

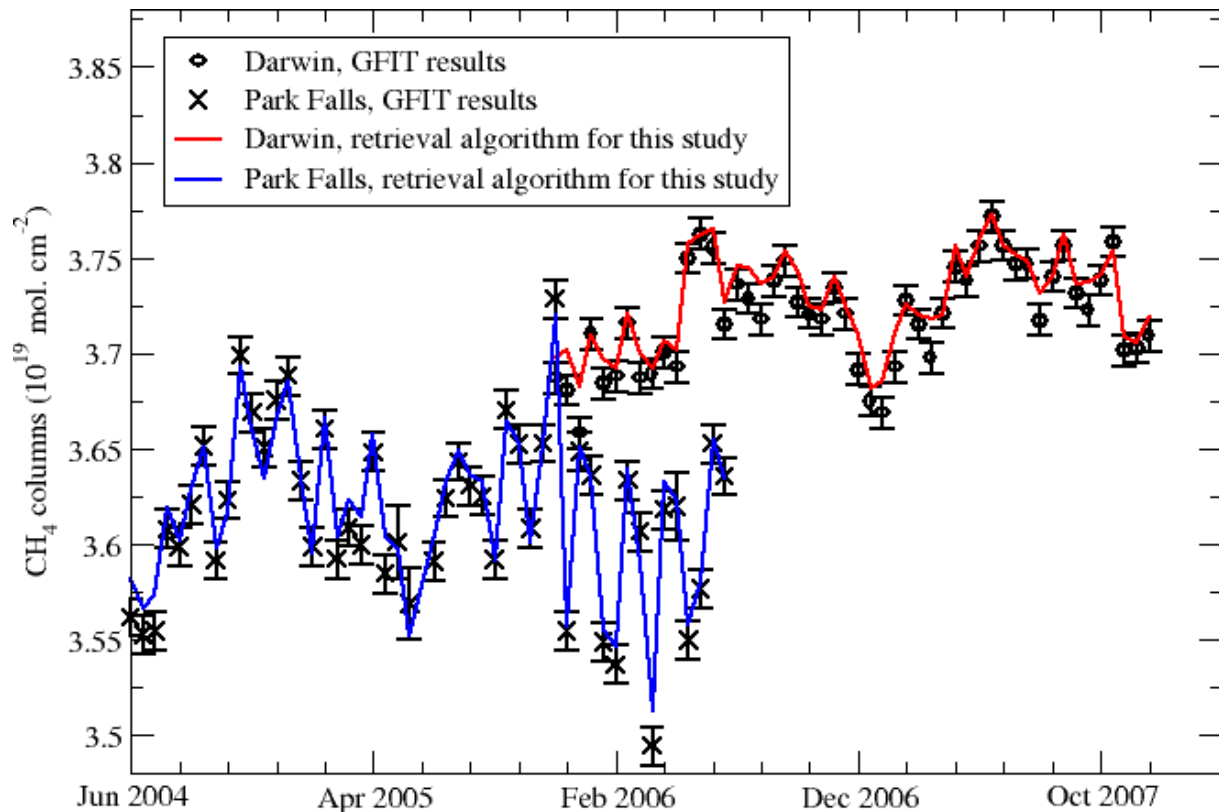
Referee #1:

<< This manuscript presents a simulation-based analysis of the effects of spectroscopic uncertainties on SWIR-based retrievals of CO and methane, in preparation for the future TROPOMI instrument. The topic of this work is certainly appropriate for AMT, and it is critical to perform studies such as this one before actually designing and building a new instrument. The paper is fairly well written overall, and I found the Conclusion section especially clear. As described below, however, there are two specific areas which I feel are relatively weak and should be revised. In addition, I have listed a number of minor comments which I feel should be addressed.>>

We appreciate the reviewer's positive perception of our work. See our response to his/her comments below.

