

1 We thank the reviewer for their positive and useful comments, and for their careful
2 reading of the paper. We have addressed their questions as follows:

3
4 **Anonymous Referee #1**

5 De Smedt and co-authors describe in great detail the theoretical basis, methods and
6 assumptions used in the operational processor of the Sentinel-5 Precursor TROPospheric
7 Monitoring Instrument (S5P TROPOMI). The algorithm described with great detail here
8 benefits from strong heritage having the co-authors lead the development of formaldehyde
9 retrievals using measurements from previous UV space sensors. Algorithm improvements
10 developed during the Quality Assurance for Essential Climate Variables (QA4ECV) project,
11 funded by the European Union, are also presented since they are the basis of future updates
12 to the operational processor. The error budget for TROPOMI formaldehyde observations
13 is derived and discussed in the context of the Copernicus requirements. Finally, validation
14 methods and goals are discussed.

15 This paper should be the reference document for TROPOMI formaldehyde. Anyone using
16 the operational product should read it to understand the meaning of the retrieved
17 quantities, and their suitability to carry on scientific studies. The content is presented in a
18 clear and sound way, it follows the logical steps of the algorithm and is well organized. The
19 paper is ready to be published with minor changes that will only add to its great value. A
20 set of recommendations to minimally expand the content of the paper is followed by some
21 technical corrections.

22
23 **Section 1. Introduction.**

24 Some extra references should be added to support the statements describing formaldehyde chemistry
25 in the atmosphere.

26 Ok. Done.

27
28 **Section 2.2 Algorithm Description**

29 Why are the HCHO Meller and Moorgat (2000) cross sections used instead of the more recent and
30 intensity corrected Chance and Orphal (2011)?

31 The Chance and Orphal is based on the Cantrell et al., 1990 cross-sections, rescaled to match the Meller
32 and Moortgat, 2000 cross-section.

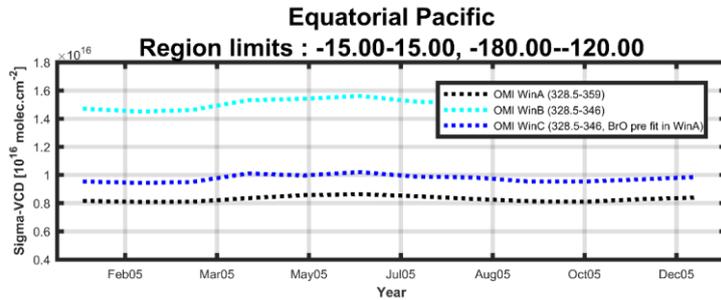
33 Cantrell et al. offers a better spectral resolution ($R = 0.01\text{nm}$), but its absolute values are biased. With
34 the 0.5 nm resolution of OMI and TROPOMI, we have chosen to use Meller and Moorgat ($R = 0.025\text{ nm}$)
35 avoiding any handmade modification. The two datasets (Chance et Orphal and Meller and Moorgat)
36 result in very consistent slant columns. As Chance and Orphal is the official HITRAN database, we will
37 consider to switch the cross-sections, but this will not affect the results.

38
39 **Section 2.1.1 HCHO SCD retrieval**

40 Page 11, line 233. It will be valuable to show evidence of the reduction in the correlation between
41 formaldehyde and bromine monoxide by using a two-step DOAS retrieval adding a new figure? Given
42 the extension of the paper it is maybe not necessary, but it will be interesting to have it here.

43 For GOME2, we refer to De Smedt et al., 2012 (figure 7).

44 For OMI, I add a figure in this review (figure 1), based on the QA4ECV dataset for which the 2 fitting
45 windows have been processed, the second without or with pre-fit of BrO (respectively WinB and WinC).
46 The figure shows the standard deviation of the HCHO columns for 2005 in the remote Equatorial Pacific.
47 The effect of pre fitting BrO columns in the small window (328.5-346 nm) can be seen by comparing
48 WinB and WinC results (50% reduction of the noise level, from $1.5\text{e}16$ to $1\text{e}16\text{ molec.cm}^{-2}$).



