSE by 20% and 4% in the ObsPack and TCCON validation datasets, respectively, compared to CT2022. While MCAS(in situ) and CT2022 perform similarly, MCAS(in situ) shows better RMSE and correlation coefficients on the ObsPack validation data, though with a slightly larger bias, whereas CT2022 performed better on the TCCON dataset. The lower panel of Fig. 2 offers a detailed view of the evaluation results among the CT2022 and

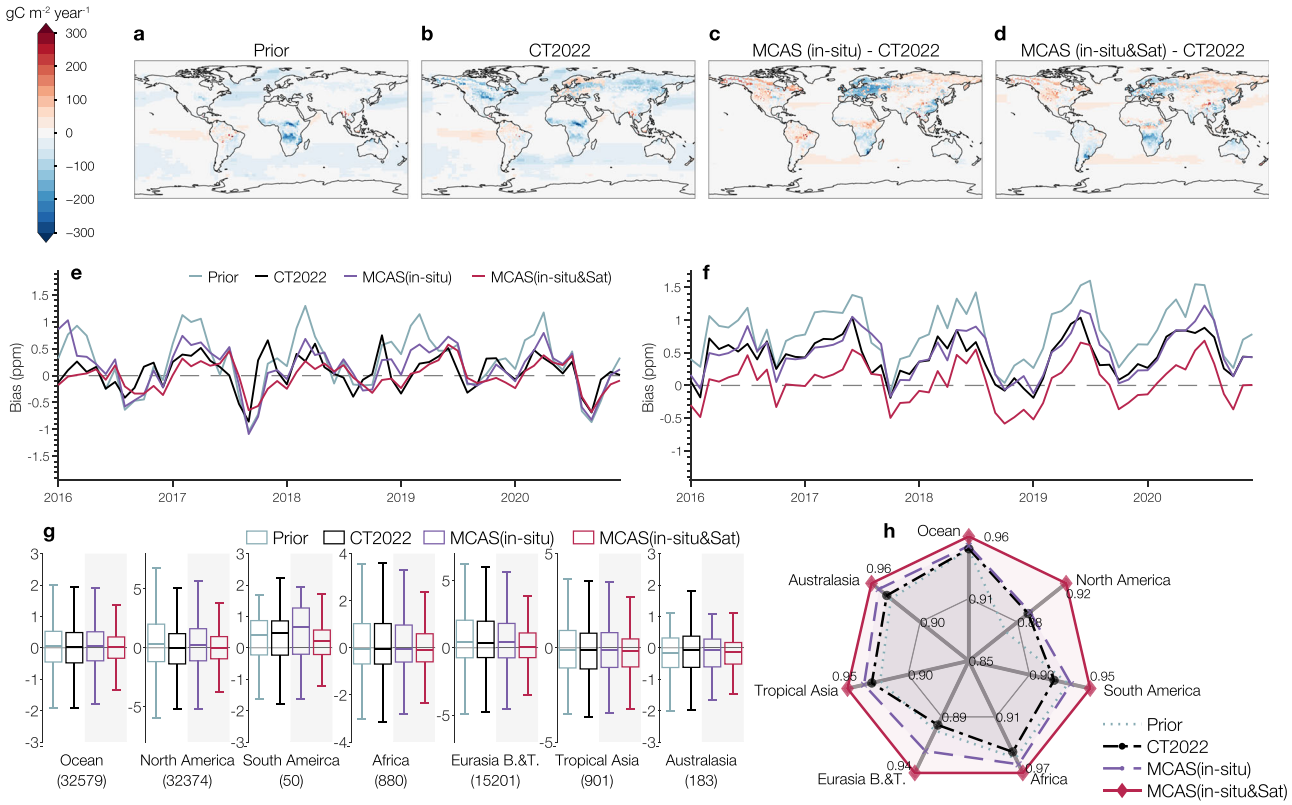

Fig. 2 | Comparison of global terrestrial ecosystem and ocean carbon flux, and evaluation of carbon flux inversion performance. Upper panel shows the average global terrestrial ecosystem and ocean carbon flux for the period 2016–2020 (PgC year^{-1}) for **a** prior, **b** CT2022, and the difference between **c** MCAS(in situ) and CT2022, **d** MCAS(in situ& Sat) and CT2022. The middle and lower panels show evaluation results for Prior, CT2022, MCAS(in situ), and MCAS(in situ& Sat): **e** average biases on the ObsPack validation set, **f** average biases on the TCCON dataset, **g** boxplots of prediction biases in seven regions (Ocean, North America, South America, Africa, Eurasia Boreal and Temperate, Tropical Asia, and Australasia) with sample sizes in parentheses, and **h** correlations between observations and forecasts in each region.

two sets of MACS. Panels (e, f) present the biases between the observed and predicted CO_2 of each product in the ObsPack validation and TCCON datasets, respectively, showing that MCAS(in situ&Sat) has smaller biases than the others in both two validation sets. The validation results on the TCCON show that after assimilating both in situ and satellite data simultaneously, MCAS(in situ&Sat) exhibits smaller biases on XCO_2 prediction, while the forecast results of the in situ inversion exhibit a systematic underestimation. Figure 2g, h further displays the evaluation results within each localization region (see section "Ensemble Kalman Filter with multiple inflation and localization") in the ObsPack validation dataset. MCAS(in situ&Sat) exhibits very low biases and shows the smallest variability across all regions (TCCON sites are very sparse, with most of them located in North America or Europe, making them unsuitable for regional displays). MCAS(in situ) shows a bias mainly in the limited observations over South America (50 observations), while its performance in other regions is similar to CT2022, but with less variation except in North America, where CT2022 has noticeably smaller variation, likely due to having more in situ observations and the higher-resolution nested grid in that region. In terms of the correlation coefficient (see Fig. 2h) between forecasts and observations of CO_2, MCAS(in situ&Sat) outperforms the other two in all regions, while MCAS(in situ) is better than CT2022 in most regions, with both performing similarly in North America.

