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

Joshua L. Laughner[1], Sébastien Roche[2*], Matthäus Kiel[1], Geoffrey C. Toon[1], Debra Wunch[2], Bianca C. Baier[3,4], Sébastien Biraud[5], Huilin Chen[6], Rigel Kivi[7], Thomas Laemmel[8**], Kathryn McKain[3,4], Pierre-Yves Quéhé[9], Constantina Rousogenous[9], Britton B. Stephens[10], Kaley Walker[2], and Paul O. Wennberg[11,12]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2]Department of Physics, University of Toronto, Toronto, Canada
[3]Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA
[4]Cooperative Institute for Research in Environmental Sciences, University of Colorado - Boulder, Boulder, CO, USA
[5]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[6]Center for Isotope Research, University of Groningen, Groningen, the Netherlands
[7]Space and Earth Observation Centre, Finnish Meteorological Institute, Sodankylä, Finland
[8]Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), UMR CEA-CNRS-UVSQ, Gif-sur-Yvette, France
[9]Climate and Atmosphere Research Centre (CARE-C), The Cyprus Institute, Nicosia, Cyprus
[10]Earth Observing Laboratory, National Center for Atmospheric Research (NCAR), Boulder, CO, USA
[11]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
[12]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA
[*]now at: School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[**]now at: Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Bern, Switzerland

Correspondence: Joshua L. Laughner (josh.laughner@jpl.nasa.gov) or Paul O. Wennberg (wennberg@caltech.edu)

Abstract. Optimal estimation retrievals of trace gas total columns require prior vertical profiles of the gases retrieved to drive the forward model and ensure the retrieval problem is mathematically well-posed. For well-mixed gases, it is possible to derive accurate prior profiles using an algorithm that accounts for general patterns of atmospheric transport coupled with measured time series of the gases in questions. Here we describe the algorithm used to generate the prior profiles for GGG2020, a new version of the GGG retrieval that is used to analyze spectra from solar-viewing Fourier transform spectrometers, including the Total Carbon Column Observing Network (TCCON). A particular focus of this work is improving the ~~description~~ accuracy of CO_2, CH_4, N_2O, HF, and CO ~~in the~~ across the tropopause and into the lower stratosphere. We show that the revised priors agree well with independent in situ and space-based measurements~~and improve the~~, and discuss the impact on the total column retrievals.

1 Introduction

The Total Carbon Column Observing Network (TCCON) has been in operation since 2004, beginning with its first dedicated instrument in Park Falls, WI, USA (Wunch et al., 2011). Since then, the network has expanded to 29 active sites located around the world. The network provides column average dry mole fractions (DMFs) of numerous gases, including carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluoric acid (HF), and carbon monoxide (CO). These observations have been used to infer or evaluate natural and anthropogenic carbon fluxes (e.g. Yang et al., 2007; Chevallier et al., 2011; Keppel-Aleks et al., 2012; Basu et al., 2013; Fraser et al., 2013; Ott et al., 2015; Peng et al., 2015; Deng et al., 2016; Wang et al., 2016; Feng et al., 2017; Hedelius et al., 2018; Crowell et al., 2019; Babenhauserheide et al., 2020; Dogniaux et al., 2020; Sussmann and Rettinger, 2020; Zhang et al., 2020; Villalobos et al., 2021), to study carbon transport (e.g. Keppel-Aleks et al., 2012; Polavarapu et al., 2016), and to provide ground-truth values for space-based measurements of CO_2 and CH_4, including the Greenhouse gas Observing Satellites (GOSAT and GOSAT-2, e.g. Butz et al., 2011; Cogan et al., 2012; Schepers et al., 2012; Boesch et al., 2013; Frankenberg et al., 2013; Liu et al., 2013; Oshchepkov et al., 2013; Yoshida et al., 2013; Dils et al., 2014; Inoue et al., 2014; Heymann et al., 2015; Ohyama et al., 2015; Parker et al., 2015; Dupuy et al., 2016; Inoue et al., 2016; Kulawik et al., 2016; Schepers et al., 2016; Liang et al., 2017a; Ohyama et al., 2017; Velazco et al., 2019), TanSat (Yang et al., 2020), the Orbiting Carbon Observatories (OCO-2 and OCO-3, e.g. Liang et al., 2017a, b; Wunch et al., 2017; Kiel et al., 2019), and the Tropospheric Monitoring Instrument (TROPOMI, e.g. Borsdorff et al., 2019; Schneising et al., 2019; Lorente et al., 2021).

The TCCON instruments are solar-viewing Bruker 125HR (high resolution) Fourier transform infrared (FT-IR) spectrometers, which record an interferogram once every few minutes. These interferograms are processed by the GGG software package to provide column average DMFs. Once the interferograms are converted to spectra, the core routine of GGG calculates the expected spectra from a forward model based on a custom linelist and a priori profiles of the absorbing gases with absorption lines in the fitting window. The retrieval calculates a posterior trace gas profile that minimizes the root mean square (RMS) fitting residuals between the forward modeled and observed spectra.

There are two common terms used to describe different approaches towards finding the optimal posterior profile: a "scaling" retrieval or a "profile" retrieval. In a scaling retrieval, the retrieval multiplies the entire prior profile by a single value, finding the scaled version that produces the best agreement with the observed spectrum. In a profile retrieval, each level of the profile can be varied, with the allowed variation constrained by a specific covariance matrix. Compared to a profile retrieval, a scaling retrieval is faster and does not alias spectroscopic or instrument line shape errors into profile shape errors. It is more sensitive to errors in the *shape* of the prior profile compared to a full profile retrieval because it cannot change the shape of the posterior solution (meaning the ratio of DMFs between levels in the profile cannot change). However, it is not affected by a *uniform multiplicative* error in the prior DMFs at all altitudes. That is, if the entire profile under- or over- estimates the true atmospheric DMFs by the same multiplicative factor, a scaling retrieval can—in theory—perfectly correct the retrieved profile. Roche et al. (2021) examines the differences between scaling and profile retrievals in the context of TCCON data in more detail.

The relationship between the shape error in the prior and the error in the retrieved column amount depends on the averaging kernels. For TCCON CO_2 retrievals, testing with synthetic spectra shows that a ~~1%~~ 4 ppm error in the profile shape (defined as the error in the prior compared to the true profile changing by ± 4 ppm between the top and bottom levels) leads to an error of $\leq 0.025\%$ in XCO_2 at solar zenith angles (SZAs) $\lesssim 60°$, and $\leq 0.125\%$ up to SZA $\approx 75°$. (Details of how this was quantified are given in Sect. S1.) This means that for typical SZAs observed by TCCON, an error of ~~1% to 2% (~~about 4 to 8 ppm ~~)~~ in the CO_2 prior results in a retrieval error well below the 0.25% ceiling required for TCCON data.

In both GGG2014 and GGG2020, the prior profiles are derived as much as possible from meteorological variables and general correlations between these variables and trace gas DMFs in the atmosphere. GGG2014 used meteorological reanalyses from the National Centers for Environmental Prediction (NCEP). GGG2020 uses the Goddard Earth Observing System Forward Product for Instrument Teams (GEOS-5 FP-IT or GEOS FP-IT) reanalysis product. The GEOS FP-IT product was chosen because it is provided on a finer temporal resolution than the NCEP product (3 hourly vs. 6 hourly), is available with a lag of one day in normal operation, and includes diagnosed potential vorticity (PV). The PV fields are of particular importance because they allow the GGG2020 priors to better represent latitudinal transport in the stratosphere, thus improving the stratospheric trace gas profiles. However, GEOS FP-IT data is only available from the year 2000 on, so the GGG package retains the capability to use NCEP meteorology as input data. This capability has been further developed since GGG2014, though we do not include those changes in this paper.

Here, we describe the algorithm used to compute the prior profiles of CO_2, N_2O, CH_4, HF, CO, H_2O, and O_3 for GGG2020. The algorithm is named "ginput" and is available through GitHub (Laughner, 2022). We begin in this paper by describing the core parts of the algorithm that are common across many of the gases (Sect. 2). We then address elements specific to individual gases in Sect. 3. Finally, we compare the GGG2014 and GGG2020 priors against a wide variety of observations in Sect. 5.

As a final note, the CO_2 priors described here are also used in the versions 10 and 11 OCO-2/3 retrievals. There are small differences in the OCO-2/3 priors compared to the TCCON priors which are discussed in Sect. 4.

2   General design

The central algorithms for the GGG2020 (CO_2, N_2O, CH_4) priors are similar to each other. Trace gas mole fractions are tied to the monthly average measurements in whole-air flasks sampled at the Mauna Loa, Hawaii (MLO) and American Samoa (SMO) sites operated by the United States National Oceanic and Atmospheric Administration's (NOAA's) Global Monitoring Laboratory. The fundamental underlying assumption of the GGG2020 priors algorithm is that the spatial variation in these gases can be largely captured by accounting for the transport lag between the location of the prior profile and the tropics (where MLO & SMO flask samples are made), and chemistry occurring during stratospheric transport.

The MLO & SMO data used to create the GGG2020 priors ends in December 2018. In order to ensure consistent priors are created with this version of GGG, these files will not be updated until the next GGG release even as NOAA releases more data in the interim. Therefore, it is necessary to extrapolate the MLO & SMO records forward in time for retrievals of spectra taken after December 2018. This is done by:

Gas	$f(t)$	n (years)
CO_2	$c_0 e^{c_1 t}$	10
CH_4	$c_0 + c_1 t + c_2 t^2$	5
N_2O	$c_0 + c_1 t + c_2 t^2$	10

Table 1. Function forms ($f(t)$) and number of years used to fit the combined MLO & SMO DMF record to extrapolate beyond 2018. In $f(t)$, the c's are the fit parameters.

1. Fitting a function, $f(t)$ to the last n years of the MLO & SMO records. Both $f(t)$ and n are chosen for each gas to best represent that gas's behavior.

2. Calculating the average seasonal cycle over the last n years as the anomaly relative to $f(t)$.

3. Extend the record to the necessary date using $f(t)$ as the baseline and applying the average seasonal cycle on top of it.

This procedure is shown graphically in Fig. 1.

Details of $f(t)$ and n are provided in Table 1. Note: this method is also used to extrapolate back in time if data prior to the start of the combined MLO & SMO record is needed to represent the distribution of ages of air in the stratosphere (see Sect. 2.3).

Errors in extrapolating the MLO & SMO DMFs will negatively impact the TCCON retrievals if the error in extrapolation introduces an error in the profile shape, due to an El Niño year, for example. In a scaling retrieval, such as the GGG algorithm used by TCCON, the posterior optimal profile is the prior profile multiplied by a scale factor, with the same scale factor applied to all levels. At its core, the algorithm we are describing here builds the priors by calculating what date to pull the MLO & SMO DMFs from for each level in the prior. If the extrapolation error caused all the MLO & SMO DMFs to be incorrect by the same percentage, this would manifest as the prior profile being incorrect by that percentage, for which a scaling retrieval can theoretically perfectly account. However, if the error in MLO & SMO DMFs is not the same for each level in the prior, that means the error in the prior cannot be represented by the same scalar multiplier for every level, and so a scaling retrieval could never completely eliminate the error in the posterior profile.

Currently, we estimate the error in the MLO & SMO DMFs due to extrapolation to be about 0.25% for CO_2, 0.15% for N_2O, and 0.6% for CH_4 (see over a five-year extrapolation (see Sect. S2 in the supplement for details). A scaling retrieval, such as the GGG algorithm used by TCCON, can theoretically perfectly account for an error in magnitude of the prior, thus we We deem this level of uncertainty acceptable for TCCON priors. However, How errors in the priors alias into the posterior state in a profile retrieval(, such as that used by OCO-2 and -3, is more complex. However, the OCO-2/3) will be more negatively impacted by such errorsretrieval uses a relatively tight covariance matrix for levels in the stratosphere (see Fig 3-15 of Crisp et al., 2021), making it important that the priors not exhibit any long-term drift in these levels. Therefore, when these priors are used for the version 11 OCO-2/3 retrievals, more recent NOAA data is ingested (see Sect. 4).
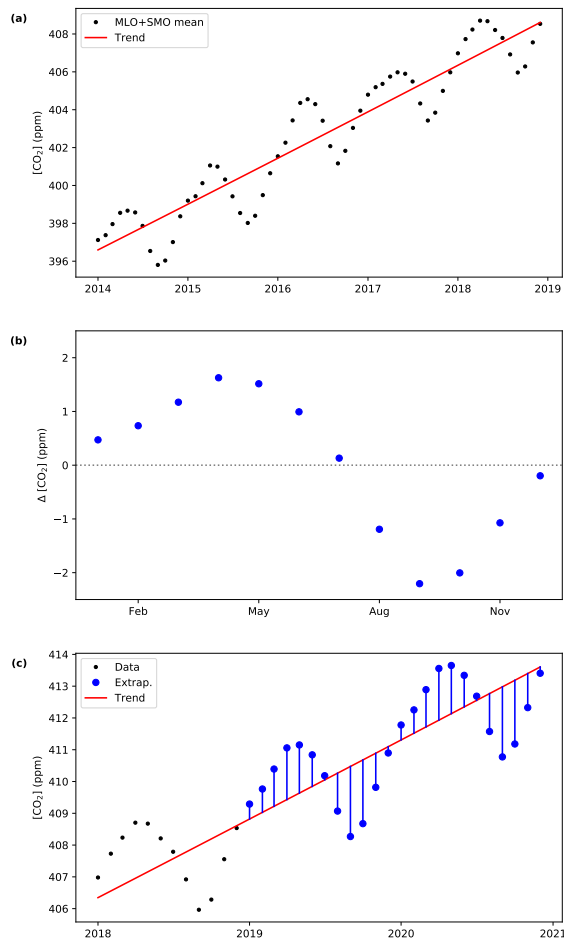
Figure 1. Process to extrapolate the combined MLO & SMO monthly average record. (**a**) Fit the last 5 or 10 years with the best function for a given gas. (**b**) Calculate the mean monthly anomaly relative to the trend over the same time period. (**c**) extend the trend in time and apply the mean monthly anomalies on top of it.

Gas	Scale (GGG2020)	Scale (OCO-2/3 v10)	Scale (OCO-2/3 v11)
CO_2	X2007	X2007	X2019
CH_4	X2004	N/A	N/A
N_2O	X2006	N/A	N/A
CO^*	X2014A	N/A	N/A

Table 2. The WMO calibration scales to which the in situ data used in the GGG2020 and OCO-2/3 priors are tied. [*]Note that, unlike for CO_2, N_2O, and CH_4 (for which this tie comes from the MLO & SMO data), for CO this is from scaling to ATom data in the troposphere.

Ingesting the MLO & SMO data as the basis for the priors effectively ties those priors to the WMO scale to which the MLO & SMO data are calibrated. Table 2 describes which scale each gas is tied to for each algorithm in which these priors are used. As these priors were developed at the same time as the X2019 CO_2 scale (Hall et al., 2021), whether the CO_2 priors are tied to the X2007 or X2019 CO_2 scale depends on which scale the MLO & SMO data are calibrated to.

Unlike the other gases in Table 2, CO is not tied to its scale through the MLO & SMO data. CO priors are created using a different approach to the other primary gases; this approach will be described in Sect. 3.6. The relevant point here is that CO is taken from the GEOS FP-IT product (Lucchesi, 2015) and, in the troposphere, is scaled to match observations from the first three Atmospheric Tomography Mission (ATom) aircraft campaigns (Thompson et al., 2022). As the ATom QCLS CO observations used were calibrated to the X2014A scale, the CO priors are considered tied to that scale.

Several gases (CO, H_2O, HDO, O_3) are contained in the GEOS-5 FP-IT meteorology product ingested by GGG2020. H_2O and O_3 are taken directly from GEOS-5 FP-IT, while CO and HDO are derived from GEOS-5 FP-IT. Details are given in Sect. 3.

Finally there are a large number of gases that must be accounted for as interfering absorbers during retrievals of primary TCCON target gases. These gases use priors derived from climatological profiles from the summer at $35°$ N. Details are given in Sect. 2.4.

2.1 Design rational

In developing the GGG2020 priors, we had two guiding principles in mind:

1. Minimize direct dependence on other measurements or models as much as possible such that retrievals using these priors are independent measurements (in the statistical sense) that other observations or models can be compared to.

2. Produce an algorithm which generates reproducible prior profiles if run at different times.

The first principle is why the GGG2020 priors only ingest MLO & SMO data, rather than more surface data or why we do not use modeled gas profiles (other than for CO). For the much shorter-lived CO, we decided that capturing the spatial variability was worth the trade off of relying on GEOS FP-IT modeled CO (especially as GGG2020 already uses GEOS FP-IT

125   meteorology). Other data used in generating the priors (e.g. latitudinal gradients of CO_2 and CH_4 from HIPPO & ATom, ACE-FTS profiles) were likewise adopted because the improvement in the priors was deemed worth the loss of statistical independence. Since these data are used to generate static values (such as lookup tables or coefficients in functions) rather than being directly ingested, we retain some independence from these sources.

    The second principle is why the GGG2020 priors and OCO-2/3 v10 priors only use MLO & SMO flask data through the

130   end of 2018, rather than updating regularly. One concern raised during development was whether such regular data updates would alter previously obtained data, such as from retrospective quality control. This would introduce a situation where we could not exactly reproduce priors generated using an old version of the input data. Given time constraints, it was not possible to engineer a solution to detect or avoid this issue for GGG2020 and OCO-2/3 v10 priors. With the additional development time for OCO-2/3 v11, we were able to update the priors algorithm to safely ingest more rapidly updated MLO & SMO data.

135 ## 2.2   Tropospheric prior

The GGG2020 tropospheric priors assume that the trend observed by MLO & SMO is driven by emissions in the northern midlatitudes, thus the measured DMF at MLO & SMO will lag behind the DMFs in the northern hemisphere and precede the DMFs in the southern hemisphere. To compute the tropospheric DMFs, we average MLO & SMO data together with equal weight, deseasonalize the MLO & SMO average to get the underlying trend, approximate the offset forward or backward in

140   time relative to MLO & SMO with an idealized distance function, apply a multiplicative and additive correction to match observed latitudinal gradients, and impose a latitudinally-dependent seasonal cycle. Mathematically, this follows Eq. (1):

$$\mathrm{DMF}(l, z, z_{\mathrm{trop}}, f_y) = s(l, z, f_y, d) \cdot [\alpha(d) \cdot \mathrm{DMF}_{\mathrm{ref}}(d) + \beta \cdot l] \tag{1}$$

The variables in this function are:

- l is latitude. In the GGG2020 TCCON priors, this is an "effective latitude" derived from mid-tropospheric potential

145   temperature (c.f. Sect. 2.2.1).

- z is altitude with the bottom half of the troposphere stretched downward slightly to treat the bottom layer as being at the surface for the purpose of this calculation (c.f. Sect. 2.2.2).

- z_{trop} is the tropopause altitude

- f_y is the fractional year (defined as 1-based day-of-year / 365.25)

150   - $\mathrm{DMF}_{\mathrm{ref}}$ is the reference DMF taken from a deseasonalized MLO & SMO trend

- d is the distance offset function, defined by Eq. (2)

- s is the seasonal cycle factor, defined by Eq. (4d)

Gas	α	β
CO_2	$-3.55 \cdot d(l,z;l_{ref},z_{trop}) \cdot \frac{\partial DMF_{ref}}{\partial t}$	0
N_2O	$\exp\left(\frac{-d(l,z;l_{ref},z_{trop})}{121\text{ yr}}\right)$	0
CH_4	$\exp\left(\frac{-d(l,z;l_{ref},z_{trop})}{12.4\text{ yr}}\right)$	$\begin{cases} 0.75\text{ ppb/}^\circ & \text{for} \quad l \geq 0 \\ 0 & \text{for} \quad l < 0 \end{cases}$

Table 3. Values of α and β coefficients in Eq. (1) for the three primary well-mixed gases. Rationales for these choices are given in the gas-specific sections (Sect. 3).
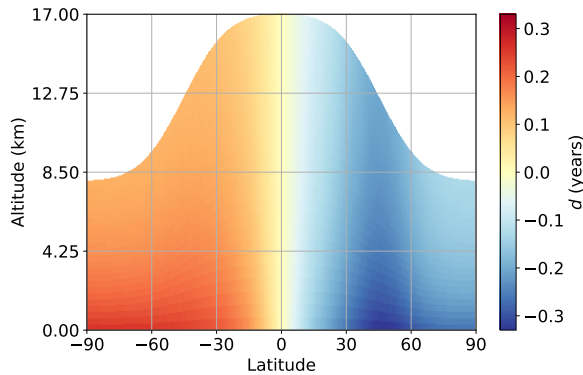
Figure 2. Form of the distance function d assuming that emissions occur at 45° N and a latitudinally-dependent tropopause height that varies smoothly from 17 km at the equator to 8 km at the poles.

  – α and β are coefficients that scale and adjust the ideal gradients assumed by d to account for differences between gases. Their values are given in Table 3 and are discussed in detail in Sect. 3.

155   The distance function d is shown in Fig. 2 (assuming a simple latitudinal dependence for the tropopause altitude). It has the mathematical form:

$$d = d'(l,z;l_{ref},z_{trop}) - d'(0^\circ, 0.01\text{ km}; l_{ref}, z_{trop}) \tag{2}$$

where

$$d'(l,z;l_{ref},z_{trop}) = 0.313 - 0.085 \cdot \exp\left(-\left[\frac{l-l_{ref}}{18}\right]^2\right) - 0.268 \cdot \exp\left(-1.42\frac{z}{z+z_{trop}}\right) \cdot \frac{l/22}{\sqrt{1+(l/22)^2}} \tag{3}$$

160   Although d has units of years, it does not represent a physical age or time. It is effectively a basis function to impose the ideal distribution of DMFs relative to MLO & SMO as shown in Fig. 2. Specifically, it assumes that surface DMFs precede

MLO & SMO DMFs in the northern hemisphere, lag MLO & SMO DMFs in the southern hemisphere, and have a smaller latitudinal gradient in the upper troposphere due to faster winds. The basic shape is modified for each gas via α and β.

DMF_{ref} in Eq. (1) is the combined MLO & SMO record, deseasonalized by taking a 12-month rolling mean. This is done because the seasonal cycle at MLO & SMO is not representative of all latitudes. We impose a latitudinally dependent seasonal cycle by multiplying the DMFs by a scaling factor s:

$$s_v = \sin\left(2\pi \cdot [f_y - 0.78]\right) \tag{4a}$$

$$s_l = \frac{s_v \cdot l/15}{\sqrt{1 + (l/15)^2}} \tag{4b}$$

$$s_a = s_l \cdot \exp\left(-d'(l, z; l_{\text{ref}}, z_{\text{trop}})/0.85\right) \tag{4c}$$

$$s = 1 + s_a \cdot c_{\text{gas}} \tag{4d}$$

for all gases but CO_2. For CO_2 the parameterization is:

$$s_v = \sin\left(2\pi \cdot [f_y - 0.834 - d]\right) \tag{5a}$$

$$s_l = s_v + 1.8 \cdot \exp\left(-\left[\frac{l - 74}{41}\right]^2\right) \cdot (0.5 - s_v^2) \tag{5b}$$

$$s_a = s_l \cdot \exp\left(-d/0.2\right) \cdot \left\{1 + 1.33 \cdot \exp\left(-\left[\frac{l - 76}{48}\right]^2\right) \cdot \frac{z + 6}{z + 1.4}\right\} \tag{5c}$$

$$s = 1 + s_a \cdot c_{\text{gas}} \tag{5d}$$

where f_y is the fraction of year passed (defined as 1-based day-of-year / 365.25), l is latitude, z is altitude (in kilometers), z_{trop} the tropopause altitude (in kilometers), l_{ref} a reference latitude (45° N), d' is the function from Eq. (3), and c_{gas} is a gas-specific constant defined in Table S5. s_v represents the basic seasonal variation, s_l the latitudinal variation, s_a the altitude variation. The form of these equations for CO_2 and CH_4 are shown in Fig. 3.

