Supporting Information for

Investigating the differences in calculating global mean surface CO_2 abundance: the impact of analysis methodologies and site selection

Zhendong Wu[1,2,*], Alex Vermeulen[2,*], Yousuke Sawa[3], Ute Karstens[1,2], Wouter Peters[4,5], Remco de Kok[4], Xin Lan[6,7], Yasuyuki Nagai[3], Akinori Ogi[3], Oksana Tarasova[8]

[1]ICOS Carbon Portal at Lund University, Department of Physical Geography and Ecosystem Sciences, Lund, Sweden
[2]ICOS ERIC, Carbon Portal, Lund, Sweden
[3]Japan Meteorological Agency (JMA), Tokyo, Japan
[4]Wageningen University, Wageningen, The Netherlands
[5]University of Groningen, Groningen, The Netherlands
[6]NOAA Global Monitoring Laboratory (GML), Boulder, USA
[7]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, USA
[8]WMO, Geneva, Switzerland

Correspondence to: Alex Vermeulen (alex.vermeulen@icos-ri.eu), and Zhendong Wu (zhendong.wu@nateko.lu.se)

Contents of this file

Text S1 to S4

Figures S1 to S4

Tables S1 to S3

Text S1. The WDCGG global analysis method

The WDCGG method consists of seven separate steps. The full documentation can be found in Tsutsumi et al. (2009).

Step 1: Station selection based on traceability to the WMO standard scale

In order to avoid the potential biases that can be introduced by using different concentration scales, WDCGG only uses data from stations that report results traceable to the most recent CO_2 scale from the GAW Central Calibration Laboratories (CCL) assigned for that parameter. The current scale is the WMO standard scale WMO-CO2-X2019.

Step 2: Integration of parallel data from the same station

The WDCGG method uses continuous (hourly averaged) observations as these better represent the average concentrations compared to the flask-air samples taking during daytime once per two weeks. For remote stations where both flask and continuous data exist, NOAA found offsets between continuous and flask based monthly averages of 0.16-0.35 ppm (Tans et al., 1990), in less remote areas this difference can be expected to be larger. For selected stations flask data are used for gap filling when continuous data is lacking.

Step 3: Selection of stations suitable for global analysis

All of station data are normalized against the South Pole and averaged for the whole observation period. The normalized and averaged data points are plotted against latitude, and a curve is fitted by using a nearest-neighbour local-quadratic regression. The stations with normalized data locate outside the 3 standard deviations of the latitudinal fitted curve are excluded from the selection. This selection procedure is repeated until all stations in the selection locating within the 3 standard deviations of the latitudinal fitted curve. This procedure results in 139 stations remaining, which have a reasonable latitudinal scatter range (Figure 1).

Step 4: Abstraction of a station's average seasonal variation expressed by the Fourier harmonics

The average seasonal variation is obtained from the longest continuous segment of data by using three Fourier harmonics. Here is loop procedure where the following processes a-d are repeated until neither the long-term trend nor the average seasonal variation changes: a). de-trend original data, b). apply the harmonics to obtain seasonality, c). de-seasonality from original data to obtain long-term trend, d) smooth the long-term trend by using low-pass filter (a cut-off frequency of 0.48 cycle / year). After reaching this condition the average seasonal variation is determined and subtracted from the full data which leaves us with deseasonalized data that still can contain gaps.

Step 5: Interpolation of data gaps

The gaps of the deseasonalized data are filled by linear interpolation. Subsequently, the CO_2 time series without gaps is the sum of the interpolated trend and the average seasonality.

Step 6: Extrapolation for synchronization of data period

Extrapolate the long-term trend to the synchronization period and then add the average seasonal variation to obtain the synchronized data. This is an optional step that is excluded in this analysis.

Step 7: Calculation of the zonal and global mean mole fractions, trends, and growth rates.

Global and hemispheric means, trends and growth rates are calculated by area-weighted averaging the zonal means over each latitudinal band (30°). The growth rate is determined by taking the first derivative of the long-term trend.