The OCO-2 MIP v10 is a comparative project administered by NOAA, involving 14 members (participants as listed in Table S1) and different versions obtained with different types of observations. Specifically, we consider three OCO-2 MIP methods: IS which uses only in situ data, LNLG which uses only OCO-2 observations over land, and LNLGOGIS which assimilates both in situ and all satellite observations.

We took the average of the 14 members for each of the three OCO-2 MIP v10 methods for comparison. Figure S4 summarizes the estimated average annual global CO_2 growth from 2016 to 2020 obtained from the three MCAS methods and the three corresponding OCO-2 MIP methods (IS, LNLG, and LNLGOGIS), as well as the benchmark CO_2 growth from NOAA sites measurements. Comparing the results from MCAS and OCO-2 MIP, shows that assimilating both in situ and satellite data improves accuracy for MCAS(in situ&Sat), but not necessarily for LNLGOGIS. In fact, LNLGOGIS performs worse than IS and LNLG. In contrast, MCAS(in situ&Sat) outperforms both MCAS(in situ) and MCAS(Sat), highlighting the MCAS algorithm's effectiveness in assimilating heterogeneous multi-source observations.

Then we considered the NBE (Net Biome Exchange) distribution of these carbon fluxes over the 11 land TransCom regions. The results are shown in Fig. 3a. Attention is first drawn to Northern Africa, where the OCO-2 MIP results indicate a significant carbon source, while the MCAS results suggest carbon neutrality. Given the global scale of carbon fluxes analyzed by both MCAS and OCO-2 MIP, we referred to more specific studies such as Gaubert et al., 2023[26]. Using in situ airborne observations from NASA's ATom mission over the tropical Atlantic Ocean during four seasons from 2016 to 2018, they concluded that this region is close to carbon neutral, aligning with the MCAS results. Additionally, they found that the OCO-2 MIP results significantly overestimate the carbon source in this region, particularly LNLG, by almost 1 PgC year^{-1}. This supports our conclusions and may be due to the effectiveness of our localization. Observations in the Northern Africa region are very sparse. Implementing localization will prevent the inversion results for this region from being dominated by observations from other areas. In Boreal North America, only MCAS(in situ) and MCAS(in situ&Sat) indicate carbon neutrality, while

