Supplement of Atmos. Chem. Phys., 24, 1249–1264, 2024
https://doi.org/10.5194/acp-24-1249-2024-supplement

Supplement of

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

Zhendong Wu et al.

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

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 (Fig. 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.

S2. The CTE station network

290 stations are evaluated in the CTE inversion, the observations come from the ObsPack data product (Schuldt et al., 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 (Fig. 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.

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.

$$m_{CO_2} = Cw_{CO_2} * m_{air} \quad \text{(S1)}$$

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:

$$Cw_{CO_2} = Cv_{CO_2} * \frac{M_{CO_2}}{M_{air}} \quad \text{(S2)}$$

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:

$$p_{air} = \frac{F_{air}}{S} \quad \text{(S3)}$$

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:

$$F_{air} = m_{air} * g \quad \text{(S4)}$$

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.

$$S = 2 * \pi * (R + gph)^2 * |\sin(lat1) - \sin(lat2)| * \frac{|lon1 - lon2|}{360} \quad \text{(S5)}$$

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:

$$m_{air} = \frac{p_{air} * S}{g} \quad \text{(S6)}$$

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_gfit.ipynb

  Apply the GFIT 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_gfit.ipynb

  Calculate global co2 mole fraction average and its growth rate, and estimate their uncertainty, using output from GAW(GFIT)
- 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(GFIT)
- 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(GFIT)
- 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(GFIT)
- 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 notebooks, it needs to download the data (GAW_data.zip) [https://doi.org/10.18160/Q788-9081] and change the data path in jupyter notebooks to where the data is unzipped. 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 GFIT method, i.e. GAW (GFIT).

Global mean：

df_co2_annual_global_NH_SH_gaw_GFIT.csv

df_co2_monthly_global_NH_SH_gaw_GFIT.csv

df_co2rate_annual_global_NH_SH_gaw_GFIT.csv

df_co2rate_monthly_global_NH_SH_gaw_GFIT.csv

Their uncertainty basing on bootstrap method:

bootstats_co2_annual_global_gaw_GFIT.csv

bootstats_co2_monthly_global_gaw_GFIT.csv

bootstats_co2rate_annual_global_gaw_GFIT.csv

bootstats_co2rate_monthly_global_gaw_GFIT.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 GFIT method, i.e. CTE_obs (GFIT).

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 GFIT method, i.e. CTE_output (GFIT).

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 GFIT method, i.e. CTE_global (GFIT)

  co2_annual_global_cte2021(level1-3)_GFIT.csv

  co2_monthly_global_cte2021(level1-3)_GFIT.csv

  co2rate_annual_global_cte2021(level1-3)_GFIT.csv

  co2rate_monthly_global_cte2021(level1-3)_GFIT.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 S3.