Text S2. The CTE station network

290 stations are evaluated in the CTE inversion, the observations come from the ObsPack data product (Kenneth N., 2022). The measurement methods at the stations include surface-based, shipboard-based, tower-based and aircraft-based. In this study, we only focus on data derived from the first three measurement types (i.e. aircraft-based measurements are excluded), and in total 230 out of 290 stations are selected (Figure 1). For the stations that have both surface-based and tower-based measurements, we used the tower-based measurements for analysis. For the stations that have tower-based measurements, we selected the highest measurement.

Text S3. Calculation of atmospheric CO_2 mass

CTE simulates 3D CO_2 mole fraction with 25 levels in the vertical direction. The CO_2 mass at each level of the atmosphere can be calculated as a function of air mass and CO_2 concentration by weight.

$$m_{CO_2} = Cw_{CO_2} * m_{air} \tag{1}$$

where m_{CO_2} is the mass of the CO_2, kg. Cw_{CO_2} is the CO_2 concentration by weight, w %. m_{air} is the mass of the air, kg. CO_2 concentration by weight is obtained by the formula below:

$$Cw_{CO_2} = Cv_{CO_2} * \frac{M_{CO_2}}{M_{air}} \tag{2}$$

where Cv_{CO_2} is the mole fraction of CO_2 in air, mol / mol. According to the ideal gas assumption, equal volume of gases at same temperature and pressure contains equal number of moles regardless of chemical nature of gases, i.e. the CO_2 concentration by mole equals the CO_2 concentration by volume. M_{CO_2} is the CO_2 molar mass (44.009 g/mol). M_{air} is the average molar mass of dry air (28.9647 g / mol).

Pressure is the force applied perpendicular to the surface of an object, therefore, air pressure can be expressed by:

$$p_{air} = \frac{F_{air}}{S} \tag{3}$$

where p_{air} is the pressure of air, Pa or N / m^2. In this case, p_{air} is the difference of air pressure between adjacent level boundaries, e.g. air pressure at level 1 is $p_1 - p_2$. F_{air} is the magnitude of the normal force of air or gravity of air, N or kg m / s^2. The gravity of air at each level can be estimated by:

$$F_{air} = m_{air} * g \tag{4}$$

where g is the gravitational field strength, about 9.81 m / s^2 or N / kg.

S is the area of the surface, m^2. Here S is the area of grid cell at each level, increasing with geopotential height (gph). It is calculated as a function of latitude and longitude on earth's surface, radius of the earth (R), and gph.

$$S = 2 * \pi * (R + gph)^2 * |\sin(lat1) - \sin(lat2)| * \frac{|lon1 - lon2|}{360} \tag{5}$$

Where, $lat1$, $lat2$, $lon1$and $lon2$ are the boundary of grid cell. R = 6378.1370 km, here we use the equatorial radius which is the distance from earth’s center to the equator.

Hence the mass of the air in Eq. 1 can be estimated by:

$$m_{air} = \frac{p_{air} * S}{g} \quad (6)$$

Text S4. File list

All code necessary to calculate the global mean surface CO_2 mole fraction and Atmospheric CO_2 mass is freely available on ICOS Carbon Portal as a zipped archive (GAW_code.zip) [https://doi.org/10.18160/Q788-9081], when unzipped, the code include:

- fit_filter_seminoaa.ipynb

  Apply the semi-NOAA method to GAW observations (139 stations), CTE observations (230 stations), CTE model output at stations (230 stations) and CTE model output (full global)
- cal_zonal_global_co2_gaw_seminoaa.ipynb

  Calculate global co2 mole fraction average and its growth rate, and estimate their uncertainty, using output from GAW(semi-NOAA)
- cal_zonal_global_co2_gaw_wdcgg.ipynb

  Calculate global co2 mole fraction average and its growth rate, and estimate their uncertainty, using output from GAW(WDCGG)
- cal_zonal_global_co2_ctracker_obs.ipynb

  Calculate global co2 mole fraction average and its growth rate, and estimate their uncertainty, using output from CTE_obs(semi-NOAA)
- cal_zonal_global_co2_ctracker_model_sample.ipynb