These parameterized seasonal cycles are the same as that used in GGG2014 priors. The amplitude and phase were derived from surface in situ data and the amplitude is assumed to decay with altitude due to mixing of airmasses with different ages.

2.2.1 Potential temperature-based effective latitude

CO_2 profiles for locations on the edge of the tropics are sometimes more "tropical" in nature than their geographic latitudes suggest. In these cases, the observed profile would be more constant versus altitude than the prior profile, which would have some drawdown at the surface.

Keppel-Aleks et al. (2012) showed that, in the extratropics, there is a correlation between 700 hPa potential temperature and CO_2 DMFs in the free troposphere, as variations in this potential temperature serve as an indicator of synoptic-scale motion,
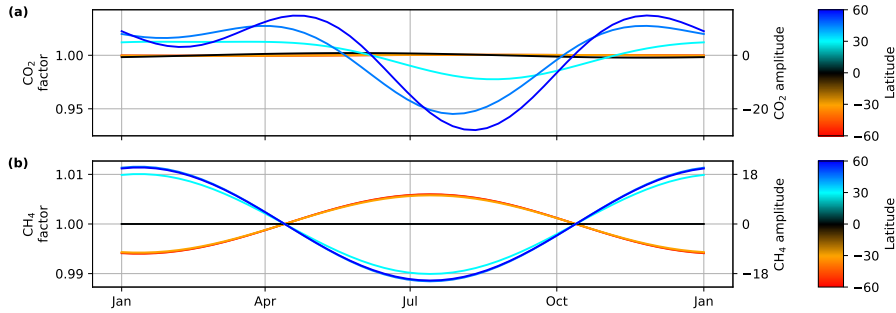
Figure 3. Parameterized seasonal cycle for **(a)** CO_2 and **(b)** CH_4. The left y-axis is the factor s in Eq. (4d) and (5d). The right y-axis gives what the seasonal cycle amplitude would be for a CO_2 DMF of 400 ppm in **(a)** and CH_4 DMF of 1800 ppb in **(b)**.

and therefore the true source latitude of the air. We use dry potential temperature, i.e. the temperature a parcel of dry air would have if brought to a pressure of 1000 hPa adiabatically. This allows us to use potential temperature to derive an "effective latitude" that better predicts the shape of the prior profile. Note that while this was originally developed to improve the CO_2 priors, it is used for all gases.

To calculate this effective latitude, we first build a climatology of mid-tropospheric potential temperature from the GEOS-5 FP-IT product by averaging potential temperature between 500 and 700 hPa (henceforth termed θ_{mid}) vs. latitude for two-week periods in 2018 (Fig. 4a). A hypothetical example is shown in Fig. 4b. For a prior in the extratropics, we select the appropriate θ_{mid}-vs.-latitude curve from the table (Fig. 4b, black line) and compare the θ_{mid} value for the prior against the tabulated mean. If the prior's θ_{mid} is greater than the mean θ_{mid} for that latitude, the effective latitude is moved equatorward until it matches, and vice versa if the prior's θ_{mid} is less.

More specifically, the implementation searches north and south of the prior's geographic latitude for the two latitudes (one north, one south) with the smallest difference between the prior's θ_{mid} and the mean θ_{mid}. If the difference between the mean θ_{mid} values at both latitudes is within 0.25 K, then the nearer latitude is used. Otherwise, the latitude with the smallest difference between its θ_{mid} and the prior's θ_{mid} is used.

There are two caveats to this approach. First, the effective and true (geographic) latitude must have the same sign—that is, both must be in the same hemisphere. Second, within the tropics (defined as $\pm20°$ of the equator), the effective latitude calculation is disabled and the geographic latitude is used. This is done because mid-tropospheric temperature gradients are weak in the tropics and largely uncorrelated with zonal advection (Sobel et al., 2001). To smoothly blend between geographic and effective latitude, a linear interpolation between them occurs in the $20°$ to $25°$ range. For example, a profile at $22°$ N would have a latitude calculated as $0.6l_g + 0.4l_e$, where l_g is the geographic latitude and l_e the effective latitude.
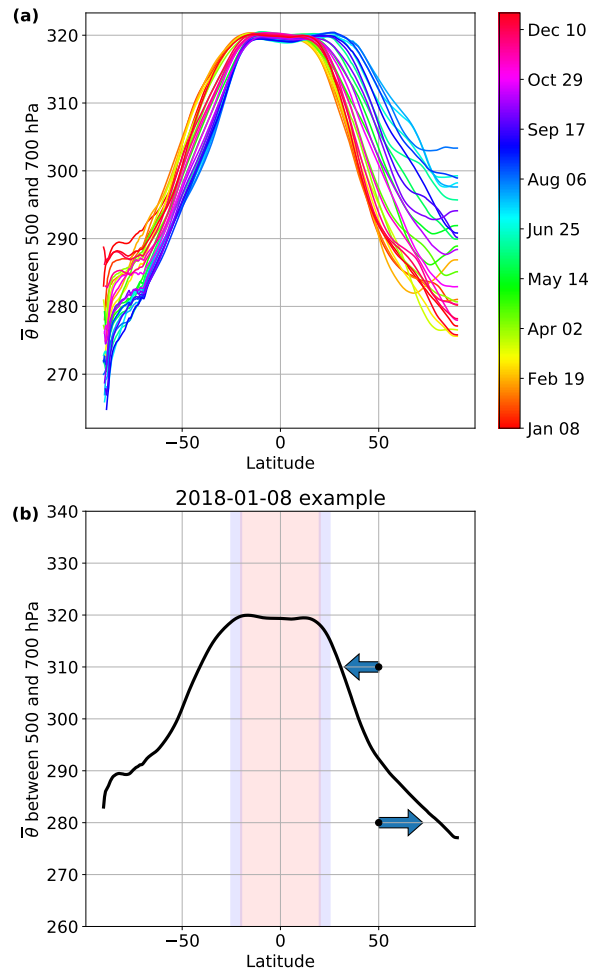
Figure 4. (a) the lookup table for θ_{mid} vs. latitude and time-of-year. **(b)** a hypothetical example of how the effective latitude calculation works. The black line represents the climatological θ_{mid} and the black points represent hypothetical θ_{mid} for individual profiles'. The arrows indicate how the effective latitude of each profile is adjusted such that the individual θ_{mid} matches the climatological θ_{mid}. The red shading indicates latitudes where this method is not applied; the blue shading indicates transitional areas between the geographic and effective latitude. (See text for additional details.)

2.2.2 Altitude grid adjustment

The seasonal cycle and distance basis function assume that the surface is at 0 km altitude. To this end, we use an adjusted altitude as z in Eqs. (1) through (5d). To compute this adjusted z, we stretch or squeeze the bottom of the altitude grid so that the bottom layer is at the surface altitude from the GEOS-5 FP-IT 2D files. The adjustment follows:

$$
z_{\mathrm{adj}} = \begin{cases} z_{\mathrm{orig}} & \text{if} & z_{\mathrm{orig}} \geq z_{\mathrm{blend}} \\ z_{\mathrm{orig}} + dz \cdot f^2 & \text{if} & z_{\mathrm{min}} \leq z_{\mathrm{orig}} \leq z_{\mathrm{blend}} \\ 0 & \text{if} & z_{\mathrm{orig}} < z_{\mathrm{min}} \end{cases} \tag{6}
$$

where z_{orig} is the original altitude, $dz = z_{\mathrm{surf}} - z_{\mathrm{min}}$, z_{min} is the original grid altitude closest to z_{surf}, z_{blend} is the original grid altitude closest to $z_{\mathrm{surf}} + \frac{1}{2} \cdot (z_{\mathrm{trop}} - z_{\mathrm{surf}})$, z_{surf} is the GEOS-5 FP-IT surface altitude, z_{trop} is the tropopause altitude, and f is:

$$
f = \frac{i_{\mathrm{blend}} - i}{i_{\mathrm{blend}} - i_{\mathrm{min}}} \tag{7}
$$

where i_{blend}, i_{min}, and i are the indices for z_{blend}, z_{min}, and z, respectively. Figure S7 shows an example of the adjustment. This adjustment is minor (typically 50 to 100 m) since the priors are generated on the terrain following levels from the GEOS FP-IT model.

2.3 Stratospheric prior

The design of the stratospheric priors draws heavily from Andrews et al. (2001a). That work showed that the profiles of CO_2 and N_2O in the lower stratosphere can be well-captured using surface in situ data from the MLO & SMO observatories to determine the trace gas mole fraction entering the stratosphere and then accounting for mixing of air during stratospheric circulation. We extend this method by using atmospheric profile measurements between February 2004 and March 2019 from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS, Bernath et al., 2005), data version 3.6 (Boone et al., 2013), to capture chemical production and/or loss of N_2O and CH_4 and production of HF.

2.3.1 Stratospheric age of air

The age of stratospheric air parcels is calculated from a climatology simulated by the Chemical Lagrangian Model of the Stratosphere (CLaMS) and scaled to match the mean midlatitude age in the Goddard Space Flight Center 2D (GSFC2D) model (Fleming et al., 2011), which provides age of air as a function of latitude, potential temperature, and day-of-year. Age of air in this context refers to the time since the air entered the stratosphere. Figure 5 shows both latitudinal and temporal slices of the CLaMS age of air. The CLaMS model is a 2-D representation of the mean dynamics of the statosphere. To account for the zonal displacements driven by large-scale Rossby waves, we compute an equivalent latitude profile. Equivalent latitude is derived from ~~potential vorticity (PV)~~ PV following Eq. (1) in Allen and Nakamura (2003).
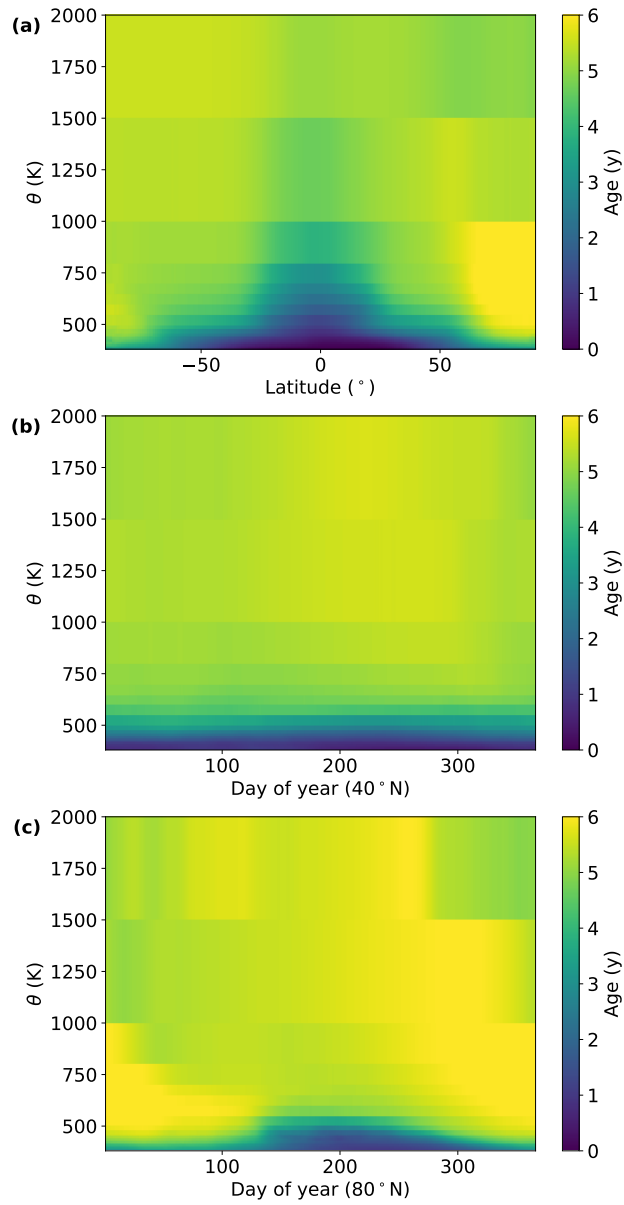
Figure 5. Mean age of air from the CLaMS climatology. **(a)** Age vs. latitude and potential temperature for Jan 1, **(b)** Age vs. day of year and potential temperature at $40°$ N, **(c)** As (b), but for $80°$ N.

Note that this equivalent latitude is not the same as the effective latitude used in the tropospheric part of the prior calculation. PV-derived equivalent latitude has been previously shown to predict stratospheric chemical fields well (e.g. Allen and Nakamura, 2003) while a coordinate derived from mid-tropospheric potential temperature predicts synoptic variation in tropospheric trace gas mixing ratios (e.g Keppel-Aleks et al., 2012). Therefore, we use the PV-derived equivalent latitude here for the stratospheric part of the priors and potential temperature-derived effective latitude in Sect. 2.2.1 for the tropospheric part of the priors.

2.3.2 Age spectra and chemistry

Once the age of air is known, we can look backwards in the combined MLO & SMO record to determine the stratosphere boundary condition (SBC), that is, the mole fraction of each gas when a parcel of air entered the stratosphere. The SBC time series is defined as the MLO & SMO average lagged by two months; Andrews et al. (2001a) and references therein show that this is a good proxy for the SBC. However, the mole fraction for a given level in the prior is not simply the mole fraction of, e.g. CO_2 when that air entered the stratosphere, but is the result of mixing of air with different ages during convective transport. This mixing can be represented by solutions to Green's function derived from CO_2 measurements (Andrews et al., 2001a), which we represent as age spectra.

Age spectra were precomputed for three regions (tropics, midlatitudes, and polar vortex) and ~ 45 different mean ages. Andrews et al. (1999) and Andrews et al. (2001b) showed that different age spectra were necessary to capture tropical and midlatitudinal behavior; likewise, the polar vortex requires its own age spectra form due to strong wintertime descent of air. Example age spectra are shown in Fig. 6. Note that spectra for the youngest mean ages are not shown.

For each stratospheric level in the priors, the mole fraction of a gas is computed as

$$\bar{c} = \overline{F}(a,\theta) \int S_{a,r}(t)c(t)\, dt \tag{8}$$

where $S_{a,r}(t)$ is the value of the age spectrum for the given mean age (a) and region (r) and $c(t)$ is the SBC, both at time t. That is, the mole fraction is a weighted average of the SBC over time with the weights set by the age spectrum. $\overline{F}(a,\theta)$ is the fraction of gas remaining after chemical loss, θ is potential temperature, which we use as a vertical coordinate. For CO_2 this fraction is always 1, but varies with mean age and potential temperature for other gases, as discussed in more detail in Sect. 3.

2.3.3 Middleworld treatment

The middleworld is defined as the part of the atmosphere between the tropopause pressure from GEOS-5 FP-IT and the 380 K isentrope. Of the three tropopause pressure estimates in GEOS-5 FP-IT, we use the blended (thermal and potential vorticity) estimate. The 380 K isentrope is the lowest potential temperature surface entirely contained within the stratosphere; therefore the stratospheric approach described in Sect. 2.3 is only applicable to levels above 380 K (the stratospheric overworld). To fill in the prior in the middleworld, we linearly interpolate mole fraction as a function of potential temperature between the tropopause and 380 K.
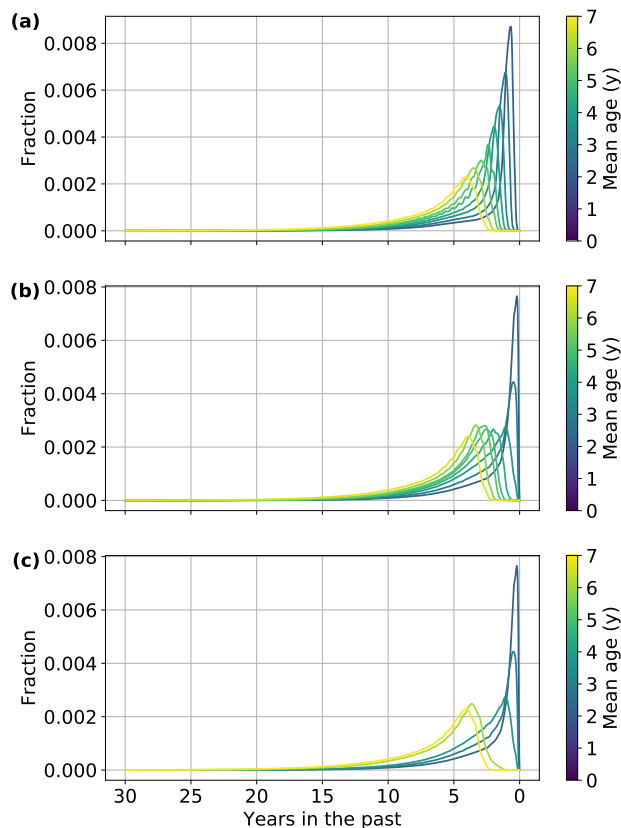
Figure 6. Example age spectra for **(a)** tropics, **(b)**, midlatitudes, and **(c)** polar vortex. The y-values represent the contribution of air from that time to the average mole fraction of the parcel as a whole. Note that age spectra for the youngest air are not shown because they are nearly delta functions.

2.4 Secondary gases

For the purpose of this paper, "secondary gases" are defined as those which are neither tied directly to MLO & SMO records nor the GEOS-5 FP-IT product. This is all gases other than CO_2, N_2O, CH_4, HF, CO, H_2O, HDO, and O_3. O_2 and HCl are the two most relevant to standard TCCON retrievals. Priors for these gases are based on climatological profiles for summer at 35° N derived from profiles measured by MkIV spectrometer balloon flights (Toon, 1991) and the ACE-FTS instrument. These climatological profiles are modified for a given location and time in four steps:

1. stretch or compress the profile vertically so that the tropopause is at the correct altitude,

2. apply a latitudinal gradient,

3. apply a secular trend,

    4. apply a seasonal cycle.

These steps require the latitude and age of air of the profiles. This approach is nearly identical to that used for all gases in the GGG2014 priors, except that for steps 2–4, the age of air and effective latitude described in Sect. 2.2 are used in the troposphere and the CLaMS age and PV-derived equivalent latitude from Sect. 2.3 are used in the stratosphere. The middleworld is filled in by linear interpolation in θ between the tropopause and 380 K, as is done for the primary gases. Details of the calculation are given in the supplement.

2.5   Conversion to number density

All trace gas quantities shown and discussed in this paper are in dry mole fractions (DMFs, i.e. moles of trace gas per moles of dry air). However, in its forward model, GGG uses gas profiles in number density (molec. cm^{-3}) for spectroscopic calculations. To convert DMF to number density, we use:

$$n_{gas} = \frac{c_{gas}}{1 + c_{H_2O}} n_{ideal} \tag{9}$$

where n_{gas} is the number density of the gas of interest, c_{gas} is the DMF of that gas, c_{H_2O} is the DMF of water (from the H_2O prior profile), and n_{ideal} the ideal gas number density. The factor $1 + c_{H_2O}$ converts n_{ideal} into number density of dry air.

3   Gas-specific design

In this section, we will discuss elements of the algorithm unique to each gas. With the exception of O_2, each section will be divided into subsections for the tropospheric and stratospheric priors.

3.1 O_2

We assume a uniform DMF of 0.2095 for O_2 at all altitudes. During the retrieval, this is converted to number density following Sect. 2.5. In the GGG2014 priors, the conversion to number density did not include a correction for water. This led to a profile shape error: as water DMFs are highest near the surface, failing to include the water correction led to an overestimate of the near-surface number density for every absorbing gas.

The impact of this error in the previous priors on the final column amounts was small because, in public TCCON data, all gas column amounts are reported as column average mole fractions (termed Xgas, e.g. XCO_2). These are calculated as:

$$Xgas = \frac{V_{gas}}{V_{O_2}/0.2095} \tag{10}$$

where V_{gas} and V_{O_2} are the total column amounts (in molec. cm^{-2}) of the target gas and O_2, respectively. The denominator represents a column of dry air inferred from the retrieved O_2 column. The advantage of this method over using a column of air

derived from surface pressure is that, because primary TCCON target gases are measured on the same detector as O_2, certain types of instrumental error cancel out in this ratio, reducing their impact on the final data product (Washenfelder et al., 2003; Wunch et al., 2011). Likewise, the shape error due to the missing water correction in GGG2014 priors largely canceled out in the column-averaged Xgas DMFs. However, the GGG2020 treatment, following Eq. (9), is more physically consistent, leads to more consistent O_2 scaling factors retrieved among TCCON stations, and yields a better shape—especially under warm, humid conditions.

3.2 CO$_2$

Troposphere: The value of α in Eq. (1) for CO_2 was derived by comparing the priors generated with $\alpha = 1$ and $\beta = 0$ against profiles from the HIPPO (Wofsy, 2011) and ATom (Wofsy et al., 2018; Thompson et al., 2022) campaigns. We used CO_2 measurements from the Harvard quantum cascade laser spectrometer (QCLS) for HIPPO and CO_2 measurements from the NOAA Picarro for ATom. Only data from invididual vertical profiles (identified as data points where the `PFP/prof.no` variable is > 0 in the merge files from https://doi.org/10.3334/ORNLDAAC/1581) with ≥ 10 valid data points were used. The differences between the priors and observations below 800 hPa were averaged over $20°$ latitude bins and converted from units of ppm to multiples of the interannual CO_2 growth rate, derived from the MLO & SMO average deseasonalized trend. The output of the distance function d (Eq. 2) was also averaged for all prior levels below 800 hPa and binned to $20°$ latitude bins.

The result is shown in Fig. 7a. The red line is a York fit (York et al., 2004) to the data using the inverse square of the standard deviations of the prior-observation differences and distance function values in the latitude bins as the weights. This fit indicates setting α equal to -3.55 times the CO_2 interannual growth rate will give a latitudinal gradient that matches observations. Figure 7b shows the mean differences vs. latitude with α set to 1 (i.e. no adjustment) and with the best fit to the data. Using the α derived from Fig. 7a and $\beta = 0$, the priors show no latitudinal bias versus observations.

Stratosphere: CO_2 follows the algorithm laid out in Sect. 2.3. No additional modifications were required. For our purposes, we assume that CO_2 DMFs are unaffected by stratospheric chemistry (e.g. CH_4 oxidation) and do not include a correction for chemistry in stratospheric CO_2.

