Supplement to "A new algorithm to generate a priori trace gas profiles for the GGG2020 retrieval algorithm"

Joshua L. Laughner, et al.

S1 Required accuracy of prior profiles

In the introduction, we asserted that a shape error of $\leq 1\%$ in the prior profiles is desired to keep the error contribution from the prior below the 0.25% precision expected from TCCON XCO_2 data. Here we will describe that derivation in more detail.

To test the effect of shape errors in the priors, we generated synthetic spectra for a flat CO_2 profile (400 ppm at all altitudes) with only O_2 and a negligible amount (1 ppm) of water in the simulated atmosphere. We generated 8 spectra covering solar zenith angles (SZAs) from 25 to 75 degrees and four months (January, April, July, September) out of the year. Temperature and pressure from the Lamont, OK, USA TCCON site (36.604 N, 97.486 W) were used. These spectra were then retrieved with the same temperature and pressure profiles, but different prior CO_2 profiles.

Figure S1 shows the different prior profiles (panel a) and the resulting change in retrieved XCO_2 compared to the true profile (panel b). We defined two types of shape error: a "jump" where the CO_2 DMF increases or decreases suddenly at a specific altitude, and a "linear" error where the CO_2 DMF varies linearly with respect to pressure. For all shape errors, we defined a 1% error to mean that the DMF changes by 1% (4 ppm) between the top and bottom of the profile. Both the "jump" and "linear" cases each have three subcases that vary whether the troposphere, stratosphere, or both have the error.

What we see in Fig. S1b is that the tests at 25° and 50° SZA all the tests are within 0.1 ppm ($\sim 0.025\%$) of the truth, and even at large SZAs, the errors are < 0.5 ppm ($\sim 0.125\%$). This is well below the maximum 0.25% uncertainty required of TCCON XCO_2 data.

Two final notes. First, different shape errors that are just offset from one another (such as linear-2to-2 and linear-4to0) have nearly identical ΔXCO_2 values because GGG uses a profile scaling retrieval, so profiles with the same shape and a different offset should converge to nearly the same posterior profile, all else being equal. Second, because this is using a profile scaling retrieval, these results are not applicable to a full profile retrieval, such as that used by OCO-2 and -3.

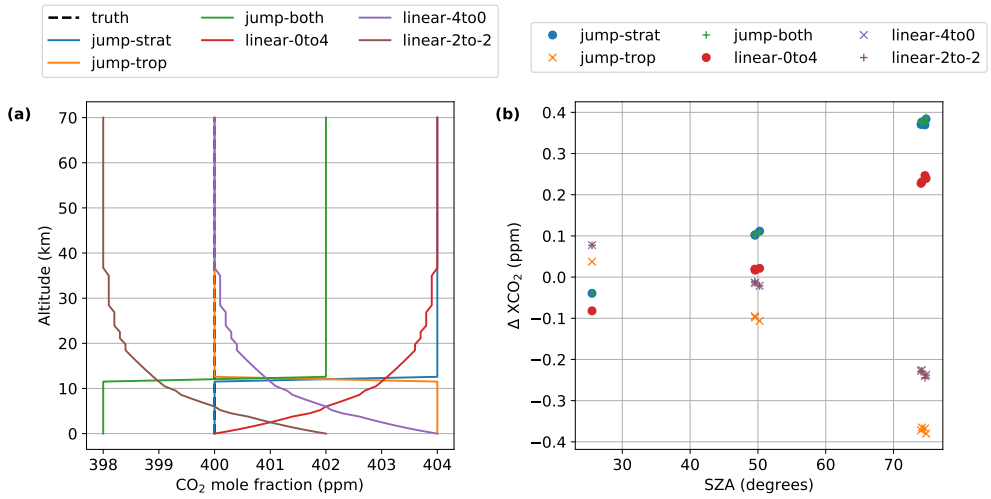

Figure S1: The effect of six different shape errors in a priori CO_2 profiles on retrieved XCO_2 for 3 different SZAs and 4 different days of year (one each in January, April, July, and September). **(a)** The six different test a priori CO_2 profiles and the "true" CO_2 profile used to generate the synthetic spectra for these tests. The true profile is the flat 400 ppm profile in black. **(b)** The change in XCO_2 compared to the truth for each of the six test a priori profiles shown in (a).

S2 MLO & SMO extrapolation accuracy

To evaluate the error caused by extrapolating the MLO & SMO data used in the GGG2020 and OCO-2/3 v10 priors past 2018, we used two methods. For CO_2, Ed Dlugokencky provided us updated flask data through the end of 2020 on the previous X2007 scale to compare our extrapolation to the truth in 2019 and 2020. This in shown in the top panel of Fig. S2. The error increases to approximately 0.8 ppm by the middle of 2019, which is more rapidly than we expected, but then remains fairly constant. This may be due in part to a weak El Ni no in 2019, evidenced by the Nino 3.4 anomaly index. Nevertheless, this was a major motivation to switch to rapidly updating CO_2 data for OCO-2/3 v11 priors.

To evaluate the other gases, we used the same MLO & SMO data that the GGG2020 and v10 OCO-2/3 priors are based on, but test extrapolating beginning in each year from 2004 to 2017. We then compare the extrapolated values to the true values. This is shown in the bottom panels of Fig. S2. Unsurprisingly, the longer period of time that the data must be extrapolated, the poorer the agreement with the true values. Interestingly, for CO_2, the test cases which begin extrapolating in more recent years (2013 and later) perform worse over the same length of extrapolation than earlier years. This may be due to the strong 2015/2016 El Ni no disrupting the long term trend. Conversely, the CH_4 tests that begin extrapolating between 2004 and 2009 have significantly worse long term performance than later years. We attribute this to the hiatus in CH_4 growth rate in the 2000s (Dlugokencky et al., 2009), which makes it difficult to extrapolate the CH_4 growth rate from data ending in the 2000s. Our CH_4 extrapolation specifically uses a shorter training period than CO_2

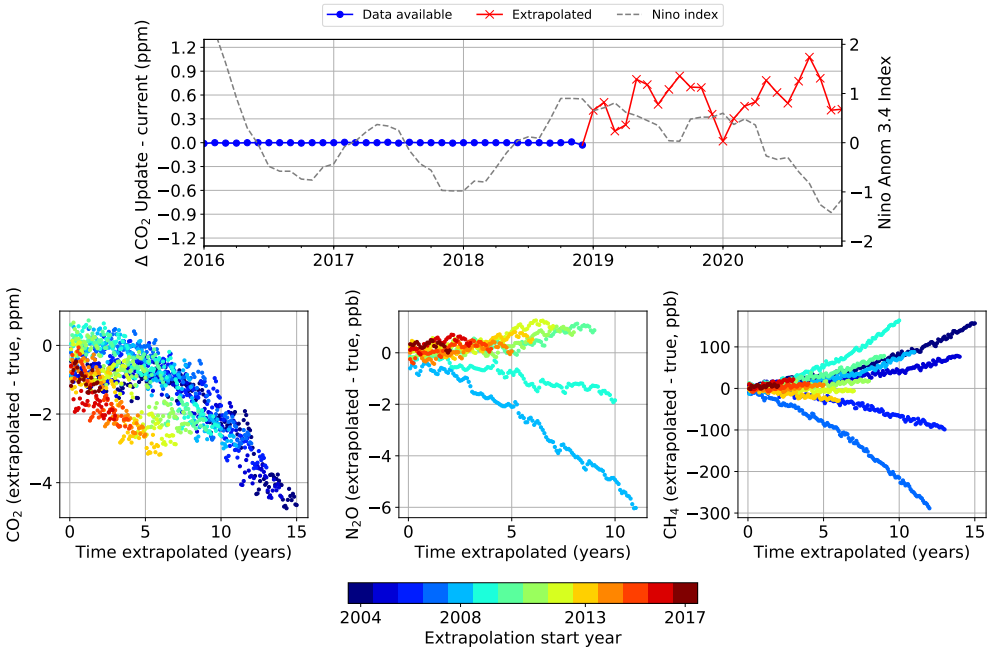

Figure S2: Top panel: Difference between the CO_2 record used in the GGG2020 and OCO-2/3 v10 priors and updated MLO & SMO mean monthly flask data that extends through 2020. Blue indicates where the GGG2020/v10 priors have MLO & SMO data available, red indicates where they must extrapolate. The Niño 3.4 anomaly index is also shown as the grey dashed line and corresponds to the right axis. Bottom panels: extrapolation error in CO_2, N_2O, and CH_4, respectively, calculated with the MLO & SMO data used in the GGG2020/v10 priors by starting extrapolation at each year between 2004 and 2007 and comparing the extrapolated values to the true values.

or N_2O (Table 1, main paper) to avoid including the hiatus in the extrapolation fit for the production GGG2020 CH_4 priors.

