Assessment of errors and biases in retrievals of $X_{\text{CO}_2}$, $X_{\text{CH}_4}$, $X_{\text{CO}}$, and $X_{\text{N}_2\text{O}}$ from a 0.5 cm$^{-1}$ resolution solar-viewing spectrometer

Jacob K. Hedelius$^1$, Camille Viatte$^2$, Debra Wunch$^{2,a}$, Coleen M. Roehl$^2$, Geoffrey C. Toon$^{3,2}$, Jia Chen$^{4,b}$, Taylor Jones$^4$, Steven C. Wofsy$^4$, Jonathan E. Franklin$^5,c$, Harrison Parker$^6$, Manvendra K. Dubey$^6$, and Paul O. Wennberg$^2$

$^1$Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA
$^2$Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
$^3$Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
$^4$School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA
$^5$Department of Physics & Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada
$^6$Earth and Environmental Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA

$^a$now at: Department of Physics, University of Toronto, Toronto, Ontario, Canada
$^b$now at: Electrical and Computer Engineering, Technische Universität München, Munich, Germany
$^c$now at: School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

Correspondence to: Jacob K. Hedelius (jhedeliu@caltech.edu)

Received: 5 February 2016 – Published in Atmos. Meas. Tech. Discuss.: 4 March 2016
Revised: 24 June 2016 – Accepted: 28 June 2016 – Published: 3 August 2016

Abstract. Bruker™ EM27/SUN instruments are commercial mobile solar-viewing near-IR spectrometers. They show promise for expanding the global density of atmospheric column measurements of greenhouse gases and are being marketed for such applications. They have been shown to measure the same variations of atmospheric gases within a day as the high-resolution spectrometers of the Total Carbon Column Observing Network (TCCON). However, there is little known about the long-term precision and uncertainty budgets of EM27/SUN measurements. In this study, which includes a comparison of 186 measurement days spanning 11 months, we note that atmospheric variations of $X_{\text{gas}}$, within a single day are well captured by these low-resolution instruments, but over several months, the measurements drift noticeably. We present comparisons between EM27/SUN instruments and the TCCON using GGG as the retrieval algorithm. In addition, we perform several tests to evaluate the robustness of the performance and determine the largest sources of errors from these spectrometers. We include comparisons of $X_{\text{CO}_2}$, $X_{\text{CH}_4}$, $X_{\text{CO}}$, and $X_{\text{N}_2\text{O}}$ respectively, with 1σ running precisions of 0.08 and 0.06 % for $X_{\text{CO}_2}$ and $X_{\text{CH}_4}$ from measurements in Pasadena. We also identify significant error caused by nonlinear sensitivity when using an extended spectral range detector used to measure CO and N$_2$O.

1 Introduction

Measurements of atmospheric mixing ratios of greenhouse gases (GHGs), including CO$_2$ and CH$_4$, are needed to aid in estimating fluxes and flux changes, and to ensure international treaties to reduce emissions are fulfilled. The Total Carbon Column Observing Network (TCCON) makes daytime column measurements of these gases. The Orbiting Carbon Observatory2 (OCO-2) and Greenhouse Gases Observing Satellite (GOSAT) missions enable column GHG measurements with global coverage. These GHG monitoring satellites make measurements at one time of day and, therefore, lack the temporal resolution that a dedicated ground site provides.
Due to cost, lack of infrastructure, and stringent network requirements, there are limited ground sites on a global scale; e.g., there are no TCCON sites currently in operation in continental Africa, South America, or central Asia (Wunch et al., 2015), and there currently is no urban area with more than one TCCON site. Cheaper, portable, solar-viewing Fourier transform spectrometers (FTSs) can make contributions in these settings provided they have long-term precision. The Bruker Optics™ EM27/SUN, with the “SUN” indicating a built-in solar tracker, is a transportable FTS that may supplement global GHG measurements made by current networks (Gisi et al., 2012). This unit is small and stable enough to easily be transported for field campaign measurements, including measurements at multiple locations in 1 day. Column-averaged dry-air mole fractions (DMFs) of gases (Xgas) are retrieved from the EM27/SUN measurement, like the TCCON. Xgas is calculated from (Wunch et al., 2010):

\[
X_{\text{gas}} = \frac{\text{column}_{\text{gas}}}{\text{column}_{\text{dry air}}} = 0.2095X_{\text{gas}}^{\text{columnO}_2},
\]

where the 0.2095 factor is the fraction of dry air that is oxygen.

Retrieved Xgas has been compared with a co-located TCCON site in Karlsruhe, Germany, in past work for 26 days of XCO2 retrievals from one EM27/SUN instrument (Gisi et al., 2012), and 6 days of both XCO2 and XCH4 retrievals from five EM27/SUN instruments (Frey et al., 2015).

Operators of these instruments have different end goals to better understand the carbon cycle. XCO2 and XCH4 retrievals from these instruments have been compared with satellite measurements in areas without a TCCON site (Klappenbach et al., 2015) as well as with satellite measurements in highly polluted areas (Shiomi et al., 2015). Emission flux estimates from the Berlin area (<30 × 30 km2) were made by combining upwind/downwind measurements from five spectrometers and were compared with a simulation (Hase et al., 2015). Chen et al. (2016) have assessed gradient strengths around a large dairy farm (~100 000 cows) in Chino, California (<12 × 12 km2), using measurements from upwind/downwind spectrometers. Weather Research and Forecast Large-Eddy Simulations (WRF-LES, 4 km resolution) were used in combination with four simultaneous measurements to estimate fluxes from specific grid boxes in a subregion of the Chino dairy farm area, which is within a larger urban area (Viatte et al., 2016).

The column measurements used in these studies provide some advantages over in situ measurements, including less sensitivity to vertical exchange, surface dynamics, and small-scale emissions (McKain et al., 2012), which are difficult to model. Though column measurements can depend on mixed layer height in highly polluted areas, generally, column measurements depend primarily on regional-scale meteorology, and regional fluxes (Wunch et al., 2011b; McKain et al., 2012). For example, Lindenmaier et al. (2014) used observations from a single TCCON site to verify 1 day of emissions from coal power plants of about 2000 MW each at ~4 and 12 km away. Because of their large spatial sensitivity, column measurements are well suited for estimation of net emissions, model comparison, and satellite validation. A single site has been used to estimate Los Angeles, California (L.A.), emissions based on a sufficiently accurate emissions inventory and the observation that Xgas anomalies within L.A. are highly correlated (Wunch et al., 2009, 2016). Generally though, a single column measurement site is insufficient to estimate emissions from an entire urban region (Kort et al., 2013). However, multiple column measurements can be combined to characterize part or all of an urban area (Hase et al., 2015; Chen et al., 2016; Viatte et al., 2016).

The main goal of this work is to quantitatively evaluate the robustness of EM27/SUN retrievals over a long period of time. This is accomplished by comparing retrievals from the EM27/SUN with a co-located standard (TCCON site) at Caltech, in Pasadena, California, United States. TCCON spectrometers make the same type of measurements (direct solar near-infrared) at high spectral resolution. Here we report XCO2, XCH4, XCO, and XN2O comparison measurements from an EM27/SUN. The XCO and XN2O measurements were made possible by a detector with an extended spectral range provided by Bruker™. The EM27/SUN XCO2 and XCH4 to TCCON comparison is the longest to date. 186 measurement days spanning 11 months. In part of January 2015, an additional three EM27/SUN instruments were at Caltech for 9 to 12 days of XCO2 and XCH4 comparisons to assess their relative biases. In Sect. 2 we briefly describe differences in instruments and the data acquisition process. In Sect. 3 we describe the retrieval software. In Sect. 4 we describe the inherent properties of EM27/SUNs such as instrument line shapes (ILSs), frequency shifts, ghosts, detector linearity, and external mirror degradation. Section 5 focuses on biases and sounding precision of different gases compared with the TCCON. Section 6 describes sources of instrumental error. We conclude with general recommendations of tests to perform on any new type of direct solar near-infrared (IR) instrument used to retrieve abundances of atmospheric constituents.

2 Instrumentation

2.1 TCCON IFS 125HR

All TCCON sites employ the high-resolution Bruker Optics™ IR FT spectrometer (IFS) 125HR that has been described in detail elsewhere (Washenfelder et al., 2006; Wunch et al., 2011b). For the Caltech TCCON site (34.1362° N, 118.1269° W, 237 m a.s.l.), the IFS 125HR uses an extended InGaAs (indium gallium arsenide) detector, covering 3800–11 000 cm−1 for detection and retrieval of all gases relevant to this study (O2, CO2, CH4, CO, and N2O). Figure 1 has example spectra from IFS 125HR and
EM27/SUN instruments, with the spectral regions where individual gases are retrieved highlighted. Oxygen (O_2) abundance is useful in calculating the DMF because it represents the column of dry air and is combined with the column of the gas of interest to yield the DMF (Wunch et al., 2010).