3.3 N$_2$O

Troposphere: We set α in Eq. (1) to

$$\exp\left(\frac{-d}{\tau}\right) \tag{11}$$

where d is the output of the distance function from Eq. (2) and $\tau = 121$ yr (the mean atmospheric lifetime of N_2O, following Myhre et al., 2013, Table 8.A.1). This imposes a slight additional north-south gradient to N_2O in the troposphere.

Stratosphere: In the stratosphere, N_2O is more complicated than CO_2 because it is removed, principally through photolysis forming nitrogen N_2 and an oxygen atom O, but also via a reaction with excited oxygen ($O(^1D)$) (Jacob, 1999). Andrews et al. (2001a) fit this loss of N_2O in the lower stratosphere versus age of air with a third-order polynomial. We examined how this
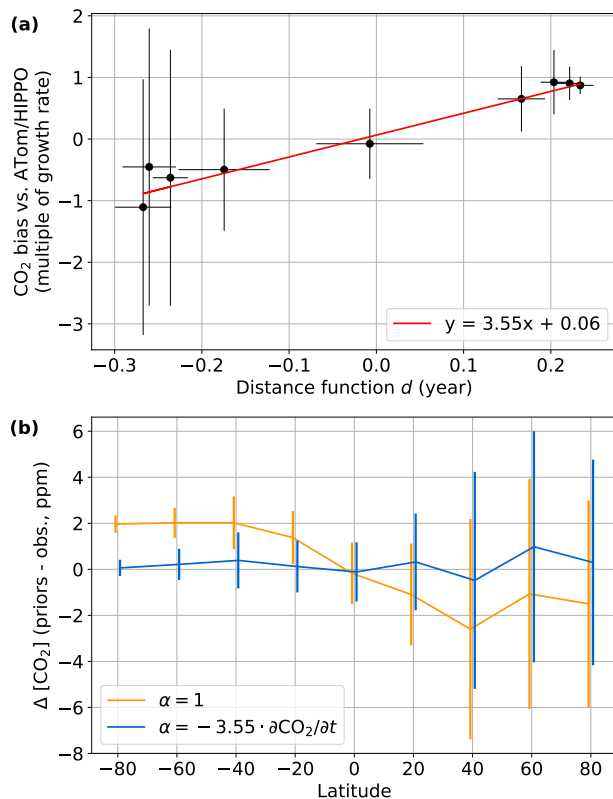
Figure 7. (a) Bias between the inital CO_2 DMFs and HIPPO/ATom profile vs. the distance function (Eq. 2) for profile levels below 800 hPa. Note that the y-axis is not in ppm, but in multiples of the interannual CO_2 growth rate. See text for details. **(b)** The mean difference between priors and observations in $20°$ latitude bins below 800 hPa, vs. latitude bin center. In both panels, error bars are 1σ standard deviations of the respective variable within the $20°$ latitude bins.

polynomial compares to N_2O data from the ACE-FTS instrument (Bernath et al., 2005) and found that the polynomial's skill in predicting the fraction of N_2O remaining relative to the SBC (F(N_2O)) decreased above approximately 25 km altitude, with the polynomial overestimating the N_2O mixing ratio by up to 150 ppb. We hypothesize this is due to different chemistry in

335   the upper stratosphere compared to the lower stratosphere. As the original polynomial was based on lower stratospheric data, it did not capture this behavior. While the fraction of the N_2O column in the upper stratosphere is small (a few percent above 20 km), our goal was to develop priors with reasonably accurate DMFs at all altitudes, not just where the bulk of the column mass is. Additionally, developing our own method to estimate F(N_2O) allows us to be consistent when calculating the same quantity for CH_4 and HF.

340   We use N_2O data from the ACE-FTS instrument to build a lookup table of the fraction of N_2O remaining as a function of age of air and potential temperature. Strong et al. (2008) validated a previous version of the ACE-FTS N_2O data and found that mean differences between ACE-FTS and other stratospheric N_2O measurements were ±10 ppbv between 18 and 30 km,

and mostly within -2 to $+1$ ppbv between 30 and 60 km. They note that these are large relative to the magnitude of N_2O mole fractions at these altitudes; however, for our purposes, these are acceptable, given that we are averaging a large number of ACE-FTS profiles and need only a climatological relationship between fraction of N_2O remaining, age of air, and potential temperature. Waymark et al. (2014) compared the version 3 ACE-FTS data (used in this work) to the version 2 evaluated by Strong et al. (2008) and note that the main difference is a 10% reduction in N_2O above 30 km. Thus the general results in Strong et al. (2008) should still hold. For ACE-FTS v3.5 data (one minor version earlier than that used in this work), Sheese et al. (2017) found biases between ACE-FTS and MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) of between -9% to 5% and between ACE-FTS and MLS (Microwave Limb Sounder) of between -18% and 4% in the altitude range of 19 to 34 km.

To build the lookup table, age of air is computed as in Sect. 2.3; for each ACE profile, the stratospheric equivalent latitude is computed for the GEOS-5 FP-IT files that bound it in time, then it is interpolated to the latitude, longitude, and time of the profile. This equivalent latitude and the potential temperature calculated from ACE-FTS temperature and pressure is used as input to the CLaMS model from Sect. 2.3 to look up the age of air.

$F(N_2O)$ is defined relative to the stratospheric boundary condition in the ACE-FTS data, not the MLO & SMO record, to ensure self-consistency and avoid introducing error from the bias between the ACE-FTS and MLO & SMO data (Fig. S9). The stratospheric boundary condition is computed from a quadratic fit in time of ACE-FTS N_2O data in the tropics (latitude within $\pm 20°$) and with $360\,\mathrm{K} < \theta < 390\,\mathrm{K}$, excluding outliers (defined as values more than five times the median deviation from the median). This definition of the stratospheric boundary condition assumes that most of the air entering the stratosphere does so in the tropics and that the tropical tropopause is in that range of potential temperature values.

Finally, to compute the $F(N_2O)$ lookup table, the ACE-FTS data are binned by age of air (0.25 year increments) and potential temperature (variable increments; 50 K in the lower stratosphere to 200 K in the upper stratosphere). ACE-FTS data are excluded if:

- $F(N_2O) < 0$,

- altitude ≥ 70.0 km (this is the top altitude in the TCCON priors),

- the profile is in the polar vortex,

- potential temperature is < 380 K (as we are only concerned with levels in the stratospheric overworld).

Additionally, $F(N_2O)$ values > 1 are limited to 1. The resulting lookup table is shown in Fig. 8. As there are large gaps in age-θ space with no ACE-FTS data, we extrapolate to fill in these gaps. We use essentially a constant-value extrapolation along age, that is, if there is no value for a given age-θ bin, the nearest point at the same θ is used. Linear extrapolation along age is done second, using the nearest two points to determine the slope. In general, points in these extrapolated regions are expected to be very infrequent, as the absence of ACE data suggests that those combinations of age and θ are rare in the atmosphere.

The need to capture how $F(N_2O)$ depends on both age and θ is apparent in Fig. 8. Consider the points in Fig. 8 at age = 5 years. Over the range of 1000 K, the $F(N_2O)$ decreases from ~ 0.5 to almost 0. This is likely because at greater θ (i.e. higher
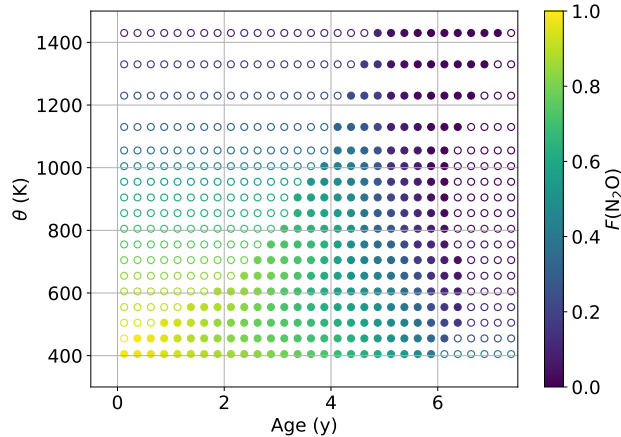
Figure 8. $F(N_2O)$ lookup table derived from ACE-FTS v3.6 data as a function of potential temperature and age of air. Unfilled circles are extrapolated points.

altitude) the N_2O photolysis ($N_2O + h\nu \rightarrow N_2 + O$) pathway proceeds more rapidly than at lower altitudes. Age of air alone cannot capture this difference.

3.4 CH$_4$

Troposphere: Similar to N_2O, the CH_4 priors use Eq. (11) as α, with a lifetime of 12.4 yr (Myhre et al., 2013, Table 8.A.1).
380 The orange line in Fig. 9 shows the mean prior vs. observation differences below 800 hPa in 20° latitude bins, as in Fig. 7b. A latitudinal bias in tropospheric methane mole fractions in the northern hemisphere remains. Therefore we set β to 0.75 ppb/° in the northern hemisphere, which removes this bias (blue line, Fig. 9).

Stratosphere: CH_4 must also include a fraction remaining term, $F(CH_4)$, to account for stratospheric chemistry, similarly to N_2O. Figure 10a shows a tight correlation between ACE-FTS N_2O and CH_4 in the stratosphere; therefore, we can use the
385 relationship between $F(N_2O)$ and age derived in Sect. 3.3 as a basis for the $F(CH_4)$ lookup table.

To compute the lookup table, we first limit the ACE-FTS data to points where $F(N_2O)$ and $F(CH_4)$ are positive, the CH_4 mole fraction is < 2000 ppb (points ≥ 2000 ppb are almost certainly tropospheric), the profile is outside the polar vortex, and the altitude is below 70 km. We bin the data by $F(N_2O)$ and θ. Within each $F(N_2O)$ bin, outliers are rejected (distance $\geq 5\times$ median absolute deviation) and the mean $F(CH_4)$ value in each $F(N_2O)$ and θ bin pair is computed. As with N_2O, we use
390 extrapolation to fill in parts of the lookup table not covered by ACE-FTS data. We use constant value extrapolation along the θ dimension first, then also along the $F(N_2O)$ dimension if necessary.

To compute the stratospheric prior profiles, Eq. (8) is used with the $F(CH_4)$ value described above. To compute the $F(CH_4)$ value, the age and θ values are first used to compute the $F(N_2O)$ value as described in Sect. 3.3, and then the $F(CH_4)$ value is determined by linearly interpolating the lookup table in Fig. 10b to the required $F(N_2O)$ and θ.
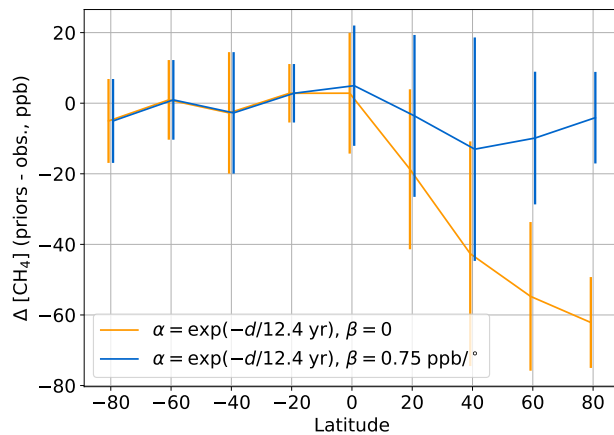
Figure 9. Differences in CH_4 between ATom and HIPPO observations and priors, binned as in Fig. 7b, with and without $\beta = 0.75\,\mathrm{ppb}/°$ correction in the northern hemisphere. Error bars are 1σ standard deviations in the $20°$ latitude bins.
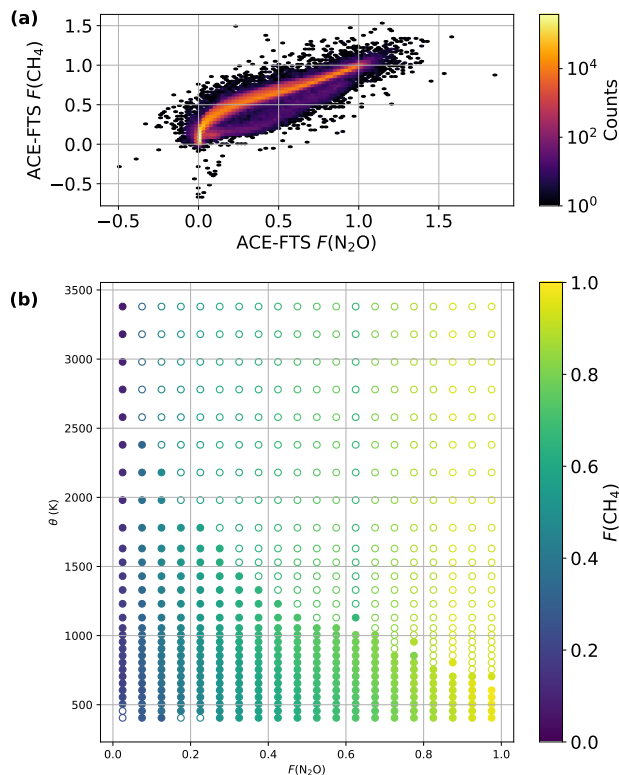
Figure 10. (a) 2D histogram of $F(N_2O)$ and $F(CH_4)$ from ACE-FTS. **(b)** The $F(CH_4)$ lookup table used in the GGG2020 algorithm. As in Fig. 8, filled circles are derived directly from data while the unfilled circles are extrapolated. F values are computed as described in Sect. 3.3.

3.5 HF

Measurements of HF DMFs in the troposphere are very rare; the most recent direct measurement of gaseous fluoride that we found in the literature was Okita et al. (1974), which reported measurements around an aluminum refinery. Their measurements near but not downwind of the refinery reported fluoride concentrations of $< 1 \, \mu g \, m^{-3}$, or a DMF of order 10 to 100 parts per trillion (ppt). Spectroscopic measurements over Antarctica (Toon et al., 1989) and Switzerland (Zander et al., 1987) found upper tropospheric HF DMFs of 1 to 10 ppt were consistent with solar-viewing spectra.

For our purposes, we assume that the tropospheric DMF of HF is negligible compared to the stratospheric component, and so imposed a small but non-zero DMF of 0.1 ppt. This is less than the previous measurements (Okita et al., 1974; Zander et al., 1987; Toon et al., 1989), but the impact on HF retrievals should be small given that TCCON HF averaging kernels are usually < 0.5 below 200 hPa.

In the stratosphere, we once again make use of tracer-tracer relationships. HF is produced by reaction of fluorine atoms from photolysis of COF_2 and $COFCl$ (which are the products of destruction of CFC-11, CFC-12, and HFC-22) with CH_4, H_2, or H_2O (Washenfelder et al., 2003). Thus, CH_4 and HF mole fractions are tightly anticorrelated in the stratosphere. Previous studies (e.g. Saad et al., 2014) have used this relationship to separate tropospheric and stratospheric CH_4 columns; here, we do the reverse, using CH_4 prior profiles to determine HF prior profiles.

We follow a similar approach to Saad et al. (2014); we determine the CH_4:HF slope (m) and directly compute the HF mole fraction from the CH_4 mole fraction as:

$$[HF] = \frac{[CH_4] - [CH_4]_{sbc}}{m} \tag{12}$$

where $[CH_4]_{sbc}$ is the CH_4 stratospheric boundary condition determined from the MLO & SMO record, as described in Sect. 2.3.2.

Because of the time dependence in the ratio of methane to the long-lived fluorine containing gases in the troposphere and because of the non-uniform ratio of the lifetime of CH_4 and the CFCs in the stratosphere, the slope m depends on both time and latitude (Washenfelder et al., 2003; Saad et al., 2014). Before the beginning of the ACE-FTS data set in 2004, we use CH_4:HF slopes reported in Washenfelder et al. (2003). From 2004 on, we bin ACE-FTS CH_4 and HF data into the same three latitude bins (tropics, midlatitudes, and polar vortex) as for the age spectra (Sect. 2.3.2). We filter for $[CH_4] \leq 2000$ ppb and $[HF] \leq 10$ ppb and limit to altitudes < 70 km. The limit on CH_4 is imposed for the same reason as in Sect. 3.4; the limit on ACE-FTS HF is imposed due to erroneously large values of ~ 200 ppb found in rare cases (despite only using data with CH_4 and HF quality flags ≤ 1). A 10 ppb upper limit was determined to only exclude these extraordinary values. The CH_4:HF slopes were fit as in Saad et al. (2014) using a robust fit with Tukey's biweighting function.

Finally we combine the ACE-FTS-derived slopes with those from Washenfelder et al. (2003) and fit the change over time with an exponential. This allows us to extrapolate forward or backward in time as needed. Each latitude bin has its own exponential fit that fits the bin-specific ACE-FTS slopes and the Washenfelder et al. (2003) slopes. (All bins used the same Washenfelder et al. (2003) data.) For consistency, we always take the slope from the exponential fit.
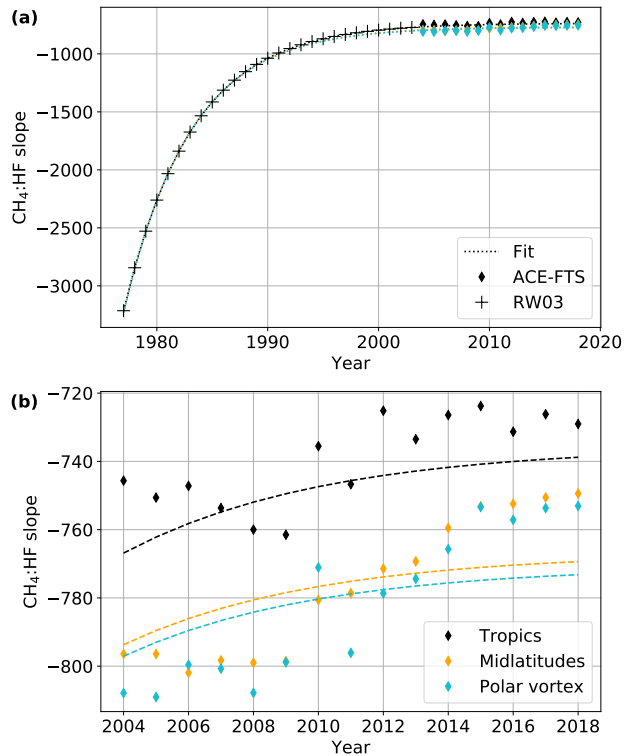
Figure 11. (a) CH_4:HF slopes and the exponential fits over the entire time period with data, be it from Washenfelder et al. (2003) (RW03) or ACE-FTS. **(b)** Similar to (a), but zoomed in on the ACE-FTS time period and colored by latitude bin.

Therefore, for each overworld level ($\theta \geq 380$ K), a CH_4 mole fraction is calculated (following Sect. 3.4) and the CH_4:HF slope for the year and latitude bin (based on equivalent latitude, Sect. 2.3) is used in Eq. (12) to compute the HF mole fraction. 430 Note that we use the slope for the year of the observation, and not the year the air entered the stratosphere, because the slopes are based on observations for specific years.

3.6 CO

Troposphere: With a shorter tropospheric lifetime (of order months) than the above gases, CO requires a custom treatment in order to adequately account for its spatial variability. The GEOS-5 FP-IT product contains a CO forecast that shows reasonable 435 skill in comparison to QCLS CO measurements taken during the ATom campaigns (Wofsy et al., 2018). We therefore adopt the GEOS-5 FP-IT CO product as the base profile for the CO priors with the following modifications.

First, our comparison against the first three ATom campaigns shows a low bias in the GEOS-5 FP-IT CO mole fractions, as seen in Fig. 12a. While there is some variation with latitude, the pattern was not sufficiently clear to lend itself to a robust correction, therefore, we multiply the troposphere CO mole fractions by 1.23 ($= 1/0.81$) to bring them in line with ATom 440 observations.
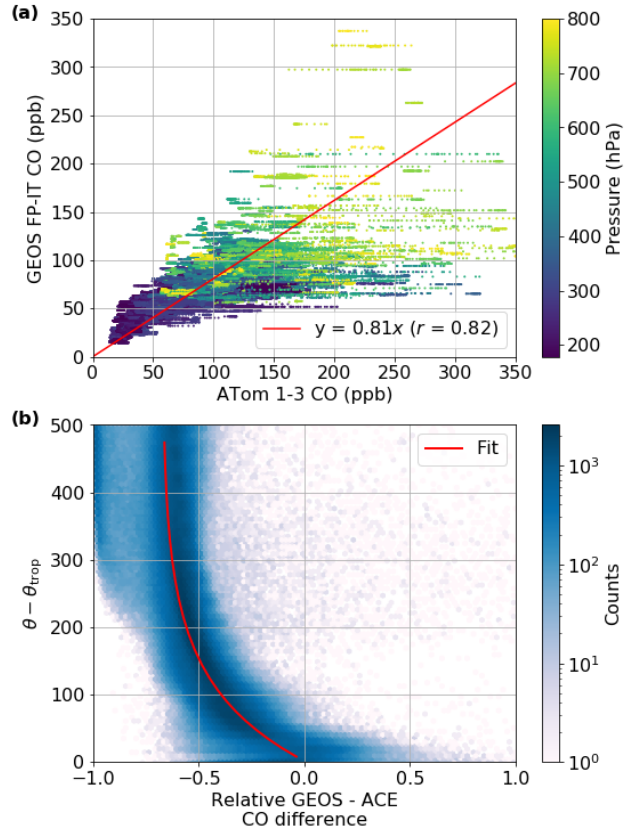
Figure 12. (a) Comparison of colocated ATom measured and GEOS-5 FP-IT forecasted CO mole fractions. GEOS-5 FP-IT CO matched to ATom observations using 4D nearest neighbor interpolation. The fit is a robust fit using a Tukey biweight function with no ~~inteceptintercept,~~ i.e. using the `RLM` linear model with `M = TukeyBiweight()` from the Python `statsmodels` package (Seabold and Perktold, 2010). Only points with pressure < 800 hPa used. **(b)** Comparison of colocated ACE-FTS and GEOS-5 FP-IT CO data. The x-axis is the unitless relative difference, (GEOS − ACE)/ACE. The y-axis is potential temperature relative to the tropopause. The background shading is a 2D histogram of the relative bias between ACE-FTS and GEOS-5 FP-IT CO as a function of θ; the red line is a fit through the mean bias.
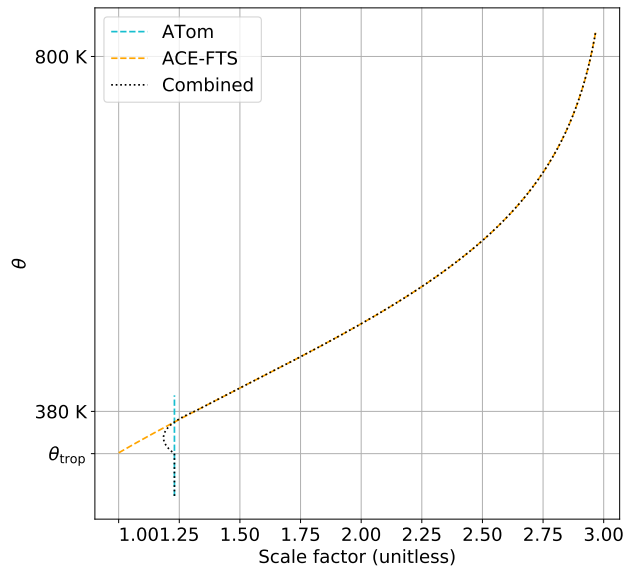
Figure 13. The form of the CO bias correction scaling factor. The blue and red lines show the form derived from ATom and ACE-FTS data, respectively while the black line shows the blending of these two corrections. Note that the ATom line is extended up to 380 K for reference, it does not imply that ATom collected data into the mid-stratosphere.