  cte2021_co2mass_2000_2020_monthly.csv

  cte2021_co2mass_2000_2020_annual.csv

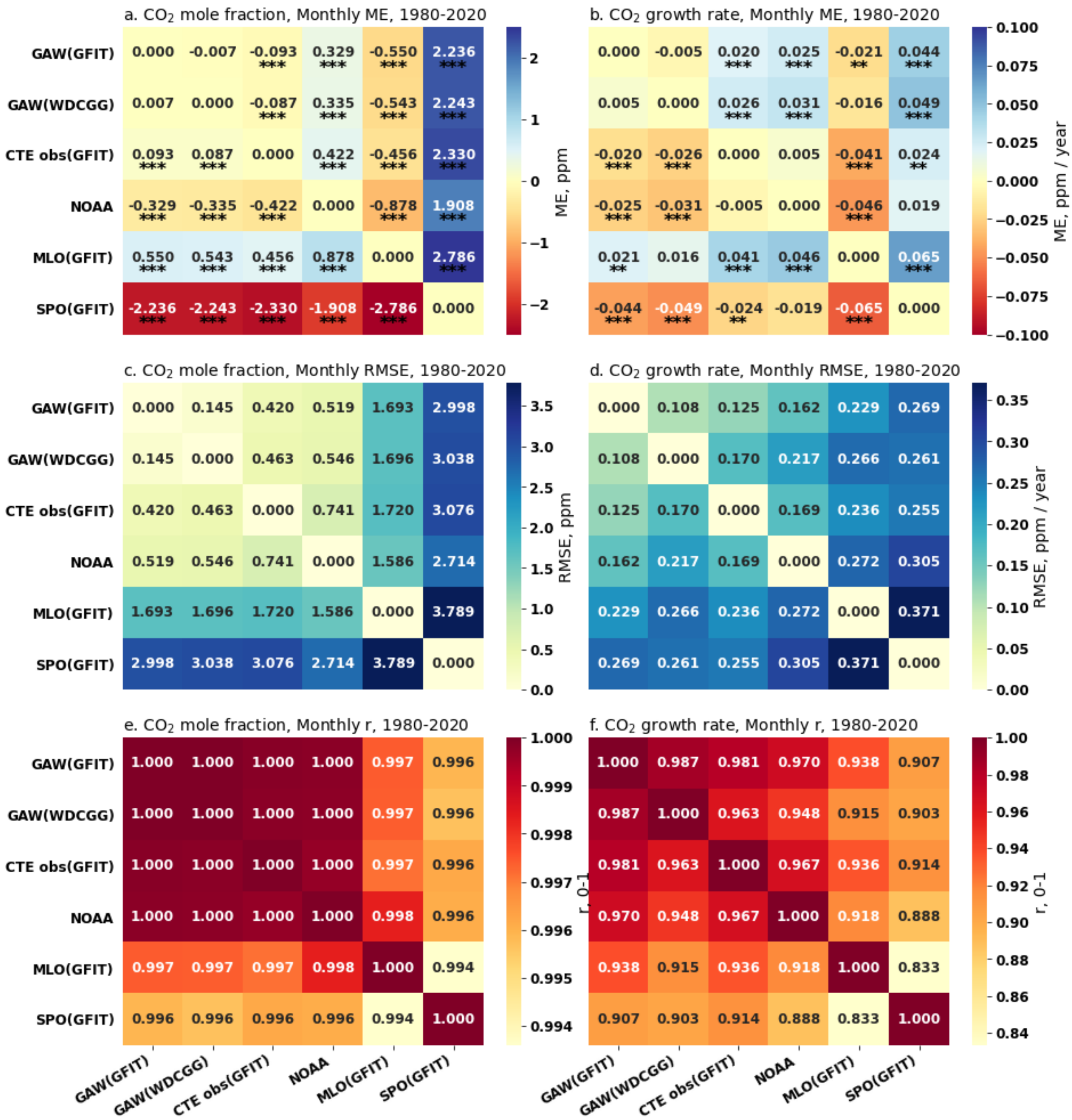

Figure S1. Pair-wise statistical metrics assess the agreement of monthly global and local CO_2 mole fraction (ppm) and its G_{ATM} (ppm yr-1) across various networks and methodologies (see Table 1 and Fig. 4) for the period 1980-2020. Panel (a) presents the Mean Error (ME) quantifying the difference for each pair, focusing on CO_2 mole fraction, while panel (b) does the same for G_{ATM}. The significance levels of paired t-test for ME are indicated as follows: * p<0.1, ** p<0.05, *** p<0.01. Panel (c) and (d) present the Root Mean Squared Error (RMSE) for CO_2 mole fraction and G_{ATM}, respectively. Panel (e) and (f) present the Pearson Correlation Coefficient (r) for CO_2 mole fraction and G_{ATM}, respectively.