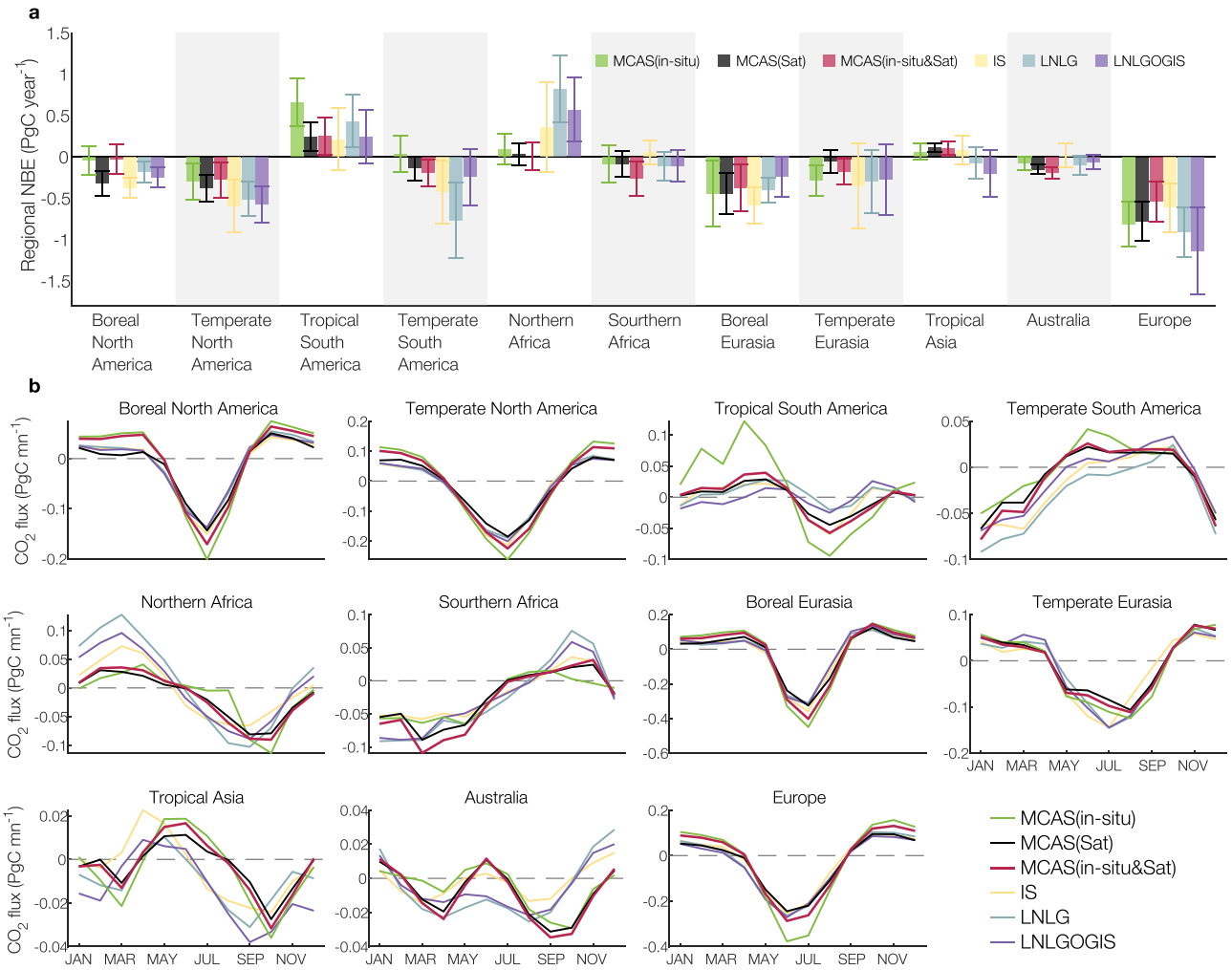

Fig. 3 | Comparison of natural land fluxes and seasonal cycle amplitudes across 11 TransCom land regions over 2016-2020. a Annual natural land fluxes derived by MCAS and OCO-2 MIP averaged over 2016 to 2020, **b** Seasonal cycle amplitudes of terrestrial ecosystem fluxes derived in the 11 TransCom land regions from 2016 to 2020. Each result is represented by a solid curve showing the average value during this period.

other results show a strong carbon sink. Recent studies also suggest that Boreal North America is transitioning from a carbon sink to a carbon source[27,28] due to the reduced growth of trees caused by atmospheric dryness[29] and the increase in wildfires[30]. The differences in these inversion results stem from factors such as the data assimilation algorithm, the error design, and the prior settings.

We presented the seasonal variations of NEE (Net Ecosystem Exchange) for each TransCom region in Fig. 3b. The peak summertime drawdown of fluxes in northern ecosystems, indicative of deeper carbon sinks during the growing season, remains consistent among the inversion products in the Boreal North American, Temperate North American, Boreal Eurasia, Temperate Eurasia, and Europe regions. However, these regions show significant variations in carbon source estimates during fall and winter, attributed to differing observational constraints and localization strategies, leading to substantial discrepancies in the annual flux estimates. For instance, MCAS indicates relatively small annual fluxes in the Boreal North American region compared to OCO-2 MIP estimates, primarily due to higher carbon sources between January and April. Additionally, we observed notable disagreements in the seasonal cycle of flux amplitudes in the African and Australian regions. The primary cause of these discrepancies is the lack of in situ observations in these areas and divergent seasonal amplitudes of the prior fluxes used in the inversion process. Additionally, in Section S2, we performed a correlation analysis between these carbon flux products and the land ecosystem model. These results demonstrate the consistency of each carbon flux product with land ecosystem indicators.

Discussion

In this work, we develop the MCAS, which incorporates both in situ and satellite CO_2 measurements into the chemical transport model to obtain a more accurate carbon flux estimation. The main innovation of this paper is to address the challenge of heterogeneity between observations from different sources through a modified ensemble Kalman filter.

Despite the abundance of CO_2 observation types, primarily in situ and satellite observations, most current carbon assimilation efforts rely on observations from a single source, leading to under-utilization of available data. Although OCO-2 MIP v10 has explored the employment of multiple data sources, the heterogeneity among observational data caused the simultaneous assimilation of in situ and satellite observations to perform worse than using in situ data alone[18]. Our findings demonstrate that, by appropriately addressing data heterogeneity, jointly assimilating both observation types yields better results than assimilating a single source alone. Specifically, based on the RMSE evaluation from an independent in situ validation dataset, the error of the MCAS when assimilating both types of data was more than 15% lower than those assimilating data from a single source. The net land flux and ocean fluxes estimated using MCAS are consistent with those of the global carbon project. These provide evidence for a new avenue for a more precise estimation of carbon flux.