  Calculate global co2 mole fraction average and its growth rate, and estimate their uncertainty, using output from CTE_output(semi-NOAA)
- cal_zonal_global_co2_ctracker_model_global.ipynb

  Calculate global co2 mole fraction average and its growth rate, and estimate their uncertainty, using output from CTE_global(semi-NOAA)
- cal_co2mass_co2ppm_cte_global.ipynb

  Calculate global co2 mole fraction and global atmospheric co2 mass, using the 3D co2 output from CTE model
- compare_co2_co2rate.ipynb

  Statistically compare the co2 mole fraction and its growth rate among different data sources and analysis methods
- plot_results.ipynb

  The script is used to analyze and plot the results in the paper.

In order to run the jupyter booknotes, it needs to download the data (GAW_data.zip) [https://doi.org/10.18160/Q788-9081].

The key results with CSV format are accessible on ICOS Carbon Portal as a zipped archive (GAW_results.zip) [https://doi.org/10.18160/Q788-9081], when unzipped, the data include:

- Global monthly and annual surface CO_2 mole fraction and its growth rate for 1980-2020 derived from the GAW observations by using the semi-NOAA method, i.e. GAW (semi-NOAA).

Global mean：

df_co2_annual_global_NH_SH_gaw_seminoaa.csv

df_co2_monthly_global_NH_SH_gaw_seminoaa.csv
df_co2rate_annual_global_NH_SH_gaw_seminoaa.csv
df_co2rate_monthly_global_NH_SH_gaw_seminoaa.csv

Their uncertainty basing on bootstrap method:

bootstats_co2_annual_global_gaw_seminoaa.csv
bootstats_co2_monthly_global_gaw_seminoaa.csv
bootstats_co2rate_annual_global_gaw_seminoaa.csv
bootstats_co2rate_monthly_global_gaw_seminoaa.csv

- Global monthly and annual surface CO_2 mole fraction and its growth rate for 1980-2020 derived from the GAW observations by using the WDCGG method without extrapolation, i.e. GAW (WDCGG).

Global mean:

df_co2_annual_global_NH_SH_gaw_wdcgg.csv
df_co2_monthly_global_NH_SH_gaw_wdcgg.csv
df_co2rate_annual_global_NH_SH_gaw_wdcgg.csv
df_co2rate_monthly_global_NH_SH_gaw_wdcgg.csv

Their uncertainty basing on bootstrap method:

bootstats_co2_annual_global_gaw_wdcgg.csv
bootstats_co2_monthly_global_gaw_wdcgg.csv
bootstats_co2rate_annual_global_gaw_wdcgg.csv
bootstats_co2rate_monthly_global_gaw_wdcgg.csv

- Global monthly and annual surface CO_2 mole fraction and its growth rate for 1980-2020 derived from the observations at the CTE 230 stations by using semi-NOAA method, i.e. CTE_obs (semi-NOAA).

Global mean:

co2obs_co2_annual_global_NH_SH_ct2021_obs.csv
co2obs_co2_monthly_global_NH_SH_ct2021_obs.csv
co2obs_co2rate_annual_global_NH_SH_ct2021_obs.csv
co2obs_co2rate_monthly_global_NH_SH_ct2021_obs.csv

Their uncertainty basing on bootstrap method:

bootstats_co2_annual_global_cal_ct2021_obs.csv
bootstats_co2_monthly_global_cal_ct2021_obs.csv
bootstats_co2rate_annual_global_cal_ct2021_obs.csv
bootstats_co2rate_monthly_global_cal_ct2021_obs.csv

- Global monthly and annual surface CO_2 mole fraction and its growth rate for 2001-2020 derived from the CTE model output sampling at the CTE 230 stations by using semi-NOAA method, i.e. CTE_output (semi-NOAA).