The Caltech IFS 125HR uses a resolution of approximately 0.02 cm^{-1} (with a maximum optical path difference (MOPD) of 45 cm). It takes about 170 s to complete one forward/backward scan pair. TCCON sites have single sounding 2σ uncertainties of 0.8 ppm (X_{CO_2}), 7 ppb (X_{CH_4}), 4 ppb (X_{CO}), and 3 ppb (X_{N_2O}) (Wunch et al., 2010). TCCON data are tied to the World Meteorological Organization (WMO) in situ trace gas measurement scale through extensive comparisons with in situ profiles obtained from aircraft and balloon flights. We use the TCCON as a standard against which to compare the EM27/SUN instruments. TCCON data from this study are publicly available from the Carbon Dioxide Information Analysis Center (Wennberg et al., 2014).

### 2.2 Caltech EM27/SUN

EM27/SUN spectrometers have been described elsewhere (Gisi et al., 2012; Frey et al., 2015; Klappenbach et al., 2015) so we focus on differences in setup and acquisition here. The standard EM27/SUN configuration uses an InGaAs detector sensitive to the spectral range spanning 5000–12 000 cm^{-1}, which permits detection of O_2, CO_2, CH_4, and H_2O (Frey et al., 2015). For this study, the Caltech EM27/SUN was delivered with an extended-band InGaAs detector sensitive to 4000–12 000 cm^{-1}, which allowed for additional measurements of CO and N_2O (Fig. 1). All EM27/SUN spectrometers used in this study (Sects. 2.2, 2.3) used the typical MOPD of 1.8 cm, corresponding to a spectral resolution of 0.5 cm^{-1}. Interferograms (ifgs) were acquired in direct-current-coupled mode to allow post-acquisition low-pass filtering of brightness fluctuations to reduce the impact of variable aerosol and cloud cover effects (Keppel-Aleks et al., 2007). Ghosts were reduced as data were acquired by employing the interpolated sampling option provided by Bruker™ (see also Sect. 4.3). A 10 KHz laser fringe rate is used to reduce scanner velocity deviations, and each forward/backward scan took 11.6 s, or 5.8 s per individual measurement.

To be more consistent with the TCCON measurements, no spectrum averaging or interferogram apodization was applied before retrieving DMFs. We recommend averaging only after retrievals if disc storage and processor speeds are sufficient, so spurious data can be filtered. To test the pre- vs. post-averaging effect we used 9 retrieval days with 26 000 forward/backward measurements and used Bruker™ OPUS software to create spectra from ifgs. We compared retrievals from using five combined backward/forward measurements averaged pre-retrieval with those averaged post-retrieval. We also compared combined forward/backward measurements using a medium Norton–Beer apodization with those using no special apodization. Results are in Table 1 and suggest that different averaging methods cause only small inconsistencies, under ∼0.02 % for X_{CO_2} and X_{CH_4}.

The EM27/SUN was placed within 5 m of the Caltech TCCON solar tracker mirrors on the roof of the Linde+Robinson building (Hale, 1935). Measurements started on 2 June 2014 and, for this study, we include 186 measurement days that end on 4 May 2015. About 800 000 individual EM27/SUN measurements and 40 000 individual TCCON measurements were acquired over this period. Of these, about 580 000 and 15 000 were considered coincident and were not screened out by our quality control filters (QCFs). Our QCFs were conservative, and they required signal>30 (Sect. 4.4), solar zenith angle (SZA)<82°, 370 ppm<X_{CO_2}<430 ppm, X_{CO_2, error}<5 ppm, X_{CO_2, error}<20 ppb, and X_{CH_4, error}<0.1 ppb. Other users may consider stricter QCFs. After averaging data into 10 min bins, there were about 6500 binned comparison points.

### 2.3 LANL and Harvard EM27/SUN instruments

Three additional EM27/SUN instruments were compared with the Caltech TCCON site in January 2015 – one owned by Los Alamos National Laboratory (LANL) and two owned by Harvard University (HU). To be consistent, all the acquisition and retrieval settings were the same as for the Caltech EM27/SUN. As opposed to the Caltech EM27/SUN (also abbr. cn), the LANL (abbr. pl) and HU instruments (abbr. ha and hb) used the original InGaAs detector type sensitive over 5500–12 000 cm^{-1} (Frey et al., 2015). The LANL instrument, however, has a different high-pass filter, allowing it to measure up to 14 500 cm^{-1}. This different filter is neither beneficial nor disadvantageous to this instrument as no
gas column amounts are retrieved in that region. The LANL instrument was first used in January 2014 and has been compared with multiple TCCON sites in the United States, including sites at Four Corners, LANL, NASA Armstrong, La- mont, Park Falls, and multiple Caltech comparisons (Parker et al., 2015). The HU instruments have been operational since May 2014 and were compared against each other at Harvard before traveling over 4100 km to Caltech. As noted by Gisi et al. (2012) and Chen et al. (2016), the ILS of these instruments is remarkably stable considering the long distances they traveled.

### Table 1. Pre-averaging and apodization effects on EM27/SUN retrievals.

<table>
<thead>
<tr>
<th></th>
<th>X(_{\text{CO}_2})</th>
<th>X(_{\text{CH}_4})</th>
<th>X(_{\text{H}_2\text{O}})</th>
<th>X(_{\text{CO}})</th>
<th>X(_{\text{N}_2\text{O}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>% error</td>
<td>Md.</td>
<td>σ</td>
<td>Md.</td>
<td>σ</td>
<td>Md.</td>
</tr>
<tr>
<td>5 fwd/bwd pre-avgd.(^{a,b})</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>−0.02</td>
<td>0.01</td>
<td>0.36</td>
</tr>
<tr>
<td>NB med. apodz.(^{a,c})</td>
<td>0.29</td>
<td>0.09</td>
<td>−0.07</td>
<td>0.10</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Measurement compared over 1-10 July 2014. Md denotes the median. NB denotes the medium Norton–Beer apodization. \(^{a}\) As compared to retrievals from 1 fwd/bwd averaged non-apodized measurement averaged over same time post-retrieval. \(^{b}\) Same apodization as standard. \(^{c}\) Same pre-averaging as standard.

### 3 Retrieval software

SFIT (Pougatchev et al., 1995), PROFFIT (“PROFile fit”, Hase et al., 2004), and GGG (Wunch et al., 2015) are the three widely used retrieval algorithms to fit direct solar spectra and obtain column abundances of atmospheric gases. PROFFIT is maintained by the Karlsruhe Institute of Technology (KIT) and has been used to obtain DMFs from EM27/SUN instruments as well as NDACC-IRWG sites (Gisi et al., 2012; Frey et al., 2015; Hase et al., 2015). GGG is maintained by the Jet Propulsion Laboratory (JPL) and has been used to obtain DMFs from other low-resolution instrument measurements (e.g., an IFS 66, see Petri et al., 2012), in addition to being used to retrieve DMFs from the MkIV spectrometer in balloon-borne measurements (Toon, 1991) and for the Atmospheric Trace Molecule Spectroscopy Experiment (ATMOS) flown on the space shuttle (Irion et al., 2002). GGG is the retrieval algorithm used by the TCCON (Wunch et al., 2011b). We chose to use GGG for our analysis because (1) we want to be consistent with the TCCON for comparison and (2) the GGG software suite containing GFIT is open-source allowing us to adapt routines if needed. We used the GGG2014 version for retrievals (Wunch et al., 2015).

All retrievals used the same p\(_Tz\) and H\(_2\)O modeled profiles as well as the same a priori profiles (Wunch et al., 2015). We also used the same meteorological surface data for retrievals from all five instruments. All retrievals also used the same 0.2 hPa surface pressure offset. This offset was determined by comparing measurements from the standard barometer with a calibrated Paroscientific Inc. 765–16B Barometric Pressure Standard that has a stated accuracy of better than 0.1 hPa.

### 3.1 Interferogram-to-spectrum – double-sided

TCCON uses an interferogram-to-spectrum subroutine part of GGG to perform fast Fourier transforms (FFTs) to create spectra from ifgs (Wunch et al., 2015). Though the Bruker™ OPUS software used to operate the spectrometer can also perform FFTs, we again chose to use GGG to maintain consistency. A developmental version of GGG was used, which was adapted to also allow FFT processing on EM27/SUN interferograms. GGG splits a raw forward/backward ifg into two different double-sided ifgs which are then FFTed to yield two spectra. GGG also corrects source brightness fluctuations (Keppel-Aleks et al., 2007).