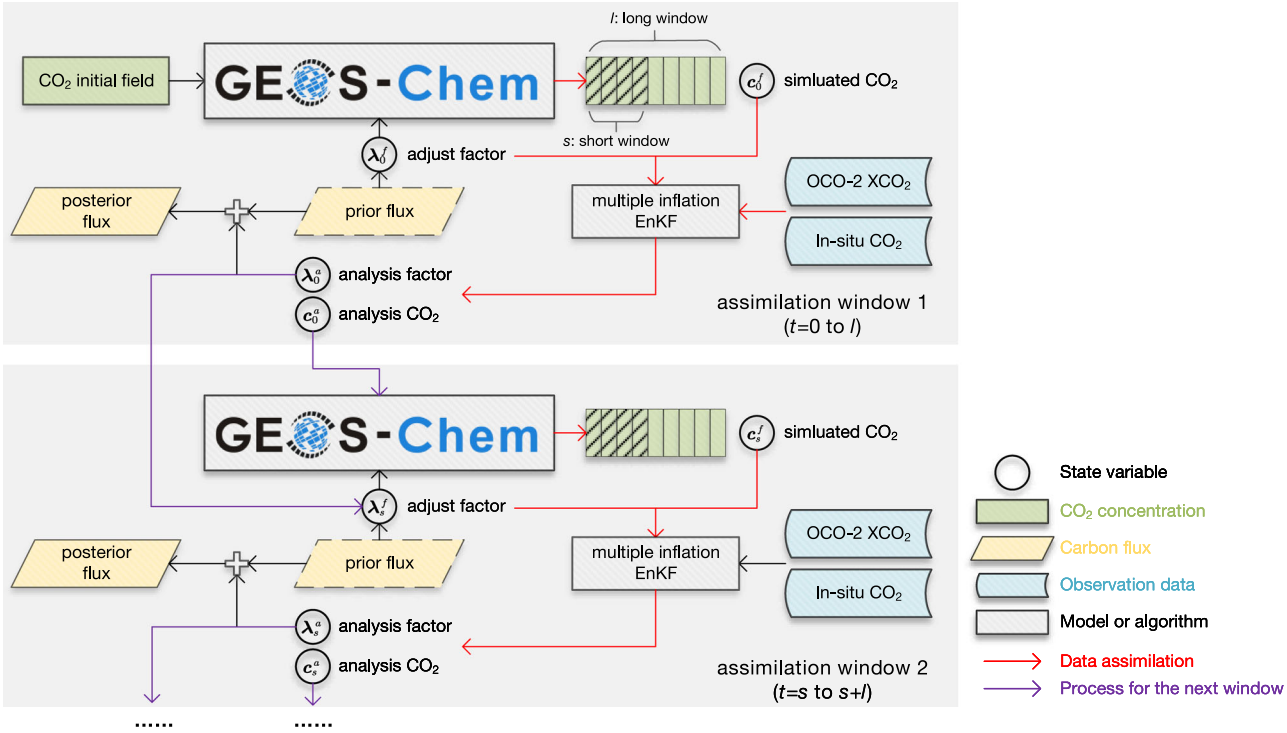

Fig. 4 | Flowchart of the MCAS. Schematic of the first two assimilation windows of MCAS. Red arrows represent the data assimilation steps, while purple arrows indicate the transfer steps between windows.

Our analysis indicates that high-quality data assimilation algorithms are crucial for accurate carbon flux inversion. We have enhanced the ensemble Kalman filter by introducing multiple inflation factors to address the heterogeneity among multi-source observations. Our framework can be easily extended to incorporate more types of observation, such as simultaneously assimilating data from different carbon satellites. In future research, we plan to use more advanced high-dimensional ensemble Kalman filter algorithms[31] to further tackle the challenges of ultra-high-dimensional state variables. We also intend to introduce the Bayesian model averaging method[32] to handle uncertainties in the prior carbon flux and atmospheric transport models, aiming to achieve better carbon assimilation results.

Methods

Transport model and prior carbon fluxes

GEOS-Chem is a global three-dimensional chemical transport model,[33] (https://geos-chem.seas.harvard.edu/, last access: Oct 2023) driven by meteorological input from Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office. GEOS-Chem has become one of the most widely utilized chemical transport models, and is widely applied in atmospheric inversion for carbon flux estimation[7,10,11,19,34]. Our GEOS-Chem simulation was performed at a spatial resolution of 2° latitude × 2. 5° longitude with 47 vertical layers, driven by MERRA-2 meteorological data.

As commonly practiced, we divide CO_2 fluxes into four distinct modules: fossil fuel, wildfire, ocean, and terrestrial biosphere. Our focus is particularly on refining the terrestrial ecosystem and oceanic carbon flux estimates to better match observational data. Hence, we treated fossil fuel and wildfire fluxes as fixed throughout the analysis. In each module, we consider two different priors, which lead to eight points (datasets) in total for the four modules. They are obtained from the NOAA Global Monitoring Laboratory's data archive (https://gml.noaa.gov/aftp/products/carbontracker/co2/CT2022/) with the following specifics.

- The fossil fuel component was derived from the Carbon Dioxide Information and Analysis Center (CDIAC,[35]) and the Open-source Data Inventory of Anthropogenic CO_2 (ODIAC) national emissions[36].
- The wildfire fluxes were sourced from the Global Fire Emissions Database (GFED) 4.1s[37,38]. Alternatively, wildfire fluxes from the NASA Carbon Monitoring System produced by the GFED_CMS team were also used.
- The first-guess ocean-atmosphere CO_2 was obtained from the *p*CO2-Clim[39] and the Ocean Inversion Fluxes (OIF,[40]).
- The terrestrial biosphere fluxes at a 3-h resolution were provided by two biosphere models (GFED 4.1s and GFED_CMS), which are versions of the Carnegie-Ames Stanford Approach (CASA) biogeochemical model[41], associated with the GFED project.

The eight prior carbon fluxes are categorized into two group combinations: (i) CDIAC for fossil fuel, GFED 4.1s for wildfire, GFED 4.1s for terrestrial ecosystem and OIF for the ocean, and (ii) ODIAC for fossil fuel, GFED_CMS for wildfire, GFED_CMS for terrestrial biosphere and *p*CO2-clim for the ocean. Each of the group combinations is used in a single inversion experiment. Specifically, for each of the experiments, fossil fuel and wildfire parts are treated as CO_2 emissions in GEOS-Chem, and they are fixed during the inversion experiment. Additionally, the prior estimates for the terrestrial biosphere and ocean serve as the initial guesses in GEOS-Chem and will be adjusted during the data assimilation to obtain posterior estimates of finer carbon fluxes for the terrestrial ecosystem and the ocean. Specifically, we consider using scaling factors $\boldsymbol{\lambda}_t = (\lambda_{\text{bios}}(t), \lambda_{\text{ocean}}(t))$, which have a mean value of 1, to adjust the prior carbon fluxes. The perturbed surface flux at location s and time t is given by:

$$F(\boldsymbol{\lambda}_t, s, t) = \lambda_{\text{bios}}(s,t)F_{\text{bios}}(s,t) + \lambda_{\text{ocean}}(s,t)F_{\text{ocean}}(s,t) + F_{\text{fossil}}(s,t) + F_{\text{fire}}(s,t), \quad (1)$$

where $F_{\text{bios}}(s, t)$, $F_{\text{ocean}}(s, t)$, $F_{\text{fossil}}(s, t)$, and $F_{\text{fire}}(s, t)$ describe the prior estimates introduced above. The scaling factors λ_{bios} and λ_{ocean} are used to adjust the carbon fluxes from the land biosphere and ocean, respectively. The allocation of adjustment factors $\boldsymbol{\lambda}_t$ is guided by a framework that takes into account 11 Transcom regions[42] and 19 Olson ecosystem types[43,44]. By combining the GEOS-Chem model and observational data, the Ensemble

Kalman Filter (EnKF) introduced in section "Ensemble Kalman Filter with multiple inflation and localization" will help us determine the optimal $\boldsymbol{\lambda}_t$.