Our estimates of 0.25% for CO_2, 0.15% for N_2O, and 0.6% for CH_4 come from considering the mean error over a five year period of extrapolation. For CO_2, about two-thirds of the tests shown in the lower panels of Fig. S2 have average errors ≤ 1 ppm; for N_2O, they are ≤ 0.4 ppb; for CH_4, about half have average errors ≤ 10 ppb. (We consider fewer test cases for CH_4 because of the impact of the growth rate hiatus.) Then we assume nominal mole fractions of 400 ppm, 300 ppb, and 1800 ppm to estimate the percent errors corresponding to each of these absolute errors.

For TCCON retrievals, an error in the growth rate of these gases should not impose a significant error on the retrieved column amount. As discussed in the main paper, TCCON uses a scaling retrieval, which can theoretically correct a constant multiplicative error in the a priori mole fractions. As long as the error in extrapolation is approximately linear in time, it should produce roughly that sort of error in the priors. However, based on Fig. S2, modifying this algorithm to use rapidly updated NOAA data should be a priority for the next major GGG version.

S3 Secondary gas calculation details

In Sect 2.4, we described in general terms how profiles for the secondary gases are derived. Here we provide the calculations in detail.

S3.1 Stretching/compressing vertically

The first step is to modify the climatological profile so that the tropopause is at the correct altitude. For each altitude, z, in the profile, we compute an effective altitude, z_{eff}, such that interpolating the climatological profile to z_{eff} gives the correct concentration for z. In the troposphere:

$$z_{\text{eff,trop}} = z \cdot \frac{z_{\text{trop,clim}}}{z_{\text{trop,met}}} \tag{S1}$$

where $z_{\text{trop,clim}}$ is the tropopause height in the climatological profile and $z_{\text{trop,met}}$ is the tropopause height from the meteorology for the profile under construction.

In the stratosphere, we only want to modify the lower stratosphere as well as account for the location of the intertropical convergence zone (ITCZ):

$$z_{\text{eff,strat}} = z + \exp\left(-\frac{z - z_{\text{trop,met}}}{10}\right) \cdot \left[z_{\text{trop,clim}} - z_{\text{trop,met}} - 3.5 \cdot z_{\text{trop,met}} \cdot \left(\frac{z}{z_{\text{trop,met}}} - 1\right)^2 \cdot \exp\left(-\left\{\frac{l - l_{\text{itcz}}}{w_{\text{itcz}} + 10}\right\}^4\right)\right] \tag{S2}$$

where l is the profile latitude, l_{itcz} is center latitude of the ITCZ, and w_{itcz} the width of the ITCZ in degrees. The ITCZ terms are interpolated from a lookup table with the spatial and temporal behavior shown in Fig. S3. This produces the relationship between z_{eff} and z shown in Fig. S4. This is designed so that outside the ITCZ, only the lower stratospheric levels are stretched or compressed to match the climatological tropopause, while inside the ITCZ the lowest levels have a more "tropospheric" altitude, to mimic the stronger vertical transport in the ITCZ.

We then interpolate the climatological profiles such that a concentration for level k at altitude z_k is taken from altitude $z_{\text{eff},k}$ in the climatological profile.

To adjust these climatological profiles to the observation time and latitude, we apply a latitudinal gradient, secular trend, and seasonal cycle (in that order). The seasonal cycle approach is the same as described in Sect. 2.2, except that it is applied to both the tropospheric and stratospheric components.

The latitudinal gradient is applied by scaling the DMFs by

$$\frac{1 + f_r \cdot x_{\text{obs}}}{1 + f_r \cdot x_{\text{ref}}} \tag{S3}$$

where

$$x_{\text{obs}} = g_{\text{gas}} \cdot \left(\frac{l/15}{\sqrt{1 + (l/15)^2}}\right) \tag{S4}$$

$$x_{\text{ref}} = g_{\text{gas}} \cdot \left(\frac{l_{\text{ref}}/15}{\sqrt{1 + (l_{\text{ref}}/15)^2}}\right) \tag{S5}$$

$$f_r = \frac{1}{1 + (z/z_{\text{trop,met}})^2} \tag{S6}$$

where g_{gas} is the latitude coefficient for each gas in Table S5, l is the tropospheric effective latitude (Sect. 2.2.1 or stratospheric equivalent latitude (Sect. 2.3.1) for that profile level, and l_{ref} is the reference latitude (35° N).

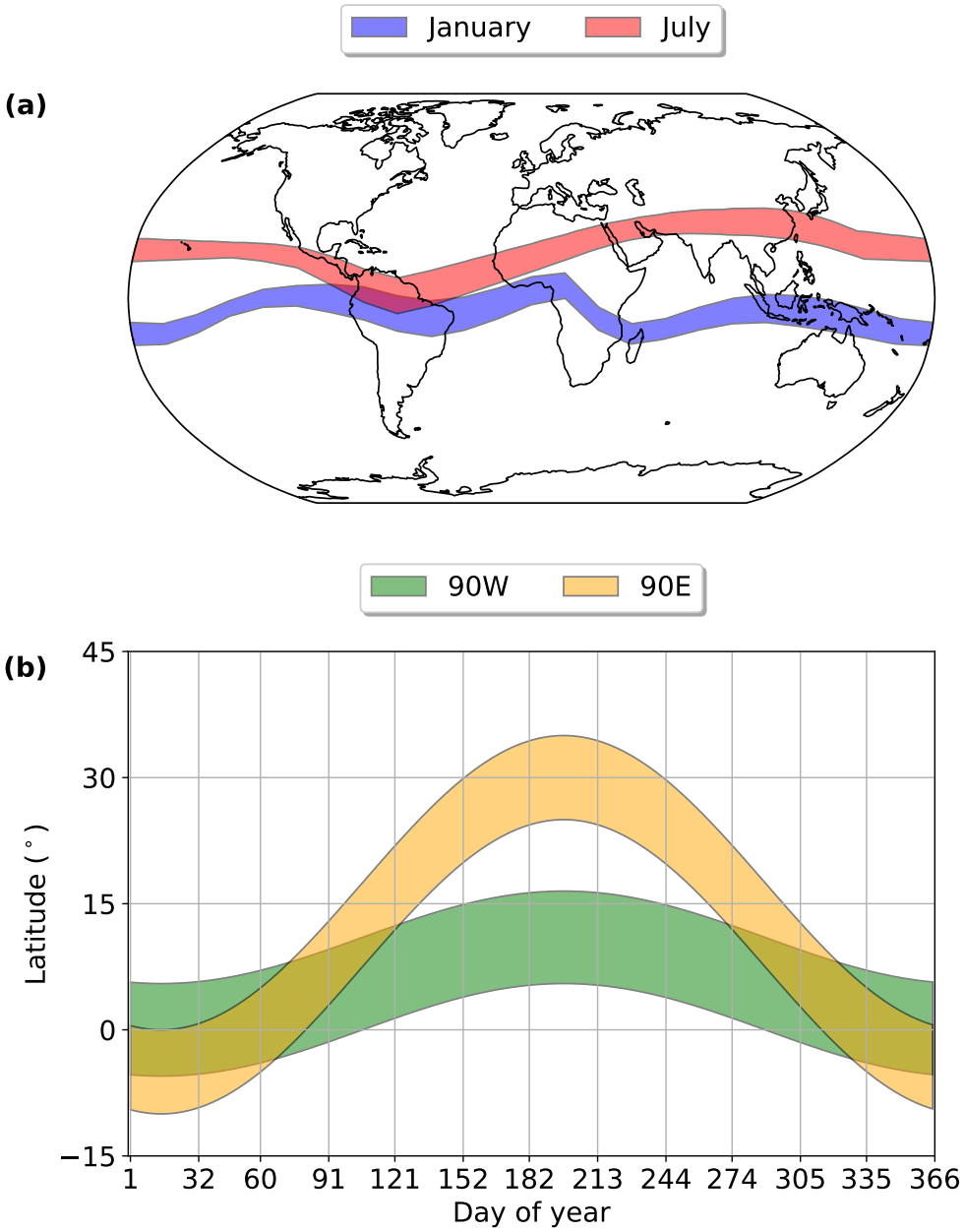

Figure S3: Assumed **(a)** longitudinal variation at two dates and **(b)** temporal variation at two longitudes of the ITCZ in GGG2020.