### 3.2 EM27/SUN GGG and interferogram processing suite (EGI)

To make GGG retrievals simpler for new EM27/SUN users, an add-in software suite (EGI) was developed at Caltech to create correctly formatted input files. This suite is opensource and can be obtained through correspondence to the email address listed. EGI can be run using MATLAB or Python. EGI runs in UNIX, Mac OS, and Linux environments and runs GGG on multiple processors. EGI centralizes settings for paths to read and write files, it coordinates separately acquired ground weather station and GPS data with EM27/SUN ifgs, and it optimizes processing order. It also provides some ancillary calculations such as a spectral signal-to-noise ratio (SNR) calculation. EGI provides a simple way to turn on and off saving of ancillary retrieval files (i.e., spectral fits and averaging kernels). EGI can run for instruments employing one or two detectors, such as the type described by Hase et al. (2016). Like the GGG software suite, EGI also includes benchmark spectra acquired under different conditions to run simple tests on. EGI is automated, reducing the learning time as well as the amount of user time needed to retrieve DMFs. After an initial setup, EGI will run from ifgs to retrieved X\(_{\text{gas}}\) with two commands. On a computer with 1400 MHz processors the code takes ~ 30 s per CPU to process each interferogram from the EM27/SUN extended InGaAs detector.
Figure 2. Frequency shifts (FS) of all four instruments vary with temperature because the lasers are not frequency-stabilized. FS for the CO$_2$ 6220 cm$^{-1}$ window are shown. FS of the Caltech (CIT) instrument are far from zero, so an empirical correction is made to correct the sample spacing number. Only every 300th CIT point and every 20th LANL point is plotted for clarity. HU EM27 1 and 2 are also referred to as ha and hb respectively by Chen et al. (2016).

4 Instrument characterizations and performance

4.1 Instrument line shape

Knowledge of the instrument line shape (ILS), or the observed shape of a spectral line from a monochromatic input, is crucial in assessing instrument performance and avoiding unknown biases in retrievals. Two parameters in the the LINEFIT algorithm (Hase et al., 1999) are used to characterize the ILS in relation to an ideal instrument, namely the modulation efficiency (ME) and phase error (PE). ME and PE both describe the interferogram and vary with OPD (Hase et al., 1999). PE has an ideal value of 0 radians, and indicates the degree of asymmetry in spectral lines. ME is a measure of the normalized observed interferogram signal compared with that of a nominal instrument with an ideal value of 1 (unitless) (Hase, 2012). At maximum OPD (MOPD), an ME < 1 causes a broadening of the measured spectral lines, while an ME > 1 at MOPD causes a narrowing. The ILS can be calculated by analyzing absorption lines measured through a low-pressure gas cell, and varies with OPD (Hase et al., 1999). Here, we use only single ME and PE values at the MOPD (Frey et al., 2015) to describe the ILS. We characterized the ILS for the EM27/SUN instruments using the method described elsewhere (Frey et al., 2015; Klappenbach et al., 2015). This method is able to characterize ME to within 0.15 % using the LINEFIT algorithm (Hase et al., 1999), with supplemental MATLAB scripts for automation purposes (Chen et al., 2016). ILS can affect retrieved column values. We note that the ME at MOPD of the cn and ha instruments in Table 2 are significantly lower than those reported by KIT on campus of ∼0.997 (Frey et al., 2015), and post-campaign of ∼0.996 (Klappenbach et al., 2015).

For this study, the ILS is used to help explain biases, to demonstrate the stability of the instruments, and gives insight into how well the EM27/SUN instruments are aligned and their optical aberrations. Though GGG2014 retrievals do not account for non-ideal ILS, future versions of GGG will. For the current study, we assume that ILS impacts using PROFFIT will be similar to impacts using GGG. This assumption will need to be tested when GGG also can account for a non-ideal ILS. Because future GGG retrievals will be revised using historical ILS measurements, a need remains to monitor the ILS both for future retrievals and as an indicator if realignment is necessary.

4.2 Frequency shifts

EM27/SUN units contain a HeNe 633 nm (15 798 cm$^{-1}$) metrology laser to sample the IR signal accurately as a function of the OPD. The laser is not frequency-stabilized (Gisi et al., 2012). This causes apparent spectral frequency to change with temperature as is shown in Fig. 2. Frequency shifts are affected by changes in the input laser wavenumber, laser alignment, and IR beam alignment. The input laser wavenumber will affect the spacing between spectral points. Since the frequency shift is furthest from zero for the Caltech EM27/SUN (on order of −100 ppm, in red Fig. 2), the spectral spacing is empirically corrected in the EGI suite based on the CO$_2$ 6220 cm$^{-1}$ window frequency shifts. This made little difference for the primary gases of interest affecting X$_{CO_2}$ by 0.015 % and X$_{CH_4}$ by −0.005 %, though it did affect X$_{H_2O}$ by 4 %.

4.3 Ghosts

Ghosts are artificial spectral features linked to the aliasing of true spectral lines that arise in FTS spectra (Learner et al., 1996). The InGaAs detectors are optically sensitive at wavenumbers greater than half the HeNe metrology laser frequency (7899 cm$^{-1}$). To fulfill the Nyquist criterion and
prevent aliasing, the IR interferogram is sampled twice each laser interferogram cycle, on the rising and falling edge. However, if the laser sampling is asymmetric – for example from a faulty electronics board – aliasing can still occur, folded across the half laser frequency (Messerschmidt et al., 2010). Because the asymmetry is typically small, the aliased signal, or ghost spectrum, is small compared with the true spectrum (Dohe et al., 2013; Wunch et al., 2015).

In EM27/SUN instruments the laser sampling error (LSE) can be minimized as data are collected by employing the interpolated sampling option provided by Bruker™. This resampling mode uses only the rising edge of the laser interferogram and assumes constant velocity in between the rising edges to interpolate the sampling (Gisi, 2014). We use a narrow band-pass filter (3 dB band width 5820–6150 cm\(^{-1}\)) in the Caltech EM27/SUN to test for LSE ghosts at 9800 cm\(^{-1}\). The ghost to parent ratio is 1.73 \(\times 10^{-4}\) at a 10 kHz acquisition rate without the interpolated sampling activated. This ghost is eliminated with the interpolated sampling turned on. In actual solar tests, turning the interpolated sampling on and off had no noticeable effect on the DMF retrievals for the Caltech EM27/SUN; however this may not hold true for all instruments. The LSE ghost also disappeared at an acquisition frequency of 20 kHz, and returned at higher acquisition frequencies. We opted for the recommended 10 kHz acquisition rate with the interpolated sampling on for all EM27/SUNs in this analysis because other instruments may be more significantly affected by LSE ghosts. A double-frequency ghost remains at \(\sim 11900\) cm\(^{-1}\) from radiation passing through the interferometer twice that is much larger than the LSE ghost, but is not in a region that will affect retrievals.

**4.4 Mirror degradation and detector linearity**

Solar tracking mirrors provided with the EM27/SUN instruments are gold with a protective coating. Gold is used because of its excellent reflectance in the near-IR and low reflectance in the visible region (Bennett and Ashley, 1965), which allows a high signal while reducing excess heating of the field stop and other optics. Through extended tests, we noted the first two mirrors (gold on plated aluminum, with a coating) degrade over time, with an e-folding degradation time of \(\sim 90\) days as is shown in Fig. 3. Arbitrary units (AUs) for signal are the maximum ordinate values of the unmodified interferograms multiplied by 6450. The AUs of signal happen to be close to the spectral SNR – a scaling factor of 1.3 applied to the arbitrary signal has an \(R^2\) of 0.63 relative to the SNR. Cleaning helped restore some signal, but never to the original values. The mirror change may not have restored full signal because the rest of the optics were not cleaned at the time of the mirror change. Below the blue 150 AU line in Fig. 3 the fitted \(O_2\) root mean square (rms) as a percentage of the continuum level dropped 26 times faster with signal intensity than above it. The instrument did come with an extra set of mirrors, but because mirrors are consumable parts, it adds recurring cost and effort to maintain these instruments long-term. After 1 year of use, the third mirror (gold coated glass) still remains completely intact. Feist et al. (2016) had success using steel mirrors under the very harsh conditions at the Ascension Island TCCON site, though at a cost of 35\% reflectivity per mirror. The JPL TCCON sites near Caltech noted no degradation on the external gold mirrors over more than 1 year of measurements. The lack of degradation on the third external mirror and the JPL TCCON mirrors is likely due to differences in how the mirrors were manufactured, including how the gold is applied to the substrate and the coatings used. Mirror degradation has likely not been a widely reported problem for most of the EM27/SUN community, perhaps because these instruments typically are stored indoors and only used for a few days for campaigns (for example, Frey et al., 2015). However, this problem may affect mirrors on other EM27/SUN instruments when mirrors are exposed outside for extended periods of time.

With signal loss, we would anticipate that gas measurements would become noisier but remain unbiased. However, with time, the Caltech EM27/SUN \(X_{CO_2}\) and \(X_{CH_4}\) DMFs decreased relative to the TCCON DMFs as mirror reflectance decreased, and \(X_{CO_2}\) and \(X_{CH_4}\) increased when the mirrors were replaced. The TCCON IFS 125HR InGaAs detectors are already known to be sufficiently linear that no correction is required (Wunch et al., 2011b). We also performed a simple test repeatedly adding mesh screens in front of the en-
Addition of band-pass filters or use of different detectors will be necessary to provide high-quality measurements of CO, CO$_2$, and CH$_4$ (Hase et al., 2016).