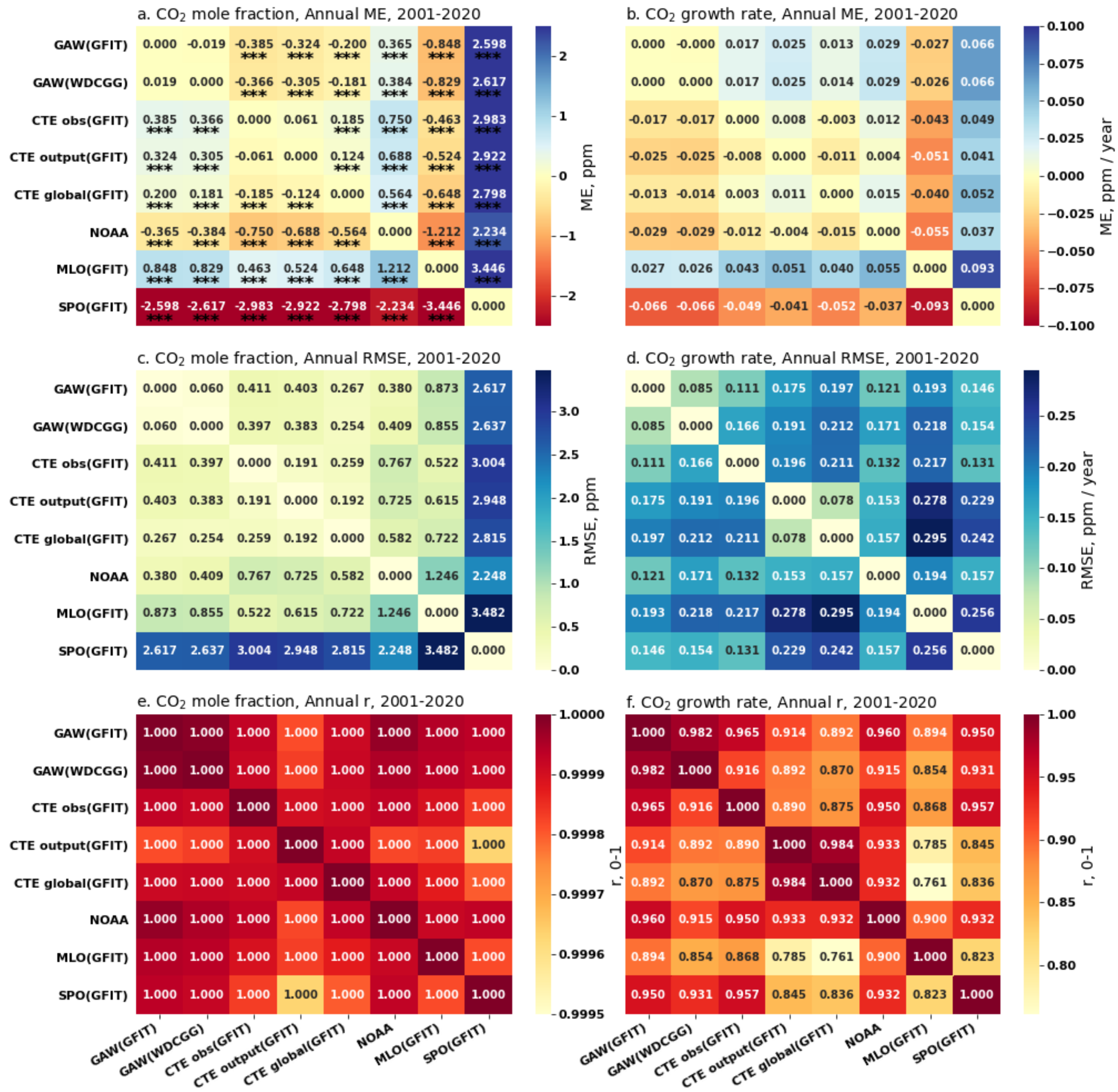

Figure S2. Pair-wise statistical metrics assess the agreement of annual global and local CO_2 mole fraction (ppm) and its G_{ATM} (ppm yr-1) across various networks and methodologies (see Table 1 and Fig. 4) for the period 2001-2020. Panel (a) presents the Mean Error (ME) quantifying the difference for each pair, focusing on CO_2 mole fraction, while panel (b) does the same for G_{ATM}. The significance levels of paired t-test for ME are indicated as follows: * p<0.1, ** p<0.05, *** p<0.01. Panel (c) and (d) present the Root Mean Squared Error (RMSE) for CO_2 mole fraction and G_{ATM}, respectively. Panel (e) and (f) present the Pearson Correlation Coefficient (r) for CO_2 mole fraction and G_{ATM}, respectively.