Global mean:

co2model_co2_annual_global_NH_SH_ct2021_modelsample.csv
co2model_co2_monthly_global_NH_SH_ct2021_modelsample.csv
co2model_co2rate_annual_global_NH_SH_ct2021_modelsample.csv
co2model_co2rate_monthly_global_NH_SH_ct2021_modelsample.csv

Their uncertainty basing on bootstrap method:

bootstats_co2_annual_global_cal_ct2021_modelsample.csv
bootstats_co2_monthly_global_cal_ct2021_modelsample.csv
bootstats_co2rate_annual_global_cal_ct2021_modelsample.csv
bootstats_co2rate_monthly_global_cal_ct2021_modelsample.csv

- Global monthly and annual surface CO_2 mole fraction and its growth rate for 2001-2020 derived from the CTE model output covers full global (averaged over the first three levels, 0 to 0.35 km Alt.) by using semi-NOAA method, i.e. CTE_global (semi-NOAA)

co2_annual_global_cte2021(level1-3)_seminoaa.csv
co2_monthly_global_cte2021(level1-3)_seminoaa.csv
co2rate_annual_global_cte2021(level1-3)_seminoaa.csv
co2rate_monthly_global_cte2021(level1-3)_seminoaa.csv

- Global monthly and annual surface CO_2 mole fraction for 2001-2020 derived from the CTE model output covers full global with different heights (i.e. level1-3 and level1-25).

cte2021(lv1-3)_co2_2000_2020_annual.csv
cte2021(lv1-3)_co2_2000_2020_monthly.csv
cte2021(lv1-25)_co2_2000_2020_annual.csv
cte2021(lv1-25)_co2_2000_2020_monthly.csv

- Global monthly and annual atmospheric CO_2 mass (up to ~200 km) for 2000-2020 derived from the CTE model output by using the method described in Text S3.