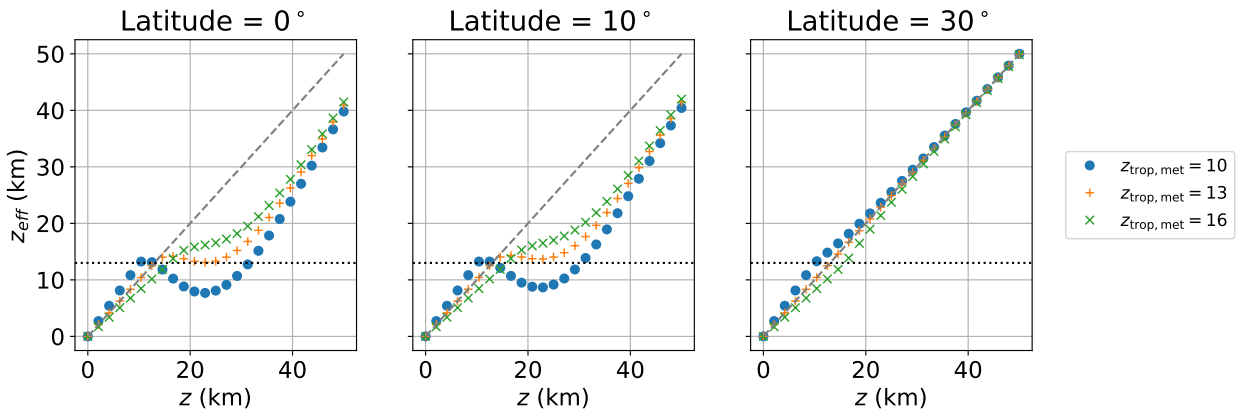

Figure S4: z_{eff} [from Eqs. (S1) and (S2)] vs. original altitude, assuming $z_{\text{trop,clim}} = 13\,\text{km}$, $l_{\text{itcz}} = 0°$, and $w_{\text{itcz}} = 10°$. Each panel demonstrates a profile for a different latitude, shown in the title, and within each panel, the different series represent profiles with different $z_{\text{trop,met}}$ values.

The secular trend is applied by scaling the DMFs with latitude gradients applied by

$$1 + r_{\text{gas}} \cdot \Delta t \cdot \gamma \tag{S7}$$

where

$$\Delta t = t - t_{\text{ref}} - a \tag{S8}$$

$$\gamma = \begin{cases} 1 + (\Delta t/155.0)^2 & \text{if } CO_2 \\ 1.004 - 0.024 \cdot \frac{\Delta t + 2.5}{\sqrt{25 + (\Delta t)^2}} & \text{if } CH_4 \\ 1 + \exp\left([-\Delta t - 16.0]/5\right) & \text{if HF} \\ 1 + \exp\left([-\Delta t - 4.0]/9\right) & \text{if } F_{113} \\ 1 & \text{otherwise} \end{cases} \tag{S9}$$

where t is the observation date (in years), t_{ref} is the reference time for the base profiles (2005), a is the tropospheric or stratospheric age-of-air (Sects. 2.2 and 2.3.1), and r_{gas} is the secular trend in Table S5.

After the latitudinal gradients, secular trends, and seasonal cycles have been applied, the middleworld levels are filled by interpolating between the tropopause and bottom overworld DMFs linear in θ.

S4 CO additional column

Because CO can have a significant mesospheric column, we add a concentration of CO to the top prior level that represents an equivalent mass of CO. To compute this additional CO, we integrate the CO column above the top prior level in the CMAM climatology (Sect. 3.6) as:

$$V_{\mathrm{CO}} = \mathbf{n}_{\mathrm{CO}}^{T} \cdot \mathbf{v} \tag{S10}$$

where V_{CO} is the vertical column of CO, $\mathbf{n}_{\mathrm{CO}}$ the profile of CO number density, and $\mathbf{v}$ an effective vertical path, whose elements v_i are given by:

$$\begin{aligned} v_i = \frac{1}{2}\left[z_i - z_{i-1}\right] \cdot \left(1 + \frac{1}{3}\ln\left(\frac{n_{i-1}}{n_i}\right) + \frac{1}{12}\left[\ln\left(\frac{n_{i-1}}{n_i}\right)\right]^2 + \frac{1}{60}\left[\ln\left(\frac{n_{i-1}}{n_i}\right)\right]^3\right) \\ + \frac{1}{2}\left[z_{i+1} - z_i\right] \cdot \left(1 - \frac{1}{3}\ln\left(\frac{n_i}{n_{i-1}}\right) + \frac{1}{12}\left[\ln\left(\frac{n_i}{n_{i-1}}\right)\right]^2 - \frac{1}{60}\left[\ln\left(\frac{n_i}{n_{i-1}}\right)\right]^3\right) \end{aligned} \tag{S11}$$

where z_i is the altitude at level i and n_i is the number density of air at level i. This represents a density-weighted path such that $\mathbf{n}^T \cdot \mathbf{v} = \int_{z_0}^{z_{\mathrm{top}}} n(z)\, dz$.

We then compute the mixing ratio of CO that would be added to the top prior level if this column were compressed into it as:

$$c_{\mathrm{CO,top}} = \frac{V_{\mathrm{CO}}}{v_{\mathrm{top}} \cdot n_{\mathrm{top}}} \tag{S12}$$

where v_{top} and n_{top} are the effective path (Eq. S11) and number density of air in that top level.

S5 Preprocessing algorithm for hourly surface NOAA data

The OCO-2/3 version 11 product uses CO_2 priors that ingest hourly in situ surface data from the Mauna Loa and American Samoa NOAA observatories, instead of the monthly average flask data used in GGG2020 and OCO-2/3 version 10. These hourly in situ data are converted into monthly average files before being ingested by the priors code. The steps are:

1. Based on the creation date in the hourly data file, select only complete months. For a creation date in month M, only take data from month $M-1$ or earlier.

2. Apply site-specific background selection. See Sects. S5.1 and S5.2 for details.

3. Group remaining data by month and average.

For use in the OCO-2/3 V11 algorithm, once a monthly average is computed it is not updated, even if future hourly data includes changes in quality control or other factors that would lead to a different monthly average. This ensures that retrievals can be reprocessed at different times without introducing changes in the a priori profiles.

Thus, a record of monthly CO_2 is updated periodically by appending new monthly averages from the latest hourly data, while alread extant monthly averages are left as-is. V11 input monthly CO_2 averages from 2010 on are derived by applying the above algorithm to the NOAA hourly data; averages before 2010 are taken from the NOAA in situ monthly averages (Thoning et al., 2021, downloaded from `https://gml.noaa.gov/dv/data/` on 7 Jul 2021).

S5.1 Mauna Loa background selection

An hourly data point from Mauna Loa is selected as background if:

- the first two characters in the "flag" column are periods. The third character in the flag is ignored, as for our purposes it usually indicates that data is preliminary, which is acceptable.
- the hourly uncertainty is ≤ 0.2 ppm
- the difference in dry mole fraction with either the preceeding or following point is ≤ 0.25 ppm *or* the difference in time to the preceeding or following point is > 1 hour (due to removal by QC flags or hourly uncertainty).
  - This selection is applied month-by-month, so the first and last data points remaining after removal by QC flags and uncertainty only consider the following and preceeding points, respectively.
- The local hour is between 0 and 7 (midnight and 7 AM, inclusive)

This approach mimics the simpler selection described in Thoning et al. (1989) but does not try to reproduce it exactly. The most notable difference is that we chose to limit data based on local time to remove local influence, rather than use the iterative method described on page 8551 of Thoning et al. (1989). We did this to avoid the possibility of the iterative method failing to converge, which would undesireably delay delivery of OCO-2/3 data if it occurred. Differences in monthly averages computed by our method compared to NOAA are usually less than 0.1 ppm, though do reach 0.3 ppm (Fig. S5).

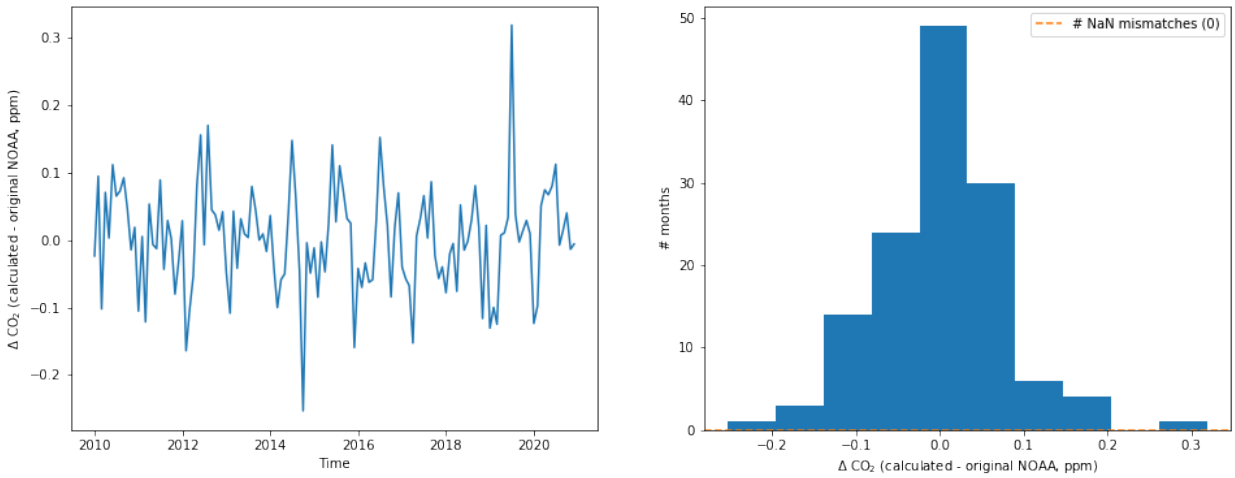

Figure S5: Timeseries (left) and histogram (right) of differences in monthly averages of CO_2 when computed by NOAA versus using the method described here. In the right panel, the "NaN mismatches (0)" indicates that no months had no data in one timeseries but not the other.