49

50 **Figure 1: Standard deviation of the OMI HCHO columns for 2005 in the remote Equatorial**
 51 **Pacific, as retrieved in WinA (328.5-359 nm), WinB (328.5-346 nm) or WinC (328.5-346 nm),**
 52 **with pre fit of BrO in WinA)**

53

54 [Page 11, line 264. The text says "\(3\) possible row-dependent biases \(stripes\) are directly corrected](#)
 55 [owing to the use of one reference per detector row." Are irradiances not recorded for each detector row?](#)
 56 [If they are, as it is done with OMI, the reason for the removal of the stripes when using radiance reference](#)
 57 [should be other than just having an irradiance reference for each row.](#)

58 We agree with this comment. The reason for the strong effect of using radiance as reference in reducing
 59 the stripes cannot be simply attributed to the use of each detector row. The reason is likely to be more
 60 complex and might rather be related to the observed improvement of the fits, involving a better
 61 cancellation of optical effects resulting in offsets (points (1) and (2)). We changed the text as:

62

63 "The main advantages of this approach are (1) an important reduction of the fit residuals (by up to 40%)
 64 mainly due to the cancellation of O₃ absorption and Ring effect present in both spectra; (2) the fitted
 65 slant columns are directly corrected for background offsets present in both spectra; (3) possible row-
 66 dependent biases (stripes) are greatly reduced by cancellation of small optical mismatches between
 67 radiance and irradiance optical channels; and (4) the sensitivity to instrument degradation affecting
 68 radiance measurements is reduced because these effects tend to cancel between the analyzed spectra
 69 and the references that are used"

70

71 [Page 12, line 284. Equation 5 only shows the shift \(Δi\) but line 284 mentions that this approach allows](#)
 72 [compensating for stretch and shift errors. How correlated are those? Is it not possible to treat them](#)
 73 [separately?](#)

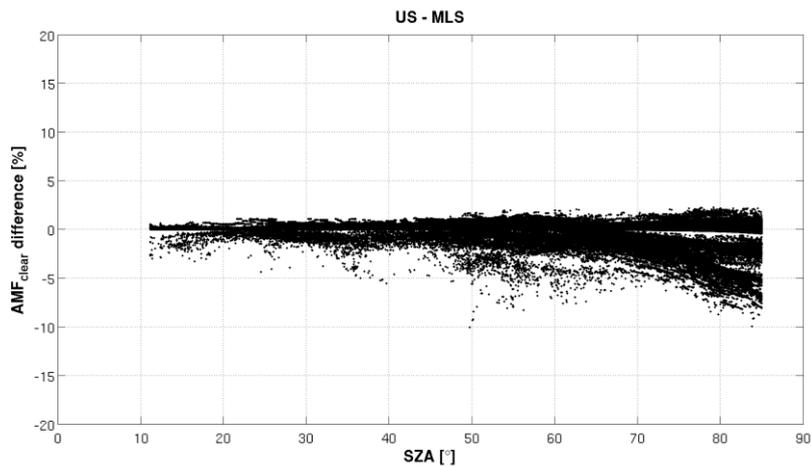
74 The stretch can be seen as a wavelength-dependent shift, therefore shift and stretch are always related.
 75 In our approach, the overall stretch results from the use of several sub-windows in which shifts are fitted
 76 separately. Note that within each sub-window, a linear stretch around the central wavelength is also
 77 applied to avoid biases in the retrieved shift values.

78

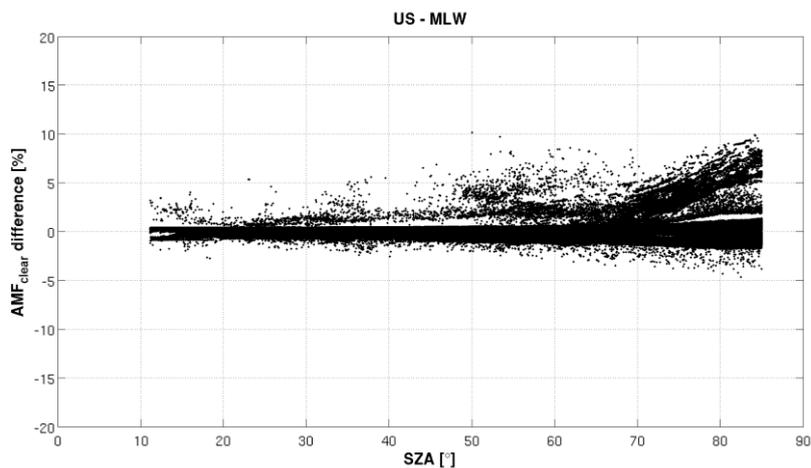
79 **Section 2.2.2 Tropospheric air mass factor**

80 [An evaluation of the effect of using only one atmospheric model \(US Standard\) for the calculation of](#)
 81 [altitude dependent air mass factors should be included. Ozone distribution can vary significantly](#)
 82 [between tropics, polar region and season.](#)

83 This effect is considered in section 3.1.2, Table 8 (structural uncertainties). The two figures hereafter
 84 show the comparison of OMI HCHO AMF for one day (20050202), using different atmospheric profiles
 85 (US atmosphere, mid-latitude summer, mid-latitude winter). Our RT simulations show that this effect is
 86 relatively small for HCHO tropospheric AMFs. It can reach 10% at large SZA, but remains below 5% for
 87 classical observation conditions of HCHO. In the future, we will consider to add an additional dimension
 88 for O₃ in the LUT of altitude resolved AMF, especially for sentinel-4 and its larger observation angles.