cte2021_co2mass_2000_2020_monthly.csv
cte2021_co2mass_2000_2020_annual.csv

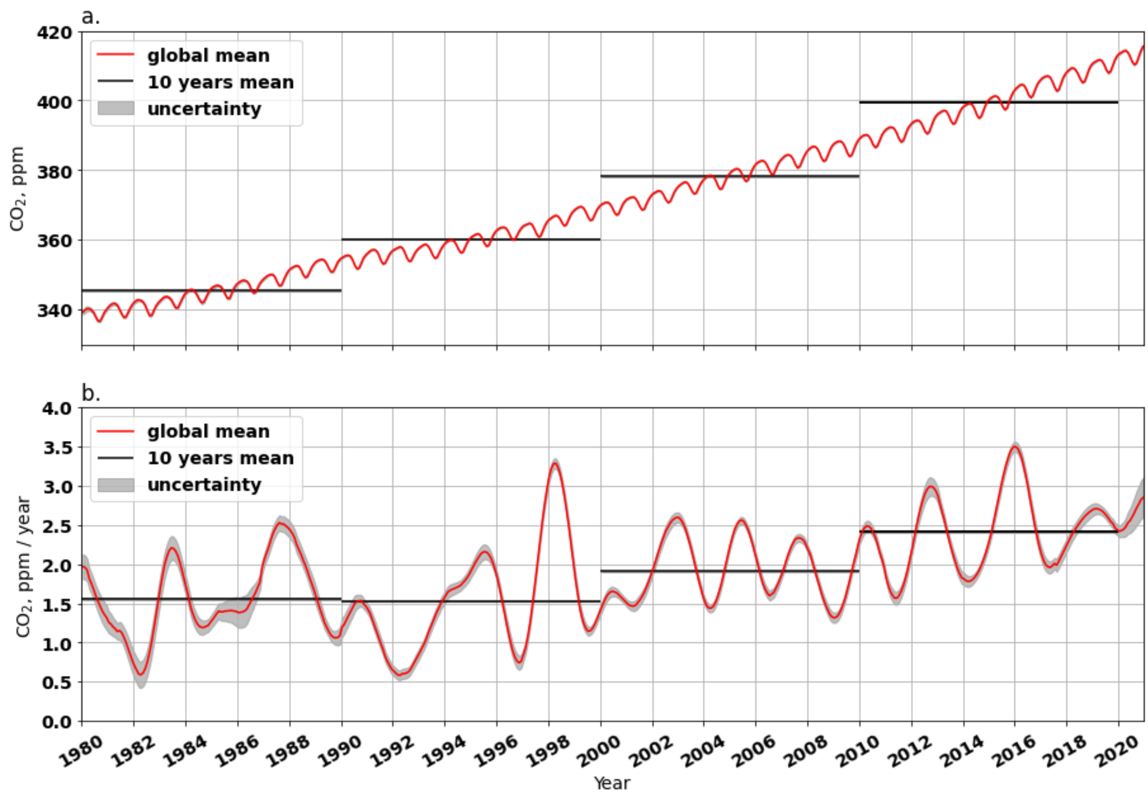

Figure S1. Globally averaged CO_2 mole fraction (a) and its G_{ATM} (b) from 1980 to 2021. In panel (a), the red line shows the mean CO_2 mole fraction, black lines show the mean CO_2 mole fraction over 10 years, the grey area shows the uncertainty derived from the 200 bootstrap networks. Similarly, panel (b) shows the G_{ATM} instead of the mole fraction. The CO_2 and its G_{ATM} results are derived from the GAW observations from 139 stations by using semi-NOAA method.

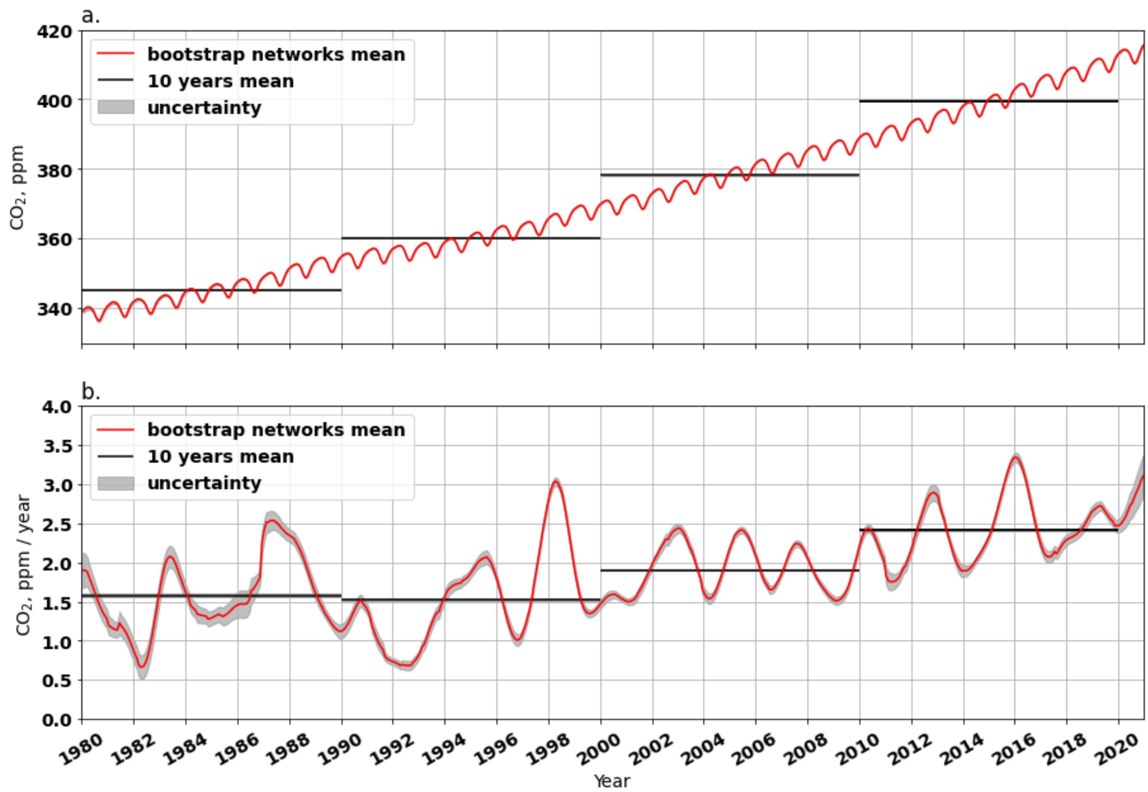

Figure S2. Globally averaged CO_2 mole fraction (a) and its G_{ATM} (b) from 1980 to 2021. In panel (a), black lines show the mean CO_2 mole fraction over 10 years, the grey lines show the 200 bootstrap networks, the red line

shows the mean of the 200 bootstrap networks. Similarly, panel (b) shows the G_{ATM} results instead of CO_2 mole fraction. This result is derived from the GAW observations from 139 stations by using WDCGG method.