S5.2 American Samoa background selection

The background selection for American Samoa data follows the same first three criteria as Mauna Loa (Sect. S5.1), but instead of using local time, uses wind direction during that hour. We limit to hours with wind originating in the north-facing arc between 330° and 160° (degrees clockwise from north) as in Waterman et al. (1989). We use the 10 meter surface wind from the 2D GEOS FP-IT files for this filter. As the GEOS FP-IT data is provided every 3 hours, we interpolate the U and V wind vectors to the time of each hourly data point.

S6 In situ data used for validation

Source	Campaign or ID	Providers	TCCON sites
CO_2 ObsPack	Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS)	Gao Chen & Joshua P. DiGangi (NASA LaRC)	eu
CO_2 ObsPack	CO_2 Budget and Regional Airborne Study (COB2004)	Steve Wofsy (Harvard U.)	pa
CO_2 ObsPack	Deep Convective Clouds & Chemistry (DC3), DC8 aircraft	Gao Chen (NASA LaRC), Joshua DiGangi (NASA LaRC), & Andreas Beyersdorf (CSUSB)	oc
CO_2 ObsPack	Goddard Space Flight Center (GSFC)	Stephan Randolph Kawa, James Brice Abshire, & Haris Riris (NASA GSFC)	df,fc,pa
CO_2 ObsPack	HIAPER Pole-to-Pole Observations (HIPPO)	Steve Wofsy (Harvard U.), & Britton Stephens (NCAR)	db,fc,ll,wg
CO_2 ObsPack	Intercontinental Chemical Transport Experiment - North America (INTEX-NA)	Gao Chen & Joshua P. DiGangi (NASA LaRC)	pa
CO_2 ObsPack	Korea-United States Air Quality Study (KORUS-AQ)	Gao Chen, Joshua P. DiGangi, & Michael Shook (NASA LaRC)	an,df,js,rj
CO_2 ObsPack	O2/N2 Ratio and CO2 Airborne Southern Ocean Study (ORCAS)	Britton Stephens (NCAR), Colm Sweeney (NOAA ESRL), Kathryn McKain (NOAA ESRL), Eric Kort (U. Michigan)	oc
CO_2 ObsPack	Studies of Emissions and Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC4RS), ER-2 aircraft	Gao Chen (NASA LaRC), Joshua DiGangi (NASA LaRC), & Andreas Beyersdorf (CSUSB)	df
CO_2 ObsPack	Studies of Emissions and Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC4RS), DC8 aircraft	Gao Chen (NASA LaRC), Joshua DiGangi (NASA LaRC), & Andreas Beyersdorf (CSUSB)	fc,oc
CO_2 ObsPack	Stratosphere-Troposphere Analyses of Regional Transport (START08)	Steve Wofsy (Harvard U.)	pa
CO_2 ObsPack	Atmospheric Tomography Mission (ATom)	Kathryn McKain (NOAA ESRL), Colm Sweeney (NOAA ESRL), Steve Wofsy (Harvad U.), Bruce Daube (Harvard U.), Roisin Commane (Harvard U.)	ae,ci,df,eu,ll,oc,pa
CH_4 ObsPack	HIAPER Pole-to-Pole Observations (HIPPO), QCLS instrument	Steve Wofsy, Greg Santoni, & Jasna Pittman (Harvard U.)	db,fc,ll,oc,pa,wg
CH_4 ObsPack	Stratosphere-Troposphere Analyses of Regional Transport (START08)	Steve Wofsy (Harvard U.)	pa
CH_4 ObsPack	Atmospheric Tomography Mission (ATom)	Kathryn McKain (NOAA ESRL)	ae,ci,df,eu,ll,oc,pa
IMECC Repository (CO_2, CH_4, CO)	Infrastructure for Measurement of the European Carbon Cycle (IMECC)	Various	bi,br,gm,je,ka,or
NOAA AirCores (CO_2, CH_4, CO)	N/A	Bianca Baier & Colm Sweeney (NOAA ESRL)	df,oc,pa,so
Sodankylä AirCores (CO_2, CH_4, CO)	N/A	Huilin Chen (RUG) & Rigel Kivi (FMI)	so
Nicosia AirCores (CO_2, CH_4, CO)	N/A	Pierre-Yves Quéhé (CARE-C) & Thomas Laemmel (LSCE/IPSL)	ni

Table S1: Airborne profile data used to validate the priors. "CO_2 Obspack" is the CO_2 GLOBALVIEWplus v5.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2019) and "CH_4 ObsPack" the CH_4 GLOBALVIEWplus v2.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2020). The "TCCON sites" column indicates which sites profile were used at, the IDs are mapped to locations in Table S3. In the "Providers" column, affiliations are given in parentheses. If only one affiliation is listed, it applies to all individuals named. Abbrevations: NASA = National Aeronautics and Space Administration; LaRC = Langley Research Center; Harvard U. = Harvard University; CSUSB = California State University San Bernadino; GSFC = Goddard Space Flight Center; NCAR = National Center for Atmospheric Research; NOAA = National Oceanic and Atmospheric Administration; ESRL = Earth System Research Laboratories; FMI = Finnish Meteorological Institure; CARE-C = Climate and Atmosphere Research Center; LSCE/IPSL = Laboratoire des Sciences du Climat et de l'Environnement.

Source	Measurement type	Providers/partners	Location	TCCON site
CO_2 ObsPack	Programmable flask packages	Arlyn Andrews (NOAA ESRL), Ed Dlugokencky (NOAA ESRL), Ankur Desai (U. of WI), & Dan Baumann (USGS)	Park Falls, WI, USA	pa
CO_2 ObsPack	Li-cor NDIR on tower	Arlyn Andrews (NOAA ESRL), Ed Dlugokencky (NOAA ESRL), Ken Davis (PSU), Ankur Desai (U. of WI), & Dan Baumann (USGS)	Park Falls, WI, USA	pa
CO_2 ObsPack	CRDS on tower	Sebastien Biraud & Margaret Torn (LBNL)	Southern Great Plains ARM site, OK, USA	oc
CH_4 ObsPack	Programmable flask packages	Arlyn Andrews (NOAA ESRL), Ed Dlugokencky (NOAA ESRL), Ankur Desai (U. of WI), & Dan Baumann (USGS)	Park Falls, WI, USA	pa
CH_4 ObsPack	CRDS on tower	Arlyn Andrews (NOAA ESRL), Ed Dlugokencky (NOAA ESRL), & Dan Baumann (USGS)	Park Falls, WI, USA	pa
CH_4 ObsPack	Flask	Ed Dlugokencky (NOAA ESRL), Sebastien Biraud (LBNL), & Margaret Torn (LBNL)	Southern Great Plains ARM site, OK, USA	oc
CH_4 ObsPack	CRDS on tower	Sebastien Biraud & Margaret Torn (LBNL)	Southern Great Plains ARM site, OK, USA	oc
NIWA (direct)	Licor 7000 NDIR (CO_2), in situ GHG FTS (CH_4)	Dan Smale (NIWA)	Lauder, New Zealand	ll

Table S2: Ground in situ data used in validating the priors. "CO_2 Obspack" is the CO_2 GLOBALVIEWplus v5.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2019) and "CH_4 ObsPack" the CH_4 GLOBALVIEWplus v2.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2020). The "TCCON sites" column indicates which sites profile were used at, the IDs are mapped to locations in Table S3. In the "Providers" column, affiliations are given in parentheses. If only one affiliation is listed, it applies to all individuals named. Abbrevations: NDIR = Nondispersive infrared; NOAA ESRL = National Oceanic and Atmospheric Administration Earth System Research Laboratories; U. of WI = University of Wisconsin; USGS = United States Geological Survey; LBNL = Lawrence Berkeley National Laboratory; ARM = Atmospheric Radiation Measurement; CRDS = cavity ring-down spectroscopy; NIWA = National Institute of Water & Atmospheric Research Ltd.