89



90

91 **Figure 2: Differences of OMI HCHO AMF for one day (20050202), using different**
 92 **atmospheric profiles (US atmosphere - mid-latitude summer (upper panel), US**
 93 **atmosphere - mid-latitude winter (lower panel)).**

94

95 Another subsection could be added to discuss the role of the surface reflectance properties to complete
 96 the description since all other AMF parameters have their own (LUT of altitude dependent AMFs, cloudy
 97 scenes, aerosols, and a priori vertical profile shapes).

98 The role of surface reflectance on HCHO AMFs is discussed in section 3.1.2. In section 2.2.2, our aim is to
 99 describe the algorithm, and for that, albedo is an auxiliary dataset to which we do not apply any
 100 modification.

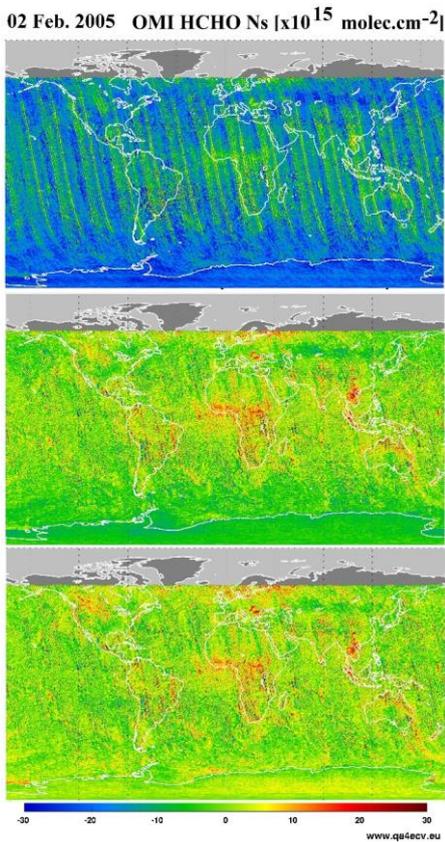
101

102 **Section 2.2.3. Across-track and zonal reference sector correction**

103 Page 20, line 470. "The natural background level of HCHO is well estimated from chemistry model
 104 simulations of CH₄". Actually there is evidence that models underpredict HCHO in the Pacific Ocean
 105 (<http://onlinelibrary.wiley.com/doi/10.1002/2016JD026121/abstract>). Add a sentence discussing this
 106 situation.

107 Thank you for this interesting paper. We added the reference in section 3.1.3 about the error on the
108 reference sector correction. Our estimate of this error contribution is in the range of 1 to 2 x 10e15
109 molec./cm2. This is compatible with the differences between model and observations shown in figure
110 14 of Anderson et al., 2017. Please note that for this part, only the error on the HCHO total columns in
111 the reference sector plays a role (not the profile). Note also that our reference sector is the Pacific Ocean,
112 more remote from the continent than the sector presented in Anderson et al., that is closer to Indonesia,
113 and therefore more influenced by continental sources.
114

115 As said above, despite the paper being fairly long, it could benefit of a plot showing the process of the
116 background correction for one orbit illustrating the changes for each step.
117 We added this figure in section 2.2.3:



118

119 **Figure 3: Illustration of the across-track and zonal reference sector correction steps applied to**
120 **one day of OMI HCHO slant columns (02/02/2005). The upper panel shows the uncorrected**
121 **slant columns obtained using as DOAS reference spectrum the solar irradiance. The center**
122 **panel shows the same slant columns after the first across-track correction step or when row**
123 **averaged radiances selected in the Pacific Ocean are used as reference. The lower panel shows**
124 **the final background corrected slant columns ΔN_s .**

125 **Section 3.1.1. Errors on the slant columns**
126 Please add O4 uncertainty to table 7.
127

128 We added 2% for an uncertainty of O4 cross-section. We tested the fit of an alternative cross-section
129 (Herman et al.); and the fit of O4 in a dedicated window (339-364nm). The impact is pretty small on
130 the absolute HCHO SCDs of one orbit.

131
132 **Section 3.1.2. Errors on air mass factors**

133 Page 26, Figure 9: Would it be possible to specify the geometry, surface pressure, and the rest of
134 parameters kept fix in the calculation for each panel.