<< Major revision 1: To quantify retrieval errors due to spectroscopy, simulated retrievals are compared to 'reference retrievals' produced with the GFIT package. Ideally, simulated retrievals (using the authors' retrieval algorithm) and GFIT-based retrievals would be identical for retrievals based on the same spectral range and resolution. Such tests were performed and are reported in Sec. 2.4. For CO, algorithm-related differences (biases and standard deviations) are on the order of a few percent. This seems acceptable, given that the natural variability of CO is much larger (CO concentrations can vary by a factor of two or more). However, for methane, the algorithm-related retrieval differences are up to 0.26% (the reported standard deviation at Park Falls). Since the natural variability of methane is only a few percent, this algorithm error seems very large, and will limit the interpretation of spectroscopic errors. The authors' statement on this matter that 'The small differences between our algorithm and GFIT may be explained by differences in averaging kernels and a priori profile for CH₄, CO, and H₂O (smoothing and interference errors), and differences in the H₂O line list' is not supported by any analysis and therefore seems speculative. This part of the paper should be supported by stronger analysis, more conclusive evidence regarding the source of the differences, and should include comparisons of the a priori and averaging kernels for the two types of retrieval algorithms. >>

The referee is right that CH₄ retrieval errors of 0.2% are not negligible. But since the retrieval uncertainty of the GFIT algorithm is as large, we cannot expect that the retrieval differences should be smaller. The retrieval uncertainty of CH₄ columns attributed to the GFIT retrieval, averaged over the 50 Park Falls spectra, is 0.28%. For Darwin, it is 0.22%. See below a plot of our retrieval algorithm results versus the GFIT results with their error bars.



Apart from a positive bias of 0.1%, the time-series from the two different retrieval algorithms (coloured line vs. black symbols) are statistically indistinguishable. Concerning this bias, we found that the different CH_4 prior and the different H_2O spectroscopic line list (GFIT uses one global CH_4 prior profile, whereas our algorithm assumes a different prior for Park Falls and Darwin) accounted for CH_4 retrieval differences of up to 0.2%. This is consistent with our stated retrieval uncertainty. The TCCON science team found that exclusion of the first CH_4 band in the GFIT retrieval introduces a constant bias of +0.2% CH_4 . We also ran our retrieval algorithm, using the same H_2O spectroscopy and priors as GFIT, and also included the first CH_4 band. However, the standard deviation of retrieval differences did not decrease further, whereas the spectral fit residuals were larger than for our default settings. This indicates that the 0.2% retrieval uncertainty, derived by the TCCON science team and derived from the comparison of the two different retrieval algorithms is the intrinsic retrieval accuracy of a single CH_4 column retrieval from a TCCON spectrum.

The retrieval uncertainty of CH_4 columns of roughly 0.2% in the 6000 cm^{-1} range of course limits our possibilities to assess spectroscopic errors. On the other hand, we do not know of any IR spectroscopic measurement and inversion that yields CH_4 total columns at a better accuracy. This is the reason why we caution the reader in the conclusion that “the intrinsic retrieval accuracy of CH_4 from TCCON spectra is a few tenths of a percent. This means we can only identify spectroscopic deficiencies that result in column errors with the same order of magnitude as the intended accuracy for the S5P mission.”

To keep the paper from getting much longer, we propose to revise the paragraph in Sect. 2.4, but not to add the above figure unless the editor and the referees feel that this would greatly benefit the paper. In the revised manuscript, we explain:

“Comparing our retrieved CH₄ values from windows 2 and 3 to the GFIT values, we found a bias between the two datasets of 0.1 % for both Parkfalls and Darwin and standard deviations of 0.26 % and 0.18 % for Park Falls and Darwin, respectively. These differences are not larger than the average uncertainty of the CH₄ column retrieved with the GFIT algorithm. The latter is 0.28 % for the Park Falls time-series and 0.22 % for the Darwin time-series. For CO, we found biases of 1.6 % and -2.8 %, and standard deviations of 1.1 % and 1.8 % for Park Falls and Darwin, respectively. The small differences between our algorithm and GFIT can be explained by the differences in averaging kernels and a priori profiles for CH₄, CO, and H₂O (smoothing and interference errors), and the differences in the H₂O line list. Therefore, we considered our retrieved columns from the 5880 – 6174 cm⁻¹ range (for CH₄) and from 4209 – 4319 cm⁻¹ range (for CO) as reference values and the standard deviations between the algorithms as the associated error σ_{ref} .”