Site ID	Site location	Latitude	Longitude
ae	Ascension Island	7.916° S	14.332° W
an	Anmyeondo, South Korea	36.538° N	126.331° E
bi	Bialystok, Poland	53.23° N	23.025° E
br	Bremen, Germany	53.10°	8.85° E
ci	Pasadena, CA, USA (Caltech)	34.136° N	118.127° W
db	Darwin, Australia	12.425° S	130.892° E
df	Dryden, CA, USA (Armstrong AFB)	34.958° W	117.882° W
eu	Eureka, Canada	80.05° N	86.42° W
fc	Four Corners, USA	36.797° N	108.480° W
gm	Garmisch, Germany	47.476° N	11.063° E
je	Jena, Austria	50.91° N	11.57° E
js	Saga, Japan	33.241° N	130.288° E
ka	Karlsruhe, Germany	49.100°N	8.439° E
ll	Lauder, New Zealand	45.038° S	169.684° E
ni	Nicosia, Cyprus	35.141° N	33.381° E
oc	Lamont, OK, USA	36.604° N	97.486° W
or	Orléans, France	47.97° N	2.113° E
pa	Park Falls, WI, USA	45.945° N	90.273° W
rj	Rikubetsu, Japan	43.457° N	143.766° E
so	Sodankylä, Finland	67.367°N	26.631° E
wg	Wollongong, Australia	34.406° S	150.879° E

Table S3: List of TCCON sites and their locations referenced in Table S1. Also note that here we use "ll" to represent Lauder, New Zealand; however, in the TCCON data, Lauder uses three IDs ("lh,", "ll," and "lr") for different instruments operated at different times.

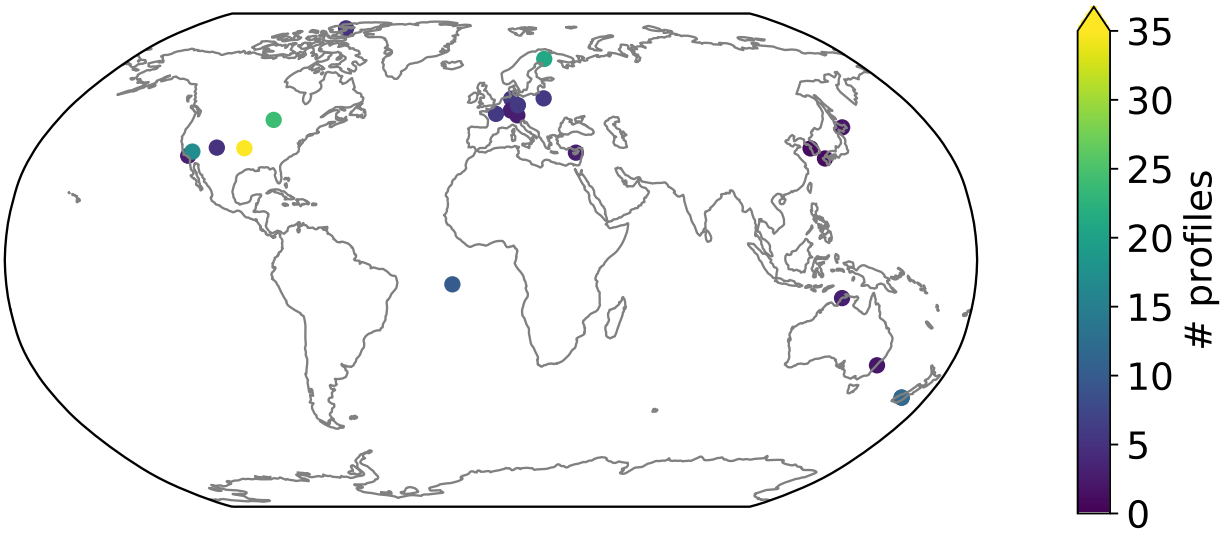

Figure S6: Locations of in situ profiles used to validate the TCCON priors, colored by number of profiles at that location. Note that if a single profile provided more than one gas it is only counted once.

S7 Additional figures and tables

Gas	NOAA ObsPack	NOAA AirCores (v20201223)	FMI AirCores	Nicosia AirCores	IMECC
CO_2	67	33	19	3	10
CH_4	30	33	19	3	10
CO	0	33	19	3	10

Table S4: Breakdown of the number of profiles used to validate the priors by source and gas. NOAA ObsPack indicates either the CO_2 GLOBALVIEWplus v5.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2019) and "CH_4 ObsPack" the CH_4 GLOBALVIEWplus v2.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2020). IMECC are profiles from the Infrastructure for Measurement of the European Carbon Cycle campaign. AirCore profiles are balloon launches by FMI (at Sodankylä, Finland), LSCE (Nicosia, Cyprus) and NOAA (various locations).

Gas	Seasonal cycle coefficient	Latitude coefficient	Secular trend
CO_2	0.007	N/A	N/A
N_2O	0.0	N/A	N/A
CH_4	0.012	N/A	N/A
HF	0.0	N/A	N/A
NO_2	0.0	0.25	0.00
NH_3	0.0	0.20	0.00
HNO_3	0.0	0.10	0.00
H_2CO	0.0	0.20	0.00
HCN	0.0	0.10	0.00
CH_3F	0.0	0.20	0.00
CH_3Cl	0.0	0.20	0.00
CF_4	0.0	0.20	0.00
CCl_2F_2	0.0	0.20	0.00
CCl_3F	0.0	0.20	0.00
CH_3CCl_3	0.0	0.20	0.00
CCl_4	0.0	0.20	0.00
C_2H_6	0.0	0.30	0.00
C_2H_4	0.0	0.30	0.00
C_2H_2	0.0	0.30	0.00
$CHClF_2$	0.0	0.20	0.05
CH_3Br	0.0	0.20	0.00
HCOOH	0.0	0.20	0.00
$CHCl_2F$	0.0	0.20	0.00
SF_6	0.0	0.30	0.00
F_{113}	0.0	0.30	0.00
$F_{142}b$	0.0	0.20	0.00
CH_3OH	0.0	0.20	0.00
CH_3CHO	0.0	0.20	0.00
CH_3CN	0.0	0.20	0.00
NF_3	0.0	0.30	0.00
CHF_3	0.0	0.20	0.00
$F_{141}b$	0.0	0.20	0.00
CH_3COOH	0.0	0.20	0.00
C_3H_8	0.0	0.50	0.00

Table S5: Seasonal cycle, latitude gradient, and secular trend coefficients for each gas considered in the GGG TCCON retrieval. Gases not listed have all 0 values or do not use these values.

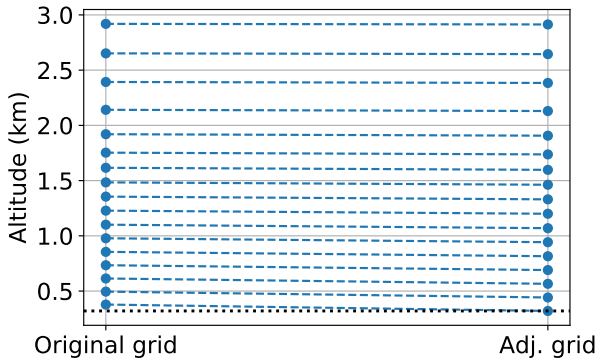

Figure S7: Example of original and adjusted z-grid for Lamont, OK (36.6° N, 97.49° W).