Stratosphere: Comparison with ACE-FTS data in the lower stratosphere also demonstrates a low bias, which varies with altitude. However, the general structure is consistent as a function of potential temperature relative to the tropopause, as seen in Fig. 12b. This can be represented by an exponential function.

Therefore, the overall CO correction has the form shown in Fig. 13. Below the tropopause, the 1.23 factor derived from ATom is used, while above 380 K (i.e. the stratospheric overworld) the exponential form derived from ACE-FTS is used. In the middleworld, we linearly blend between the two functions in order to provide a smooth transition.

The second correction required concerns the intrusion of mesospheric CO into the stratosphere. In the mesosphere, very large mixing ratios of CO are produced through photolysis of CO_2. As this descends (especially in the polar vortex), it can lead to very large CO mole fractions at altitudes as low as 40 km. This process is not captured in the GEOS-5 FP-IT product, but is represented in the Canadian Middle Atmosphere Model (CMAM), which compares well with ACE-FTS and MLS data (Jin et al., 2009; Kolonjari et al., 2018). We use here output from a version of CMAM run with dynamics specified (see Sect. 2.2 of Kolonjari et al., 2018, and references therin).

Comparison of GEOS-5 FP-IT with ACE-FTS data shows the mesospheric CO impact beginning around 30 hPa and becoming dominant by 10 hPa. Therefore, we replace the GEOS-5 FP-IT CO with CMAM CO above 10 hPa (i.e. at pressure < 10 hPa) and linearly interpolate from GEOS-5 FP-IT to CMAM in pressure-log space between 30 and 10 hPa. The CMAM CO is drawn from a monthly climatology constructed from the monthly-averaged CO DMFs in the 30-year CMAM model run (available at http://climate-modelling.canada.ca/climatemodeldata/cmam/output/CMAM/CMAM30-SD/mon/atmosChem/

vmrco/index.shtml, last accessed 24 Jul 2019). CMAM model data before 2000 is not used in the climatology because there is not a trend present after 2000.

460   The third and final correction accounts for the mesospheric CO itself. While the priors used in TCCON retrievals have a 70 km ceiling, the CO above that altitude in the CMAM model can comprise up to $\sim 2.5\%$ of the total column, particularly in the polar regions. To account for this in the prior, we add an equivalent mass of CO to the top level of the priors. This is detailed in Sect. S4 of the supplement.

3.7 H_2O and HDO

465   The H_2O profile is computed directly from the GEOS-5 FP-IT specific humidity. The HDO profile is directly computed from the H_2O profile as:

$$c_{HDO} = c_{H_2O} \cdot 0.14 \cdot [8 + \log_{10}(c_{H_2O})] \tag{13}$$

where c_{H_2O} and c_{HDO} are the DMFs of H_2O and HDO, respectively. In the GGG retrieval, the line intensities of isotopologs are multiplied by the isotope abundance. This form therefore does not need to reproduce the abundance of HDO, but instead

470 just the decrease of HDO relative to H_2O with altitude due to Rayleigh fractionation (Kuang et al., 2003). While reading the priors, GGG takes the absolute value of the HDO DMF to eliminate negative DMFs resulting from $H_2O < 10^{-8}$. In versions of ginput after 1.1.4, the absolute value of the HDO DMF is output.

4 Use as OCO-2/3 priors

The Orbiting Carbon Observatory 2 (OCO-2) and OCO-3 retrievals use these CO_2 priors starting in their respective version

475 10 products. The version 10 products use this algorithm exactly as described above except for one small change: in Eq. (1), l is geographic, rather than effective, latitude. This difference ensures a smooth latitudinal variation in CO_2. Using effective latitude introduced discontinuities near the equator (Fig. S17a).

    The specific structure of the discontinuities in Fig. S17a arise because version 10 of the OCO-2/3 algorithm uses an earlier version of the priors algorithm than GGG2020; in this earlier version, rather than transition between geographic latitude and

480 effective latitude between $20°$ and $25°$, effective latitude was used for profiles at all latitudes but disallowed from crossing the equator. (That is, a profile in the northern hemisphere could not have an effective latitude in the southern hemisphere and vice versa.)

    Switching the version 10 priors to use geographic latitude for all soundings trades some ability to capture day-to-day variation in the troposphere for guaranteed spatially smooth priors (Fig. S17b), which is well worth it for nadir viewing instruments

485 such as OCO-2 and OCO-3. In contrast, for discrete measurement sites such as TCCON, the ability to capture day-to-day variations is preferred.

The OCO-2/3 version 11 priors introduced an additional change to allow more frequent updating of the input in situ data. GGG2020 and OCO-2/3 version 10 use a static file of MLO & SMO data as input that contains monthly averages of flask data prepared by NOAA (Dlugokencky et al., 2019) up through the end of 2018. These records are extended by extrapolation (see Sect. 2) as needed. This has the virtue of simplicity, but cannot capture anomalies in the trend of CO_2 such as those cause by El Niños.

The OCO-2/3 version 11 algorithm switched to using hourly in situ data from the continuous trace gas analyzers stationed at MLO & SMO NOAA observatories (Thoning et al., 2021) that has undergone preliminary quality control, but not full background selection by NOAA personnel. These hourly in situ data are preprocessed by the priors code to produce monthly averages, allowing the main algorithm to use either monthly flask or hourly in situ data as needed. The preprocessing algorithm is described in Sect. S5 of the supplement.

5 Validation

5.1 Comparison with aircraft and AirCore observations

To directly validate the GGG2020 priors, we use aircraft data from the NOAA CO_2 GLOBALVIEWplus v5.0 Obspack (Co-operative Global Atmospheric Data Integration Project, 2019; Masarie et al., 2014), NOAA CH_4 GLOBALVIEWplus v2.0 ObsPack (Cooperative Global Atmospheric Data Integration Project, 2020; Masarie et al., 2014), and the Infrastructure for Measurement of the European Carbon Cycle (IMECC) campaign (Geibel et al., 2012), as well as AirCore (Tans, 2009; Karion et al., 2010) profiles from NOAA routine and campaign balloon flights (v20201223, Baier et al., 2021) and selected Air-Core balloon flights from FMI/~~LSCE/~~RUG at the Sodankylä, Finland (Kivi and Heikkinen, 2016) ~~and~~ TCCON site and from CARE- C/LSCE/LMD/IPSL at the Nicosia, Cyprus ~~(Messerschmidt et al., 2012) TCCON sites~~TCCON site. Data from tower measurements at Park Falls, WI, USA (Andrews et al., 2014; Desai et al., 2015), the Southern Great Plains Atmospheric Radiation Measurement facility near Lamont, OK, USA, and at the National Intitute of Water & Atmospheric Research Ltd. site in Lauder, New Zealand were used to extend airborne profiles in these locations to the surface as needed. The data used, and which gases are provided by each, are tabulated in Tables S1 and S2.

Figure 14 shows the root mean squared error (RMSE) for each vertical level of both the GGG2014 and GGG2020 priors. Mean and individual profile errors are given in Fig. S10. A breakdown of the number of profiles by gas and source is given in Table S4.

For CO_2, the RMSE is noticeably smaller at all altitudes for the GGG2020 priors compared to the GGG2014 priors (Fig. 14a). This results from removing a small but clear negative bias throughout the troposphere arising from an underestimate of the CO_2 secular growth rate in GGG2014. Using the MLO & SMO data eliminates that as a source of uncertainty for profiles before 2019. (2019 is the first year that the MLO & SMO trend is extrapolated for GGG2020 as we chose to use a static file to avoid the complications of updating the input data in a reliable, reproducible manner, as discussed in Sect. 2.) In the stratosphere (above 200 hPa), the improved representation of stratosphere dynamics (Sect. 2.3) better captures the gradient of CO_2 in the lower stratosphere, reducing the previous overestimate of lower stratospheric CO_2 in the GGG2014 priors.
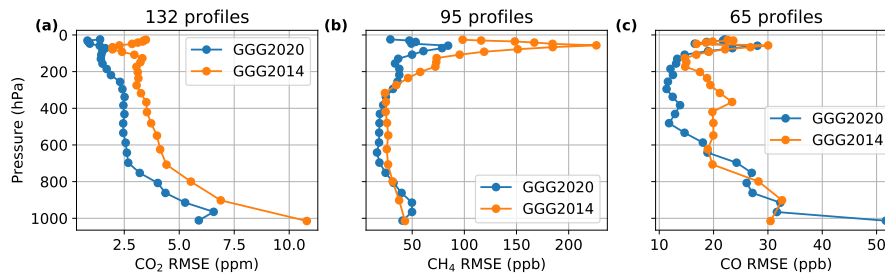
Figure 14. Root mean squared error (RMSE) of **(a)** CO_2, **(b)** CH_4, and **(c)** CO priors versus combined AirCore and aircraft observations. Data sources are listed in S1 and S2. In each panel, both the GGG2020 and GGG2014 priors' RMSE is shown. The number of profiles contributing to each panel is printed above the panel. FMI/RUG Sodankylä AirCore data above 20 km altitude are not included due to anomalously high mixing ratios in CO_2. CO_2 and CH_4 data above 20 km also excluded for consistency.

520 The CO_2 RMSE for the GGG2020 priors is still greater near the surface than at higher altitudes. This may be due to the simplified seasonal cycle (Sect. 2.2). Comparing the priors to ATom and HIPPO observations in different seasons (Fig. S8) shows large differences near the northern hemisphere surface in spring and summer. As the seasonal cycle has latitudinal dependence, revising its parameterization will require adjustment to the distance function (Eq. 1) and the α and β coefficients (Table 3). This area will be revisited in a future version of the GGG priors.

525 CH_4 shows a small improvement in RMSE throughout most of the troposphere (Fig. 14b, 800 to 200 hPa). Above 200 hPa, the RMSE shows a greater improvement, again due to the improved representation of stratospheric dynamics. However, near the surface (below 800 hPa), the RMSE increases somewhat in the GGG2020 priors compared to the GGG2014 priors. This increase in RMSE is driven by near-surface CH_4 emissions not accounted for in the priors. Figure 15a shows differences of the CH_4 priors vs. AirCore data (which has frequent sampling of areas with high emissions), colored by which TCCON

530 site the prior represents. The bias in CH_4 below 800 hPa is clearly due to underestimated CH_4 in the Lamont, OK profiles. The Lamont TCCON site is situated near a region of significant oil and natural gas production (Karion et al., 2015), and thus experiences enhanced CH_4 mole fractions of 100 to 200 ppb near the surface (Fig. S13). Neither the GGG2014 nor GGG2020 priors attempt to account for local anthropogenic emissions. The increase in RMSE near the surface in the GGG2020 priors is due to the removal of a compensating error in assumed vertical gradients—introducing the tropospheric effective latitude

535 (Sect. 2.2.1) accounts for times when Lamont has a profile that varies less with altitude due to the influence of tropical air.

  The GGG2020 CO priors' RMSE improves throughout the free troposphere (600 to 200 hPa). Unlike CO_2 and CH_4, RMSE is similar between GGG2014 and GGG2020 in the stratosphere (above 200 hPa). Near the surface, GGG2020 priors' RMSE is ~ 20 ppb greater than GGG2014. Figure 15b shows that this is driven by overestimated CO at the Armstrong Air Force Base (AFB) TCCON site and both over- and under- estimated CO at the Lamont TCCON site.

540 The cause of the over- and under- estimates in the Lamont profiles is not clear. The GGG2020 CO profiles are based on the CO field in the GEOS-5 FP-IT product (Sect. 3.6). The underestimated CO DMFs could be due to changes in energy economies in the region in recent years (Franklin et al., 2019; Willyard and Schade, 2019). GEOS-5 FP-IT uses 2008 anthropogenic CO
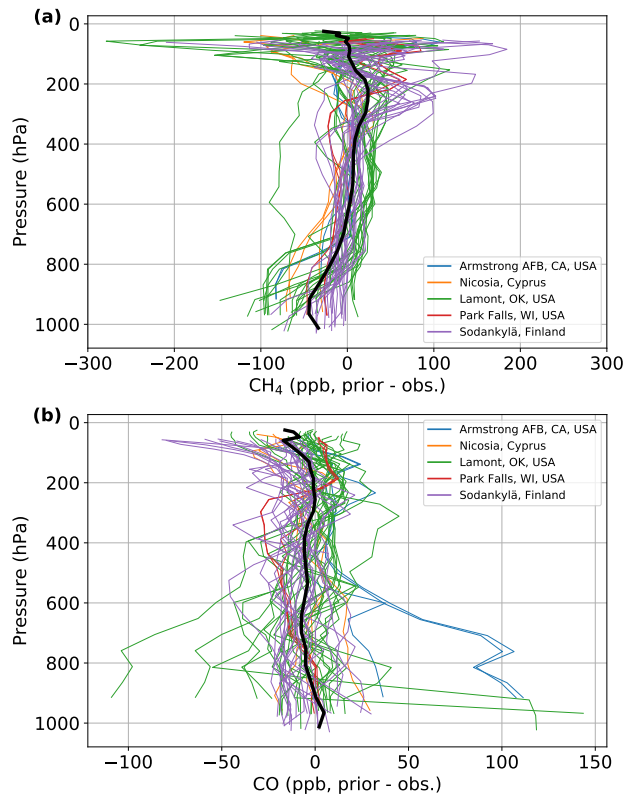
Figure 15. Difference plots for GGG2020 priors vs. **(a)** CH_4 and **(b)** CO AirCore data. The thinner, colored lines represent differences for individual profiles, the thick black line indicates the mean difference across all profiles shown. The individual differences are colored by their TCCON site.

emissions for all years after 2008 (Ott, private communication), so the CO priors would have no information on changes past 2008.

545   The overestimated CO at Armstrong AFB is due to its proximity to Los Angeles. CO emissions in Los Angeles have been decreasing (Brioude et al., 2013), a trend not captured in GEOS-5 FP-IT as 2008 emissions are repeated for all years after 2008. Additionally, given that the GEOS-5 FP-IT model resolution is $0.67° \times 0.5°$ (longitude $\times$ latitude), the complex topography of the Los Angeles Basin, and that Armstrong AFB is only $\sim 0.8°$ north of Los Angeles, the model is likely not able to capture the full separation of Los Angeles and Armstrong profiles.

550   Outside of urban or energy-intensive locations, the agreement between the new GGG2020 priors and co-located in situ profiles is much improved. Figure S15 compares RMSEs and mean prior vs. in situ differences for CO when Armstrong AFB, Lamont, and Orléans (another near-urban location) are excluded from the comparison. In that case, the RMSE reduces by about a factor of two or better at all levels except the surface in the new GGG2020 priors compared to the GGG2014 priors.

Despite the increase in RMSE near the surface, overall the CO priors demonstrate important improvement. The ~~scatter versus~~
565 ~~observations is noticeably reduced in the GGG2020 priors (Fig. S10) and the~~ reduction in error in the mid-troposphere will be very beneficial to TCCON retrievals, as the CO averaging kernels increase with altitude up to the tropopause. Therefore, the retrievals are more sensitive to errors in the upper troposphere than the surface. We performed a sensitivity test where we retrieved one year of XCO at Armstrong using two sets of priors. We found that the sensitivity of the retrieved XCO to the surface CO in the prior was small, only 0.024 ppb change XCO per 1 ppb change in surface prior CO (2.4%, Fig. S14c).

570 **5.2 Indirect validation through retrievals**

We can also evaluate the quality of the priors indirectly using the TCCON retrievals themselves. TCCON uses a scaling retrieval, in which the prior profiles are multiplied by scalar volume mixing ratio scale factors (VSFs) until the optimal match between the forward spectroscopic model and measured spectrum is found. A VSF near 1 usually indicates that the prior profile represented the true atmospheric column abundance well (provided that the forward model spectroscopy is accurate), though it
575 is also possible that compensating errors also yield a VSF near 1. However, given that the direct validation shown in Sect. 5.1 do not show compensating positive and negative biases on average, we expect such compensating errors are unlikely.

Figure 16 shows VSFs for HF and N_2O. Figure 16a shows that the median HF VSF decreased from ~ 1.25 in GGG2014 to ~ 0.94 in GGG2020, and the distribution is substantially tighter. HF is found only in the stratosphere (Washenfelder et al., 2003), therefore this result provides additional evidence that the stratosphere is well modeled by the GGG2020 priors.

580 Figure 16b shows that N_2O VSFs moved slightly closer to 1 in GGG2020 with a tighter distribution. N_2O is well mixed in the troposphere with an extremely uniform mixing ratio but varies substantially in the stratosphere due to loss via photolysis. Again, this implies improvement in the stratospheric priors and is a valuable check as we did not directly validate N_2O against aircraft or AirCore observations due to sparse N_2O profiles over TCCON stations.

Finally, we also consider the interhemispheric bias in CH_4 and N_2O VSFs. For CH_4, Saad et al. (2014) found a $\sim 1\%$
585 bias between northern and southern hemisphere CH_4 VSFs using GGG2014 data, and Saad et al. (2016) determined that this was because the GGG2014 priors assumed a smooth DMF profile across the tropopause. In fact, the gradient in the lower stratosphere is driven by stratospheric circulation and CH_4 entering through the tropics (Sect. 2.3). As the priors now correctly
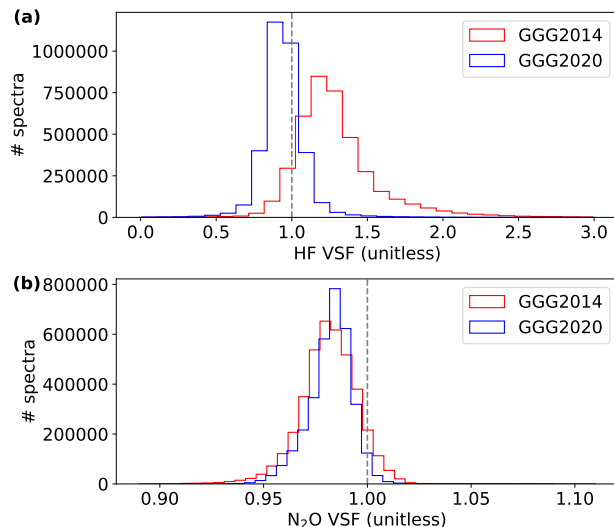
Figure 16. Volume mixing ratio (VMR) scale factors (VSFs) of **(a)** HF and **(b)**N_2O retrieved using GGG2014 and a preliminary version of GGG2020. The vertical dashed gray line marks VSF = 1.

account for this, the underlying error driving the interhemispheric bias in tropospheric XCH_4 in Saad et al. (2014) should now be eliminated and in fact the difference between median CH_4 VSFs between the northern and southern hemispheres has reduced by nearly 50% (Fig. S11).

For N_2O, the difference between median northern and southern hemisphere VSFs remains nearly the same magnitude ($\sim 0.4\%$, Fig. S12) but flips with the GGG2020 priors such that the median VSF is now greater in the southern hemisphere. Figure S12c compares the surface N_2O DMFs from 6 NOAA stations against the surface DMFs in the priors for 5 TCCON sites. While the priors' surface N_2O in the southern hemisphere is approximately correct, there is a high bias in the northern hemisphere, possibly due to an incorrect assumed tropospheric lifetime (Sect. 3.3) or a need for an additional correction to our distance function (Sect. 2.2) that was not identified during development. This will be corrected in a future version of the TCCON priors.

5.3 Impact on retrieved Xgas values

Figure 17 shows how the bias of the Xgas value retrieved by TCCON relative to in situ profiles changes between using the priors from the previous GGG2014 data version and using the new priors described in this paper. For this comparison, we used only AirCore profiles, as these profiles extend into the lower stratosphere and therefore require the least extension to produce a total column profile, suitable for comparison to TCCON. We follow Wunch et al. (2010) in applying TCCON averaging kernels and pressure-weighted integration to the AirCore profiles to produce an in situ Xgas value for comparison to TCCON.
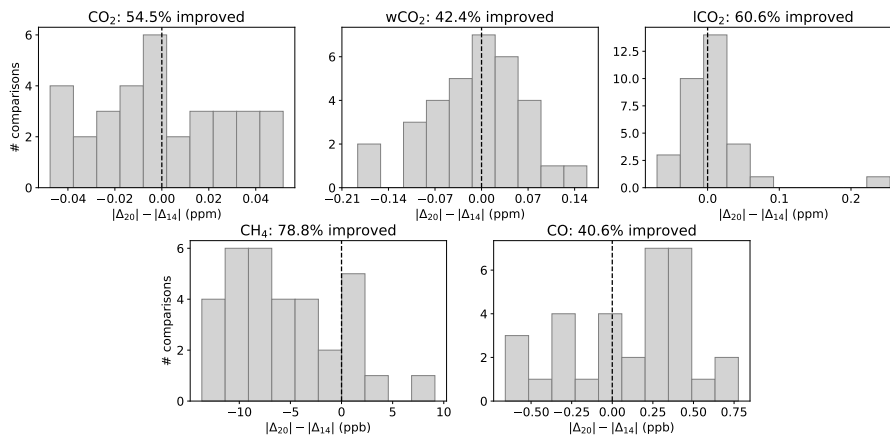
Figure 17. Impact of the new priors on the retrieved TCCON Xgas values compared to coincident AirCore profiles. The x-axis shows how the difference between the retrieved TCCON Xgas value and an averaging kernel-smoothed and integrated in situ profile changes between using the GGG2020 priors described in this paper versus the previous GGG2014 priors. A negative value indicates a reduction in bias compared to in situ with the new priors; the percentage in the title indicates what fraction of the comparisons had reduced bias. The vertical dashed black line marks 0 on the x-axis. Each panel is a different TCCON Xgas product. wCO$_2$ and lCO$_2$ are experimental TCCON CO$_2$ products added in GGG2020 that are more sensitive to the upper atmosphere and near surface, respectively, than the standard TCCON CO$_2$.

For the TCCON CO$_2$ products, the differences are of order 0.05 to 0.1 ppm. Only about half of the comparisons show improvement; this is true for both the standard TCCON CO$_2$ (Fig. 17, top left) and two experimental CO$_2$ products introduced in GGG2020 with different vertical sensitivities (wCO$_2$ and lCO$_2$, Fig. 17, top middle and right).

CO worsened on the whole (Fig. 17, bottom right), but by less than 1 ppb. However, this only includes three comparisons at the Armstrong site (most of the comparisons from Fig. 15 are from aircraft profiles and we only use AirCore profile here as mentioned above), where the new priors have a known bias (see Sect. 5.1) and none at the Pasadena site (as it is difficult to obtain profiles safely over urban sites), which is more strongly affected by the same issue. Thus we consider 1 ppb a lower bound on the bias introduced at these sites by the overestimated near-surface CO in the priors.