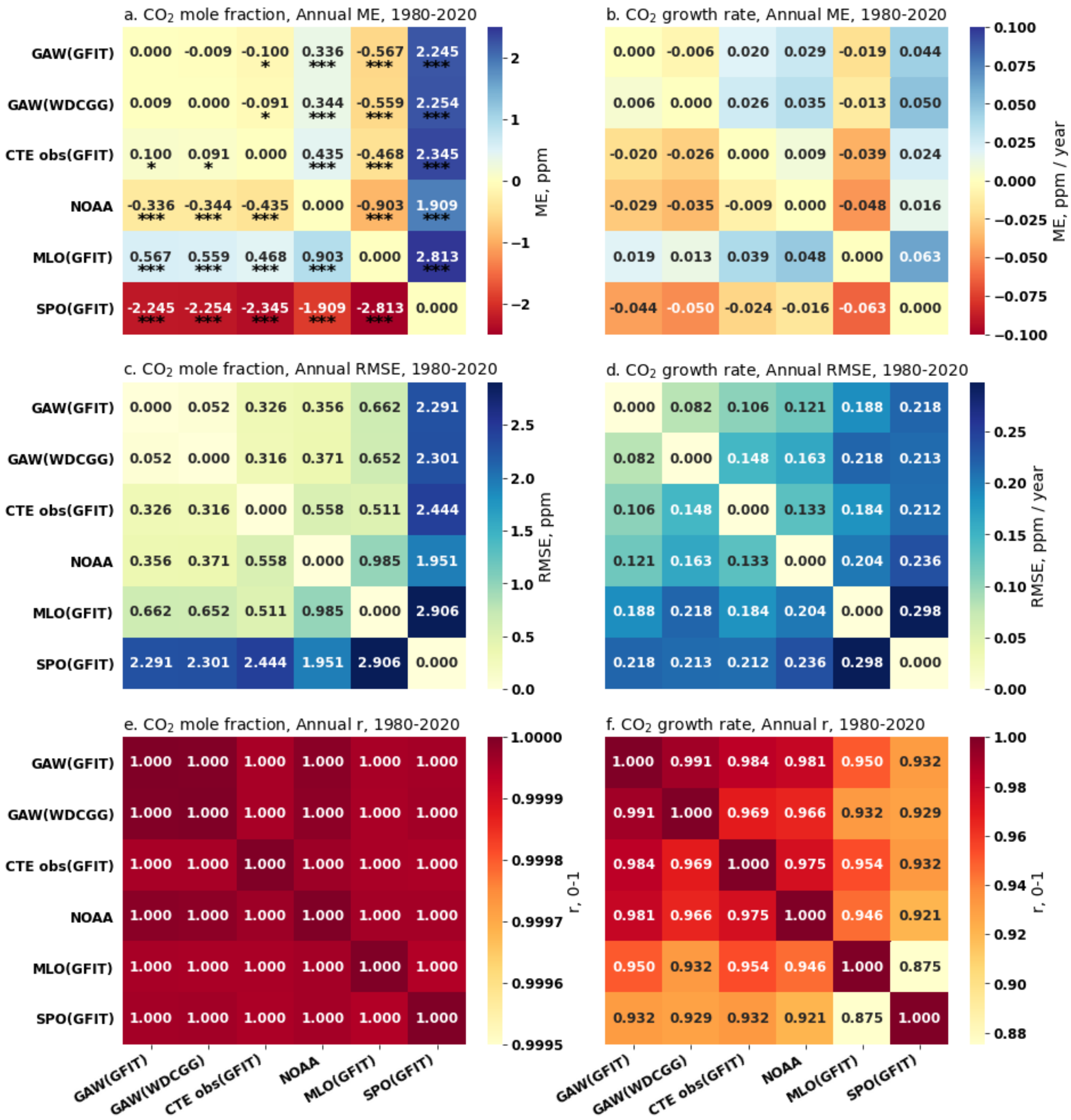

Figure S3. Pair-wise statistical metrics assess the agreement of annual global and local CO_2 mole fraction (ppm) and its G_{ATM} (ppm yr-1) across various networks and methodologies (see Table 1 and Fig. 4) for the period 1980-2020. Panel (a) presents the Mean Error (ME) quantifying the difference for each pair, focusing on CO_2 mole fraction, while panel (b) does the same for G_{ATM}. The significance levels of paired t-test for ME are indicated as follows: * p<0.1, ** p<0.05, *** p<0.01. Panel (c) and (d) present the Root Mean Squared Error (RMSE) for CO_2 mole fraction and G_{ATM}, respectively. Panel (e) and (f) present the Pearson Correlation Coefficient (r) for CO_2 mole fraction and G_{ATM}, respectively.