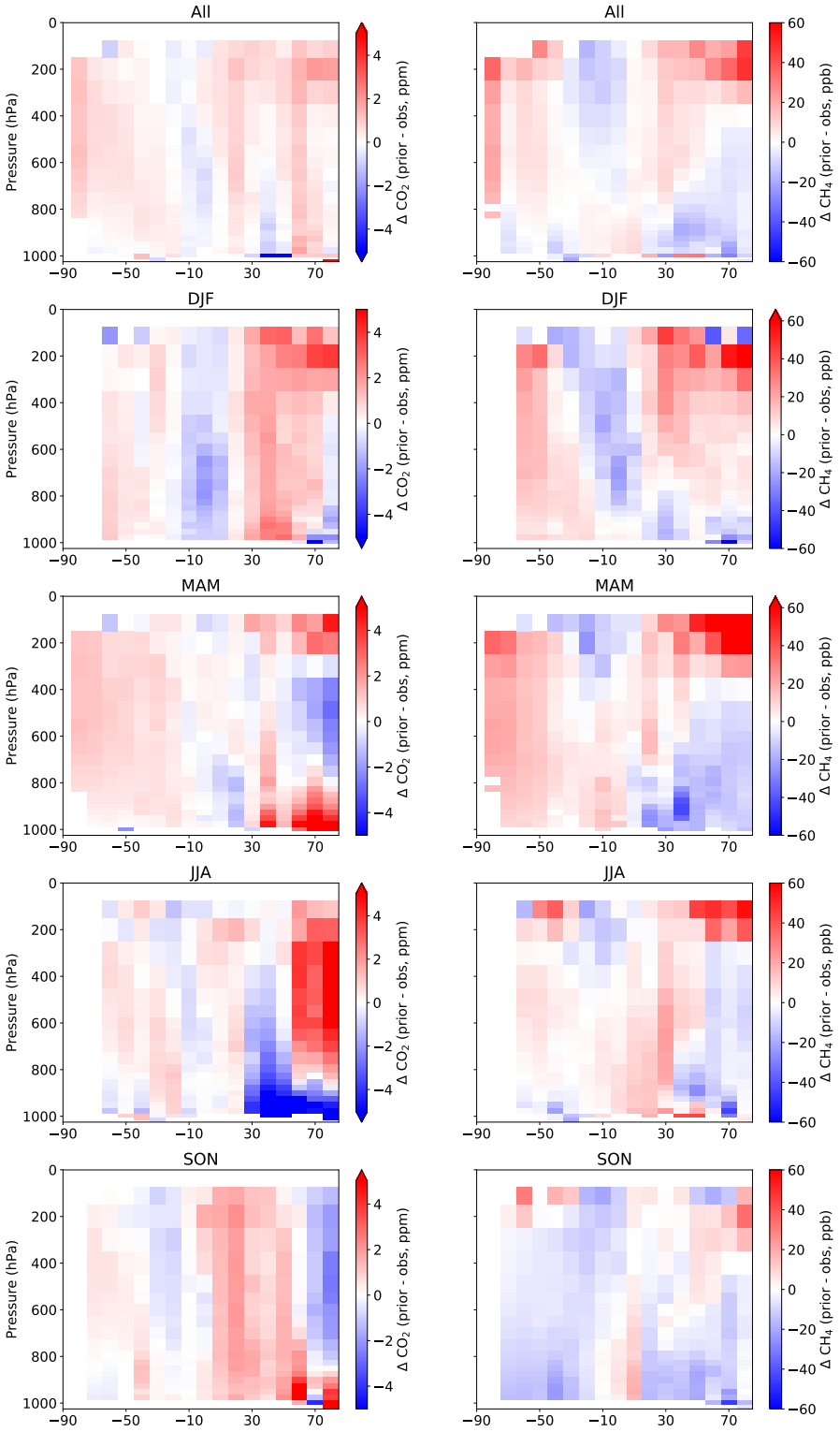

Figure S8: Curtain plots (longitudinal means plotted vs. latitude and pressure) of the difference between the priors and HIPPO + ATom observations. The left column is CO_2, the right column CH_4. The top row is the average over all seasons, the following four split the data into three month bins, indicated by the titles.

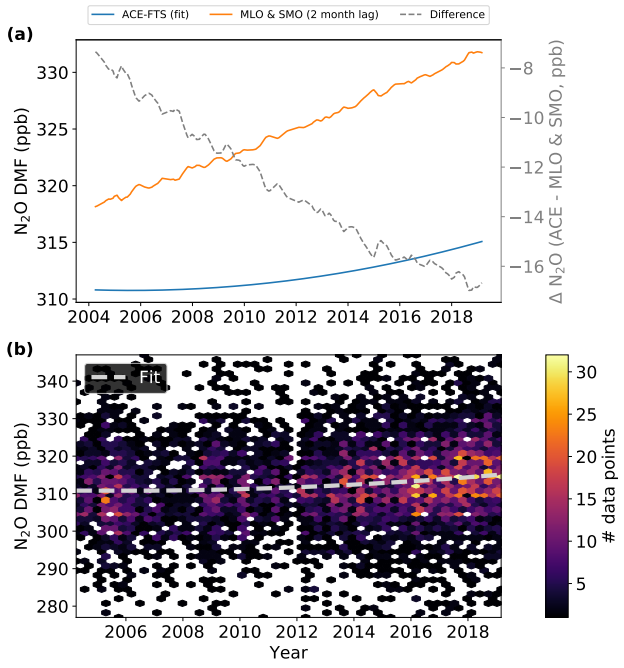

Figure S9: **(a)** Comparing time series of a quadratic fit to N_2O DMFs reported by the ACE-FTS satellite (version 3.6) against MLO & SMO mean N_2O DMFs lagged by two months. ACE-FTS data are only valid data points between latitudes 23°S to 23°N and potential temperatures 360 K and 390 K. These are both estimates of the N_2O DMFs entering the stratosphere in the tropics. The gray dashed line is the difference between the two data sets and is plotted against the right y-axis. **(b)** The fit to ACE-FTS N_2O data shown in panel (a) plotted over a 2D histogram of the individual ACE-FTS N_2O data points.

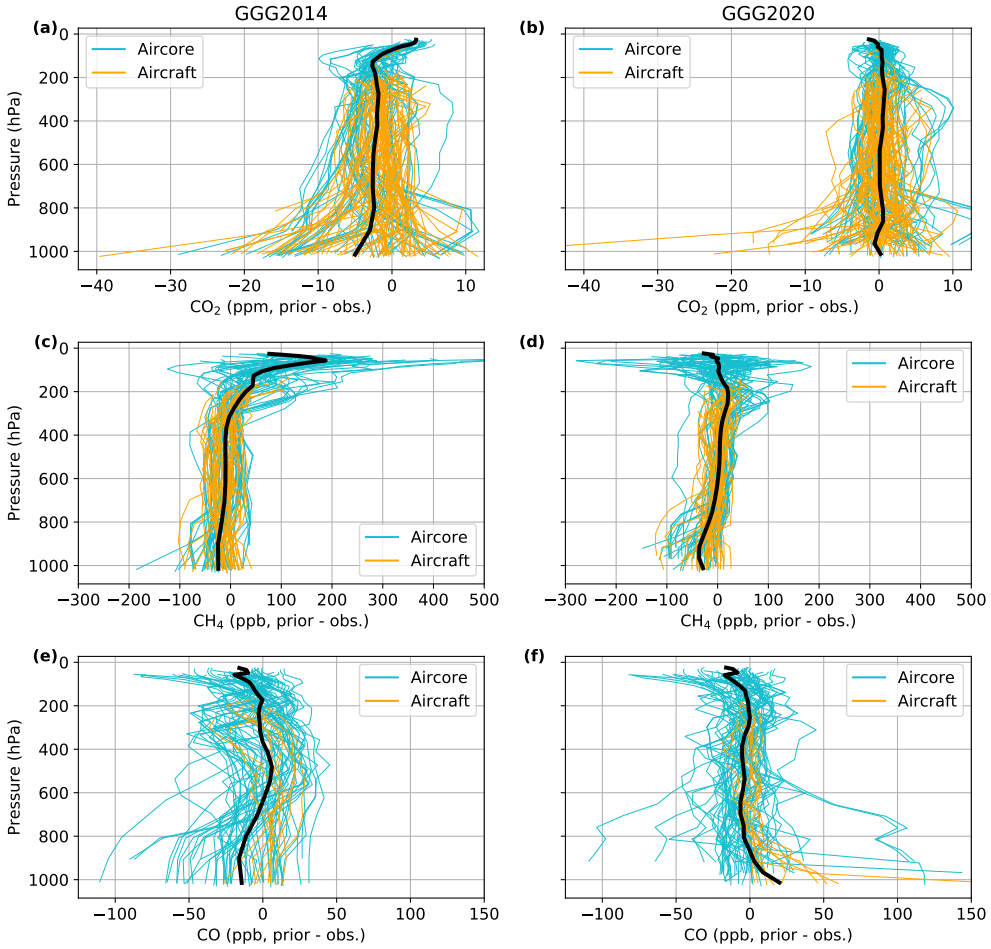

Figure S10: Spaghetti plots of CO_2, CH_4 and CO. The left column shows GGG2014 priors vs. observations, the right column GGG2020 priors vs. observations. As in the main paper, the thin lines are individual profiles' differences and the thick black line is the mean.