The data shown in Fig. 5 were divided into bins based on the signal intensity and were separated before and after the mirror change. Within each bin the relationship was treated as approximately linear. Fits using fewer than 10 points or with correlation coefficients less than 0.1 were discarded. The change with half signal was calculated. The analysis was repeated for 10 bins and again for 20 bins. The weighted mean change in X$_{CO_2}$ for halving the signal is $-1.43$ ppm in agreement with the mesh tests or

$$\Delta X_{CO_2} \left( ppm^{-1} \right) = 2.06 \ln \left( S/S_0 \right),$$

where S and $S_0$ are the final and initial signals respectively. This relationship holds for S and $S_0$ in the middle 80 %. For a similar methane analysis the mean change for half signal is $-7.25$ ppb or

$$\Delta X_{CH_4} \left( ppb^{-1} \right) = 10.5 \ln \left( S/S_0 \right).$$

5 Comparisons with $X_{gas}$

GGG2014 includes an air-mass-dependent correction factor derived for TCCON $X_{gas}$ measurements. The air mass correction factor for each gas is calculated using data obtained at a variety of relatively clean sites as described by Wunch et al. (2011b). We expect that the air mass dependence, which is due primarily to spectroscopic uncertainties, should be common for the same type of measurement. Parker et al. (2015) noted that the average EM27/SUN factors are similar compared to the TCCON for $X_{CO_2}$ at three clean sites in the United States. The $X_{CH_4}$ $\beta$ factor was different ($-0.0077$ EM27/SUN; 0.0053 TCCON) but when applied here it worsened the $R^2$ and standard deviation of the comparisons. This could be because the air mass dependence of $X_{CH_4}$ may not be solely from spectroscopic issues. Hence, we used the same air-mass-dependent correction factors as the TCCON.

Detector nonlinearity in FTS instruments can be corrected in the ifgs post-acquisition in two ways. The first option deals with artifacts around the ZPD (zero path difference point) and is already included in GGG (Keppel-Aleks et al., 2007). When the ifg is smoothed, a nonlinear detector exhibits a dip around the ZPD which can be used to diagnose and reduce detector nonlinearity effects. EM27/SUN measurements are too noisy to properly characterize or detect this dip and so this correction is insufficient. The other option is to compare detector response with radiance from a controlled external light source, such as a blackbody, with very accurate radiation flux measurements (on order of 0.01 %) (Thompson and Chen, 1994). By characterizing the response to the true flux as it is varied, the detector can be characterized and ifgs can be appropriately scaled and corrected. However, this requires extremely controlled precise measurements, as all nonlinearity is likely less than 1 %, so measurements must be more precise than 1 %.

An option to prevent nonlinearity from interfering with measurements is to only use the detector over its linear range by sufficiently attenuating the incoming sunlight. However, the SNR is already low so we opted against this method. Ultimately, we purchased the non-extended InGaAs detector at the loss of CO, and N$_2$O for future measurements for the Caltech instrument. For the historical field measurements we use a bias correction to match the TCCON for the nearest comparison days. The nonlinearity has nearly an equal effect for short times, but has a larger variation on multi-monthly scales as the mirrors degrade. In future measurements we recommend against using these extended InGaAs detectors. The change with half signal was calculated. The analysis was repeated for 10 bins and again for 20 bins. The weighted mean change in X$_{CO_2}$ for halving the signal is $-1.43$ ppm in agreement with the mesh tests or

$$\Delta X_{CO_2} \left( ppm^{-1} \right) = 2.06 \ln \left( S/S_0 \right),$$

where S and $S_0$ are the final and initial signals respectively. This relationship holds for S and $S_0$ in the middle 80 %. For a similar methane analysis the mean change for half signal is $-7.25$ ppb or

$$\Delta X_{CH_4} \left( ppb^{-1} \right) = 10.5 \ln \left( S/S_0 \right).$$

Figure 4. $X_{CO_2}$ retrievals on 11 October 2014 when mesh screens were repeatedly moved in front of and away from the EM27/SUN (with extended InGaAs detector) entrance window. Gray points are all EM27/SUN measurements. Large points are 10 min averages. Error bars are 1σ. This test was performed a few days after the mirrors were replaced.
To compare measurements between the TCCON and the EM27/SUN instruments, data were first averaged into 10 min bins to reduce the variance of binned differences (Chen et al., 2016). The median of the X\textsubscript{CO\textsubscript{2}} differences between sequential time bins is smallest (around 0.26 ppm) for 10 min bins over the entire ~11 month time period. Less averaging is more affected by noise, and more averaging starts to include instrument drift and true atmospheric variations. Averages were weighted using retrieval errors \(\hat{x}_{\text{err}}\) as in Eq. (4):

\[
\bar{x} = \frac{\sum_i \hat{x}_{i,\text{err}}^{-2} \hat{x}_i}{\sum_i \hat{x}_{i,\text{err}}^{-2}}, \tag{4}
\]

where \(\hat{x}_i\) is the retrieved value from the \(i\)th measurement in a bin, and \(\bar{x}\) is the bin average.

### 5.1 Averaging kernels

When comparing retrieved \(X_{\text{gas}}\) measurements (also denoted \(\hat{c}\)) from different remote sensing instruments, differences in their averaging kernels (AKs or \(a_i\), where \(i\) represents an instrument indicator number) and a priori profiles must be taken into account, using for example, the methods described by Rodgers and Connor (2003). Wunch et al. (2011a) compared GOSAT and TCCON total column DMFs using this method. Because GGG scales a priori profiles rather than retrieving the full profile, these AKs are vectors (i.e., column averaging kernels) rather than matrices.

Averaging kernels depend on several factors including how strong the lines are in the retrieval windows, and viewing geometry (e.g., SZA for solar-viewing instruments). Because the TCCON IFS 125HR and EM27/SUN instruments have different spectral resolutions, the apparent absorption strengths are different and so are the averaging kernels. Averaging kernels for a gas differ for each microwindow. We combined AKs of a given gas from different microwindows using an unweighted average. Averaging kernels for the Caltech EM27/SUN for the GGG retrieval windows are shown in Fig. 6. Averaging kernels from the other EM27/SUN instruments are similar. TCCON averaging kernels have been discussed by Wunch et al. (2011b) and are shown on the bottom row in Fig. 6. As a numerical example, for \(X_{\text{CO\textsubscript{2}}}\) measured at 50° SZA and 900 hPa using GGG, the AK is 1.10 for EM27/SUN instruments and 0.93 for TCCON instruments. This means EM27/SUN instruments are slightly more sensitive to a change in \(\text{CO\textsubscript{2}}\) near the surface relative to TCCON instruments. More importantly, they have the opposite sensitivity to an error in the a priori volume mixing ratio (VMR) profile at 900 hPa.

In our particular case, reducing the smoothing error using Eq. (A13) from Wunch (2011a) and using the a priori as the comparison ensemble changes little as the effect of the differences in averaging kernels from the top of the atmosphere tends to cancel out the effect of differences at the bottom. TCCON and EM27/SUN a priori profiles were the same in this comparison. However, we need to consider that the a priori profiles used in the retrieval are not representative of a highly polluted place, such as Pasadena, which is located in the same air basin as Los Angeles. Because differences in column measurements compared to background or a priori profiles occur primarily because of differences at the surface we can adjust retrievals for one instrument taking into account this knowledge using

\[
\hat{c}_1 = \frac{a_{1,a}}{a_{2,a}} \left[\hat{c}_2 - c_a\right] + c_a. \tag{5}
\]

Definitions of the terms in, as well as a discussion of assumptions needed to obtain Eq. (5) are in Appendix A. We applied Eq. (5) to the \(X_{\text{CO\textsubscript{2}}}\) and \(X_{\text{CH\textsubscript{4}}}\) retrievals.

In summary, to compare biases between two instruments, we account for diurnal dependences, then average data into comparable time bins, and take into account our prior knowl-
edge of the atmospheric profile and differences in averaging kernels.

5.2 Full comparisons of \( X_{\text{gas}} \) from extended-band InGaAs detector with a TCCON site

Gisi et al. (2012) noted that measurements taken within the first 30 min of moving the instrument to the roof and turning it on needed to be filtered out because of high scatter while waiting for the instrument to operate stably. We did not observe a similar requirement for our data. This could be because our instruments were not subjected to such fast temperature changes. It could also be because the laser frequency shift, which changes with temperature, does not seem to significantly impact our retrievals.

Examples of spectral fits from several of the retrieval windows are shown in Fig. 7 for a single spectrum. These are not necessarily representative of all the conditions under which the 800,000 spectra were acquired. The residuals are larger than those reported by Gisi et al. (2012) and Frey et al. (2015) because of the lower SNR from spectra recorded using the extended InGaAs detector.

The full time series (186 days) of the difference between the Caltech EM27/SUN and TCCON measurements is shown in Fig. 8. From this figure we see that \( X_{\text{CO}_2} \) and \( X_{\text{CH}_4} \) are the gases most affected by the mirror change in October 2014 (by about 3 ppm and 12 ppb respectively). For all gases, scatter of retrieved \( X_{\text{gas}} \) increases as signal decreases.