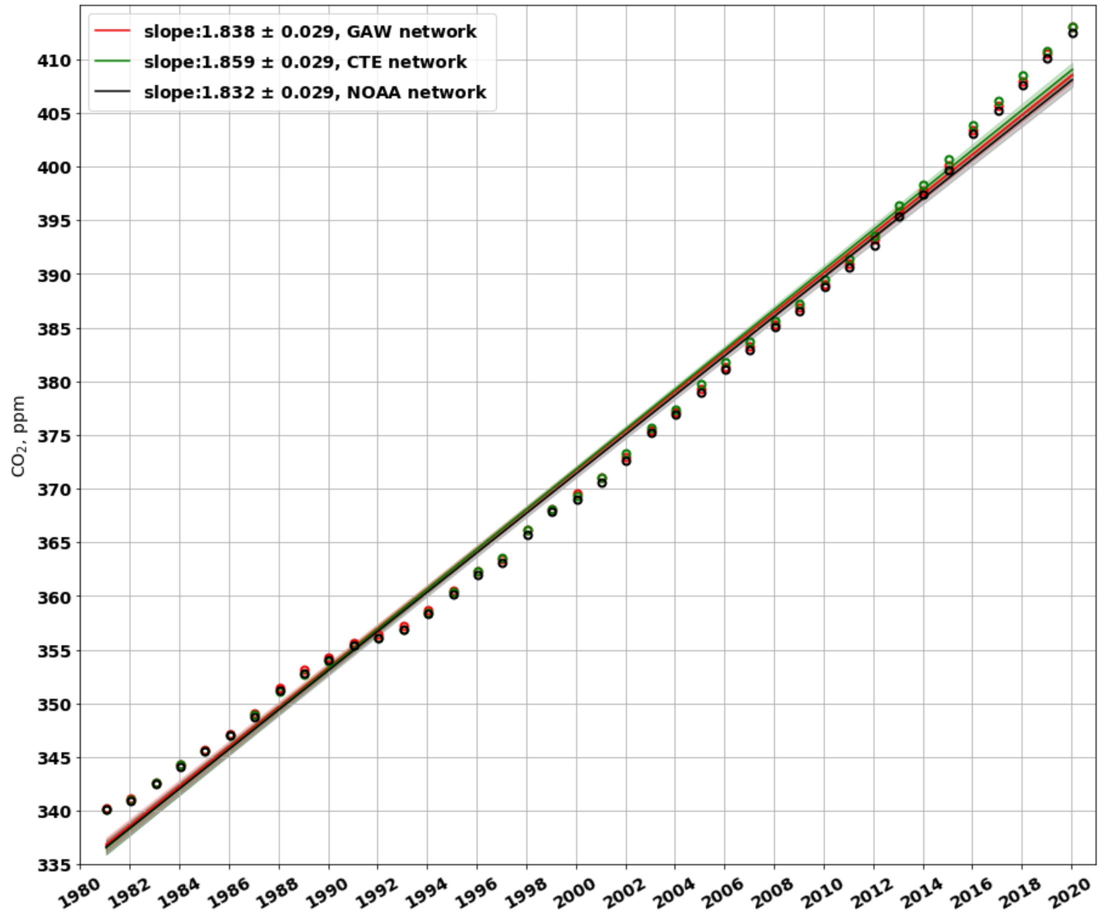

Figure S4. shows the trends of global CO_2 mole fraction for the GAW network (red line), the CTE network (green line) and the NOAA network (black line) during the whole period 1980-2020. The cycles show the annual CO_2 mole fraction, respectively.

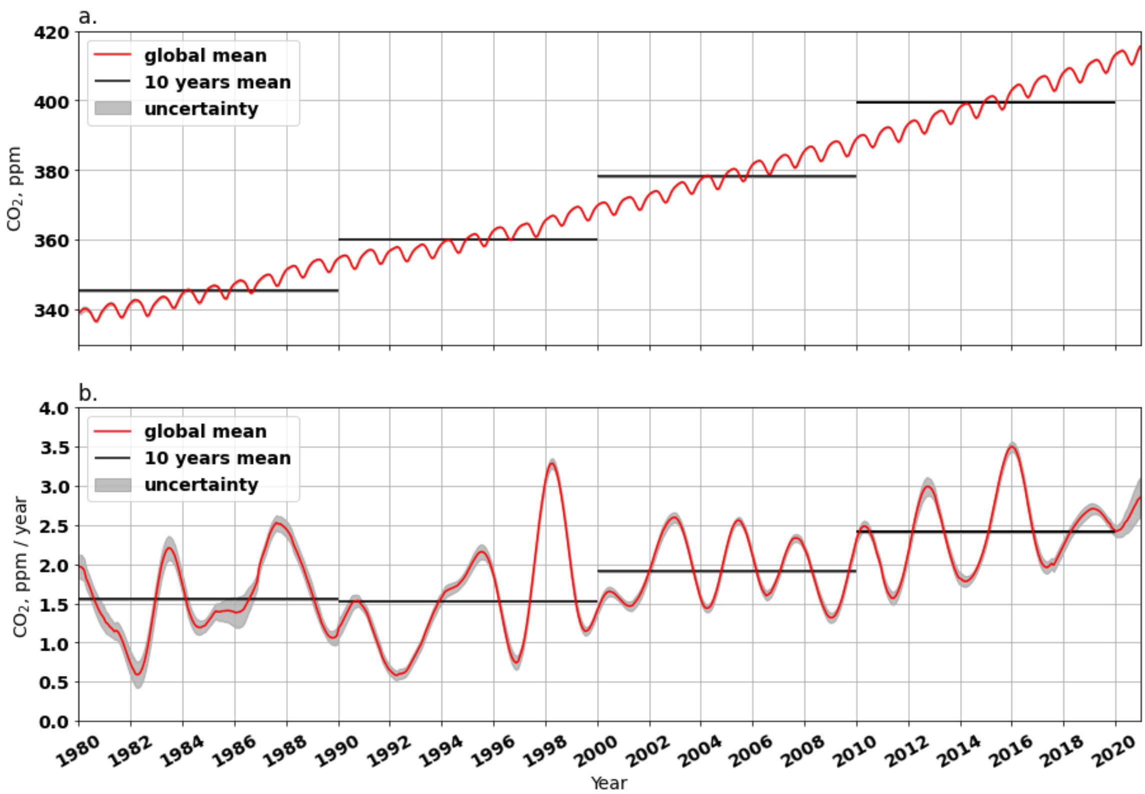

Figure S5. Globally averaged CO_2 mole fraction (a) and its G_{ATM} (b) from 1980 to 2020. In panel (a), the red line shows the mean CO_2 mole fraction, black lines show the mean CO_2 mole fraction over 10 years, the grey area shows the uncertainty derived from the 200 bootstrap networks. Similarly, 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 GFIT method.