<<Major revision 2: As discussed in Sec. 2.5, three statistics are employed to quantify how much retrieval performance is degraded by reduced spectral resolution. These statistics, presented in Tables 2 and 3, include the chi-square, standard deviation, and bias. The authors should also present the correlation coefficient as another important measure of retrieval performance. This statistic should also be presented and discussed for the analysis in the second paragraph of Sec. 3.1 (and shown in Fig. 7). The 'correlation strength' is mentioned in the fifth paragraph of Sec. 3.2 (i.e., the paragraph that starts with 'The CO bias would be ...',) but is not actually presented; the value should be given.>>

Done. Changes in the revised manuscript in detail:

- In Sect. 2.5, we added a fourth bullet: correlation coefficient ξ
- We added the column with ξ to Tables 2, 3, and 4.
- We now explain in the first paragraph of Sect. 3.1: “Table 2 lists the fit diagnostics. Those are, from left to right: average spectral fit residuals ($\langle \chi^2/\nu \rangle$), standard deviation of the differences in columns (σ_N) versus the reference error sigma_ref (see Sect. 2.4), bias b, and correlation coefficient ξ . Table 2 shows that the CH₄ columns were well reproduced for TROPOMI-like settings. The average scatter of retrieved columns increased only from 0.18 to 0.24% at Darwin, and from 0.26 to 0.28% at Park Falls. Neither the standard deviation nor the correlation changed significantly when the original spectral resolution was assumed.”
- In the second paragraph of Sect. 3.1, we now write: “Figure 8, upper panel, shows CH₄ retrieval errors against H₂O Dry air Mole Fraction (DMF) at Darwin, the air-mass being between 1.2 and 1.4 for all observations. The dashed line is the linear regression (correlation coefficient = 0.22), the dotted lines show the 1- σ uncertainty level of the slope. The slope is not significantly larger than zero at a 2- σ level (...)”
- In the third paragraph of 3.2, we note: “As for CH₄, the H₂O retrieval accuracy did not deteriorate when the input spectrum was degraded to the TROPOMI resolution; σ_N in Table 4 remained almost identical and $\xi > 0.99$ for all data sets.”
- In the fifth paragraph of Sect. 3.2, we now explain: “Figure 8 demonstrates for the Darwin observations the extent to which the CO and CH₄ retrieval errors increased with H₂O abundance. In the lower panel, the CO retrieval errors for the nominal TROPOMI range and for an optimised sub-window (see Sect. 3.3 for further explanation) are plotted against H₂O DMF. The

dashed lines show the linear fits, the correlation coefficient calculates to 0.66 for the nominal TROPOMI range and to 0.52 for the sub-window. This correlation did not change when the default H₂O line list was replaced by the HITRAN08 list.“

<< Minor comment 1: p. 2133, Abstract - Retrievals based on ground-based measurements are fundamentally different from satellite-based retrievals. Please justify the use of simulations of ground-based retrievals to the design of a satellite instrument. For example, are ground-based and satellite-based retrievals equally sensitive to spectroscopic errors? Is this an implicit assumption?>>

See our answer to the second referee (major question 1).

<< Minor comment 2: p. 2134, Introduction - Please define spectral ranges corresponding to near-infrared and short-wave infrared.>>

We specify in the revised manuscript: “Its goal is to remotely measure column abundances of CO₂, CO, CH₄, N₂O and other molecules that absorb in the short-wave infrared (SWIR) and near-infrared (4000 – 15,000 cm⁻¹).”

<< Minor comment 3: p. 2134, l. 15. Suggest adding 'polar-orbiting' before ' ... series of satellites that monitor CH4 and CO ...'>>

Done.

<< Minor comment 4: p. 2134, l. 26. 'observations' should be singular.>>

Done.

<< Minor comment 5: p. 2135, l. 4. 'CO total columns ...' should be 'CO total columns and vertical profiles ...'>>

Done.

<< Minor comment 6: p. 2135, last paragraph. In discussion of SWIR-based products, mention that TIR-based products typically exhibit poor sensitivity to CO in the lower troposphere.>>

We added the sentence in the revised manuscript: “Measurements in the thermal-infrared are usually less sensitive to CO in the lower troposphere.”