The so-called "Prior" carbon flux in section "Assimilation of CO2 observations from in-situ and satellite" is the average of the two distinct prior group combinations. And the final results of a specific product, e.g., MCAS(in situ), is the average outcomes of the product using the two group combinations. The main results of the inversions using two sets of priors are documented in Table S3.

Observational data

- In situ observation: In this study, we use atmospheric CO_2 concentration measurements collected and preprocessed by Observation Package (ObsPack) data products, as detailed in[45]. The specific datasets employed in our research are obspack_co2_1_GLOBALVIEWplus_v9.0_2023-09-09[46], (https://gml.noaa.gov/ccgg/obspack/, last access: Oct 2023), and the ninth major release of the GLOBALVIEWplus (GV+) collaborative data products. The latter extensive data product comprises 612 datasets contributed by 74 laboratories from 28 countries.
- Satellite observation: The Orbiting Carbon Observatory-2 (OCO-2) is a satellite for environmental science research launched on July 2 2014. The NASA mission serves as a replacement for the earlier Orbiting Carbon Observatory. OCO-2 operates in a near-polar, Sun-synchronous orbit, allowing it to cover the entire Earth over a period of 16 days. We use the bias-corrected XCO_2 retrievals obtained from OCO-2 Level 2 data (https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_11.1r/summary, last access: Oct 2023) as the satellite observation in this study.

The original observational data were preprocessed and thinned to a lower spatial resolution to match the resolution of GEOS-Chem and reduce the excessive amount of samples. Please refer to Section S3 for the data preprocessing procedures and see Fig. S5 for the distribution of the in situ data we used.

Validation data

To assess the performance of the carbon flux inversion methods, we use independent observations as validation samples to objectively assess the performance of the methods. They include a 5% randomly selected data sample from ObsPack, which was not used in the data assimilation process. A similar approach to CT2022 is adopted to ensure the independence between the validation dataset and the assimilation dataset. For aircraft flasks collected during flights, the basic sampling unit is a flight so that CO_2 profiles during flights are selected randomly for inclusion in the validation sample. Similarly, for in situ analyzers with quasi-continuous readings, we randomly choose 5% daily recordings of 24-h to the validation sample.

Another validation dataset is from the Total Carbon Column Observing Network (TCCON,[47], https://tccondata.org/, last access: Oct 2023), a network of ground-based Fourier Transform Spectrometers that record direct solar spectra in the near-infrared spectral region. The column-average XCO_2 data were pre-processed as OCO-2 Level 2 data and the entire TCCON dataset after processing was used as the second validation sample. All TCCON sites used in the evaluation section are listed in Table S6.

Ensemble Kalman filter with multiple inflation and localization

Since accurately estimating the optimal parameters $\boldsymbol{\lambda}_t$ in Eq. (1) is challenging, there exists considerable uncertainty in the final calculation of the surface carbon flux. The emergence of multi-source CO_2 observations presents an opportunity to enhance the assessment of carbon fluxes through optimized parameter estimation.

To effectively utilize heterogeneous data from different sources, we have developed a novel global carbon flux estimation framework, the MCAS, which simultaneously integrates multi-source data. Within the MCAS framework, a modified Ensemble Kalman Filter (EnKF) algorithm is employed to combine model forecasts with observational data, thereby identifying optimized values for the model state variables.

Assuming m ensemble members, the model forecast state for the i-th member is defined as $\boldsymbol{x}_{t,i}^f = (\boldsymbol{c}_{t,i}^f, \boldsymbol{\lambda}_{t,i}^f)$ for $i = 1, 2, \ldots, m$, where $\boldsymbol{c}_{t,i}^f$ represents the CO_2 concentration field at time t, forecasted by GEOS-Chem based on the analyzed CO_2 concentration field at time $t-1$ (denoted as $\boldsymbol{c}_{t-1,i}^a$) and a given perturbed carbon flux $F(\boldsymbol{\lambda}_{t,i}^f)$ (see Eq. (1)). This process is expressed as follows.

$$\boldsymbol{c}_{t,i}^f = \mathcal{M}_t(\boldsymbol{c}_{t-1,i}^a; F(\boldsymbol{\lambda}_{t,i}^f)), \tag{2}$$

where $\mathcal{M}_t$ represents the GEOS-Chem model. The observation state at time t is expressed as $\boldsymbol{y}_t = (\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)})$, where $\boldsymbol{y}_t^{(1)}$ and $\boldsymbol{y}_t^{(2)}$ represent in situ and satellite observations, respectively. For each ensemble member, we can use an observation operator $\mathbf{H}_t = (\mathbf{H}_t^{(1)}, \mathbf{H}_t^{(2)})$ to interpolate the model grid to the observation locations, i.e., $\hat{\boldsymbol{y}}_{t,i}^f = \mathbf{H}_t \boldsymbol{x}_{t,i}^f$, and calculate the corresponding innovation $\boldsymbol{d}_{t,i} = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t,i}^f$. Here, $\mathbf{H}_t^{(1)}$ and $\mathbf{H}_t^{(2)}$ correspond to the observation operators for in situ data and satellite data, respectively. The former is implemented using linear interpolation, while the latter follows the form used in the ACOS retrieval algorithm[48]. For details, please refer to Section S4.