CH$_4$ shows the clearest improvement (Fig. 17, bottom left). Almost 80% of comparisons show reductions in bias relative to the AirCore profiles of up to 13.6 ppb. This likely comes from a combination of the new priors' improved representation of the CH$_4$ gradient around the tropopause and the general reduction in bias through the free troposphere (Fig. 14).

6 Conclusions

GGG2020 introduces an improved algorithm to generate the prior profiles of CO$_2$, N$_2$O, CH$_4$, HF, CO and other gases needed for TCCON retrievals. The versions 10 and 11 OCO-2 and OCO-3 retrievals also use these CO$_2$ profiles. This approach is specifically designed to account for variations in vertical profiles due to synoptic-scale latitudinal motion of airmasses. Direct validation against aircraft and AirCore observations shows consistent reduction of error in the free troposphere and lower

stratosphere, and indirect validation by examining the magnitude of retrieved TCCON VSFs gives further evidence that the accuracy of the priors in the stratosphere has improved.

The column-average mole fractions retrieved by TCCON shift relative to in situ column averages by up to 0.2 ppm for CO_2, 13 ppb for CH_4, and 1 ppb for CO. For the standard TCCON CO_2, CH_4, and experimental lCO_2 (CO_2 with stronger sensitivity to the surface) products the new priors produce an overall improvement compared to the in situ column averages. The CO and experimental wCO_2 (stronger sensitivity to the upper atmosphere) products compare slightly worse overall to in situ data using the new priors. For CO, this is likely due to overestimated anthropogenic CO emissions in the source model. Finding a way to correct this, either by using a different model run or applying a geographically-varying correction, will be a high priority for the next version of the TCCON priors. The reason for the slight worsening of the wCO_2 retrievals is not yet clear.

An important guiding principle for the GGG2020 priors algorithm was to limit, as much as possible, dependence on ongoing measurements or models. Doing so means that retrievals using these priors produce data that can be treated as statistically independent with most existing and future measurements and models. Only CO_2, CH_4, and N_2O measurements from the Mauna Loa and American Samoa observatories and CO from the GEOS FP-IT model system are directly ingested, so direct comparisons of TCCON GGG2020 or OCO-2/3 data with these data sources would not be ~~not~~ fully independent. As latitudinal gradients from the HIPPO and ATom campaigns and correlations of N_2O, CH_4, and HF from the ACE-FTS instrument are used as well, comparisons between TCCON or OCO-2/3 and HIPPO, ATom, or ACE-FTS data should note that correlations of these specific characteristics (i.e. latitudinal gradients, N_2O/CH_4/HF correlations) are correlated by design.

There remain areas for improvement. The age of air parameterization used in the troposphere is known to underestimate the age of air compared to SF_6 measurements and anthropogenic emissions are not accounted for except in the CO priors. Addressing these issues is planned for a future version of GGG; at that time, we will evaluate whether incorporating additional data from measurements or models produces worthwhile improvements in the priors' accuracy. Nevertheless, ~~this represents~~ these new priors represent a significant improvement for the GGG2020 TCCON retrieval.

Code and data availability. The code to generate GGG2020 prior profiles is the "ginput" package, available from GitHub (Laughner, 2022). GGG2020 TCCON data uses ginput version 1.0.6, which is scientifically identical to the publicly archived 1.0.7 version (Laughner et al., 2021). HIPPO data was obtained from https://data.eol.ucar.edu/cgi-bin/codiac/fgr_form/id=112.123/agree. ATom data was obtained from https://doi.org/10.3334/ORNLDAAC/1581. Obspack aircraft data (CO_2 GLOBALVIEWplus v5 and CH_4 GLOBALVIEWplus v2 were obtained from https://www.esrl.noaa.gov/gmd/ccgg/obspack/. NOAA AirCore data (v20201223) was provided by Bianca Baier and Colm Sweeney. Sodankylä AirCore data was provided by Huilin Chen and Rigel Kivi. Nicosia AirCore data was provided by Pierre-Yves Quehe. The CMAM model data use in the CO priors was downloaded from https://climate-modelling.canada.ca/climatemodeldata/cmam/output/CMAM/CMAM30-SD/mon/atmosChem/vmrco/index.shtml (last access 24 Jul 2019). ACE-FTS v3.6 data are available fromhttps://database.scisat.ca/level2/; access to these products requires registration. GEOS FP-IT data was downloaded from the Goddard Earth Sciences Data Information Services Center (GES-DISC) with a data subscription (see https://gmao.gsfc.nasa.gov/GMAO_products/, accessed 2022-12-07). CAMS chemical forecast data was downloaded from https://ads.atmosphere.copernicus.eu/ (last access 2022-11-28).

Author contributions. JLL created the priors code, carried out the validation, and led the writing of the manuscript. SR developed the code to read GEOS-FPIT meteorology and interpolate to TCCON locations. MK assisted with development of the CO_2 priors. GCT developed the original GGG priors, of which the climatological profiles used for the secondary gases (Sect. 2.4), seasonal cycle parameterization, and tropospheric distance function are retained in this work. POW guided the overall project. Other authors contributed data for validation of the priors. All authors reviewed the manuscript.

Competing interests. One coauthor is a member of the AMT editorial board.

Acknowledgements. The authors are deeply grateful to Arlyn Andrews from NOAA for providing the approach to generate the stratospheric CO_2 which was adapted and expanded upon for the other stratospheric priors, as well as the tables of stratospheric age-of-air used throughout this work. JLL and POW acknowledge funding from NASA grant NNX17AE15G. We thank all TCCON partners for carrying out GGG2014 and GGG2020 retrievals that provided the VSFs for Sect. 5.2. The AirCore campaign in Cyprus received support by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 856612 and the Cyprus Government and additional support from the Service National d'Observation ICOS-France-Atmosphere coordinated by LSCE. The TCCON Nicosia site has received additional support by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 856612 and the Cyprus Government, and the University of Bremen. The CH_4 tower measurements at Park Falls were supported by the DOE Ameriflux Network Management project award to the ChEAS core site cluster, NSF #0845166 and NSF #1822420. NOAA/GML AirCore profiles were supported by NASA grant 80NSSC18K0898. The authors are grateful to Peter Bernath for his leadership of the ACE-FTS project, which was invaluable to this work. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency. A portion of this research was carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with NASA (80NM0018D0004). Government sponsorship acknowledged.

References

Allen, D. R. and Nakamura, N.: Tracer Equivalent Latitude: A Diagnostic Tool for Isentropic Studies, Am. Meteorol. Soc., 60, 287–304, 2003.

Andrews, A., Boering, K., Daube, B., Wofsy, S., Lowenstein, M., Jost, H., Podolske, J., Webster, C., Herman, R., Scott, D., Flesch, G., Moyer, E., Elkins, J., Dutton, G., Hurst, D., Moore, F., Ray, E., Romashkin, P., and Strahan, S.: Mean ages of stratospheric air derived from in situ observations of CO_2, CH_4, and N_2O, J. Geophys. Res. Atmos., 106, 32,295–32,314, 2001a.

Andrews, A. E., Boering, K. A., Daube, B. C., Wofsy, S. C., Hintsa, E. J., Weinstock, E. M., and Bui, T. P.: Empirical age spectra for the lower tropical stratosphere from in situ observations of CO_2: Implications for stratospheric transport, J. Geophys. Res. Atmos., 104, 26 581–26 595, https://doi.org/10.1029/1999jd900150, 1999.

Andrews, A. E., Boering, K. A., Wofsy, S. C., Daube, B. C., Jones, D. B., Alex, S., Loewenstein, M., Podolske, J. R., and Strahan, S. E.: Empirical age spectra for the midlatitude lower stratosphere from in situ observations of CO_2: Quantitative evidence for a subtropical "barrier" to horizontal transport, J. Geophys. Res. Atmos., 106, 10 257–10 274, https://doi.org/10.1029/2000jd900703, 2001b.

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., Wekker, S. F. J. D., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, Atmos. Meas. Tech., 7, 647–687, https://doi.org/10.5194/amt-7-647-2014, 2014.

Babenhauserheide, A., Hase, F., and Morino, I.: Net CO_2 fossil fuel emissions of Tokyo estimated directly from measurements of the Tsukuba TCCON site and radiosondes, Atmos. Meas. Tech., 13, 2697–2710, https://doi.org/10.5194/amt-13-2697-2020, 2020.

Baier, B., Sweeney, C., Tans, P., Newberger, T., Higgs, J., Wolter, S., and NOAA Global Monitoring Laboratory: NOAA AirCore atmospheric sampling system profiles (Version 20201223) [Data set]. NOAA GML, https://doi.org/10.15138/6AV0-MY81, 2021.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO2 fluxes estimated from GOSAT retrievals of total column CO2, Atmos. Chem. Phys., 13, 8695–8717, https://doi.org/10.5194/acp-13-8695-2013, 2013.

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMazière, M., Drummond, J. R., Dufour, D., Evans, W. F. J., Fast, H., Fussen, D., Gilbert, K., Jennings, D. E., Llewellyn, E. J., Lowe, R. P., Mahieu, E., McConnell, J. C., McHugh, M., McLeod, S. D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C. P., Rochon, Y. J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J. J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K. A., Walkty, I., Wardle, D. A., Wehrle, V., Zander, R., and Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview, Geophysical Research Letters, 32, https://doi.org/https://doi.org/10.1029/2005GL022386, 2005.

Boesch, H., Deutscher, N. M., Warneke, T., Byckling, K., Cogan, A. J., Griffith, D. W. T., Notholt, J., Parker, R. J., and Wang, Z.: HDO/H2O ratio retrievals from GOSAT, Atmos. Meas. Tech., 6, 599–612, https://doi.org/10.5194/amt-6-599-2013, 2013.

Boone, C. D., Walker, K. A., and Bernath, P. F.: The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, chap. Version 3 retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), pp. 103–127, A. Deepak Publishing, Hampton, Virginia, U.S.A., 2013.

Borsdorff, T., aan de Brugh, J., Schneider, A., Lorente, A., Birk, M., Wagner, G., Kivi, R., Hase, F., Feist, D. G., Sussmann, R., Rettinger, M., Wunch, D., Warneke, T., and Landgraf, J.: Improving the TROPOMI CO data product: update of the spectroscopic database and destriping of single orbits, Atmos. Meas. Tech., 12, 5443–5455, https://doi.org/10.5194/amt-12-5443-2019, 2019.

Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S.-W., Evan, S., McKeen, S. A., Hsie, E.-Y., Frost, G. J., Neuman, J. A., Pollack, I. B., Peischl, J., Ryerson, T. B., Holloway, J., Brown, S. S., Nowak, J. B., Roberts, J. M., Wofsy, S. C., Santoni, G. W., Oda, T., and Trainer, M.: Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NO_x and CO_2 and their impacts, Atmos. Chem. Phys., 13, 3661–3677, https://doi.org/10.5194/acp-13-3661-2013, 2013.

Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P., Deutscher, N., Griffith, D., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T.: Toward accurate CO_2 and CH_4 observations from GOSAT, Geophys. Res. Lett., 38, n/a–n/a, https://doi.org/10.1029/2011gl047888, 2011.

Chevallier, F., Deutscher, N. M., Conway, T. J., Ciais, P., Ciattaglia, L., Dohe, S., Fröhlich, M., Gomez-Pelaez, A. J., Griffith, D., Hase, F., Haszpra, L., Krummel, P., Kyrö, E., Labuschagne, C., Langenfelds, R., Machida, T., Maignan, F., Matsueda, H., Morino, I., Notholt, J., Ramonet, M., Sawa, Y., Schmidt, M., Sherlock, V., Steele, P., Strong, K., Sussmann, R., Wennberg, P., Wofsy, S., Worthy, D., Wunch, D., and Zimnoch, M.: Global CO_2 fluxes inferred from surface air-sample measurements and from TCCON retrievals of the CO_2 total column, Geophys. Res. Lett., 38, n/a–n/a, https://doi.org/10.1029/2011gl049899, 2011.

Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J.-F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, J. Geophys. Res. Atmos., 117, n/a–n/a, https://doi.org/10.1029/2012jd018087, 2012.

Cooperative Global Atmospheric Data Integration Project: 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, https://doi.org/10.25925/20190812, 2019.

Cooperative Global Atmospheric Data Integration Project: 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, https://doi.org/10.25925/20200424, 2020.

Crisp, D., O'Dell, C., Eldering, A., Fisher, B., Oyafuso, F., Payne, V., Drouin, B., Toon, G., Laughner, J., Somkuti, P., McGarragh, G., Merrelli, A., Nelson, R., Gunson, M., Frankenberg, C., Osterman, G., Boesch, H., Brown, L., Castano, R., Christi, M., Connor, B., McDuffie, J., Miller, C., Natraj, V., O'Brien, D., Polonski, I., Smyth, M., Thompson, D., and Granat, R.: Orbiting Carbon Observatory (OCO-2) Level 2 Full Physics Algorithm Theoretical Basis Document, Version 3.0 - Rev 1, https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO_L2_ATBD.pdf, 2021.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C., Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, Atmos. Chem. Phys., 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.

Deng, F., Jones, D. B. A., O'Dell, C. W., Nassar, R., and Parazoo, N. C.: Combining GOSAT XCO_2 observations over land and ocean to improve regional CO_2 flux estimates, J. Geophys. Res. Atmos., 121, 1896–1913, https://doi.org/10.1002/2015jd024157, 2016.

745   Desai, A. R., Xu, K., Tian, H., Weishampel, P., Thom, J., Baumann, D., Andrews, A. E., Cook, B. D., King, J. Y., and Kolka, R.: Landscape-level terrestrial methane flux observed from a very tall tower, Agr. Forest Meteorol., 201, 61–75, https://doi.org/10.1016/j.agrformet.2014.10.017, 2015.

Dils, B., Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Parker, R., Guerlet, S., Aben, I., Blumenstock, T., Burrows, J. P., Butz, A., Deutscher, N. M., Frankenberg, C., Hase, F., Hasekamp, O. P., Heymann, J., Mazière, M. D., Notholt, J., Sussmann, R., Warneke,

750   T., Griffith, D., Sherlock, V., and Wunch, D.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO2 and CH4 retrieval algorithm products with measurements from the TCCON, Atmos. Meas. Tech., 7, 1723–1744, https://doi.org/10.5194/amt-7-1723-2014, 2014.

Dlugokencky, E., Mund, J., Crotwell, A., Crotwell, M., and Thoning, K.: Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1968-2018, Version: 2019-07, https://doi.org/10.15138/wkgj-

755   f215, 2019.

Dogniaux, M., Crevoisier, C., Armante, R., Capelle, V., Delahaye, T., Cassé, V., De Mazière, M., Deutscher, N. M., Feist, D. G., Garcia, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Pollard, D. F., Roehl, C. M., Shiomi, K., Strong, K., Té, Y., Velazco, V. A., and Warneke, T.: The Adaptable 4A Inversion (5AI): Description and first XCO_2 retrievals from OCO-2 observations, Atmos. Meas. Tech. Discuss., 2020, 1–37, https://doi.org/10.5194/amt-2020-403, 2020.

760   Dupuy, E., Morino, I., Deutscher, N., Yoshida, Y., Uchino, O., Connor, B., Mazière, M. D., Griffith, D., Hase, F., Heikkinen, P., Hillyard, P., Iraci, L., Kawakami, S., Kivi, R., Matsunaga, T., Notholt, J., Petri, C., Podolske, J., Pollard, D., Rettinger, M., Roehl, C., Sherlock, V., Sussmann, R., Toon, G., Velazco, V., Warneke, T., Wennberg, P., Wunch, D., and Yokota, T.: Comparison of XH2O Retrieved from GOSAT Short-Wavelength Infrared Spectra with Observations from the TCCON Network, Remote Sens., 8, 414, https://doi.org/10.3390/rs8050414, 2016.

765   Feng, L., Palmer, P. I., Bösch, H., Parker, R. J., Webb, A. J., Correia, C. S. C., Deutscher, N. M., Domingues, L. G., Feist, D. G., Gatti, L. V., Gloor, E., Hase, F., Kivi, R., Liu, Y., Miller, J. B., Morino, I., Sussmann, R., Strong, K., Uchino, O., Wang, J., and Zahn, A.: Consistent regional fluxes of CH4 and CO2 inferred from GOSAT proxy XCH4:XCO2 retrievals, 2010–2014, Atmos. Chem. Phys., 17, 4781–4797, https://doi.org/10.5194/acp-17-4781-2017, 2017.

Fleming, E. L., Jackman, C. H., Stolarski, R. S., and Douglass, A. R.: A model study of the impact of source gas changes on the stratosphere

770   for 1850–2100, Atmos. Chem. Phys., 11, 8515–8541, https://doi.org/10.5194/acp-11-8515-2011, 2011.

Frankenberg, C., Wunch, D., Toon, G., Risi, C., Scheepmaker, R., Lee, J.-E., Wennberg, P., and Worden, J.: Water vapor isotopologue retrievals from high-resolution GOSAT shortwave infrared spectra, Atmos. Meas. Tech., 6, 263–274, https://doi.org/10.5194/amt-6-263-2013, 2013.

Franklin, M., Chau, K., Cushing, L. J., and Johnston, J. E.: Characterizing Flaring from Unconventional Oil and Gas Operations in South

775   Texas Using Satellite Observations, Environ. Sci. Technol., 53, 2220–2228, https://doi.org/10.1021/acs.est.8b05355, 2019.

Fraser, A., Palmer, P. I., Feng, L., Boesch, H., Cogan, A., Parker, R., Dlugokencky, E. J., Fraser, P. J., Krummel, P. B., Langenfelds, R. L., O'Doherty, S., Prinn, R. G., Steele, L. P., van der Schoot, M., and Weiss, R. F.: Estimating regional methane surface fluxes: the relative importance of surface and GOSAT mole fraction measurements, Atmos. Chem. Phys., 13, 5697–5713, https://doi.org/10.5194/acp-13-5697-2013, 2013.

780   Geibel, M. C., Messerschmidt, J., Gerbig, C., Blumenstock, T., Chen, H., Hase, F., Kolle, O., Lavrič, J. V., Notholt, J., Palm, M., Rettinger, M., Schmidt, M., Sussmann, R., Warneke, T., and Feist, D. G.: Calibration of column-averaged CH4 over European TCCON FTS sites with airborne in-situ measurements, Atmos. Chem. Phys., 12, 8763–8775, https://doi.org/10.5194/acp-12-8763-2012, 2012.

Hall, B. D., Crotwell, A. M., Kitzis, D. R., Mefford, T., Miller, B. R., Schibig, M. F., and Tans, P. P.: Revision of the World Meteorological Organization Global Atmosphere Watch (WMO/GAW) CO_2 calibration scale, Atmospheric Measurement Techniques, 14, 3015–3032, https://doi.org/10.5194/amt-14-3015-2021, 2021.

Hedelius, J. K., Liu, J., Oda, T., Maksyutov, S., Roehl, C. M., Iraci, L. T., Podolske, J. R., Hillyard, P. W., Liang, J., Gurney, K. R., Wunch, D., and Wennberg, P. O.: Southern California megacity CO2, CH4, and CO flux estimates using ground- and space-based remote sensing and a Lagrangian model, Atmos. Chem. Phys., 18, 16 271–16 291, https://doi.org/10.5194/acp-18-16271-2018, 2018.

Heymann, J., Reuter, M., Hilker, M., Buchwitz, M., Schneising, O., Bovensmann, H., Burrows, J. P., Kuze, A., Suto, H., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kawakami, S., Kivi, R., Morino, I., Petri, C., Roehl, C., Schneider, M., Sherlock, V., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: Consistent satellite XCO2 retrievals from SCIAMACHY and GOSAT using the BESD algorithm, Atmos. Meas. Tech., 8, 2961–2980, https://doi.org/10.5194/amt-8-2961-2015, 2015.

Inoue, M., Morino, I., Uchino, O., Miyamoto, Y., Saeki, T., Yoshida, Y., Yokota, T., Sweeney, C., Tans, P. P., Biraud, S. C., Machida, T., Pittman, J. V., Kort, E. A., Tanaka, T., Kawakami, S., Sawa, Y., Tsuboi, K., and Matsueda, H.: Validation of XCH4 derived from SWIR spectra of GOSAT TANSO-FTS with aircraft measurement data, Atmos. Meas. Tech., 7, 2987–3005, https://doi.org/10.5194/amt-7-2987-2014, 2014.

Inoue, M., Morino, I., Uchino, O., Nakatsuru, T., Yoshida, Y., Yokota, T., Wunch, D., Wennberg, P. O., Roehl, C. M., Griffith, D. W. T., Velazco, V. A., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Hase, F., Blumenstock, T., Rettinger, M., Sussmann, R., Kyrö, E., Kivi, R., Shiomi, K., Kawakami, S., Mazière, M. D., Arnold, S. G., Feist, D. G., Barrow, E. A., Barney, J., Dubey, M., Schneider, M., Iraci, L. T., Podolske, J. R., Hillyard, P. W., Machida, T., Sawa, Y., Tsuboi, K., Matsueda, H., Sweeney, C., Tans, P. P., Andrews, A. E., Biraud, S. C., Fukuyama, Y., Pittman, J. V., Kort, E. A., and Tanaka, T.: Bias corrections of GOSAT SWIR XCO2 and XCH4 with TCCON data and their evaluation using aircraft measurement data, Atmos. Meas. Tech., 9, 3491–3512, https://doi.org/10.5194/amt-9-3491-2016, 2016.

Jacob, D. J.: Introduction to Atmospheric Chemistry, chap. Ch. 10: Stratospheric Ozone, Princeton University Press, 1999.

Jin, J. J., Semeniuk, K., Beagley, S. R., Fomichev, V. I., Jonsson, A. I., McConnell, J. C., Urban, J., Murtagh, D., Manney, G. L., Boone, C. D., Bernath, P. F., Walker, K. A., Barret, B., Ricaud, P., and Dupuy, E.: Comparison of CMAM simulations of carbon monoxide (CO), nitrous oxide (N_2O), and methane (CH_4) with observations from Odin/SMR, ACE-FTS, and Aura/MLS, Atmos. Chem. Phys., 9, 3233–3252, https://doi.org/10.5194/acp-9-3233-2009, 2009.

Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System, J. Atmos. Oceanic Technol., 27, 1839–1853, https://doi.org/10.1175/2010jtecha1448.1, 2010.

Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., Conley, S. A., Davis, K., Deng, A., Hardesty, M., Herndon, S. C., Lauvaux, T., Lavoie, T., Lyon, D., Newberger, T., Pétron, G., Rella, C., Smith, M., Wolter, S., Yacovitch, T. I., and Tans, P.: Aircraft-Based Estimate of Total Methane Emissions from the Barnett Shale Region, Environmental Science & Technology, 49, 8124–8131, https://doi.org/10.1021/acs.est.5b00217, 2015.

Keppel-Aleks, G., Wennberg, P., Washenfelder, R., Wunch, D., Schneider, T., Toon, G., Andres, R., Blavier, J.-F., Connor, B., Davis, K., Desai, A., Messerschmidt, J., Notholt, J., Roehl, C., Sherlock, V., Stephens, B., Vay, S., and Wofsy, S.: The imprint of surface fluxes and transport on variations in total column carbon dioxide, Biogeosci., 9, 875–891, https://doi.org/10.5194/bg-9-875-2012, 2012.