<< Minor comment 7: p. 2137, l. 15. Some discussion about the relative effects of scattering for TCCON-type ground-based retrievals and satellite-based retrievals would be helpful. Scattering from aerosols is a significant source of error for SWIR-based satellite retrievals of methane (and CO₂), as shown by work done for the SCIAMACHY and GOSAT instruments.>>

Observations of direct sunlight are hardly affected by any scattering processes, whereas retrievals of backscattered sunlight must take these processes into account. Please see our answer to the second referee (major question 1).

<< Minor comment 8: p. 2139, l. 8. What would be the potential benefit (decreased residual) for retrieving CO and H₂O as profiles, as is done for methane?>>

In principle, retrieving a CO and H₂O profile should decrease the fit residual and possibly also the interference errors (Sussmann and Borsdorff, 2007). We found, however, that the spectra do not contain enough information to retrieve profiles both of CH₄ and of another absorber species. Since H₂O and CO profiles exhibit a lot more vertical variation than CH₄, the H₂O or CO profiles often diverged during the retrieval.

<< Minor comment 9: p. 2139, l. 11. What do the authors consider 'a meaningful profile'? DFS greater than 1, 2, 3 ...?>>

The DFS should be larger than roughly 2, and we usually obtained a DFS smaller than 1.9. In the revised manuscript, we specify: "As for all other species, the degrees of freedom of the inversion (between 1.8 and 1.9) were insufficient to retrieve a meaningful profile."

<< Minor comment 10: p. 2139, l. 21. How is the value of gamma determined? What is the value?>>

We refer those readers interested in the technical details of our retrieval algorithm to the publication by Butz et al. (2012) (cited at the beginning of the subsection). There, the choice of γ is explained as follows: "For determining γ , we use the L-curve criterion (Hansen, 1998). The L-curve is a (double-logarithmic) parametric plot of the least squares norm against the side constraint (norm of the state vector) with γ as parameter. Typically, the L-curve shows a distinct corner for the optimal choice of γ . Too small γ yields a large norm of the state vector, too large γ yields a large least-squares norm."

<< Minor comment 11: p. 2140, l. 9. Please include a brief description (one sentence) about the type of retrieval algorithm used in GFIT.>>

We explain in the revised manuscript: „These spectral ranges are also used by the TCCON science team for their retrievals with the GFIT nonlinear least squares spectral fitting algorithm. GFIT is a profile scaling retrieval which fixes the a priori profile shape of the absorber and scales the profile to produce a calculated spectrum that best matches the measured spectrum. The surface pressure is measured at each TCCON station with a barometer (Wunch et al., 2011a).“

<< Minor comment 12: p. 2140, l. 15. The statement that 'Similarly, the accuracy of retrieved CO columns from the 4209–4319 cm⁻¹ spectral range by GFIT has also been confirmed by validation with aircraft measurements' should include a supporting reference.>>

We added the reference to Wunch et al., 2010.

<< Minor comment 13: p. 2140, l. 22. What are the expected effects (on the retrieval results) of discarding window 1? Does this have any effect on the retrieval averaging kernels?>>

See our response to the major revision 1.

<< Minor comment 14: p. 2141, l. 8. The phrase 'consistent with' is not precise; what are the typical reported retrieval uncertainties for GFIT? >>

See our response to the major revision 1.

<< Minor comment 14: p. 2143, l. 16. Figure 7 appears out of sequence; shouldn't this be Figure 1?>>

"Figure 7, upper panel" in line 17 is correct. Figure 7 (Figure 8 in the revised manuscript) includes two panels of the CH₄ error and the CO errors versus H₂O. Since this figure is more important to the discussion of CO retrieval accuracy, it was inserted after the CO and H₂O time-series.

<< Minor comment 15: p. 2143, l. 25. The meanings of 'pressure' and 'total pressure' in this paragraph (and the following two paragraphs on p. 2144) are not clear. Should both of these terms be replaced with 'surface pressure'?>>

We always assumed a fixed vertical pressure profile and then fitted one scaling constant to be multiplied to this constant profile. In the three paragraphs on pp. 2143 and 2144 of the revised manuscript, we replaced "pressure" and "total pressure" with "surface pressure" at those occurrences where confusion might arise.

Referee #2:

<< This manuscript analyzed the uncertainties of CH₄, CO and H₂O by using similar observations of TCCON, which is important before the launch of a new space instrument. The experimental examples were carefully selected by covering different environment. The work is definitely suitable to publish in AMT. Although this manuscript is very well written, I still have some questions and suggestions for revision.>>

We appreciate the reviewer's positive perception of our work. See our response to the remaining questions below.