By combining ensemble forecasts and observational data, we solve the analysis state $\boldsymbol{x}_{t,i}^a$ using a modified EnKF update formula:

$$\boldsymbol{x}_{t,i}^a = \boldsymbol{x}_{t,i}^f + \begin{pmatrix} \sqrt{\theta_t}\tilde{\mathbf{C}}_t^f \\ \tilde{\boldsymbol{\Lambda}}_t^f \end{pmatrix} \left[(m-1)\mathbf{I} + \theta_t \tilde{\mathbf{Y}}_t^\top \begin{pmatrix} \mu_{t,1}\mathbf{R}_t^{(1)} & \\ & \mu_{t,2}\mathbf{R}_t^{(2)} \end{pmatrix}^{-1} \tilde{\mathbf{Y}}_t \right]^{-1} \sqrt{\theta_t}\tilde{\mathbf{Y}}_t^\top \begin{pmatrix} \mu_{t,1}\mathbf{R}_t^{(1)} & \\ & \mu_{t,2}\mathbf{R}_t^{(2)} \end{pmatrix}^{-1} \boldsymbol{d}_{t,i}, \tag{3}$$

for $i = 1, 2, \cdots, m$, where $\tilde{\mathbf{C}}_t^f$, $\tilde{\boldsymbol{\Lambda}}_t^f$, and $\tilde{\mathbf{Y}}_t$ are the de-centralized matrices of the CO_2 concentration field $\boldsymbol{c}_{t,i}^f$, scales $\boldsymbol{\lambda}_{t,i}^f$, and forecasted observations $\hat{\boldsymbol{y}}_{t,i}^f$, respectively. For example,

$$\tilde{\mathbf{Y}}_t = \left[\hat{\boldsymbol{y}}_{t,1}^f - \bar{\boldsymbol{y}}_t^f, \hat{\boldsymbol{y}}_{t,2}^f - \bar{\boldsymbol{y}}_t^f, \cdots, \hat{\boldsymbol{y}}_{t,m}^f - \bar{\boldsymbol{y}}_t^f \right], \bar{\boldsymbol{y}}_t^f = \frac{1}{m}\sum_{i=1}^m \hat{\boldsymbol{y}}_{t,i}^f. \tag{4}$$

The effective implementation of EnKF relies on the accurate estimation of forecast and observation error covariance matrices. Since the forecast error can be decomposed into ensemble error and model error, i.e.,

$$\underbrace{\mathrm{Var}\left(\boldsymbol{c}_{t,i}^f - \boldsymbol{c}_t^*\right)}_{\text{forecast error}} = \underbrace{\mathrm{Var}\left(\boldsymbol{c}_{t,i}^f - \bar{\boldsymbol{c}}_t^f\right)}_{\text{ensemble error}} + \underbrace{\mathrm{Var}\left(\bar{\boldsymbol{c}}_t^f - \boldsymbol{c}_t^*\right)}_{\text{model error}}, \tag{5}$$

where $\boldsymbol{c}_t^*$ is the unknown true CO_2 concentration field, and $\bar{\boldsymbol{c}}_t^f = \frac{1}{m}\sum_{i=1}^n \boldsymbol{c}_{t,i}^f$ is the forecast mean. Since the true state is unknown, the standard EnKF uses $\bar{\boldsymbol{c}}_t^f$ instead, which omits the model error term. Given that geophysical models typically have model errors, this omission can lead to a significant underestimation of forecast error statistics, resulting in substantial analysis errors. MCAS compensates for this omission by introducing an inflation factor θ_t. Systems like CarbonTracker do not use the chemical transport model as a forecast operator but rather as an observation operator, leading to a quite different methodology from the MCAS. The need for inflation is further explained in Section S5.

The observation error in our system arises from interpolation errors between model grid points and observation locations, as well as instrument errors. Although observational data provide uncertainty estimates based on "model-data mismatch" (e.g., "CT_MDM" in ObsPack and "xco2_uncertainty" in OCO-2 Level 2 data), the accuracy of these estimates varies significantly depending on the physical models used and the data provider.

In Eq. (3), we modified the standard EnKF by introducing inflation factors θ_t, $\mu_{t,1}$, and $\mu_{t,2}$ to adjust the forecast and observation errors. These

factors help balance different error sources and enhance assimilation performance. Unlike previous works that used fixed inflation factors[10,11], our method employs dynamic inflation factors that account for spatiotemporal variations. These dynamic factors are automatically estimated by maximizing the likelihood function. For detailed implementation of MCAS, please refer to the SI.

To mitigate biases introduced by global-scale data assimilation, particularly in sparsely observed areas, we implemented a global localization approach similar to ref. 32. The Earth was divided into seven regions: Ocean, South America, North America, Africa, Eurasia, Tropical Asia, and Australasia (see Fig. S6). Each region was extended by 800 km to create overlaps with adjacent areas.

Observations were assimilated within regions encompassing measurement sites located within the extended areas and those that passed a correlation test. This test evaluated Pearson's correlation coefficient between the ensemble of parameters $\boldsymbol{\lambda}_t$ for a specific region and the ensemble of forecasts at a measurement site outside that region. Observations with an absolute correlation coefficient greater than 0.7 were also included in the data assimilation.

The extension of regional boundaries ensured the continuity of analysis of states along borders. For a grid point g within the overlap of regions A and B, the final result was calculated using inverse-distance weighting:

$$\boldsymbol{x}_t^a(g) = \frac{\boldsymbol{x}_t^a(g,\mathrm{A}) \cdot d(g,\mathrm{A}) + \boldsymbol{x}_t^a(g,\mathrm{B}) \cdot d(g,\mathrm{B})}{d(g,\mathrm{B}) + d(g,\mathrm{B})}, \tag{6}$$

where $d(g,\mathrm{A})$ is the Euclidean distance from grid point g to the nearest point in extended region A, and $\boldsymbol{x}_t^a(g,\mathrm{A})$ is the analysis state of grid point g within region A.

MCAS configurations

The proposed MCAS was used to estimate carbon fluxes for ecosystems and the ocean-spanning the years 2016 to 2020. To underscore the efficacy of leveraging multi-source observations, we conducted three inversion experiments: (1) MCAS(in situ): Inversion using only in situ observations; (2) MCAS(Sat): Inversion using only satellite observations. (3) MCAS(in situ&Sat): Inversion using both in situ and satellite observations.