Figure 9 shows the retrieved \( X_{\text{CO}_2} \) and \( X_{\text{CH}_4} \) from all four EM27/SUN instruments for 9–12 days in January 2015 plotted against those from TCCON. We report biases for January 2015 as scaling factors to approximate to the TCCON, or scaling factors compared to 1. Biases were calculated using a linear least squares fit forced through the origin. A summary of the biases for all gases as compared to the TCCON is provided in Table 3.

5.3 \( X_{\text{CO}_2} \)

We note a smaller bias in \( X_{\text{CO}_2} \) with respect to the TCCON (+0.03 %, see Table 3) compared to previous EM27/SUN studies (Gisi et al., 2012; Frey et al., 2015; Klappenbach et al., 2015). These previous studies retrieved \( X_{\text{gas}} \) from EM27/SUN spectra using PROFFIT. When compared with the TCCON \( X_{\text{CO}_2} \) retrievals, Gisi et al. (2012) noted a +0.12 % bias, Frey et al. (2015) noted a +0.49 % bias, and Klappenbach et al. (2015) noted a +0.43 % bias. Reasons for these differences could be from (1) spectroscopy differences between PROFFIT and GGG2014 used for EM27/SUN \( X_{\text{gas}} \) retrievals, (2) because Gisi et al. (2012) used an earlier version of GGG for TCCON retrievals, and (3) because Frey et al. (2015) and Klappenbach et al. (2015) applied empirical corrections before comparing with the TCCON. In this
Table 3. EM27/SUN to Caltech TCCON biases.

<table>
<thead>
<tr>
<th></th>
<th>Caltech, January</th>
<th>LANL, January</th>
<th>Harvard 1</th>
<th>Harvard 2</th>
<th>weighted % bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>285</td>
<td>241</td>
<td>187</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>n days</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(X_{\text{CO}_2})</td>
<td>0.9999 (0.16)</td>
<td>1.0006 (0.14)</td>
<td>1.0009 (0.15)</td>
<td>0.9998 (0.15)</td>
<td>0.03</td>
</tr>
<tr>
<td>(X_{\text{CH}_4})</td>
<td>1.0069 (0.19)</td>
<td>1.0066 (0.20)</td>
<td>1.0103 (0.14)</td>
<td>1.0066 (0.14)</td>
<td>0.75</td>
</tr>
<tr>
<td>(X_{\text{H}_2\text{O}})</td>
<td>0.9840 (1.27)</td>
<td>0.9791 (1.44)</td>
<td>0.9886 (1.12)</td>
<td>0.9791 (1.01)</td>
<td>−1.73</td>
</tr>
<tr>
<td>(X_{\text{CO}})</td>
<td>0.9988 (2.30)</td>
<td></td>
<td></td>
<td></td>
<td>−0.12</td>
</tr>
<tr>
<td>(X_{\text{N}_2\text{O}})</td>
<td>1.0243 (0.42)</td>
<td></td>
<td></td>
<td></td>
<td>2.43</td>
</tr>
</tbody>
</table>

Italicized values in parentheses are percent standard deviations as compared to the TCCON over the dataset for January 2015.

Figure 8. Full time series of EM27/SUN measurements as compared to TCCON from June 2014 to May 2015. Thin vertical gray lines represent mirror cleaning. The thick line represents the mirror change. To the right are TCCON means over time to get a sense of percent deviations.

section, we investigate two possible causes of bias: spectral resolution and instrument line shape.

Following Gisi et al. (2012), we attempted to determine whether the cause of the bias is due to the difference in spectral resolution between the EM27/SUN and TCCON instruments. Petri et al. (2012) also considered resolution bias in their study using a 0.11 cm\(^{-1}\) resolution instrument and an older version of GGG. They did not report a bias in \(X_{\text{CO}_2}\) retrievals, but noted that \(X_{\text{CO}_2}\) decreased by \(\sim 0.12\%\) as interferograms were truncated to obtain spectra with resolutions of 0.02 to 0.5 cm\(^{-1}\). Most of the change occurred as the resolution changed from 0.1 to 0.5 cm\(^{-1}\) (see Fig. 11 therein). In contrast, Gisi et al. (2012) noted a 0.13 % increase in \(X_{\text{CO}_2}\) as the resolution changed from 0.02 to 0.5 cm\(^{-1}\) in PROFFIT. Here we find a 0.08 % ± 0.16 % (1\(\sigma\)) decrease in \(X_{\text{CO}_2}\) when the resolution is decreased from 0.02 to 0.49 cm\(^{-1}\) in GGG, though part of this change would be offset by considering the differences in averaging kernels.

Previous studies noted an increase in \(X_{\text{CO}_2}\) of 0.15 % for a 1 % increase in modulation efficiency at max OPD (Gisi et al., 2012; Frey et al., 2015). Using PROFFIT we performed a similar test for spectra taken under various conditions at various times of day and obtained a similar result of a 0.10 % ± 0.02 % (1\(\sigma\)) increase in \(X_{\text{CO}_2}\) for a 1 % increase in ME at the MOPD. For this study we assume that impacts of the ILS on retrievals will be similar in GGG and PROFFIT. Though we report a single value, there is an air mass dependence of \(\sim 0.05\%\) increase in EM27/SUN PROFFIT retrievals for a 1 % increase in ME and air mass change of 1.

For instruments using the standard InGaAs detectors, the \(X_{\text{CO}_2}\) 10 min running 1\(\sigma\) precision is 0.075 % [0.034 to 0.18 %, 95 % CI]. The wide confidence interval (CI) is from a combination of atmospheric variability being aliased into the running standard deviation as well as different SNRs among instruments. The spectral SNRs for measurements using this detector were in the range 1000–5000 and their pre-
An additional study we have not performed that could help in reducing bias would be to omit all or part of a CO₂ window with strong water lines. Because of the low resolution of these spectrometers (see inset Fig. 1), water lines and CO₂ lines often overlap. This can lead to inaccurate retrievals despite a good overall fit because H₂O and CO₂ can both be wrong, but in compensating ways. Reducing the size of a window would reduce precision but would decrease water and temperature sensitivity. This adjustment could also be performed for CH₄, which is retrieved over three windows in GGG.

### 5.4 XCH₄

The EM27/SUN XCH₄ retrievals are 0.75 % higher than those of TCCON (see Table 3). In previous work, high biases of 0.47 % for a 0.11 cm⁻¹ instrument (Petri et al., 2012), and 0.49 % (Frey et al., 2015) and 1.87 % (Klappenbach et al., 2015) for EM27/SUNs, were noted. Petri et al. (2012) attributed most (0.26 %) of their bias to differences in resolution and noted for a single day that the bias increased as resolution decreased. In our simulations we find a 0.28 % ± 0.20 % (1σ) increase in XCH₄ when the resolution is reduced from 0.02 to 0.49 cm⁻¹. Using PROFFIT the impact of a 1 % decrease in ME is a 0.15 % ± 0.01 % (1σ) increase in XCH₄. Again, although we report a single value there is an air mass dependence of about a 0.12 % decrease in XCH₄ using PROFFIT retrievals for an air mass change of 1, and a 1 % decrease in ME. Resolution and ME combined account for only half of the observed methane bias. Petri et al. (2012) suggested improper dry air mixing ratio and pT profiles, or spectroscopy as sources of error. Improper surface pressure, error in the calculated Observer-Sun Doppler Stretch (OSDS) due to pointing errors coupled with solar rotation, or error in the assumed field of view (FOV) may also contribute to the bias (see Sect. 6).

Chen et al. (2016) found that the 1σ XCH₄ precision among 10 min binned EM27/SUNa-EM27/SUNb differences is 0.01 %. These data were acquired in a way that about 67 spectra were acquired every 10 min, and because two instruments were used, the single sounding precision is \( \sim 0.01 \% \times \sqrt{67/2} \approx 0.058 \% \), which falls in our measured running 1σ precision range. Comparing to the TCCON, Gisi et al. (2012) reported that the 1σ daily precision is 0.08 %. The extended InGaAs detector naturally has a lower spectral SNR, in the range 100–1000, with a median of 400 over the full time series. Most of the variation in the SNR is due to loss of mirror reflectivity, but even with non-degraded gold mirrors, it is \( \sim 5 \) times lower because of the different detector. The median running 1σ precision over the full time series is 0.26 % for the XCO₂ product from the extended InGaAs detector. Because the SNR changed with time due to loss of mirror reflectivity, so did the precision. The correlation between \( 1/\sqrt{\text{SNR}} \) and running 1σ XCO₂ precision was strong \( (R^2 = 0.75) \) for retrievals from this detector and followed

\[
\sigma_{\text{XCO}_2} = 0.17 + \frac{8.4}{\sqrt{\text{SNR}} - 57},
\]