<< Major question 1: TROPOMI is a satellite based nadir viewing instrument, while TCCON is a ground based and up-looking instrument. The radiance sources, transfer path and the strength may not be the same. The authors have adjusted the spectra resolution, but it may not have been enough to make up the differences. The authors should explain how they can use up-looking radiances as a way to simulate down-looking radiances.>>

The referee is right. The observation geometry for TCCON (direct sunlight, scattering processes can be neglected) and for TROPOMI (backscattered sunlight) are different. Butz et al. (2012) inverted synthetic spectra of backscattered sunlight where the forward model included aerosol scattering. Three parameters describing the aerosol properties were retrieved simultaneously to the CH₄ columns. As was to be expected, they found that the retrievals with a low aerosol optical thickness were most accurate in terms of retrieved CH₄.

For these retrievals, the spectroscopic errors (for direct sunlight spectra) found in this work are still representative, since the Jacobians are identical except for a constant scaling factor (see answer to question 3). The spectroscopic error derived in this study should be interpreted as a minimum retrieval error we cannot evade with TROPOMI even for an optimal observation scenario with very few aerosols and cirrus. For TROPOMI inversions where the aerosol optical thickness in the forward model is high, we do not know how the spectroscopic errors will combine with the forward model errors due to scattering. We expect, however, the forward model errors due to the treatment of aerosols/cirrus to be dominant. For TROPOMI simulations, we found retrieval errors on the order of 1% when the retrieved aerosol optical thickness > 0.2.

In the revised manuscript, we elaborate: "Butz et al. (2012) cautioned, however, that a few other, unrelated, error sources had not yet been assessed. One of them, which is hard to assess by means of synthetic input spectra alone, is the inaccuracy of the assumed molecular absorption lines of CH₄, CO, and H₂O. This error source is the topic of this study. It is representative for the spectroscopy-related errors in TROPOMI retrievals with a low density of scatterers in the forward model."

<< Major question 2: Page 2137, line 14-15, TROPOMI observes the back-scattered sunlight, but the scattering processes are neglected in the forward model of this study, it is better to add a reference or proof to support "TCCON spectrometers measure direct sun-light, which is, to good approximation".>>

We added the reference to Wunch et al. (2011a). Both GFIT and our retrieval forward model neglect scattering.

<< Major question 3: Could you add Jacobians or weighting matrix for the 3 species? I would like to see the Jacobian differences between the back scattering sunlight (TROPOMI) and directly sunlight (TCCON).>>

The exact Jacobian of course varies for each spectrum and each iteration step even for the forward model without scattering.

In formulaic terms, the derivatives of the modelled reflectance for the absorber species entry j in the state vector x , $K_{ij} = \frac{\partial F_i}{\partial x_j}(x_n)$, read:

$$K_{ij} = \frac{\sum -F_i f_n(\mu_0) \sigma_n}{natm}$$

with $natm$ = number of atmospheric layers and the modelled intensity = forward model $F(k)$ from Equation 1 in the paper.

For the TROPOMI geometry, in the case of negligible aerosol optical thickness and for wavelengths > 2000 nm (Rayleigh scattering negligible), the forward model (in analogy to Eq. 1 in the paper) simplifies to the product of two exponentials:

$$F(k) = RSa \exp\left(-\int dh n(h) \sigma(f(\mu_0) + f(\mu_v))\right)$$

where a is the ground albedo (instead of instrument gain g), $f(\mu_0)$ is the air mass for the cosine of the solar zenith angle, and $f(\mu_v)$ is the air mass for the cosine of the viewing angle between the ground and the satellite. Thus, the Jacobian calculates to:

$$K_{ij} = \frac{\sum -F_i (f_n(\mu_0) + f_n(\mu_v)) \sigma_n}{natm}$$

In the case of $\mu_0 \approx \mu_v$, the Jacobians of the back-scattered sunlight case then just are a scaled version of the direct-sunlight case. This is the reason why the impact of spectroscopy errors for a retrieval of direct sunlight should be similar to a retrieval of backscattered sunlight, so long as the forward model is not dominated by particle scattering. We now state this explicitly (see answer to major question 1).