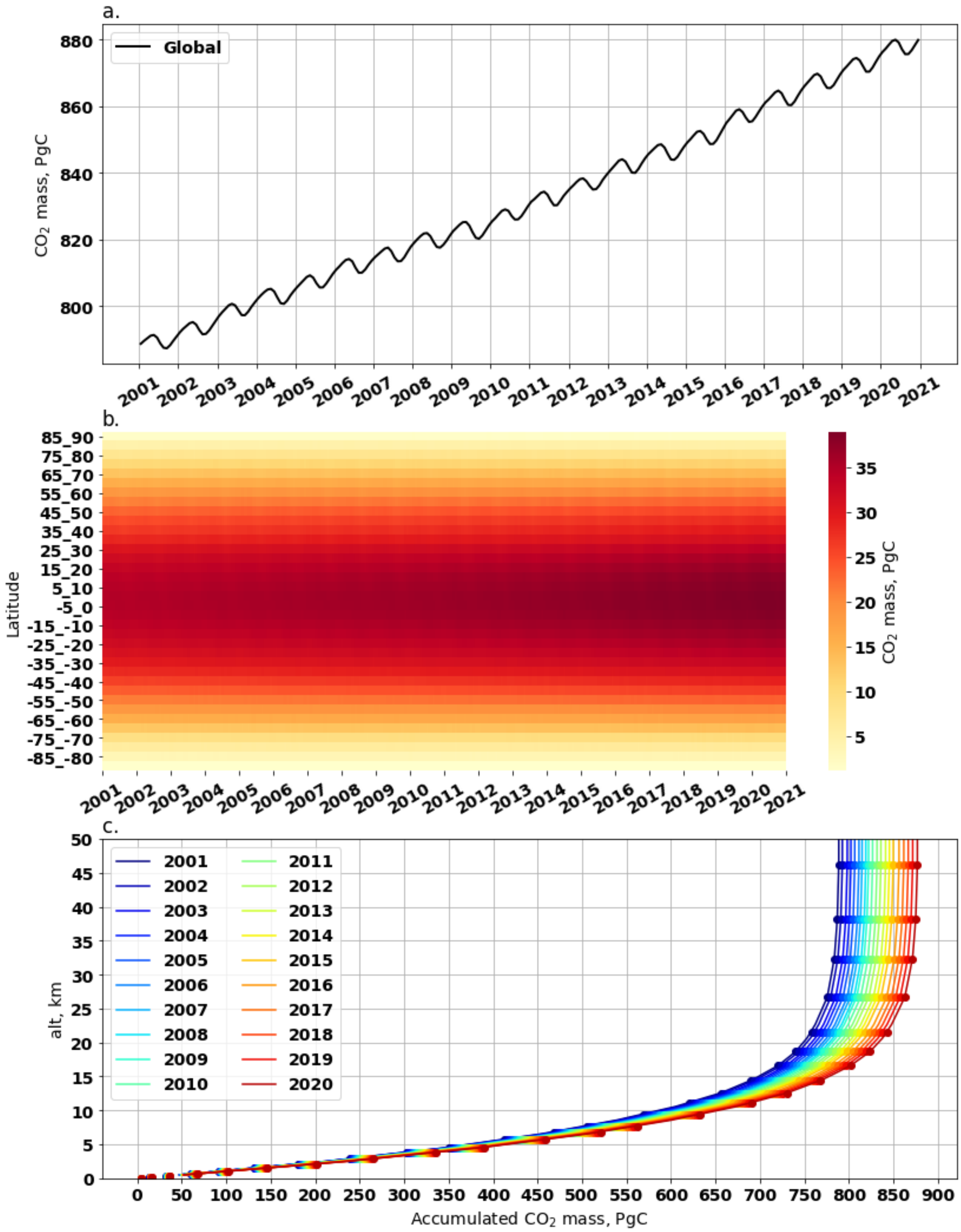

Figure S3. Atmospheric CO_2 mass derived from CTE output. Panel (a) shows the global monthly CO_2 mass in atmosphere (from surface up to 200 km altitude). Panel (b) shows the zonal (5°) average of monthly CO_2 mass. Panel (c) shows accumulated CO_2 mass with altitudes from 2001 to 2020, the dots mark CTE vertical level altitudes and lines are the linear interpolation between the altitudes.

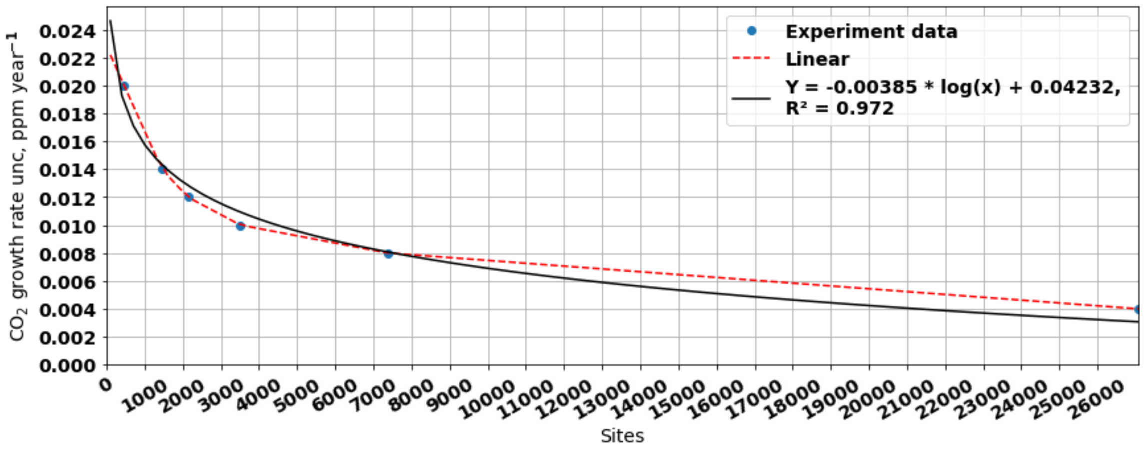

Figure S4. The relationship between the uncertainty of the global CO_2 growth rate and the number of observation sites. The relationship is estimated using CTE_global (all global grids excluding ocean grids) with different resolutions (1x1, 2x2, 3x3, 4x4, 5x5, and 10x10 degrees) to estimate the uncertainty of the global CO_2 growth rate. The bootstrap method mentioned in the main text is used to estimate the uncertainty, and the results are represented as blue dots. The red dashed line shows the linear interpolation between the experimental results, while the black line shows an exponential curve fitting.