### 5.5 XCO and XN₂O

XN₂O and XCO were also measured using an EM27/SUN spectrometer in this study. Hase et al. (2016) have also reported on XCO measurements using an EM27/SUN modified to include a second InGaAs detector with optical filters. Column CO measurements are desirable because CO is a tracer of combustion. Here these measurements were made possible...
because the extended detector is sensitive to the region 4200–4800 cm\(^{-1}\), which contains useful windows where \(\text{N}_2\text{O}\) and \(\text{CO}\) molecules absorb IR radiation. Both the \(\text{X}_{\text{CO}}\) and \(\text{X}_{\text{N}_2\text{O}}\) retrievals are highly sensitive to changes in the modeled temperature profile. The nonlinearity of the detector had a less pronounced effect on \(\text{X}_{\text{CO}}\) and \(\text{X}_{\text{N}_2\text{O}}\) retrievals than it had on \(\text{X}_{\text{CH}_4}\) and \(\text{X}_{\text{CH}_2\text{Cl}}\) retrievals (Fig. 8). \(\text{X}_{\text{CO}}\) and \(\text{X}_{\text{N}_2\text{O}}\) also have poorer precision than \(\text{X}_{\text{CO}_2}\) and \(\text{X}_{\text{CH}_4}\), so any nonlinearity effect could be less than the noise. The 4200–4800 cm\(^{-1}\) spectral region is also affected differently from the nonlinearity than the 5000–7000 cm\(^{-1}\) region where column \(\text{CH}_4\) and \(\text{CO}_2\) are retrieved; the continuum levels changed more for the latter region. This may also explain in part why there is no noticeable change in \(\text{X}_{\text{CO}}\) and \(\text{X}_{\text{N}_2\text{O}}\) with signal. For \(\text{X}_{\text{CO}}\) the median 1\(\sigma\) precision is 3.7\%. In our simulations reducing the spectral resolution from the TCCON (0.02 cm\(^{-1}\)) to near the EM27/SUN (\(\sim 0.5\) cm\(^{-1}\)) decreased \(\text{X}_{\text{N}_2\text{O}}\) by 1.5\% \(\pm 0.6\%\) (1\(\sigma\)) in low-resolution spectra, and at Caltech this change varies with time.

In general, as is seen in Fig. 8, \(\text{X}_{\text{N}_2\text{O}}\) retrievals were highly scattered and had a large offset from TCCON. In our simulations, reducing the resolution from TCCON (0.02 cm\(^{-1}\)) to EM27/SUN (0.5 cm\(^{-1}\)) decreased \(\text{X}_{\text{N}_2\text{O}}\) by 1.5\% \(\pm 0.6\%\) (1\(\sigma\)). Retrievals from the 4430 cm\(^{-1}\) window were low (\(\sim 6\)\%), while the 4719 and 4395 cm\(^{-1}\) regions were biased slightly high (\(\sim 1\)\%). The retrievals from the 4719 cm\(^{-1}\) region additionally had some long-term trends for reasons we do not understand. For \(\text{X}_{\text{N}_2\text{O}}\) the median 1\(\sigma\) precision is 1.9\%.

### 5.6 \(\text{X}_{\text{H}_2\text{O}}\)

Because of the significantly lower spectral resolution of the EM27/SUN spectrometers, the spectral band widths for the \(\text{H}_2\text{O}\) retrievals were increased as compared to the standard TCCON approach (Wunch et al., 2010). For lower resolution spectra, the \(\text{H}_2\text{O}\) lines appear much broader and the observed transmittance is much lower at the edges of standard TCCON spectral window. Thus, the spectral ranges of the low-resolution windows were expanded. Some of the standard TCCON windows used to retrieve \(\text{H}_2\text{O}\) had too few spectral points from the low-resolution instrument for good fits and were omitted. When expanding the windows, we ensured that no lines were admitted that made the effective ground-state energy \(E''\) greater than \(\sim 400\) cm\(^{-1}\). This reduces the temperature sensitivity to the modeled temperature profiles. As with the TCCON windows, we tried to keep a wide range of \(\text{H}_2\text{O}\) line strengths to accommodate large seasonal and site-to-site variations of the \(\text{H}_2\text{O}\) column. Windows were kept as wide as possible without encountering large spectral fitting residuals.

For \(\text{X}_{\text{H}_2\text{O}}\), we find a median 1\(\sigma\) precision of 1.9\% from the instrument using the extended InGaAs detector. For instruments using the standard InGaAs detectors, the \(\text{X}_{\text{H}_2\text{O}}\) 1\(\sigma\) precision is 0.81\% [0.36 to 2.12\%, 95\% CI].

6 Sensitivity tests on retrievals

As with the TCCON, EM27/SUN retrievals require modeled atmospheric pressure, temperature, altitude (pTz), and water profiles (Wunch et al., 2015). Here atmospheric profiles are generated from the NCEP/NCAR 2.5\(^{\circ}\) reanalysis product (Kalnay et al., 1996) by interpolating to the correct location at local noon of the desired day. These profiles also include the tropopause height which is used to vertically shift a priori profiles, as tropopause height can significantly affect column DMFs such as \(\text{X}_{\text{CH}_4}\) and \(\text{X}_{\text{HF}}\) (Saad et al., 2014). Selecting a profile for an incorrect location or day could lead to errors.

We ran test retrievals for the July 2014 period with incorrect profile information derived separately at latitudes north (1, 2, and 5\(^{\circ}\)) and longitudes west (1, 2, and 5\(^{\circ}\)) of our observation site, and well as from profiles derived 1, 5, 10, and 100 days prior to the measurement dates. In general, the profiles generated from a more distant location in space and time caused larger retrieval errors. For \(\text{X}_{\text{CH}_4}\) and \(\text{X}_{\text{CO}}\), the main variability from the standard retrievals was in daily offsets (standard deviation of daily medians \(\sigma\) (Md\(\sigma\) daily)) which had values of 3 and 4 ppb respectively for the 100 day prior model. The medians of daily standard deviations Md\(\sigma\) daily were 0.5 ppb for both \(\text{X}_{\text{CH}_4}\) and \(\text{X}_{\text{CO}}\) for the 100 day prior model. \(\text{X}_{\text{N}_2\text{O}}\) and \(\text{X}_{\text{H}_2\text{O}}\) also had more errors from \(\sigma\) (Md\(\sigma\) daily), except for profiles within 2\(^{\circ}\), which more strongly affected diurnal variability Md\(\sigma\) daily). For these two species, the 100 day prior model \(\sigma\) (Md\(\sigma\) daily) were 2 ppb and 50 ppm and Md\(\sigma\) daily were 1 ppb and 20 ppm respectively. These values are shown for \(\text{X}_{\text{CO}_2}\) in Fig. 10 for all tested models. The 100 day prior model had \(\sigma\) (Md\(\sigma\) daily) = 0.16 ppm and Md\(\sigma\) daily = 0.4 ppm, as well as a 1.2 ppm bias when using these models for \(\text{X}_{\text{CO}_2}\).
Various user, instrumental, and measurement errors can reduce the accuracy and precision of retrievals. GGG uses retrieved O\textsubscript{2} column amount with the average DMF of O\textsubscript{2} (0.2095) to calculate the dry pressure column of air. However, to calculate the O\textsubscript{2} absorption coefficients, GGG takes into account the surface pressure, which can lead to measurement inaccuracies if the wrong surface pressure is used. Wunch et al. (2011b) reported a 0.04 % X\textsubscript{CO\textsubscript{2}} bias for a +1 hPa surface pressure offset in the TCCON. Similarly, we find a 0.032 % X\textsubscript{CO\textsubscript{2}} bias per +1 hPa surface pressure offset, with a 0.004 % \( \sigma \) variation on average throughout a day. Because the pressure offset affects O\textsubscript{2} retrievals, the other species are also affected (Table 4). X\textsubscript{CO} may be particularly affected by a pressure bias because such a large fraction of the column CO is near the surface.

Using the same July 2014 dataset used to test the sensitivity of the retrievals to error in the pTz profile and surface pressure, we further estimated the sensitivity to error in the temperature in the lower atmosphere (surface – 700 hPa). GGG uses a single temperature profile per day that represents the local-noon temperatures, and the surface temperature is extracted from that profile. Such temperature error can arise in particular at the beginning and end of the day when the temperature is typically cooler than at noon. Here we derived the sensitivity of the retrievals to a +10 K error in the lower atmosphere (Table 4). X\textsubscript{CO} has a significantly larger bias than the other species, likely because water absorption lines are the strongest spectral features in the CO retrieval window and water absorption lines are highly sensitive to changes in temperature. Water lines are also much stronger than N\textsubscript{2}O lines in the N\textsubscript{2}O windows. These tests suggest that offsets under 1 hPa and 1 K would cause small (\( \sim 0.1 \) ppm) biases on X\textsubscript{CO\textsubscript{2}}, but a 4 K difference in near-surface (ground – 700 hPa) temperature could cause \( \sim 0.4 \) ppm bias in X\textsubscript{CO\textsubscript{2}}, which is larger than our reported \( \sigma \) precision. For other studies using multiple spectrometers and multiple meteorological measurements for X\textsubscript{gas} retrievals, we recommend cross-comparing meteorological measurements to eliminate bias – preferably to a standard.