Kiel, M., O'Dell, C. W., Fisher, B., Eldering, A., Nassar, R., MacDonald, C. G., and Wennberg, P. O.: How bias correction goes wrong: measurement of XCO2 affected by erroneous surface pressure estimates, Atmos. Meas. Tech., 12, 2241–2259, https://doi.org/10.5194/amt-12-2241-2019, 2019.

Kivi, R. and Heikkinen, P.: Fourier transform spectrometer measurements of column CO_2 at Sodankylä, Finland, Geoscientific Instrumentation, Methods and Data Systems, 5, 271–279, https://doi.org/10.5194/gi-5-271-2016, 2016.

Kolonjari, F., Plummer, D. A., Walker, K. A., Boone, C. D., Elkins, J. W., Hegglin, M. I., Manney, G. L., Moore, F. L., Pendlebury, D., Ray, E. A., Rosenlof, K. H., and Stiller, G. P.: Assessing stratospheric transport in the CMAM30 simulations using ACE-FTS measurements, Atmos. Chem. Phys., 18, 6801–6828, https://doi.org/10.5194/acp-18-6801-2018, 2018.

Kuang, Z., Toon, G. C., Wennberg, P. O., and Yung, Y. L.: Measured HDO/H_2O ratios across the tropical tropopause, Geophys. Res. Lett., 30, https://doi.org/10.1029/2003gl017023, 2003.

Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., Chevallier, F., Sherlock, V., Buchwitz, M., Osterman, G., Miller, C. E., Wennberg, P. O., Griffith, D., Morino, I., Dubey, M. K., Deutscher, N. M., Notholt, J., Hase, F., Warneke, T., Sussmann, R., Robinson, J., Strong, K., Schneider, M., Mazière, M. D., Shiomi, K., Feist, D. G., Iraci, L. T., and Wolf, J.: Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON, Atmos. Meas. Tech., 9, 683–709, https://doi.org/10.5194/amt-9-683-2016, 2016.

Laughner, J.: WennbergLab/py-ginput: ginput v1.1.6 release, https://doi.org/10.22002/D1.20285, 2022.

Laughner, J., Andrews, A., Roche, S., Kiel, M., and Toon, G.: ginput v1.0.7b: GGG2020 prior profile software, https://doi.org/10.22002/D1.1880, 2021.

Liang, A., Gong, W., Han, G., and Xiang, C.: Comparison of Satellite-Observed XCO2 from GOSAT, OCO-2, and Ground-Based TCCON, Remote Sens., 9, 1033, https://doi.org/10.3390/rs9101033, 2017a.

Liang, A., Han, G., Gong, W., Yang, J., and Xiang, C.: Comparison of Global XCO2 Concentrations From OCO-2 With TCCON Data in Terms of Latitude Zones, IEEE J. Sel. Top. App., 10, 2491–2498, https://doi.org/10.1109/jstars.2017.2650942, 2017b.

Liu, Y., Yang, D., and Cai, Z.: A retrieval algorithm for TanSat XCO2 observation: Retrieval experiments using GOSAT data, Chinese Sci. Bull., 58, 1520–1523, https://doi.org/10.1007/s11434-013-5680-y, 2013.

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velazco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., and Landgraf, J.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements, Atmos. Meas. Tech., 14, 665–684, https://doi.org/10.5194/amt-14-665-2021, 2021.

Lucchesi, R.: File Specification for GEOS-5 FP-IT (forward processing for instrument teams), Tech. rep., NASA Goddard Space Flight Center, Greenbelt, MD, USA,, https://gmao.gsfc.nasa.gov/pubs/docs/Lucchesi865.pdf, last access: 13 October 2020, 2015.

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth System Science Data, 6, 375–384, https://doi.org/10.5194/essd-6-375-2014, 2014.

Messerschmidt, J., Chen, H., Deutscher, N. M., Gerbig, C., Grupe, P., Katrynski, K., Koch, F.-T., Lavrič, J. V., Notholt, J., Rödenbeck, C., Ruhe, W., Warneke, T., and Weinzierl, C.: Automated ground-based remote sensing measurements of greenhouse gases at the Białystok site in comparison with collocated in situ measurements and model data, Atmospheric Chemistry and Physics, 12, 6741–6755, , 2012.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, chap. 8: Anthropogenic and Natural Radiative Forcing., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Ohyama, H., Kawakami, S., Tanaka, T., Morino, I., Uchino, O., Inoue, M., Sakai, T., Nagai, T., Yamazaki, A., Uchiyama, A., Fukamachi, T., Sakashita, M., Kawasaki, T., Akaho, T., Arai, K., and Okumura, H.: Observations of XCO2 and XCH4 with ground-based high-resolution

FTS at Saga, Japan, and comparisons with GOSAT products, Atmos. Meas. Tech., 8, 5263–5276, https://doi.org/10.5194/amt-8-5263-2015, 2015.

Ohyama, H., Kawakami, S., Shiomi, K., Morino, I., and Uchino, O.: Intercomparison of XH2O Data from the GOSAT TANSO-FTS (TIR and SWIR) and Ground-Based FTS Measurements: Impact of the Spatial Variability of XH2O on the Intercomparison, Remote Sens., 9, 64, https://doi.org/10.3390/rs9010064, 2017.

Okita, T., Kaneda, K., Yanaka, T., and Sugai, R.: Determination of gaseous and particulate chloride and fluoride in the atmosphere, Atmospheric Environment (1967), 8, 927–936, https://doi.org/https://doi.org/10.1016/0004-6981(74)90082-1, 1974.

Oshchepkov, S., Bril, A., Yokota, T., Wennberg, P. O., Deutscher, N. M., Wunch, D., Toon, G. C., Yoshida, Y., O'Dell, C. W., Crisp, D., Miller, C. E., Frankenberg, C., Butz, A., Aben, I., Guerlet, S., Hasekamp, O., Boesch, H., Cogan, A., Parker, R., Griffith, D., Macatangay, R., Notholt, J., Sussmann, R., Rettinger, M., Sherlock, V., Robinson, J., Kyrö, E., Heikkinen, P., Feist, D. G., Morino, I., Kadygrov, N., Belikov, D., Maksyutov, S., Matsunaga, T., Uchino, O., and Watanabe, H.: Effects of atmospheric light scattering on spectroscopic observations of greenhouse gases from space. Part 2: Algorithm intercomparison in the GOSAT data processing for CO2retrievals over TCCON sites, J. Geophys. Res. Atmos., 118, 1493–1512, https://doi.org/10.1002/jgrd.50146, 2013.

Ott, L. E., Pawson, S., Collatz, G. J., Gregg, W. W., Menemenlis, D., Brix, H., Rousseaux, C. S., Bowman, K. W., Liu, J., Eldering, A., Gunson, M. R., and Kawa, S. R.: Assessing the magnitude of CO_2 flux uncertainty in atmospheric CO_2 records using products from NASA's Carbon Monitoring Flux Pilot Project, J. Geophys. Res. Atmos., 120, 734–765, https://doi.org/10.1002/2014jd022411, 2015.

Parker, R. J., Boesch, H., Byckling, K., Webb, A. J., Palmer, P. I., Feng, L., Bergamaschi, P., Chevallier, F., Notholt, J., Deutscher, N., Warneke, T., Hase, F., Sussmann, R., Kawakami, S., Kivi, R., Griffith, D. W. T., and Velazco, V.: Assessing 5 years of GOSAT Proxy XCH4 data and associated uncertainties, Atmos. Meas. Tech., 8, 4785–4801, https://doi.org/10.5194/amt-8-4785-2015, 2015.

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., Li, X., Liu, Y., Lomas, M., Poulter, B., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zhao, F., and Zhao, H.: Benchmarking the seasonal cycle of CO2fluxes simulated by terrestrial ecosystem models, Global Biogeochem. Cy., 29, 46–64, https://doi.org/10.1002/2014gb004931, 2015.

Polavarapu, S. M., Neish, M., Tanguay, M., Girard, C., de Grandpré, J., Semeniuk, K., Gravel, S., Ren, S., Roche, S., Chan, D., and Strong, K.: Greenhouse gas simulations with a coupled meteorological and transport model: the predictability of CO_2, Atmos. Chem. Phys., 16, 12 005–12 038, https://doi.org/10.5194/acp-16-12005-2016, 2016.

Roche, S., Strong, K., Wunch, D., Mendonca, J., Sweeney, C., Baier, B., Biraud, S. C., Laughner, J. L., Toon, G. C., and Connor, B. J.: Retrieval of atmospheric CO_2 vertical profiles from ground-based near-infrared spectra, Atmospheric Measurement Techniques, 14, 3087–3118, https://doi.org/10.5194/amt-14-3087-2021, 2021.

Saad, K. M., Wunch, D., Toon, G. C., Bernath, P., Boone, C., Connor, B., Deutscher, N. M., Griffith, D. W. T., Kivi, R., Notholt, J., Roehl, C., Schneider, M., Sherlock, V., and Wennberg, P. O.: Derivation of tropospheric methane from TCCON CH4 and HF total column observations, Atmos. Meas. Tech., 7, 2907–2918, https://doi.org/10.5194/amt-7-2907-2014, 2014.

Saad, K. M., Wunch, D., Deutscher, N. M., Griffith, D. W. T., Hase, F., De Mazière, M., Notholt, J., Pollard, D. F., Roehl, C. M., Schneider, M., Sussmann, R., Warneke, T., and Wennberg, P. O.: Seasonal variability of stratospheric methane: implications for constraining tropospheric methane budgets using total column observations, Atmos. Chem. Phys., 16, 14 003–14 024, https://doi.org/10.5194/acp-16-14003-2016, 2016.

Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., Blavier, J.-F., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., and Aben, I.: Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT)

shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, J. Geophys. Res. Atmos., 117, n/a–n/a, https://doi.org/10.1029/2012jd017549, 2012.

Schepers, D., Butz, A., Hu, H., Hasekamp, O. P., Arnold, S. G., Schneider, M., Feist, D. G., Morino, I., Pollard, D., Aben, I., and Landgraf, J.: Methane and carbon dioxide total column retrievals from cloudy GOSAT soundings over the oceans, J. Geophys. Res. Atmos., 121, 5031–5050, https://doi.org/10.1002/2015jd023389, 2016.

Schneising, O., Buchwitz, M., Reuter, M., Bovensmann, H., Burrows, J. P., Borsdorff, T., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Hermans, C., Iraci, L. T., Kivi, R., Landgraf, J., Morino, I., Notholt, J., Petri, C., Pollard, D. F., Roche, S., Shiomi, K., Strong, K., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor, Atmos. Meas. Tech., 12, 6771–6802, https://doi.org/10.5194/amt-12-6771-2019, 2019.

Seabold, S. and Perktold, J.: statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference, 2010.

Sheese, P. E., Walker, K. A., Boone, C. D., Bernath, P. F., Froidevaux, L., Funke, B., Raspollini, P., and von Clarmann, T.: ACE-FTS ozone, water vapour, nitrous oxide, nitric acid, and carbon monoxide profile comparisons with MIPAS and MLS, J. Quant. Spectrosc. Radiat. Transfer, 186, 63–80, https://doi.org/https://doi.org/10.1016/j.jqsrt.2016.06.026, satellite Remote Sensing and Spectroscopy: Joint ACE-Odin Meeting, October 2015, 2017.

Sobel, A. H., Nilsson, J., and Polvani, L. M.: The Weak Temperature Gradient Approximation and Balanced Tropical Moisture Waves, J. Atmos. Sci., 58, 3650–3665, https://doi.org/10.1175/1520-0469(2001)058<3650:twtgaa>2.0.co;2, 2001.

Strong, K., Wolff, M. A., Kerzenmacher, T. E., Walker, K. A., Bernath, P. F., Blumenstock, T., Boone, C., Catoire, V., Coffey, M., De Mazière, M., Demoulin, P., Duchatelet, P., Dupuy, E., Hannigan, J., Höpfner, M., Glatthor, N., Griffith, D. W. T., Jin, J. J., Jones, N., Jucks, K., Kuellmann, H., Kuttippurath, J., Lambert, A., Mahieu, E., McConnell, J. C., Mellqvist, J., Mikuteit, S., Murtagh, D. P., Notholt, J., Piccolo, C., Raspollini, P., Ridolfi, M., Robert, C., Schneider, M., Schrems, O., Semeniuk, K., Senten, C., Stiller, G. P., Strandberg, A., Taylor, J., Tétard, C., Toohey, M., Urban, J., Warneke, T., and Wood, S.: Validation of ACE-FTS N_2O measurements, Atmos. Chem. Phys., 8, 4759–4786, https://doi.org/10.5194/acp-8-4759-2008, 2008.

Sussmann, R. and Rettinger, M.: Can We Measure a COVID-19-Related Slowdown in Atmospheric CO_2 Growth? Sensitivity of Total Carbon Column Observations, Remote Sensing, 12, https://doi.org/10.3390/rs12152387, 2020.

Tans, P.: System and method for providing vertical profile measurements of atmospheric gases., U.S. Patent 7,597,014, filed 15 Aug 2006, issued 6 Oct 2009, 2009.

Thompson, C. R., Wofsy, S. C., Prather, M. J., Newman, P. A., Hanisco, T. F., Ryerson, T. B., Fahey, D. W., Apel, E. C., Brock, C. A., Brune, W. H., Froyd, K., Katich, J. M., Nicely, J. M., Peischl, J., Ray, E., Veres, P. R., Wang, S., Allen, H. M., Asher, E., Bian, H., Blake, D., Bourgeois, I., Budney, J., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C., Chin, M., Commane, R., Correa, G., Crounse, J. D., Daube, B., Dibb, J. E., DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Fiore, A. M., Flynn, C. M., Guo, H., Hall, S. R., Hannun, R. A., Hills, A., Hintsa, E. J., Hodzic, A., Hornbrook, R. S., Huey, L. G., Jimenez, J. L., Keeling, R. F., Kim, M. J., Kupc, A., Lacey, F., Lait, L. R., Lamarque, J.-F., Liu, J., McKain, K., Meinardi, S., Miller, D. O., Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., Nault, B. A., Neuman, J. A., Nguyen, L., Gonzalez, Y., Rollins, A., Rosenlof, K., Sargent, M., Schill, G., Schwarz, J. P., Clair, J. M. S., Steenrod, S. D., Stephens, B. B., Strahan, S. E., Strode, S. A., Sweeney, C., Thames, A. B., Ullmann, K., Wagner, N., Weber, R., Weinzierl, B., Wennberg, P. O., Williamson, C. J., Wolfe, G. M., and Zeng, L.: The NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere, Bulletin of the American Meteorological Society, 103, E761–E790, https://doi.org/10.1175/bams-d-20-0315.1, 2022.

Thoning, K., Crotwell, A., and Mund, J.: 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, https://doi.org/10.15138/yaf1-bk21, 2021.

Toon, G. C.: The JPL MkIV interferometer, Opt. Photonics News, 1991.

Toon, G. C., Farmer, C. B., Lowes, L. L., Schaper, P. W., Blavier, J.-F., and Norton, R. H.: Infrared aircraft measurements of stratospheric composition over Antarctica during September 1987, Journal of Geophysical Research: Atmospheres, 94, 16 571–16 596, https://doi.org/https://doi.org/10.1029/JD094iD14p16571, 1989.

Velazco, V. A., Deutscher, N. M., Morino, I., Uchino, O., Bukosa, B., Ajiro, M., Kamei, A., Jones, N. B., Paton-Walsh, C., and Griffith, D. W. T.: Satellite and ground-based measurements of XCO_2 in a remote semiarid region of Australia, Earth Sys. Sci. Data, 11, 935–946, https://doi.org/10.5194/essd-11-935-2019, 2019.

Villalobos, Y., Rayner, P. J., Silver, J. D., Thomas, S., Haverd, V., Knauer, J., Loh, Z. M., Deutscher, N. M., Griffith, D. W. T., and Pollard, D. F.: Was Australia a sink or source of CO_2 in 2015? Data assimilation using OCO-2 satellite measurements, Atmos. Chem. Phys. Discuss., 2021, 1–44, https://doi.org/10.5194/acp-2021-16, 2021.

Wang, Y., Deutscher, N. M., Palm, M., Warneke, T., Notholt, J., Baker, I., Berry, J., Suntharalingam, P., Jones, N., Mahieu, E., Lejeune, B., Hannigan, J., Conway, S., Mendonca, J., Strong, K., Campbell, J. E., Wolf, A., and Kremser, S.: Towards understanding the variability in biospheric CO2 fluxes: using FTIR spectrometry and a chemical transport model to investigate the sources and sinks of carbonyl sulfide and its link to CO2, Atmos. Chem. Phys., 16, 2123–2138, https://doi.org/10.5194/acp-16-2123-2016, 2016.

Washenfelder, R. A., Wennberg, P. O., and Toon, G. C.: Tropospheric methane retrieved from ground-based near-IR solar absorption spectra, Geophys. Res. Lett., 30, https://doi.org/10.1029/2003GL017969, 2003.

Waymark, C., Walker, K., Boone, C. D., and Bernath, P. F.: ACE-FTS version 3.0 data set: validation and data processing update, Annals of Geophysics, 56, https://doi.org/10.4401/ag-6339, 2014.

Willyard, K. A. and Schade, G. W.: Flaring in two Texas shale areas: Comparison of bottom-up with top-down volume estimates for 2012 to 2015, Science of The Total Environment, 691, 243–251, https://doi.org/10.1016/j.scitotenv.2019.06.465, 2019.

Wofsy, S., Afshar, S., Allen, H., Apel, E., Asher, E., Barletta, B., Bent, J., Bian, H., Biggs, B., Blake, D., Blake, N., Bourgeois, I., Brock, C., Brune, W., Budney, J., Bui, T., Butler, A., Campuzano-Jost, P., Chang, C., Chin, M., Commane, R., Correa, G., Crounse, J., Cullis, P. D., Daube, B., Day, D., Dean-Day, J., Dibb, J., DiGangi, J., Diskin, G., Dollner, M., Elkins, J., Erdesz, F., Fiore, A., Flynn, C., Froyd, K., Gesler, D., Hall, S., Hanisco, T., Hannun, R., Hills, A., Hintsa, E., Hoffman, A., Hornbrook, R., Huey, L., Hughes, S., Jimenez, J., Johnson, B., Katich, J., Keeling, R., Kim, M., Kupc, A., Lait, L., Lamarque, J.-F., Liu, J., McKain, K., Mclaughlin, R., Meinardi, S., Miller, D., Montzka, S., Moore, F., Morgan, E., Murphy, D., Murray, L., Nault, B., Neuman, J., Newman, P., Nicely, J., Pan, X., Paplawsky, W., Peischl, J., Prather, M., Price, D., Ray, E., Reeves, J., Richardson, M., Rollins, A., Rosenlof, K., Ryerson, T., Scheuer, E., Schill, G., Schroder, J., Schwarz, J., St.Clair, J., Steenrod, S., Stephens, B., Strode, S., Sweeney, C., Tanner, D., Teng, A., Thames, A., Thompson, C., Ullmann, K., Veres, P., Vieznor, N., Wagner, N., Watt, A., Weber, R., Weinzierl, B., Wennberg, P., Williamson, C., Wilson, J., Wolfe, G., Woods, C., and Zeng, L.: ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, https://doi.org/10.3334/ORNLDAAC/1581, 2018.

Wofsy, S. C.: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369, 2073–2086, https://doi.org/10.1098/rsta.2010.0313, 2011.

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, Atmos. Meas. Tech., 3, 1351–1362, https://doi.org/10.5194/amt-3-1351-2010, 2010.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Philos. T R Soc. A, 369, 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., Maziere, M. D., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON, Atmos. Meas. Tech., 10, 2209–2238, https://doi.org/10.5194/amt-10-2209-2017, 2017.

Yang, D., Boesch, H., Liu, Y., Somkuti, P., Cai, Z., Chen, X., Di Noia, A., Lin, C., Lu, N., Lyu, D., Parker, R. J., Tian, L., Wang, M., Webb, A., Yao, L., Yin, Z., Zheng, Y., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Pollard, D. F., Shiomi, K., Sussmann, R., Té, Y., Velazco, V. A., Warneke, T., and Wunch, D.: Toward High Precision XCO_2 Retrievals From TanSat Observations: Retrieval Improvement and Validation Against TCCON Measurements, J. Geophys. Res. Atmos., 125, e2020JD032 794, https://doi.org/https://doi.org/10.1029/2020JD032794, e2020JD032794 2020JD032794, 2020.

Yang, Z., Washenfelder, R. A., Keppel-Aleks, G., Krakauer, N. Y., Randerson, J. T., Tans, P. P., Sweeney, C., and Wennberg, P. O.: New constraints on Northern Hemisphere growing season net flux, Geophys. Res. Lett., 34, https://doi.org/10.1029/2007gl029742, 2007.

York, D., Evensen, N. M., Martínez, M. L., and De Basabe Delgado, J.: Unified equations for the slope, intercept, and standard errors of the best straight line, American Journal of Physics, 72, 367–375, https://doi.org/10.1119/1.1632486, 2004.

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., Roehl, C. M., Wennberg, P. O., Griffith, D. W. T., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Connor, B., Rettinger, M., Sussmann, R., Ahonen, P., Heikkinen, P., Kyrö, E., Mendonca, J., Strong, K., Hase, F., Dohe, S., and Yokota, T.: Improvement of the retrieval algorithm for GOSAT SWIR XCO2 and XCH4 and their validation using TCCON data, Atmos. Meas. Tech., 6, 1533–1547, https://doi.org/10.5194/amt-6-1533-2013, 2013.

Zander, R., Roland, G., Delbouille, L., Sauval, A., Marché, P., Karcher, F., Amoudei, M., and Dufour, B.: Concentrations of hydrogen chloride and hydrogen fluoride measured during the MAP/GLOBUS campaign of September 1983, Planetary and Space Science, 35, 665–672, https://doi.org/https://doi.org/10.1016/0032-0633(87)90133-4, 1987.

Zhang, Y., Jacob, D. J., Lu, X., Maasakkers, J. D., Scarpelli, T. R., Sheng, J.-X., Shen, L., Qu, Z., Sulprizio, M. P., Chang, J., Bloom, A. A., Ma, S., Worden, J., Parker, R. J., and Boesch, H.: Attribution of the accelerating increase in atmospheric methane during 2010–2018 by inverse analysis of GOSAT observations, Atmos. Chem. Phys. Discuss., 2020, 1–42, https://doi.org/10.5194/acp-2020-964, 2020.

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. These various profiles represent different errors in both the shape *and* prior XCO_2 values. This was deliberate to test how the retrieval is sensitive to not only the error in shape but the total amount of prior CO_2.