	a. GAW (WDCGG), 1980-2020				b. NOAA, 1980-2020			
	Annual		Monthly		Annual		Monthly	
Statistic	CO_2	G_{ATM}	CO_2	G_{ATM}	CO_2	G_{ATM}	CO_2	G_{ATM}
r	0.999	0.991	0.999	0.987	0.999	0.980	0.999	0.970
RMSE	0.053	0.081	0.145	0.108	0.352	0.121	0.519	0.162
MAE	0.043	0.070	0.114	0.086	0.329	0.094	0.449	0.129
ME	0.007	0.005	0.007	0.005	-0.329***	-0.025	-0.329***	-0.025***
	c. CTE_obs (semi-NOAA), 1980-2020				**d. CTE_obs (semi-NOAA), 2001-2020**			
r	0.999	0.984	0.999	0.981	0.999	0.963	0.999	0.961
RMSE	0.324	0.104	0.420	0.125	0.401	0.115	0.487	0.136
MAE	0.275	0.081	0.340	0.100	0.370	0.086	0.398	0.107
ME	0.093*	-0.020	0.093***	-0.020***	0.368***	-0.007	0.368***	-0.007
	e. CTE_output (semi-NOAA), 2001-2020				**f. CTE_global (semi-NOAA), 2001-2020**			
r	0.999	0.917	0.999	0.904	0.999	0.903	0.999	0.896
RMSE	0.395	0.174	0.476	0.214	0.261	0.192	0.347	0.230
MAE	0.348	0.131	0.389	0.174	0.220	0.158	0.279	0.195
ME	0.299***	-0.015	0.299***	-0.015	0.186***	-0.012	0.186***	-0.012