Finally, we perform a sensitivity study following the methodology of Wunch et al. (2015). The magnitudes of the applied perturbations are in Table 5. The results of this uncertainty budget study are presented for a day for X\textsubscript{CO\textsubscript{2}} and X\textsubscript{CH\textsubscript{4}} in Fig. 11. We do not include a sum in quadrature because we do not have an exhaustive list of sources of uncertainty. This uncertainty budget indicates that the low-resolution instruments are especially sensitive to biases in a priori pressures and a priori volume mixing ratio (VMR) profiles. Some of these errors may partially account for the unexplained long-term drifts we noted compared to TC-

---

### Table 4. Meteorological sensitivity tests on EM27/SUN retrievals.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X\textsubscript{CO\textsubscript{2}}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X\textsubscript{CH\textsubscript{4}}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X\textsubscript{CO}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X\textsubscript{N\textsubscript{2}O}</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Perturbations used in uncertainty budget.

<table>
<thead>
<tr>
<th>Perturbation</th>
<th>Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>ap\textsuperscript{a} VMR</td>
<td>downshift by 1 km\textsuperscript{b}</td>
</tr>
<tr>
<td>ap temperature</td>
<td>+1 K all altitudes</td>
</tr>
<tr>
<td>ap pressure</td>
<td>+1 hPa all altitudes</td>
</tr>
<tr>
<td>Pointing offset (po)</td>
<td>increased by 0.05\textsuperscript{c}</td>
</tr>
<tr>
<td>Surface pressure</td>
<td>+1 hPa</td>
</tr>
<tr>
<td>Calculated OSDS\textsuperscript{c}</td>
<td>+2 ppm</td>
</tr>
<tr>
<td>Field of view (FOV)</td>
<td>+7 %</td>
</tr>
</tbody>
</table>

See also Fig. 11. \textsuperscript{a} ap denotes a priori. \textsuperscript{b} ap VMRs were shifted independently. For X\textsubscript{H\textsubscript{2}O} and X\textsubscript{H\textsubscript{2}DO\textsubscript{3}}, concentrations were decreased by 50 % at all levels. \textsuperscript{c} OSDS = observer sun Doppler stretch.

---

**Figure 11.** Uncertainty budget for EM27/SUN instruments using GGG2014. See Table 5 for magnitudes of perturbations.
### Table 6. Tests for assessing biases and sensitivities of solar-viewing, remote sensing instruments.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Test/observation</th>
<th>Type</th>
<th>Accepted correctiona</th>
<th>Root cause</th>
<th>Similar instr. effect</th>
<th>EM27/SUN test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incoming radiation attenuation effect</td>
<td>Gray filter after solar tracker &amp; before interferometer</td>
<td>M</td>
<td>Recom’d replace detector.</td>
<td>Detector nonlinearity</td>
<td>Consistent for same detectors</td>
<td>Sect. 4.4</td>
</tr>
<tr>
<td>ILS</td>
<td>Measure with low-p gas cell (preferred), stable laser, or ambient air (least recom’d)</td>
<td>M</td>
<td>Retrievals with non-ideal ILS</td>
<td>Instrument misalignment; in-built</td>
<td>Potentially large differences</td>
<td>Gisi et al. (2012); Chen et al. (2014); Frey et al. (2015); Sects. 4.1 (measured), 5.3, 5.4 (impacts)</td>
</tr>
<tr>
<td></td>
<td>Adjust FOV (if ILS is measured but not accounted for in retrieval)</td>
<td>RIA</td>
<td>Not recom’d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghost to parent ratio</td>
<td>Use blackbody source &amp; narrow band filter post-interferometer</td>
<td>M</td>
<td>Laser mis-sampling</td>
<td>Likely similar, potentially large diffs</td>
<td>Gisi et al. (2012); Frey et al. (2015); Sect. 4.3</td>
<td></td>
</tr>
<tr>
<td>Ghost effects</td>
<td>Measurements with &amp; without ghost correction (e.g., XSM, or ifg resampling before FFT)</td>
<td>M or RIA</td>
<td>Recom’d interpol. during acq or post-resampling</td>
<td>Laser mis-sampling</td>
<td>Likely similar, potentially large diffs</td>
<td>Sect. 4.3</td>
</tr>
<tr>
<td>Frequency shifts</td>
<td>Changes or large 0 offset</td>
<td>O &amp; RIA</td>
<td>Input spectral spacing</td>
<td>Improper wavenumber, mis-alignment of laser or NIR beam</td>
<td>Shifts differ, effect similar</td>
<td>Sect. 4.2</td>
</tr>
<tr>
<td>Solar gas stretch</td>
<td>Changes or large 0 offset</td>
<td>O &amp; RIA</td>
<td>OSDS</td>
<td>Poor spectral fits of solar lines; SE or res.</td>
<td>Similar for same detector &amp; res.</td>
<td>Sect. 6</td>
</tr>
<tr>
<td>Spectral fitting windows</td>
<td>Width, locations</td>
<td>RIA</td>
<td></td>
<td>Instrument resolution requires adaptation</td>
<td>Same for similar res., microwindows &amp; VG</td>
<td>Gisi et al. (2012); Sect. 5.7 (H2O) Sect. 5.3 (discussion)</td>
</tr>
<tr>
<td>Averaging kernels</td>
<td>Used when comparing with a different instrument type</td>
<td>O</td>
<td>Rodgers and Connor (2003) and prior info.</td>
<td>Diff. sensitivity at atm. layers from differing resolutionsb &amp; VG</td>
<td>Same for similar res., microwindows &amp; VG</td>
<td>Sect. 5.1</td>
</tr>
<tr>
<td>SZA artifacts</td>
<td>Multi-day measurements in clean location</td>
<td>O</td>
<td>Empiricalb (Wunch et al. 2011b)</td>
<td>ILS, or SE</td>
<td>See ILS entry</td>
<td>Frey et al. (2015); Parker et al. (2015)</td>
</tr>
<tr>
<td>Long-term artifacts</td>
<td>Preferred co-location with accepted measurements (e.g., TCCON)</td>
<td>O</td>
<td>Various (e.g., instrument settling, changing alignment, other)</td>
<td>May widely differ</td>
<td>Herein – for extended InGaAs only</td>
<td></td>
</tr>
<tr>
<td>Region/zone dependence</td>
<td>Co-location with spatially distributed accepted measurements</td>
<td>O/M</td>
<td>A priori insufficiencies</td>
<td>Likely similar</td>
<td>Parker et al. (2015)</td>
<td></td>
</tr>
<tr>
<td>Surface pressure effects</td>
<td>Manually adjust pressure inputs.</td>
<td>RIA</td>
<td>Accurate barometer pres. calibr.</td>
<td>Poor calculation of O2 column, directly or by poor fitting</td>
<td>Similar effects for similar resolutions</td>
<td>Sect. 6</td>
</tr>
<tr>
<td>pTz &amp; H2O model profile sensitivity</td>
<td>Adjust modeled meteorological profiles</td>
<td>RIA</td>
<td>Improve met. profiles</td>
<td>Non-representative pTz+H2O profile</td>
<td>Similar effects for similar resolutions</td>
<td>Sect. 6</td>
</tr>
<tr>
<td>A priori VMR surface sensitivity</td>
<td>Adjust a priori VMR near surface</td>
<td>RIA</td>
<td>Improve a priori profiles; reduce effect with AKs</td>
<td>Non-representative VMR profile (e.g., polluted mixed layer)</td>
<td>Similar effects for similar res. &amp; true VMR profile</td>
<td>Parker et al. (2015)</td>
</tr>
<tr>
<td>Opt. avg. time</td>
<td>Allan type plot; e.g., Chen et al. (2016)</td>
<td>O</td>
<td>Empirical</td>
<td>SNR &amp; true atmospheric variation</td>
<td>Depends on SNR &amp; location</td>
<td>Chen et al. (2016) Sect. 5</td>
</tr>
<tr>
<td>Resolution effects</td>
<td>Truncate high resolution ifg</td>
<td>RIA</td>
<td>Apply offset</td>
<td>Inst. res.</td>
<td>Similar for all solar-viewing instrs.</td>
<td>Gisi et al. (2012); Petri et al. (2012) Sects. 5.3–5.6</td>
</tr>
<tr>
<td>Uncertainty budget for current fitting algorithm</td>
<td>Various, test on each new algorithm (Wunch et al. 2015)</td>
<td>RIA</td>
<td>Informative</td>
<td>Various</td>
<td>Similar effects for similar resolutions</td>
<td>Sect. 6</td>
</tr>
</tbody>
</table>

M denotes measurement (setups/adjustments required before acquisition), RIA denotes retrieval input adjustment (post-data acquisition, pre-retrieval), O denotes observation post-acquisition (may require prior planning of locations of measurements or longer term measurements), SE denotes spectroscopy errors, VG denotes viewing geometry, res denotes resolution.

a Though empirical corrections are occasionally accepted, it is always recommended to correct the underlying problem(s) if possible.
b XSM is Bruker™ code for interpolation during acquisition.

GGG can provide ifg resampling if two detectors are on instrument. Note that the preferred correction is always of the root cause.
Figure 12. Time series comparison of EM27/SUN retrievals to retrievals from the 0.5 cm⁻¹ resolution IFS 125HR spectra.