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
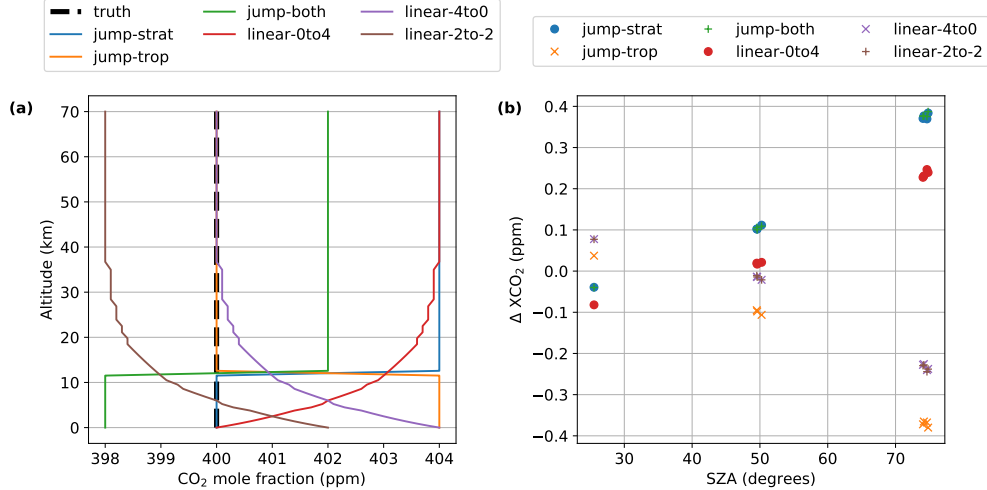
Figure S1: The effect of six different shape errors in a priori CO_2 profiles on retrieved XCO_2 for 3 different SZAs and 4 different days of year (one each in January, April, July, and September). **(a)** The six different test a priori CO_2 profiles and the "true" CO_2 profile used to generate the 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).

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.

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
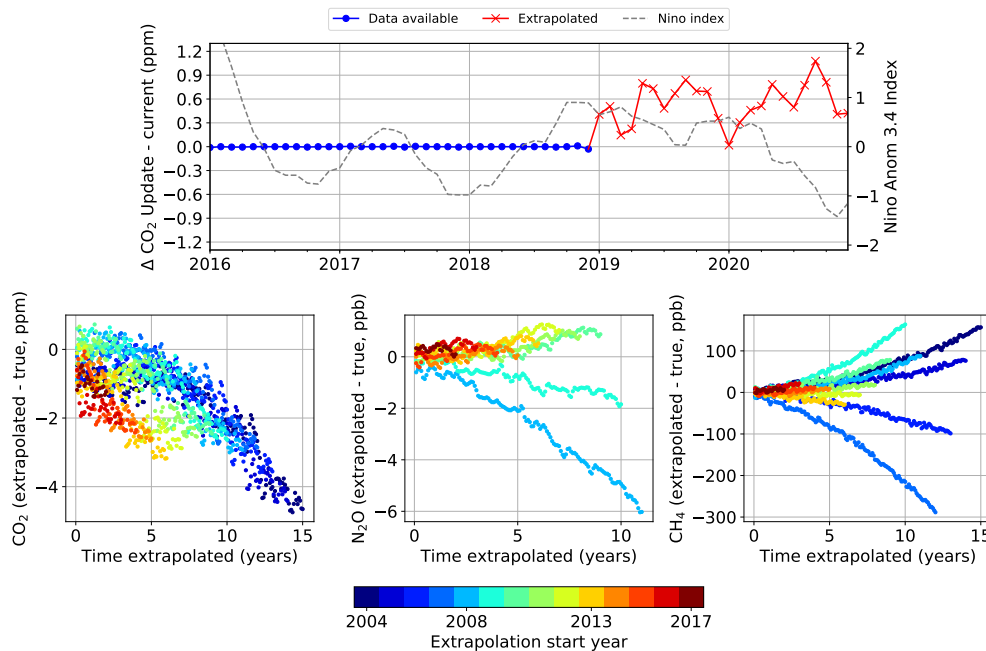
Figure S2: Top panel: Difference between the CO_2 record used in the GGG2020 and OCO-2/3 v10 priors and updated MLO & SMO mean monthly flask data that extends through 2020. Blue indicates where the GGG2020/v10 priors have MLO & SMO data available, red indicates where they must extrapolate. The Niño 3.4 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.

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 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):

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

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
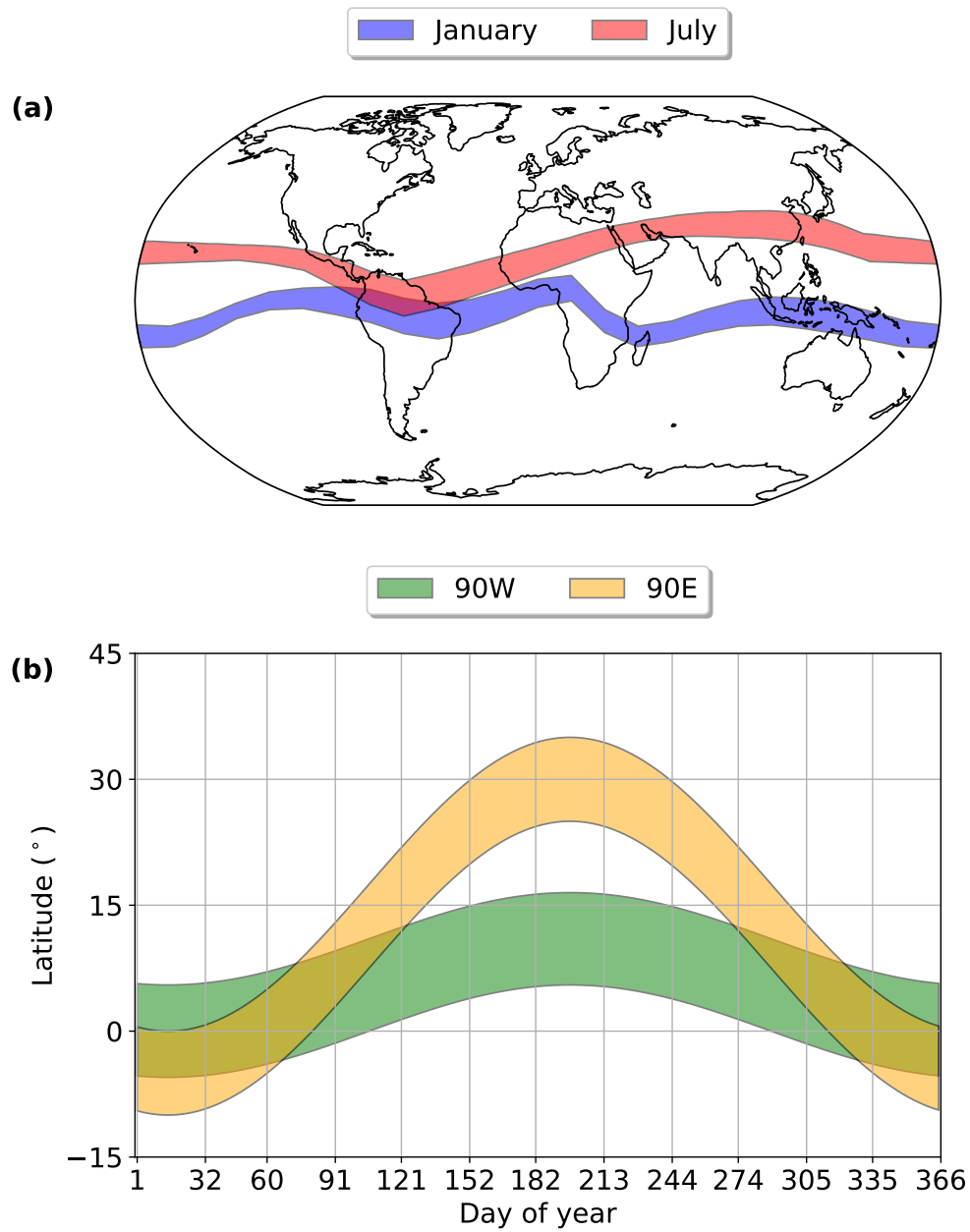
Figure S3: Assumed **(a)** longitudinal variation at two dates and **(b)** temporal variation at two longitudes of the ITCZ in GGG2020.
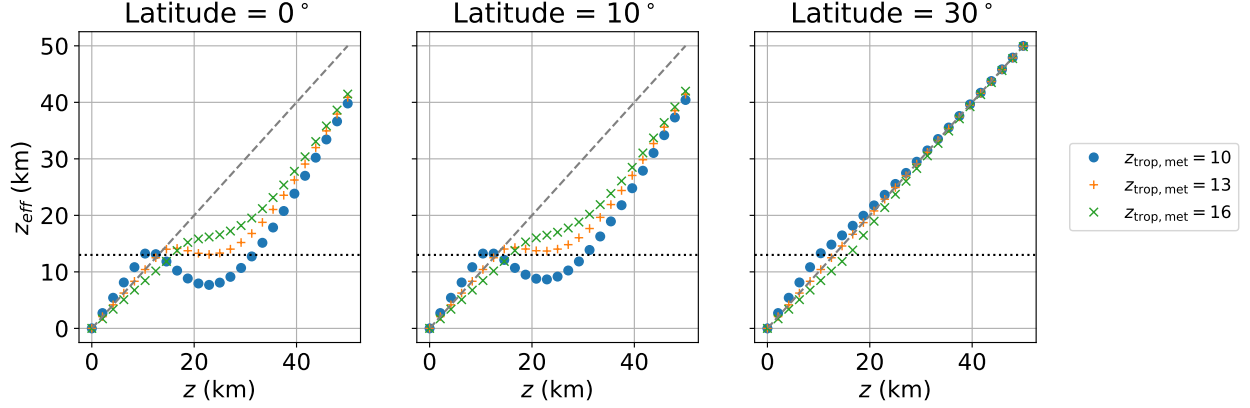
Figure S4: z_{eff} [from Eqs. (S1) and (S2)] vs. original altitude, assuming $z_{\text{trop,clim}} = 13\,\text{km}$, $l_{\text{itcz}} = 0°$, and $w_{\text{itcz}} = 10°$. Each panel demonstrates a profile for a different latitude, shown in the title, and within each panel, the different series represent profiles with different $z_{\text{trop,met}}$ values.

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

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

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

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).

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:

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

where V_{CO} is the vertical column of CO, $\mathbf{n}_{\text{CO}}$ 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 $\mathbf{n}^T \cdot \mathbf{v} = \int_{z_0}^{z_{\text{top}}} n(z)\, dz$.

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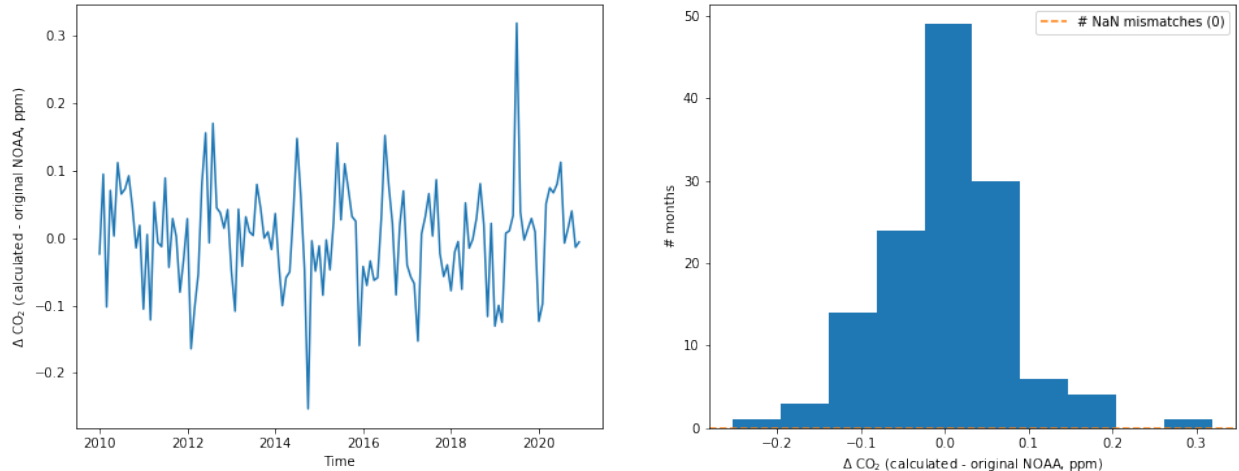
Figure S5: Timeseries (left) and histogram (right) of differences in monthly averages of CO_2 when computed by NOAA versus using the method described here. In the right panel, the "NaN mismatches (0)" indicates that no months had no data in one timeseries but not the other.

- 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).

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

Table S1

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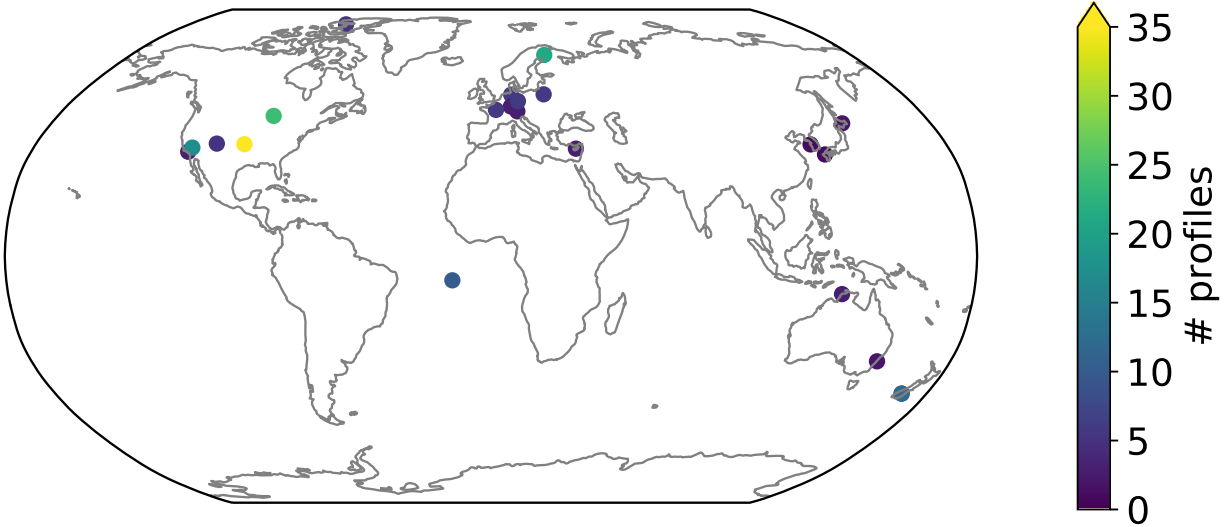
Figure S6: Locations of in situ profiles used to validate the TCCON priors, colored by number of profiles at that location. Note that if a single profile provided more than one gas it is only counted once.