Note paired t-test significant level for ME: * p<0.1, ** p<0.05, *** p<0.01

Table S1. Statistic metrics assessing the agreement of the global CO_2 mole fraction (ppm) and its G_{ATM} (ppm yr^{-1}) from GAW observations (139 sties) using the semi-NOAA method (GAW (semi-NOAA)) with, a. GAW (WDCGG), GAW observations using the WDCGG method without extrapolation (1980-2020), b. NOAA analysis for observations from the NOAA 43 sites (1980-2020), c. CTE_obs (semi-NOAA), CTE observations (230 sites) using the semi-NOAA method (1980-2020), d. CTE observations (230 sites) using the semi-NOAA method (2001-2020), e. CTE_output(semi-NOAA), CTE output at the 230 sites using the semi-NOAA method (2001-2020), f. CTE_global (semi-NOAA), CTE full global grids (averaged over the first three levels, 0 to 0.35 km Alt.) using the semi-NOAA method (2001-2020). The statistical metrics include: Pearson Correlation Coefficient (r), which ranges from -1 to 1, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Error (ME). The negative sign on ME means that the GAW (semi-NOAA) has higher values, vice versa.

GAW (WDCGG+) vs GAW (WDCGG), 1984-2020				
	Annual		Monthly	
Statistic	CO_2	G_{ATM}	CO_2	G_{ATM}
r	0.999	0.994	0.999	0.992
RMSE	0.130	0.062	0.180	0.076
MAE	0.115	0.037	0.151	0.042
ME	0.096***	-0.011	0.096***	-0.011***

Note paired t-test significant level for ME: * p<0.1, ** p<0.05, *** p<0.01

Table S2. Statistic metrics assessing the agreement of the global CO_2 mole fraction (CO_2, ppm) and its G_{ATM} (ppm yr^{-1}) from GAW (WDCGG) and GAW (WDCGG+) during common period 1984-2020. GAW (WDCGG) is GAW observations (139 sites) analysed by using the WDCGG method without extrapolation. GAW (WDCGG+) is GAW observations (139 sites) analysed by using the WDCGG method with extrapolation. The statistical metrics include: Pearson Correlation Coefficient (r), which ranges from -1 to 1, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Error (ME). The negative values in ME means the GAW (WDCGG) has higher values, vice versa.

CTE output (semi-NOAA) vs CTE obs (semi-NOAA), 2001-2020				
	Annual		Monthly	
Statistic	CO_2	G_{ATM}	CO_2	G_{ATM}
r	0.999	0.896	0.999	0.881
RMSE	0.192	0.191	0.270	0.235
MAE	0.153	0.143	0.212	0.195
ME	-0.069	-0.008	-0.069***	-0.008

Note paired t-test significant level for ME: * p<0.1, ** p<0.05, *** p<0.01

Table S3. Statistic metrics assessing the agreement of the global CO_2 mole fraction (CO_2, ppm) and its G_{ATM} (ppm yr^{-1}) from CTE_output (semi-NOAA) and CTE_obs (semi-NOAA) during common period 2001-2020. CTE_obs (semi-NOAA) is CTE observations (230 sites) analysed by using the semi-NOAA method. CTE_output (semi-NOAA) is CTE output at the 230 sites analysed by using the semi-NOAA method. The statistical metrics include: Pearson Correlation Coefficient (r), which ranges from -1 to 1, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Error (ME). The negative values in ME means the CTE_obs (semi-NOAA) has higher values, vice versa.