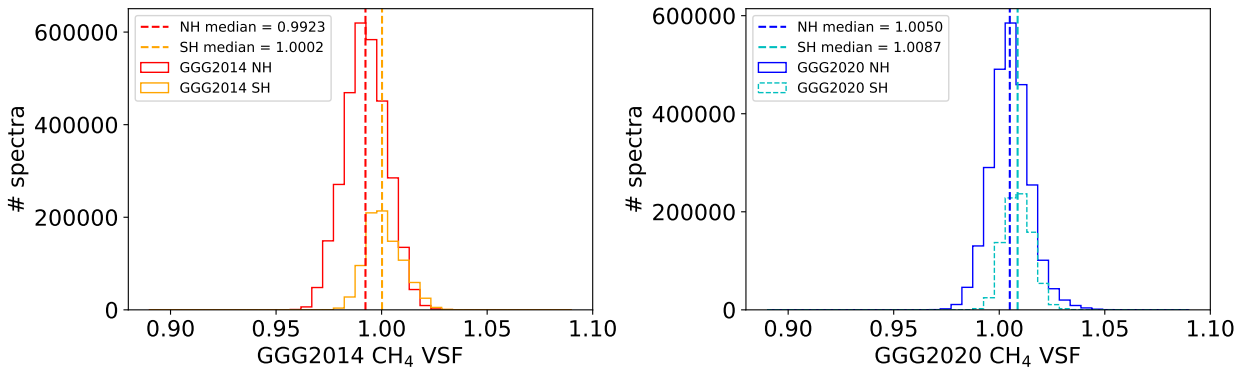

Figure S11: Histograms of volume scale factors (VSFs) for CH_4 for good quality GGG2014 (left) and GGG2020 (right) retrievals, divided into northern and southern hemisphere sites. Medians for each distribution are marked with the dashed vertical line of the same color.

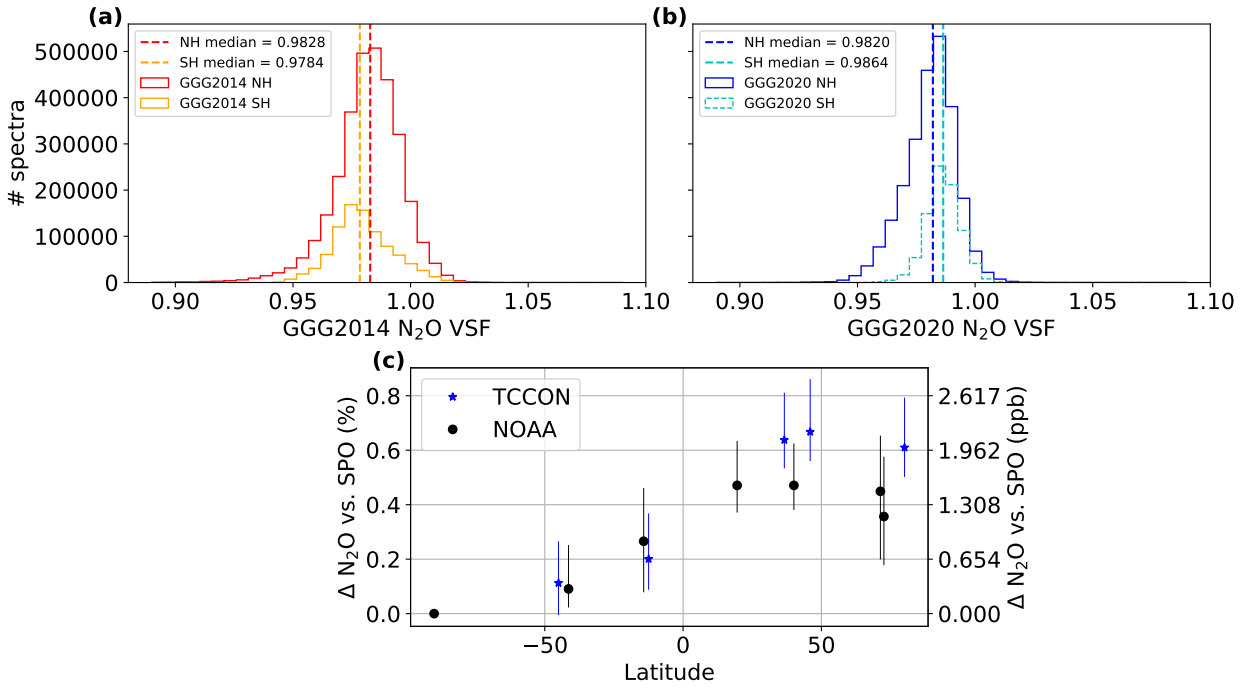

Figure S12: **(a)** and **(b)** are similar to the corresponding panels in Fig. S11, except for N_2O. **(c)** shows latitudinal gradients in surface N_2O from NOAA surface data and the TCCON priors, using the South Pole NOAA station as the baseline. The left axis gives the percent difference, the right axis the absolute difference. (All points line up with both axes.) The other 6 NOAA stations used (from south to north) are Baring Head, New Zealand; Tutuila, American Samoa; Mauna Loa, Hawaii; Niwot Ridge, Colorado; Barrows, Alaska; and Summit, Greenland. TCCON priors are from Lauder, New Zealand; Darwin, Australia; Lamont, Oklahoma; Park Falls, Wisconsin; and Eureka, Canada. Monthly averages from 2011, 2015, and 2019 are used. Each point represents the median of those 36 months for one TCCON site or NOAA station, the error bars give the 5th to 95th percentile range for the 36 monthly values.

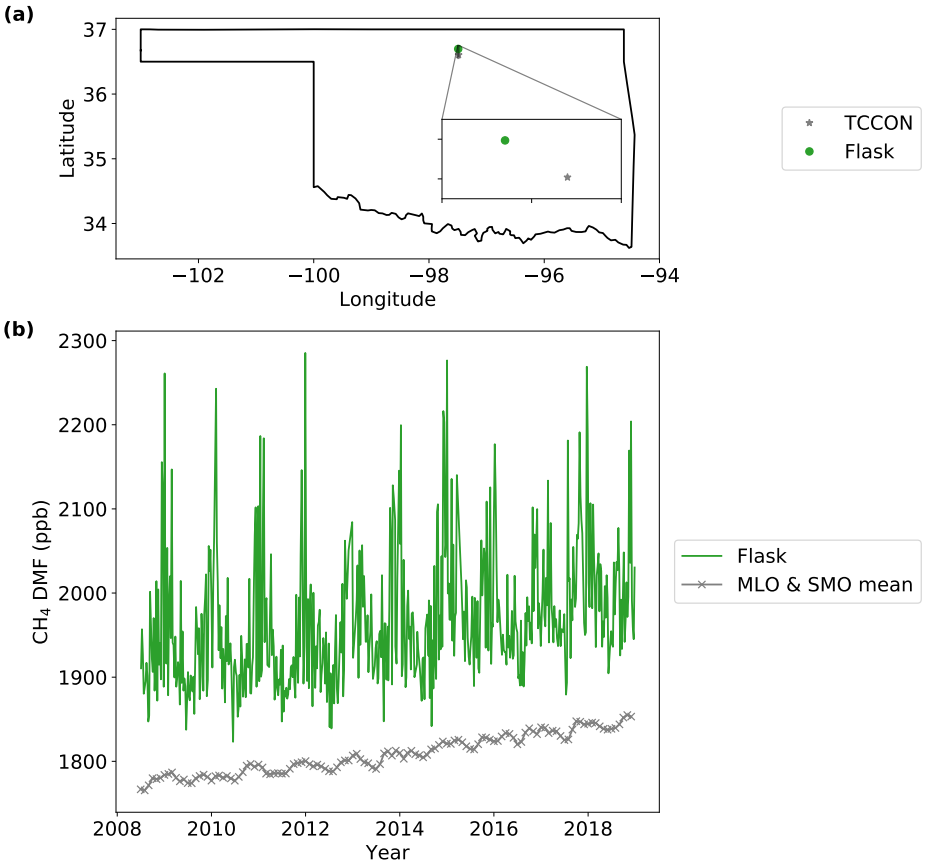

Figure S13: CH_4 enhancement near the Lamont TCCON site. **(a)** Location of the Lamont TCCON site and the Southern Great Plains (SGP) flask CH_4 measurements. **(b)** Timeseries of surface CH_4 dry mole fractions measured by the SGP flask and the MLO & SMO average mole fractions. The difference of 100 to 200 ppb between the MLO & SMO background and the flask measurements is similar to the enhancement seen downwind of wells by Karion et al. (2015), giving us a reasonable estimate of the surface CH_4 enhancement at the Lamont TCCON site due to oil and gas production. The SGP flask data source is listed in Table S2 (CH_4 Obspack - Flask).