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 GLOB-ALVIEWplus 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.
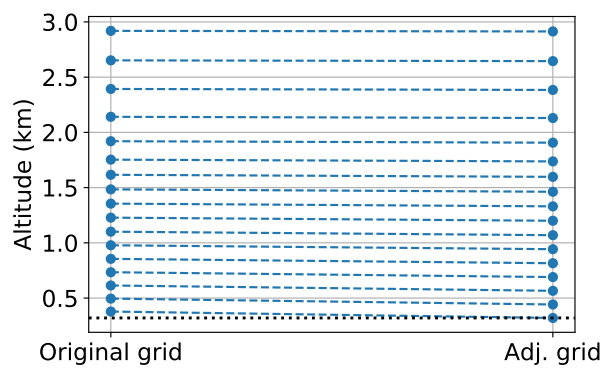
Figure S7: Example of original and adjusted z-grid for Lamont, OK (36.6° N, 97.49° W).
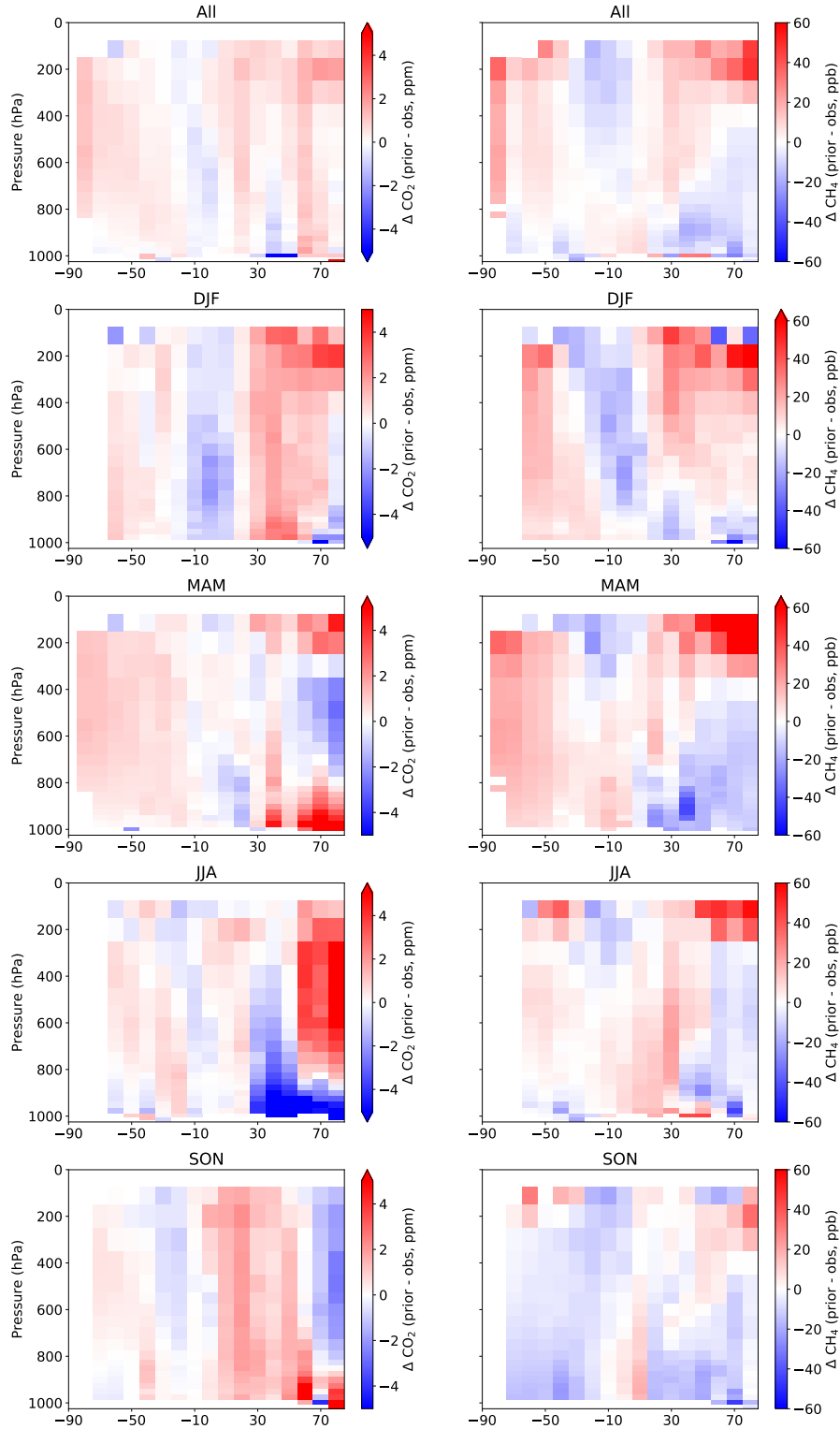
Figure S8: Curtain plots (longitudinal means plotted vs. latitude and pressure) of the difference between the priors and HIPPO + ATom observations. The left column is CO_2, the right column CH_4. The top row is the average over all seasons, the following four split the data into three month bins, indicated by the titles.
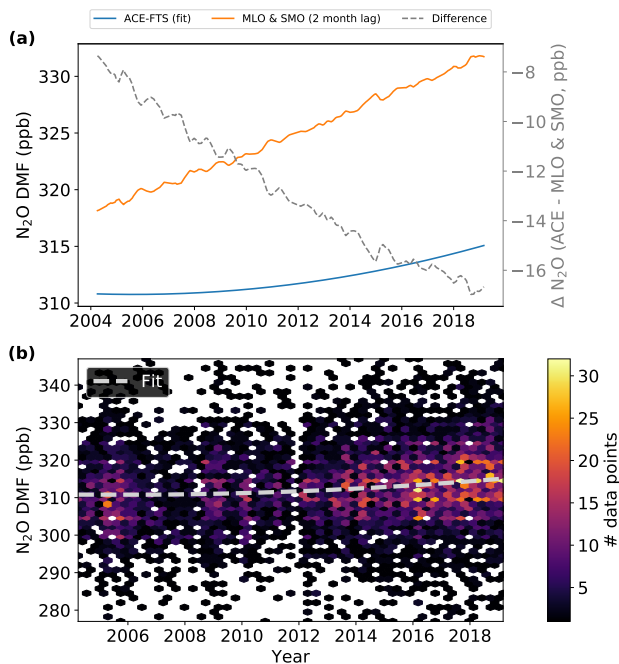
Figure S9: **(a)** Comparing time series of a quadratic fit to N$_2$O DMFs reported by the ACE-FTS satellite (version 3.6) against MLO & SMO mean N$_2$O DMFs lagged by two months. ACE-FTS data are only valid data points between latitudes 23°S to 23°N and potential temperatures 360 K and 390 K. These are both estimates of the N$_2$O 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$_2$O data shown in panel (a) plotted over a 2D histogram of the individual ACE-FTS N$_2$O data points.
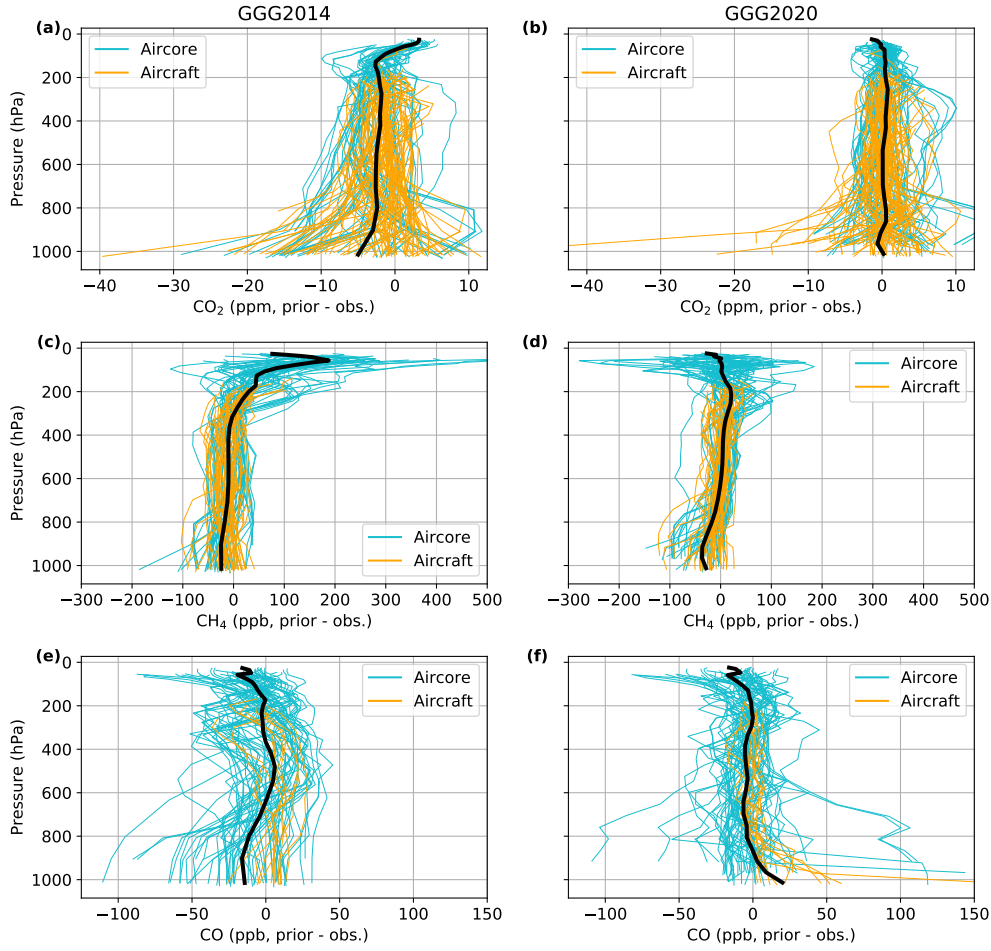
Figure S10: Spaghetti plots of CO_2, CH_4 and CO. The left column shows GGG2014 priors vs. observations, the right column GGG2020 priors vs. observations. As in the main paper, the thin lines are individual profiles' differences and the thick black line is the mean.
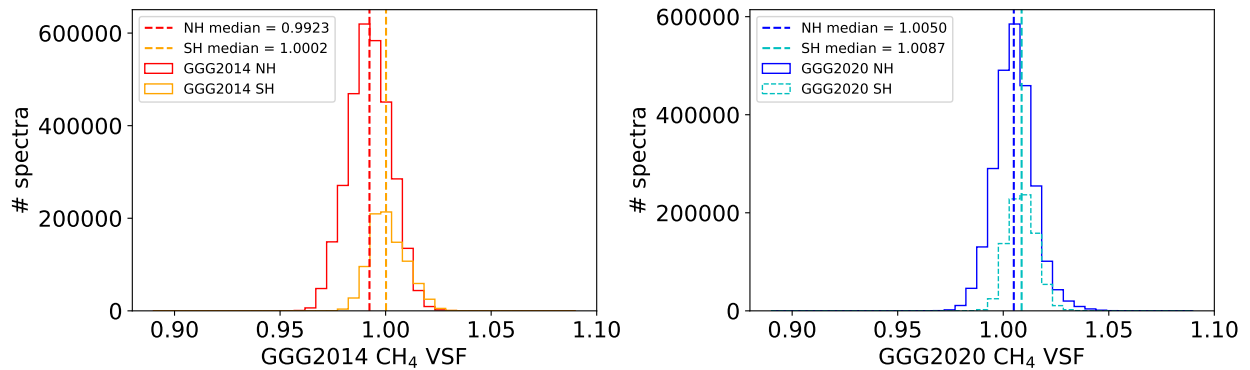
Figure S11: Histograms of volume scale factors (VSFs) for CH_4 for good quality GGG2014 (left) and GGG2020 (right) retrievals, divided into northern and southern hemisphere sites. Medians for each distribution are marked with the dashed vertical line of the same color.
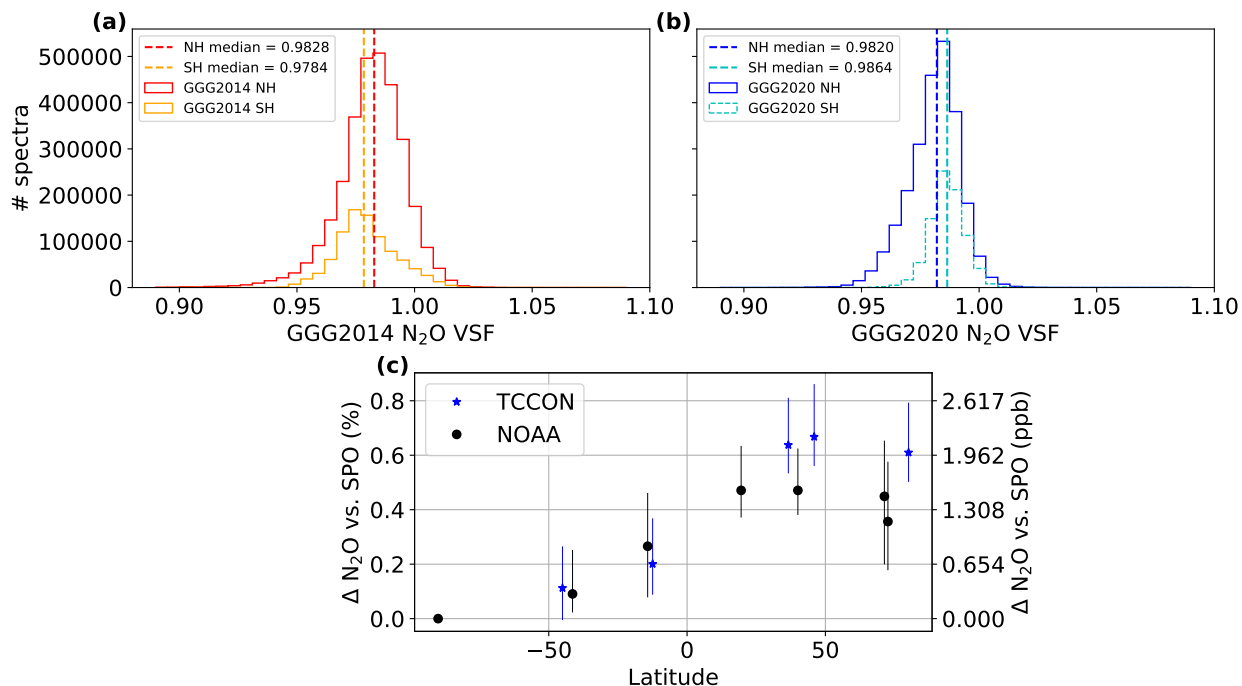
Figure S12: **(a)** and **(b)** are similar to the corresponding panels in Fig. S11, except for N_2O. **(c)** shows latitudinal gradients in surface N_2O from NOAA surface data and the TCCON priors, using the South Pole NOAA station as the baseline. The left axis gives the percent difference, the right 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.
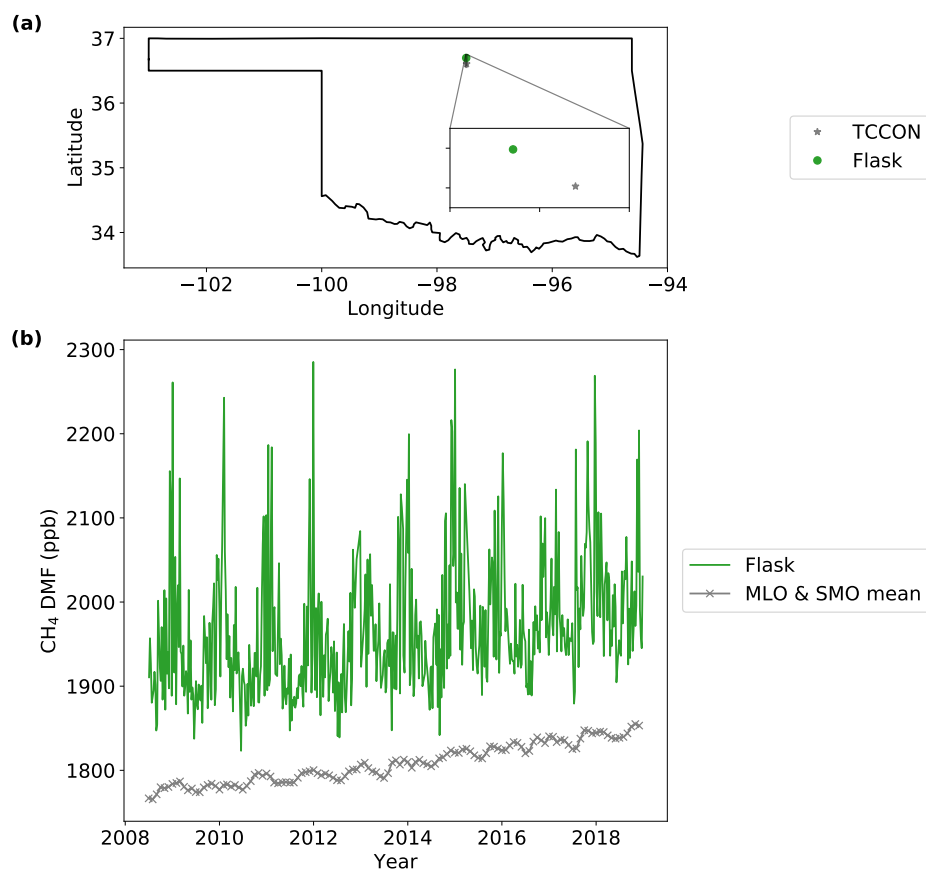
Figure S13: CH_4 enhancement near the Lamont TCCON site. **(a)** Location of the Lamont TCCON site and the Southern Great Plains (SGP) flask CH_4 measurements. **(b)** Timeseries of surface CH_4 dry mole fractions measured by the SGP flask and the MLO & SMO average mole fractions. The difference of 100 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).
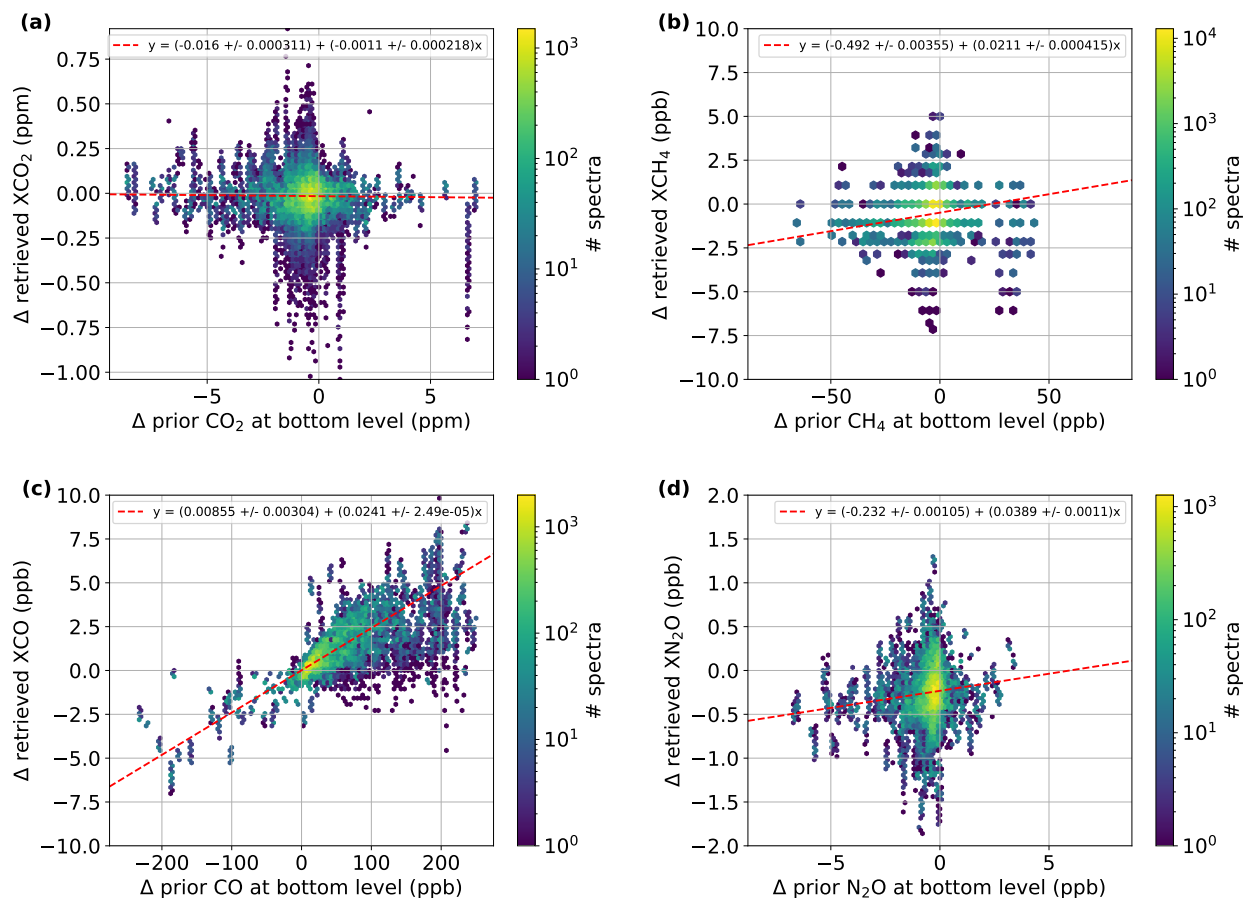
Figure S14: Plot of a sensitivity test at the Armstrong TCCON site using priors generated at the actual site latitude and longitude (34.96 N, 117.88 W) compared to priors generated for a location approximately 70 km NE (35.49 N, 117.51 W). The y-axis shows the change in retrieved Xgas and the x-axis 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.
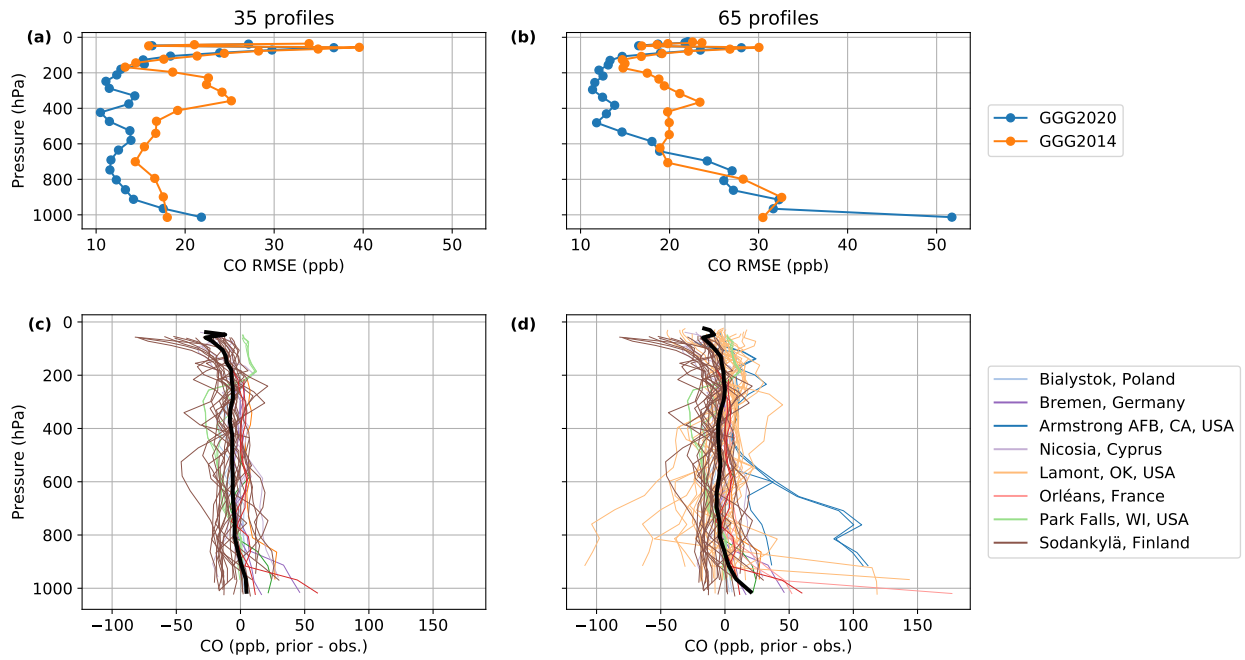
Figure S15: Plots of the RMSE of the previous GGG2014 and new GGG2020 CO priors vs. in situ profile (a, b) and spaghetti plots of the differences of the new GGG2020 CO priors vs. in situ profiles (c, d). Panels (a) and (c) exclude the Armstrong AFB, Lamont, and Orléans profiles as those are locations which seem to be affected by the incorrect CO emissions in GEOS FP-IT. Panels (b) and (d) include all sites for easy comparison.
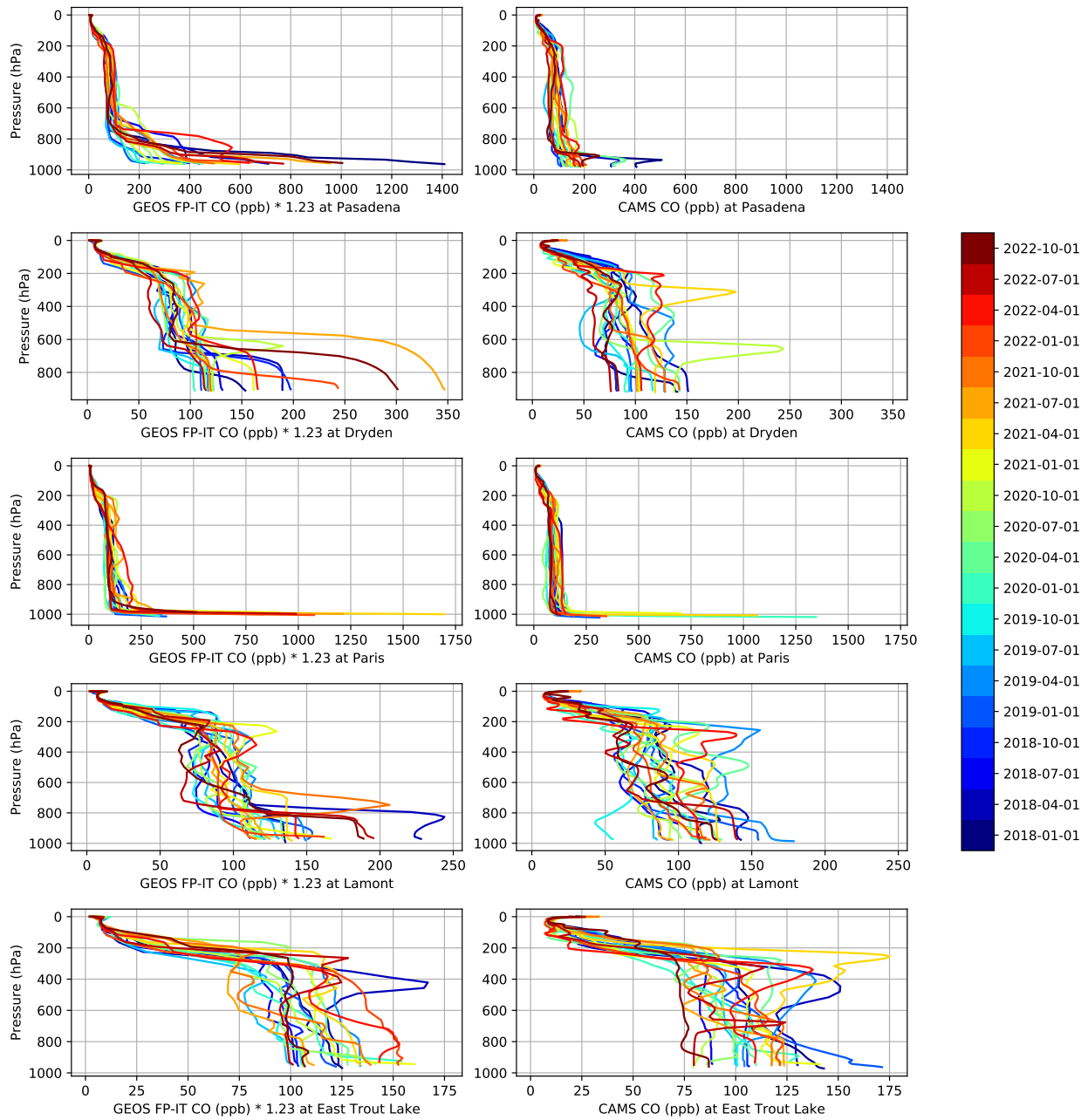
Figure S16: Comparison of CO profiles from GEOS FP-IT (left panels, scaled by 1.23 as described in Sect. 3.6 of the main paper) and the Copernicus Atmospheric Monitoring Service (CAMS) model (right panels). Each row contains profiles interpolated to the location of one TCCON site (note, "Dryden" is the same as "Armstrong AFB"). The color indicates the date of the profile; all profiles are for midnight UTC.
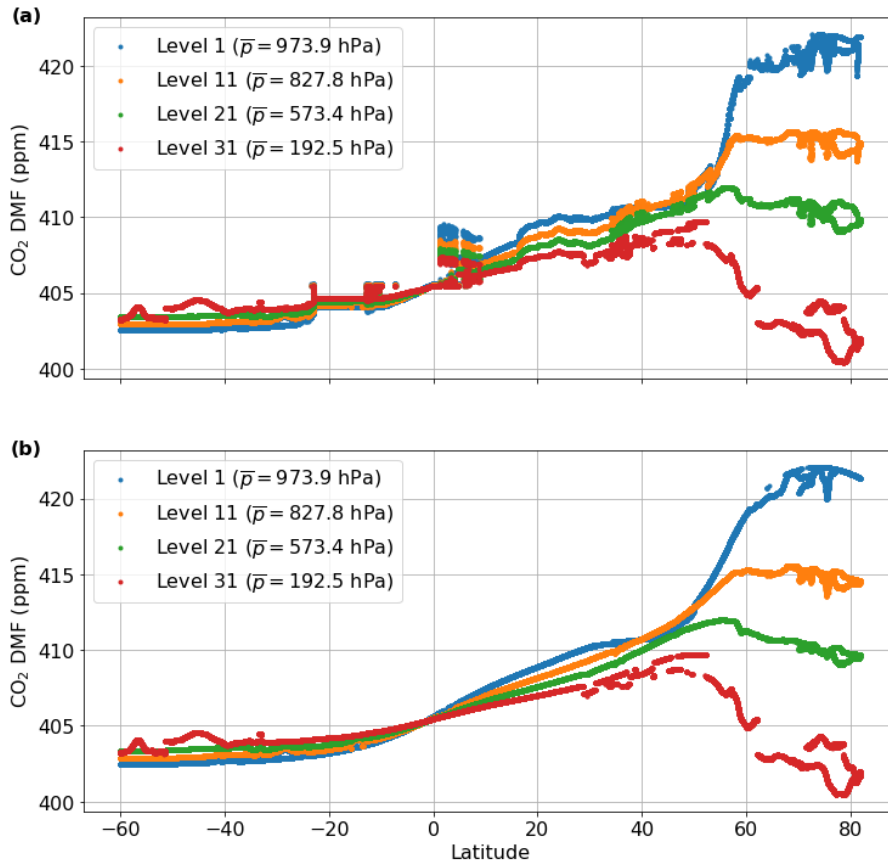
Figure S17: Effect of tropospheric effective latitude on prior CO_2 VMRs for four levels in the priors using an OCO-2 granule spanning 2017-05-14 times 18:22 UTC to 19:09 UTC. Panel **(a)** uses effective latitude, panel **(b)** uses geographic latitude. The legend gives the 1-based index of the vertical 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.