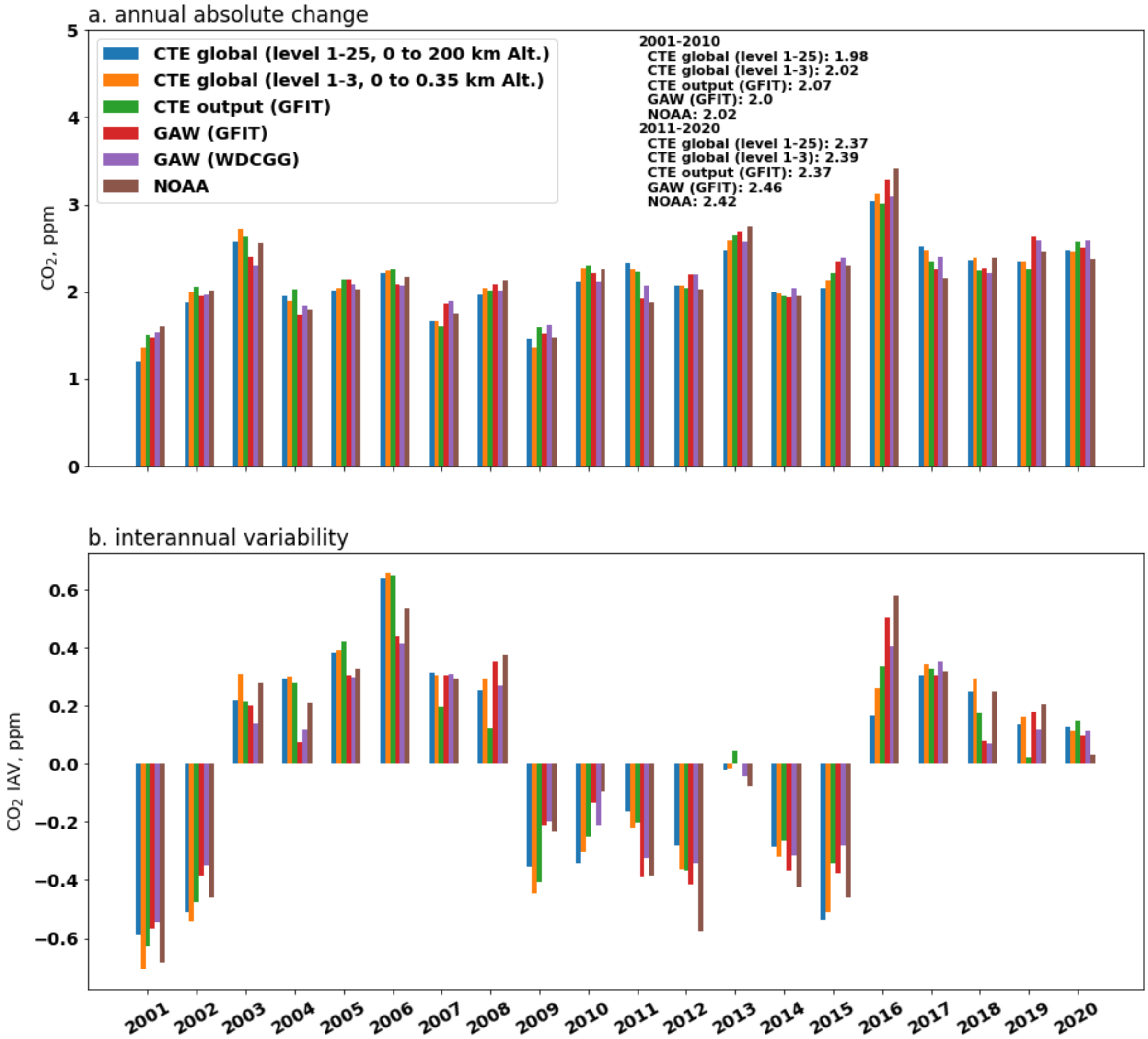

Figure S6. Annual absolute change and interannual variability of global CO_2 mole fraction derived from different data (CTE model, GAW observation and NOAA observation) and analysis methods (GFIT method, WDCGG method and NOAA method) for 2000-2020. Panel (a) shows the annual absolute change which is the difference between annal mean. Averages over 2001-2010 and 2011-2020 are also shown. Panel (b) shows the IAV which is calculated as the anomaly departure from a quadratic trend.