CON that are unrelated to signal (e.g., Fig. 8, October–November 2014). For example, surface pressure and calculated Observer-Sun Doppler Stretch (OSDS) were correlated with EM27/SUN to TCCON XCO₂ differences in the long-term measurement. However, there was no apparent trend in the spectral residuals from fitting solar lines as the OSDS changed so these correlations may not indicate cause.

Differences in Xgas between different instruments are due to a combination of differences in resolution, and real instrumental imperfections and instabilities. To attempt to distinguish between resolution causing differences (e.g., by limitations in the forward model) or instrumental issues, we repeat the test performed by Gisi et al. (2012, Fig. 11 therein) of truncating IFS 125HR interferograms for the full time series. Results are shown in Fig. 12. Mean values for XCO₂ are slightly lower because of differences from retrievals on spectra of different resolutions, as described in Sect. 5.3. When comparing 10 min averaged TCCON data with lower resolution IFS 125HR retrievals we note monthly standard deviations on order of 0.15 % for XCO₂ and XCH₄. This suggests the standard deviations of comparing retrievals from the EM27/SUN with the TCCON (Table 3) on these timescales are close to the current precision limits for directly comparing XCO₂ and XCH₄ retrieved from spectra of these different resolutions. Results in Fig. 12 are slightly more scattered than in Fig. 8 and have different offsets. The data still show an increase in XCO₂ and XCH₄ in October–November 2014 for reasons we do not understand, and unfortunately we have no ILS characterizations over this period.

Long-term drifts may or may not affect instruments employing the standard InGaAs detector and may be eliminated by future retrieval updates. They may also arise in part from how the comparison was made, e.g., the assumptions to derive Eq. (A4) may not be valid for CH₄ and N₂O. As a follow-up study, brief 5–6 day comparisons using a standard InGaAs detector were made for the months of August, September, and November 2015. Scaling factors varied from 0.99905 to 1.00001 for XCO₂ and from 1.01228 to 1.00893 for XCH₄, with larger day-to-day variability. Long-term (1 year or more) comparisons of these instruments employing the standard-InGaAs detector are needed before claims of long-term accuracy can be made or the full magnitude of drift can be quantized. Errors that could lead to drifts likely would be correlated amongst all EM27/SUN instruments, so the comparison would need to be against a standard such as the TCCON. Future studies may also benefit from comparing results using different retrieval algorithms, as the magnitude of errors that may lead to drifts in Xgas may vary among algorithms. Meanwhile, operators have already found many purposeful ways to use these instruments that require only short-term (about 1 month) precision without any assumptions about precision for longer time periods (for example Hase et al., 2015; Chen et al., 2016; Viatte et al., 2016). Studies using these spectrometers independently longer term can also be performed depending on the degree of precision required. Limits on precision described herein are likely to only improve in future work.

7 Conclusions

Despite the challenge associated with the extended InGaAs detector and mirror degradation, the EM27/SUN instruments perform well on short timescales with 1σ running 10 min precisions of 0.075 % for XCO₂ and 0.057 % for XCH₄ retrieved from measurements using the standard InGaAs detectors. These instruments perform well in terms of mobility and
stability, maintaining alignment despite frequent movement and jostling – an ideal characteristic of mobile FTS instruments. Measurements from the standard detector are precise enough to be used for campaigns of up to a few months and to provide useful supplementary X\textsubscript{gas} measurements to established networks like TCCON. However, we recommend regular – 6 months to 1 year depending on use – comparison with established measurements (e.g., a TCCON site) to account for long-term drift. The frequency of comparison with established measurements may need to be reevaluated when more long-term comparison data become available. Simultaneous use of several EM27/SUN instruments may also help characterize drift. We also recommend regular – about monthly depending on use – ILS characterization. Our experience also suggests that use of the extended InGaAs detector without limiting the spectral band-pass in the EM27/SUN is incompatible with X\textsubscript{CO\textsubscript{2}} and X\textsubscript{CH\textsubscript{4}} retrievals that are precise long-term.

In general, we recommend all new ground-based, solar-viewing, remote sensing FTS instruments to undergo some or all tests listed in Table 6 to evaluate their performance. We also recommend comparisons of retrieval outputs to those of existing instrumentation (e.g., TCCON or NDACC-IRWG). These tests assume that one of the three widely used and accepted retrieval algorithms (GGG, PROFFIT, and SFIT), known to provide accurate spectral fitting, is used. New retrieval algorithms should be subjected to additional comparisons with currently accepted algorithms. Some of the results of these tests will be similar across all instruments of a given type, and so do not need to be repeated if they have been performed on another instrument elsewhere.
Appendix A: Assumptions and limitations in the AK correction

To derive Eq. (5), we begin with Eq. (22) in Rodgers and Connor (2003):

$$\hat{c}_i = c_a + \sum_k h_{i,k} \cdot (x_{t,k} - x_{a,k}) + \epsilon_i. \quad (A1)$$

To include the pressure-weighting function $h$ (Connor et al., 2008), we have used summation notation. The “hat” represents a retrieved value, $c$ represents a column (scalar) value, and $\epsilon$ is the error. Subscript $i$ is for a particular instrument, subscript $a$ represents the a priori, subscript $k$ is for a particular atmospheric layer, and subscript $t$ represents the true atmosphere. The vectors $a$ and $x$ represent the column averaging kernel and atmospheric VMR profile respectively. This equation is derived from Eq. (1) in Rodgers and Connor (2003) using a Taylor series expansion about the a priori profile, and assuming linearity about it.

To compare retrievals from remote sounding instruments, a comparison profile (also called the comparison ensemble mean, denoted $x_c$) is used. Here, we have used the daily a priori profiles, which were the same for all instruments, as the comparison profiles. We note, however, that the comparison profiles should describe the real atmosphere as far as possible (Rodgers, 2000). Though the a priori profile has a drawdown in CO$_2$ from the biosphere near the surface, the real atmosphere in Pasadena is polluted near the surface. Thus this choice of comparison profiles is not ideal in our situation.

If we ignore retrieval error Eq. (A1), and further assume that $x_t = x_a$ except at the surface, it can be rewritten as

$$\frac{1}{a_{i,s}} (\hat{c}_i - c_a) = h_s (x_{t,s} - x_{a,s}), \quad (A2)$$

where the subscript $s$ represents a surface value. If we are comparing measurements from two different instruments, $i = 1$ and $i = 2$, in the same location, $x_{t,s}$ and $h_s$ are the same. Because the a priori profiles are also the same,

$$\frac{1}{a_{1,s}} (\hat{c}_1 - c_a) = \frac{1}{a_{2,s}} (\hat{c}_2 - c_a), \quad (A3)$$

which can be rewritten as

$$\hat{c}_1 = \frac{a_{1,s}}{a_{2,s}} (\hat{c}_2 - c_a) + c_a. \quad (A4)$$

Even in the absence of error, retrievals from instruments with different averaging kernels will still differ.

We adjust the EM27/SUN XCO$_2$ and XCH$_4$ retrievals using Eq. (A4) before comparison with the TCCON, which adjusts XCO$_2$ by up to $\sim 1.2$ ppm and XCH$_4$ by up to $\sim 8$ ppb. Future work could improve on this methodology using a better comparison ensemble or more representative a priori profiles for retrievals from measurements in Pasadena. This correction is not applied to XH$_2$O because the AKs vary more among spectra because of larger variations in absorption strengths. It is also not applied to XCO and XN$_2$O because using $x_t = x_a$ is too poor of an assumption and makes the comparison worse between the TCCON and EM27/SUN retrievals in terms of $R^2$.
Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements. We thank Frank Hase and Michael Gisi for helpful discussions on ghost reduction, detector nonlinearity, and ILS measurements. We further thank Michael Gisi and Bruker Optics™ for loaning us a standard InGaAs detector for testing and for instructions on realigning the EM27/SUN. We thank Dietrich Feist for discussions on mirror degradation. We also thank Nicholas Jones, David Griffith, Frank Hase, and Sabrina Arnold for sharing their experience with mirror degradation. This work is supported in part by the W. M. Keck Institute for Space Studies. Jacob Hedelius was also partially supported by a Caltech Chemistry and Chemical Engineering Division Fellowship funded by the Dow Chemical Graduate Fellowship, and expresses thanks to them. The authors gratefully acknowledge funding from the NASA Carbon Cycle Science program (grant number NNX14AI60G) and the Jet Propulsion Laboratory. Manvendra K. Dubey acknowledges funding from the NASA-CMS program for field observations and from the LANL-LDRD for the acquisition of the LANL EM27/SUN. Jia Chen, Taylor Jones, Jonathan E. Franklin, and Steven C. Wofsy acknowledge funding provided by NSF MRI Award 1337512.

The authors thank the referees for their comments.

Edited by: F. Hase
Reviewed by: M. K. Sha and one anonymous referee

References


J. K. Hedelius et al.: Assessment of errors in Xgas from a 0.5 cm$^{-1}$ spectrometer


www.atmos-meas-tech.net/9/3527/2016/ Atmos. Meas. Tech., 9, 3527–3546, 2016