Since the impact of carbon flux on the spatial distribution of CO_2 is not instantaneous, the moving window method[10,49], a common practice in data assimilation, was adopted. Given that satellite observations of column-wise concentrations exhibit a longer response time to carbon flux impacts, different window lengths were designed for different observation sources. We extensively discuss the configuration of MCAS and sensitivity analysis in the Section S6. Based on this analysis, we establish the following settings: In Experiment (1), each assimilation step moves 7 days, with an additional 4 days (total of 11 days) for in situ observations considered. In Experiments (2) and (3), each assimilation step also moves 7 days, with an additional 4 days for in situ observations and an extra 9 days (total of 16 days) for satellite observations. The inversion experiments started in July 2015 and ended on December 31, 2020, with the first six months as the breeding period.

To more clearly illustrate the workflow and features of MCAS, we provide a flowchart of MCAS in Fig. 4. Note that although we only describe MCAS for two data sources, our method can be easily extended to more data sources. We discussed the setting of different inflation factors for satellite observations over the ocean and land in Section S6.

Data availability

The MCAS inversion carbon fluxes are available on Zenodo (https://doi.org/10.5281/zenodo.13937731). The ninth major release of GLOBALVIEWplus(GV+) in the Observation Package product (ObsPack) is available on the NOAA/ESRL Global Monitoring Laboratory website[46], (https://gml.noaa.gov/ccgg/obspack/). The OCO-2 Level 2 bias-corrected XCO_2 retrievals are available on the NASA Earth Science Data Systems (ESDS) Program (https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_11.1r/summary). CT2022 products are available on the official website https://gml.noaa.gov/ccgg/carbontracker/. TCCON data are available on the official website of the Total Carbon Column Observing Network (https://tccondata.org/). BEPS NEP data is provided by the National Ecosystem Science Data Center, National Science & Technology Infrastructure of China. (http://www.nesdc.org.cn).

Received: 12 May 2024; Accepted: 25 October 2024;
Published online: 21 November 2024

References

1. Zhao, C. L. & Tans, P. P. Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air. *J. Geophys. Res. Atmos.* **111**, D08S09 (2006).
2. Eldering, A. et al. The orbiting carbon observatory-2 early science investigations of regional carbon dioxide fluxes. *Science* **358**, eaam5745 (2017).
3. Crowell, S. et al. The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network. *Atmos. Chem. Phys.* **19**, 9797–9831 (2019).
4. Peylin, P. et al. Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions. *Biogeosciences* **10**, 6699–6720 (2013).
5. Zhang, S. et al. A global carbon assimilation system using a modified ensemble Kalman filter. *Geosci. Model Dev.* **8**, 805–816 (2015).
6. Philip, S. et al. Prior biosphere model impact on global terrestrial CO_2 fluxes estimated from OCO-2 retrievals. *Atmos. Chem. Phys.* **19**, 13267–13287 (2019).
7. Liu, Z. et al. Improving the joint estimation of CO_2 and surface carbon fluxes using a constrained ensemble Kalman filter in COLA (v1.0). *Geosci. Model Dev.* **15**, 5511–5528 (2022).
8. Jiang, F. et al. Regional CO_2 fluxes from 2010 to 2015 inferred from GOSAT XCO_2 retrievals using a new version of the global carbon assimilation system. *Atmos. Chem. Phys.* **21**, 1963–1985 (2021).
9. Philip, S. et al. OCO-2 satellite-imposed constraints on terrestrial biospheric CO2 fluxes over south Asia. *J. Geophys. Res. Atmos.* **127**, e2021JD035035 (2022).
10. Kong, Y., Zheng, B., Zhang, Q. & He, K. Global and regional carbon budget for 2015–2020 inferred from OCO-2 based on an ensemble Kalman filter coupled with GEOS-Chem. *Atmos. Chem. Phys.* **22**, 10769–10788 (2022).
11. Kou, X. et al. The carbon sink in China as seen from GOSAT with a regional inversion system based on the community multi-scale air quality (CMAQ) and ensemble Kalman smoother (EnKS). *Atmos. Chem. Phys.* **23**, 6719–6741 (2023).
12. Liang, A., Gong, W., Han, G. & Xiang, C. Comparison of satellite-observed XCO2 from GOSAT, OCO-2, and ground-based TCCON. *Remote Sens.* **9**, 1033 (2017).
13. Chen, Y. et al. Global-scale evaluation of xco2 products from gosat, oco-2 and carbontracker using direct comparison and triple collocation method. *Remote Sens.* **14**, 5635 (2022).
14. Yang, H., Li, T., Wu, J. & Zhang, L. Inter-comparison and evaluation of global satellite XCO2 products. *Geo-spatial Inf. Sci.* https://doi.org/10.1080/10095020.2023.2252017 (2023).
15. Yi, Y. et al. Measuring and comparing in-situ CO_2 and CO profiles with satellite observations and model data. *Atmos. Ocean. Sci. Lett.* **12**, 444–450 (2019).
16. Yuan, Y. et al. Comparison of continuous in-situ CO_2 measurements with co-located column-averaged XCO2 TCCON/satellite observations and carbontracker model over the zugspitze region. *Remote Sens.* **11**, 2981 (2019).
17. Mustafa, F. et al. Validation of GOSAT and OCO-2 against in situ aircraft measurements and comparison with carbontracker and GEOS-chem over Qinhuangdao, China. *Remote Sens.* **13**, 899 (2021).