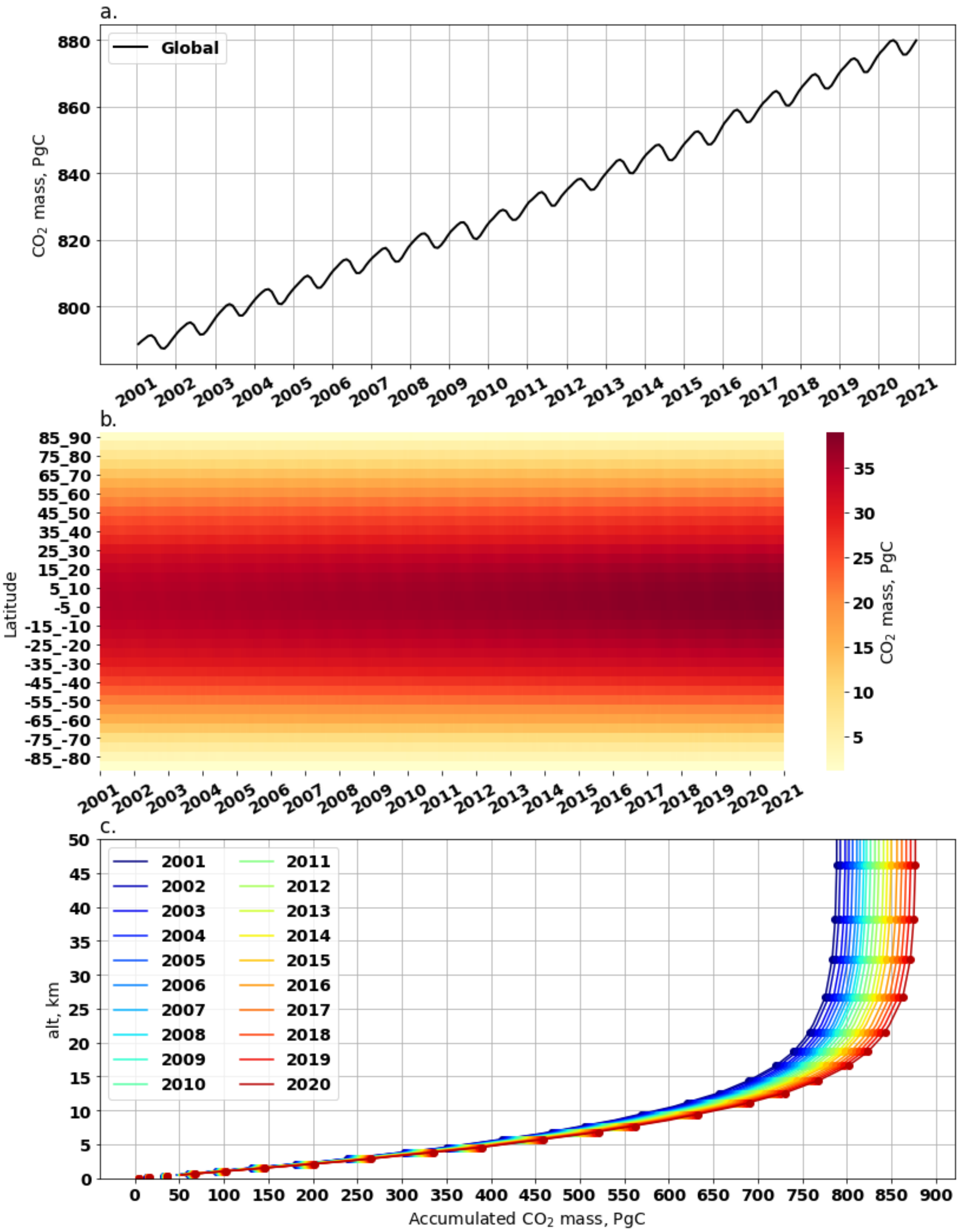

Figure S7. Atmospheric CO_2 mass derived from CTE output. Panel (a) shows the global monthly CO_2 mass in atmosphere (from surface up to 200 km altitude). Panel (b) shows the zonal (5°) average of monthly CO_2 mass. Panel (c) shows accumulated CO_2 mass with altitudes from 2001 to 2020, the dots mark CTE vertical level altitudes and lines are the linear interpolation between the altitudes.

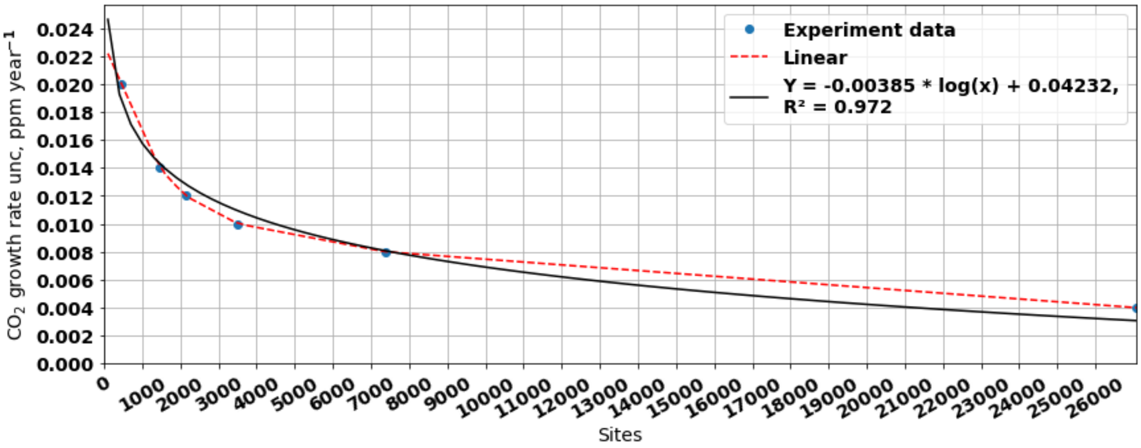

Figure S8. The relationship between the uncertainty of the global CO_2 growth rate and the number of observation sites. The relationship is estimated using CTE_global (all global grids excluding ocean grids) with different resolutions (1x1, 2x2, 3x3, 4x4, 5x5, and 10x10 degrees) to estimate the uncertainty 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.