<< Minor question 1: Page 2137, line 20, please add an explanation for p , T and μ_0 .>>

The sentence in the revised manuscript reads: "The other terms in Eq. 1 are the instrument gain g , the wavenumber of the spectrum k , the pressure and temperature p and T , the molecular cross-sections $\sigma(p, T, k)$, the particle density of the considered absorber molecules $n(h)$, and finally $f(\mu_0)$, the air-mass as a function of $\mu_0 = \cos(\text{SZA})$ (Kasten and Young, 1989)."

<< Minor question 2: Page 2139, line 10, "the degree of freedom of the inversion were insufficient to retrieve a meaningful profile", is the reason due to TCCON radiance or the TROPOMI originally design?>>

The degree of freedom is insufficient (i.e., DFS of the CH_4 profile smaller than 2) both for the TCCON and for the TROPOMI viewing geometry. In both cases the averaging kernel is most sensitive to CH_4 in the lower troposphere (which is fine for detecting terrestrial sources and sinks) where the CH_4 profile does not vary a lot anyway. Also see our answer to the first referee (minor comment 9).

<< Minor question 3: Page 2139, line 18, Eq(2), how do you decide gamma?>>

See our answer to the first referee (minor comment 10).

<< Minor question 4: Page 2139, line 18, Eq(2), will H₂O, CH₄ and CO be retrieved simultaneously? If not, is there any sequence of 3 species? I mean, for example, were the retrieved H₂O, CH₄ used as inputs when doing CO retrieval?>>

Both our algorithm and the GFIT algorithm retrieve all absorber species simultaneously.

<< Minor question 5: Page 2139, line 20, signal-to-noise ratio was assumed as 1000. While in Butz et al. (2012), it is found 500 in the NIR channel and 100 in the SWIR, in Wunch et al., (2011a), the signal-to-noise ratio (SNR) for a single spectrum ... is approximately 750 near 5000cm⁻¹. Why the signal-to-noise here is much larger than other studies?>>

The SNR of 100 mentioned by Butz et al. (2012) applies to the TROPOMI instrument and TROPOMI viewing geometry, the much higher SNR mentioned by Wunch et al., (2011a) applies to TCCON spectra. SNR = 1000 for the short-wave infrared was chosen according to the instrument log files of TCCON. It is defined at the continuum level where no absorption is present. At absorption bands the actual signal-to-noise ratio is lower, depending on the specific spectral range (Wunch et al., 2011a).

For the present study we were looking for the spectroscopic error, which is independent of the instrument noise. We experimented once with using SNR = 120 and adding additional random noise to the TCCON spectra to simulate a TROPOMI-like SNR. However, because the TROPOMI SNR and its dependence on wavelength is only known and defined for backscattering geometry and because using a lower SNR plus additional noise did not have a notable impact on the retrieved absorber columns, we decided to always use the TCCON SNR for this study.

<< Minor question 6: Page 2140, line 9, Can you add a brief introduction for the GFIT algorithm and the difference between your algorithm and GFIT.>>

See our answer to the first referee (minor comment 11).

<< Minor question 7: Page 2141, line 20, the meaning of χ^2/ν should be explained clearly.>>

In the revised manuscript, we explain: "The reduced fit residual χ^2/ν : the sum of the residuals between modelled and measured spectrum, divided by the degrees of freedom ν (number of spectral pixels minus the degrees of freedom of the fit): (Equation)

The fit residual χ^2 equals the least-squares part of the cost function in Eq. 2. In the results section, we will list the average χ^2/ν over all spectra."

<< Minor question 8: Page 2143, line 28, what's the meaning of pressure vs total pressure?>>

See our answer to the first referee (minor comment 15).

<< Minor question 9: Page 2144, line 20, fig 8 is referred before fig.5>>

We changed the figure sequence in the revised manuscript.

<< Minor question 10: Page 2148, line 3, "One reason might be", did you do tests to add a noise in your H₂O prior and see if you conclusion correct? Here I believe it is consistent with Page 2139 line 10-11, since your retrieval here are "insufficient to retrieve a meaningful profile".>>

We investigated different initial conditions for H₂O but these tests did not lead to reduced interference errors.