18. Byrne, B. et al. National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake. *Earth Syst. Sci. Data* **15**, 963–1004 (2023).
19. Feng, L. et al. Consistent regional fluxes of CH_4 and CO_2 inferred from GOSAT proxy XCH_4:XCO_2 retrievals, 2010–2014. *Atmos. Chem. Phys.* **17**, 4781–4797 (2017).
20. Liu, S. et al. Integration of surface-based and space-based atmospheric CO2 measurements for improving carbon flux estimates using a new developed 3-GAS inversion model. *Atmos. Res.* **307**, 107477 (2024).
21. Liang, X. et al. Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Q. J. R. Meteorol. Soc.* **138**, 263–273 (2012).
22. Jacobson, A. R. et al. *CarbonTracker CT2022*. https://doi.org/10.25925/Z1GJ-3254 (2023).
23. Friedlingstein, P. et al. Global carbon budget 2023. *Earth Syst. Sci. Data* **15**, 5301–5369 (2023).
24. Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P. & White, J. W. C. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature* **488**, 70–72 (2012).
25. Buchwitz, M. et al. Computation and analysis of atmospheric carbon dioxide annual mean growth rates from satellite observations during 2003–2016. *Atmos. Chem. Phys.* **18**, 17355–17370 (2018).
26. Gaubert, B. et al. Neutral tropical African CO2 exchange estimated from aircraft and satellite observations. *Glob. Biogeochem. Cycles* **37**, e2023GB007804 (2023).
27. Walker, X. J. et al. Increasing wildfires threaten historic carbon sink of boreal forest soils. *Nature* **572**, 520–523 (2019).
28. Zhao, B. et al. North American boreal forests are a large carbon source due to wildfires from 1986 to 2016. *Sci. Rep.* **11**, 7723 (2021).
29. Mirabel, A., Girardin, M. P., Metsaranta, J., Way, D. & Reich, P. B. Increasing atmospheric dryness reduces boreal forest tree growth. *Nat. Commun.* **14**, 6901 (2023).
30. Zheng, B. et al. Record-high CO_2 emissions from boreal fires in 2021. *Science* **379**, 912–917 (2023).
31. Sun, H.-X., Wang, S., Zheng, X. & Chen, S. X. High-dimensional ensemble Kalman filter with localization, inflation, and iterative updates. *Q. J. R. Meteorol. Soc.* https://doi.org/10.1002/qj.4846 (2024).
32. Zhang, S. et al. Global carbon assimilation system using a local ensemble Kalman filter with multiple ecosystem models. *J. Geophys. Res. Biogeosci.* **119**, 2171–2187 (2014).
33. Bey, I. et al. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res. Atmos.* **106**, 23073–23095 (2001).
34. Wang, J. et al. Large chinese land carbon sink estimated from atmospheric carbon dioxide data. *Nature* **586**, 720–723 (2020).
35. Hefner, M. & Marland, G. Global, regional, and national fossil-fuel CO_2 emissions: 1751–2020. https://energy.appstate.edu/research/work-areas/cdiac-appstate (2023).
36. 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.* **11**, 543–556 (2011).
37. Giglio, L., Randerson, J. T. & van der Werf, G. R. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *J. Geophys. Res. Biogeosci.* **118**, 317–328 (2013).
38. van der Werf, G. R. et al. Global fire emissions estimates during 1997–2016. *Earth Syst. Sci. Data* **9**, 697–720 (2017).
39. Takahashi, T. et al. Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans. *Deep Sea Res. Part II Topical Stud. Oceanogr.* **56**, 554–577 (2009).
40. Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L. & Gloor, M. A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes. *Global Biogeochem. Cycles.* https://doi.org/10.1029/2006GB002703 (2007).
41. Potter, C. S. et al. Terrestrial ecosystem production: a process model based on global satellite and surface data. *Glob. Biogeochem. Cycles* **7**, 811–841 (1993).
42. Gurney, K. R. et al. Transcom 3 inversion intercomparison: model mean results for the estimation of seasonal carbon sources and sinks. *Global Biogeochem. Cycles* **18**, GB1010 (2004).
43. Olson, J. S., Watts, J. A. & Allison, L. J. Major world ecosystem complexes ranked by carbon in live vegetation: a database (1985).
44. Gibbs, H. K. *Olson's Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: An Updated Database Using the GLC2000 Land Cover Product* (NDP-017b, a 2006 update of the original 1985 and 2001 data file) (2006).
45. Masarie, K. A., Peters, W., Jacobson, A. R. & Tans, P. P. ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements. *Earth Syst. Sci. Data* **6**, 375–384 (2014).
46. Schuldt, K. N. et al. Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2022; obspack_co2_1_GLOBALVIEWplus_v9.0_2023-09-09 [Dataset] (2023).
47. Laughner, J. L. et al. The total carbon column observing network's ggg2020 data version. *Earth Syst. Sci. Data* **16**, 2197–2260 (2024).
48. O'Dell, C. W. et al. The ACOS CO_2 retrieval algorithm—part 1: description and validation against synthetic observations. *Atmos. Meas. Tech.* **5**, 99–121 (2012).
49. Peters, W. et al. An atmospheric perspective on North American carbon dioxide exchange: carbontracker. *Proc. Natl. Acad. Sci. USA* **104**, 18925–18930 (2007).

Acknowledgements

This research was supported by the National Natural Science Foundation of China grant nos. 12292980, 12292983, and 92358303. We acknowledge the High-Performance Computer resources from the National Key Scientific and Technological Infrastructure project "Earth System Numerical Simulation Facility" (EarthLab).

Author contributions

S.X.C. and X.Z. designed the study. W.S. built and ran the data assimilation system. W.S. analyzed the results and wrote the manuscript, and B.W. and H.C. participated in it. L.Z. and W.S. analyzed the OCO-2 satellite data. All authors participated in the discussions and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-024-00824-w.

Correspondence and requests for materials should be addressed to Song Xi Chen.