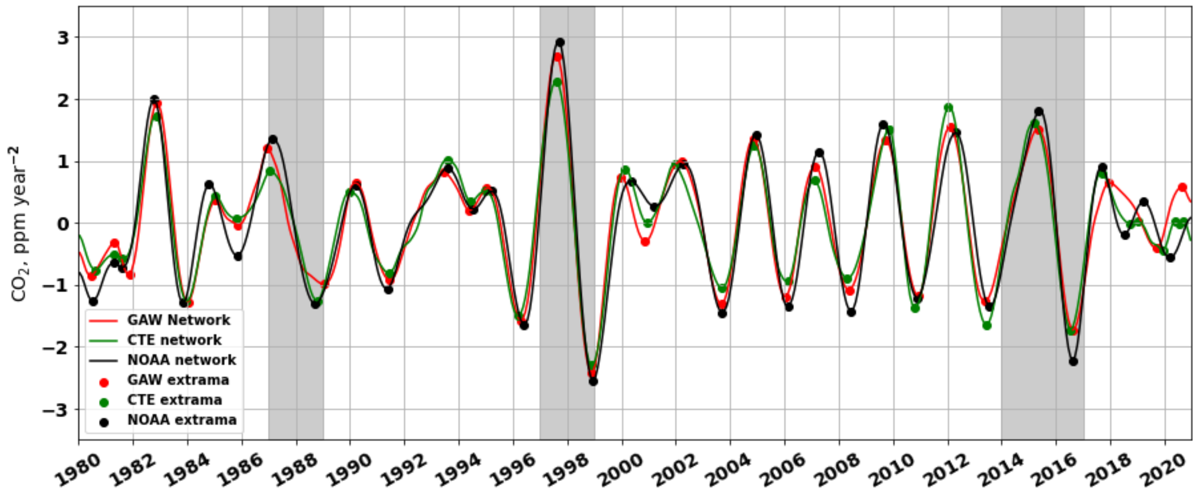

Figure S9. presents the smoothed trend of CO_2 growth rate for each month during 1980-2020. The trends (depicted in Figure 6b) are smoothed by using a Gaussian filter (with sigma=1.96). The dots represent the local extrema, which aid in identifying the start of CO_2 growth rate increase/decrease.

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 significance 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 (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.

El Niño 1987-1988				
	Trough (G_{ATM} starts increasing)		Peak (G_{ATM} starts decreasing)	
Date	Decimal year	Days of year	Decimal year	Days of year
CTE	1985.791635	289	1987.041665	15
GAW	1985.874965	319	1986.958295	350
NOAA	1985.874965	319	1987.124995	46
El Niño 1997-1998				
CTE	1996.208325	76	1997.624975	228
GAW	1996.291655	106	1997.624975	228
NOAA	1996.374985	137	1997.708305	259
El Niño 2014-2016				
CTE	2013.458315	167	2015.208325	76
GAW	2013.374985	137	2015.374985	137
NOAA	2013.541645	198	2015.374985	137

Table S2. displays the estimates of CO_2 growth rate increase/decrease for the three strong El Niño events (i.e 1987-1988, 1997-1998 and 2014-2016). These estimates are calculated from the smoothed trend of CO_2 growth rate based on CTE, GAW and NOAA networks (Fig. S9).

References

Schuldt, K., Mund, J., and Luijkx, I.: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957–2021; obspack_co2_1_GLOBALVIEWplus_v8. 0_2022-08-27, NOAA Global Monitoring Laboratory, http://doi.org/10.25925/20220808, 2022.

Tans, P. P., Thoning, K. W., Elliott, W. P., and Conway, T. J.: Error estimates of background atmospheric CO2 patterns from weekly flask samples, J. Geophys. Res. Atmos., 95, 14063-14070, https://doi.org/10.1029/JD095iD09p14063, 1990.

Tsutsumi, Y., Mori, K., Hirahara, T., Ikegami, M., and Conway, T. J.: Technical Report of Global Analysis Method for Major Greenhouse Gases by the World Data Center for Greenhouse Gases (WMO/TD-No. 1473). GAW Report No. 184. Geneva, WMO,, 1-23, https://library.wmo.int/index.php?lvl=notice_display&id=12631, 2009.