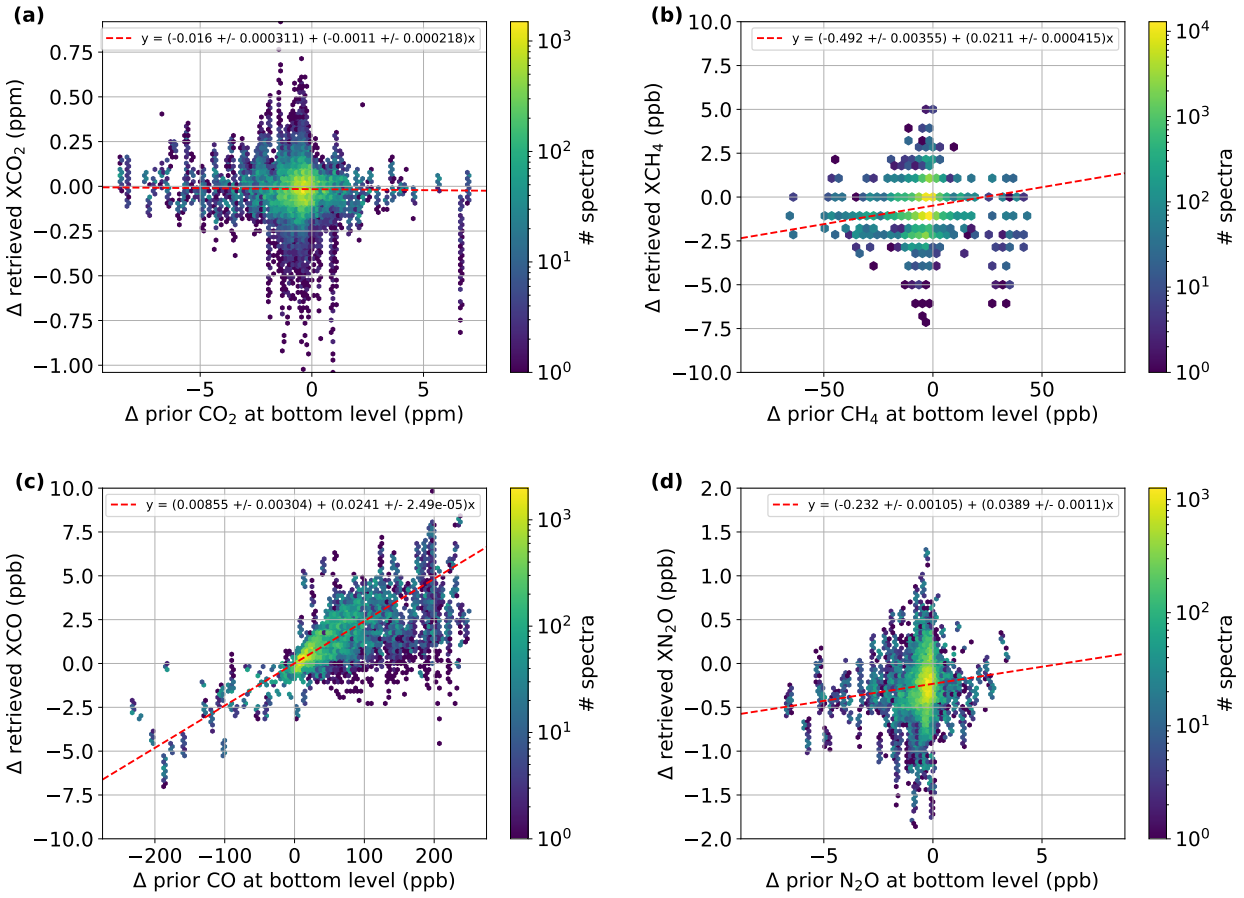

Figure S14: Plot of a sensitivity test at the Armstrong TCCON site using priors generated at the actual site latitude and longitude (34.96 N, 117.88 W) compared to priors generated for a location approximately 70 km NE (35.49 N, 117.51 W). The y-axis shows the change in retrieved Xgas and the x-axis shows the change in the DMF in the bottom level of the priors for **(a)** CO_2, **(b)** CH_4, **(c)** CO and **(d)** N_2O. The fit line is a robust fit. Note that the differences here reflect changes in temperature and pressure profiles, as well as the trace gas prior profiles.

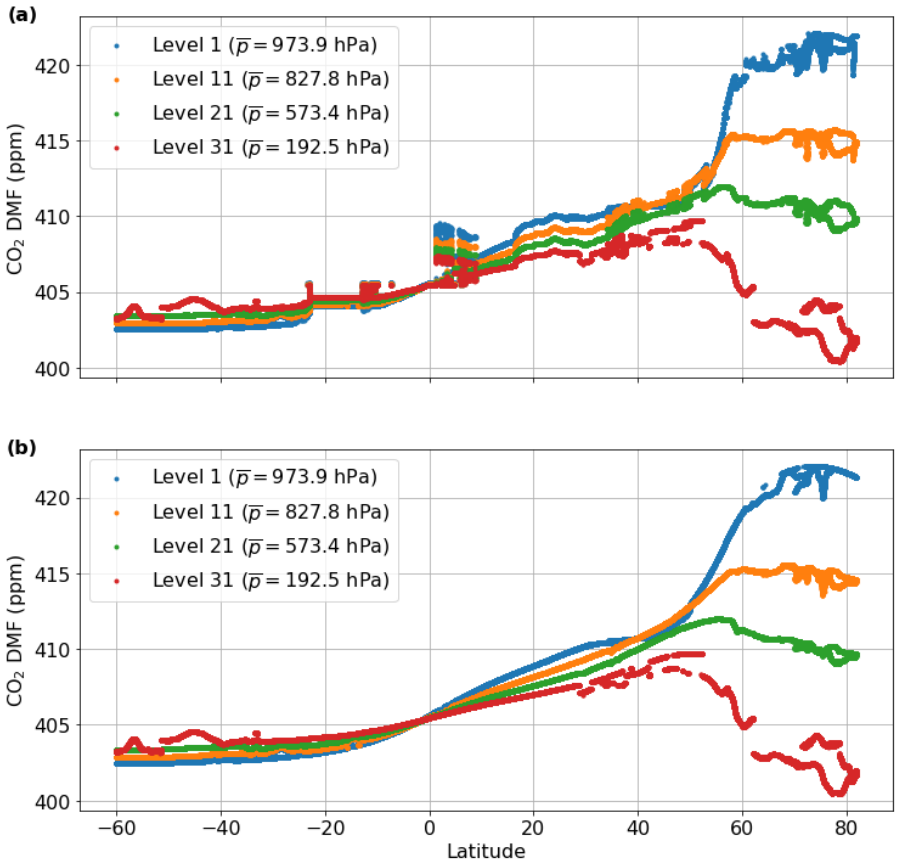

Figure S15: Effect of tropospheric effective latitude on prior CO_2 VMRs for four levels in the priors using an OCO-2 granule spanning 2017-05-14 times 18:22 UTC to 19:09 UTC. Panel **(a)** uses effective latitude, panel **(b)** uses geographic latitude. The legend gives the 1-based index of the vertical level (starting at the surface) and the mean pressure of that level in the granule.

References

Cooperative Global Atmospheric Data Integration Project (2019). *Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2018; obspack_co2_1_GLOBALVIEWplus_v5.0_2019_08_12; NOAA Earth System Research Laboratory, Global Monitoring Division*. DOI: 10.25925/20190812.

— (2020). *Multi-laboratory compilation of atmospheric methane data for the period 1957-2018; obspack_ch4_1_GLOBALVIEWplus_v2.0_2020-04-24; NOAA Earth System Research Laboratory, Global Monitoring Division*. DOI: 10.25925/20200424.

Dlugokencky, E. J. et al. (2009). "Observational constraints on recent increases in the atmospheric CH4 burden". In: *Geophys. Res. Lett.* 36.18. DOI: https://doi.org/10.1029/2009GL039780. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009GL039780. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL039780.

Karion, Anna et al. (July 2015). "Aircraft-Based Estimate of Total Methane Emissions from the Barnett Shale Region". In: *Environmental Science & Technology* 49.13, pp. 8124–8131. DOI: 10.1021/acs.est.5b00217. URL: https://doi.org/10.1021/acs.est.5b00217.

Thoning, K.W., A.M. Crotwell, and J.W. Mund (2021). *Atmospheric Carbon Dioxide Dry Air Mole Fractions from continuous measurements at Mauna Loa, Hawaii, Barrow, Alaska, American Samoa and South Pole. 1973-2020, Version 2021-08-09 National Oceanic and Atmospheric Administration (NOAA), Global Monitoring Laboratory (GML), Boulder, Colorado, USA*. DOI: 10.15138/yaf1-bk21.

Thoning, K.W., P.P. Tans, and W.D. Komhyr (1989). "Atmospheric Carbon Dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC Data, 1974–1985". In: *J. Geophys. Res. Atmos.* 94.D6, pp. 8549–8565.

Waterman, L.S. et al. (1989). "Atmospheric Carbon Dioxide Measurements at Cape Matatula, American Samoa, 1976–1987". In: *J. Geophys. Res. Atmos.* 94.D12, pp. 14817–14829.