135 **Figure 9: AMF uncertainty related to profile shape, cloud pressure and surface albedo errors, as**
136 **a function of different observation conditions. In all cases, we consider a nadir viewing and a**
137 **solar zenith angle of 30°. By default, fixed values have been used. The surface pressure is**
138 **1063hPa, the albedo is 0.05, the effective cloud fraction is 0.5, and the profile height and cloud**
139 **pressure are 795 hPa.**

140
141 **Page 26, Surface albedo:** Kleipool et al, surface climatology has a coarse resolution (0.5x0.5.)
142 compared with TROPOMI pixels. It would be interesting to incorporate in the error analysis the
143 uncertainties associated with subpixel inhomogeneity in the Kleipool database or at least discuss them
144 in the text.

145 It is true that the resolution of the Kleipool database is rather coarse in comparison to the size of Tropomi
146 pixels. We hope to switch to a new database based on Tropomi measurements, as soon as possible (this
147 has been added in Table 5). To our knowledge, there is no other LER database in the UV with a better
148 spatial resolution.

149 In the error budget, it is shown that the impact of albedo uncertainty is significantly lower than profile
150 shape uncertainty. This is because the albedo at UV wavelengths is generally small over regions where
151 VOCs are emitted. Although significantly improved, the 1° resolution of the TM-5 can introduce
152 uncertainty along coastlines for example. We have not done a detailed study using different model
153 resolutions to quantify those errors. It is currently the subject of several studies. We added a note about
154 the spatial resolution of auxiliary data in the paper.

155
156 **Page 27, Clouds and aerosols:** In section 2.2.2 “Tropospheric air mass factor” page 17, line 403 it is
157 said that pixels with cloud fractions below 10% are considered clear-sky pixels “to avoid unnecessary
158 error propagation through the retrievals” given the unstable cloud retrieval for such conditions. Under
159 that assumption, how are the AMF errors due to cloud parameters calculated?

160 Our approach is to consider an error on the cloud fraction (0.05), even for very low cloud fraction up to
161 0. However, we neglect the error on the cloud pressure, since it is weighted by a low cloud fraction
162 (equation 8).

163
164 **Page 27, Profile shape:** As for the surface climatology, would it be possible to estimate the uncertainty
165 derived from subpixel model inhomogeneity given that model information is available in 1° grid and
166 TROPOMI pixels can be as small as 7x7 km2.

167 See answer about albedo. It is currently difficult to provide a general and quantitative estimate of this
168 uncertainty since we do not have access to global models running at a resolution finer than 1°x1°. This
169 will be considered for future studies. However, we believe that our error estimate is conservative enough
170 to include these finer effects.

171
172 **Section 4. Verification**

173 **Building on the work by Lorente et al., 2017 was the AMF calculation tested using**
174 **harmonized parameters.**

175 The scattering weighting functions have been compared using harmonized parameters, showing
176 excellent agreement. Each group used its own auxiliary data.

177
178 **Page 34, Table 12:** The number of xx in the Earth coverage column for MAX-DOAS and
179 **Direct Sun, should it not be similar given that most MAX-DOAS instruments can also carry**
180 **on Direct Sun measurements?**

181 This statement is incorrect. Currently only a minor fraction of the operated static MAX-DOAS
182 instruments also feature direct-sun pointing capabilities (essentially Pandora systems).

183

184 Page 35, after paragraph devoted to MAX-DOAS could add a little paragraph describing Direct Sun
185 capabilities.

186 Some MAX-DOAS systems (e.g. Pandora instruments) include a direct-sun pointing capability allowing
187 for accurate total column measurements. It must be noted however that due to the faintness of the HCHO
188 spectral signatures and the small geometrical enhancement in direct-sun geometry in comparison to
189 MAX-DOAS, direct-sun measurements of HCHO are relatively difficult and not standard.

190

191 It will be nice to add a map of current ground-based measurements sites lined up for validation.

192 We rather refer to the ESA S5PVT project NIDFORVal, in which such a map is included
193 ([https://sentinel.esa.int/documents/247904/2474724/Sentinel-5P-Science-Validation-
194 Implementation-Plan](https://sentinel.esa.int/documents/247904/2474724/Sentinel-5P-Science-Validation-Implementation-Plan)).

195

196 **Section 5.4 Satellite-satellite intercomparisons**

197 Page 36, line 861: For completeness about current and future instruments it will be good to add
198 mentions to OMPS, GEMS, and TEMPO.

199 Ok, done.

200

201 **Technical Corrections:**

202 Page 1, line 33: "Its lifetime being of the order of a few hours, ..." is grammatically incorrect.
203 What about, "With its lifetime of the order of a few hours, HCHO concentrations in the boundary
204 layer..."

205 Ok, Thanks.

206

207 Page 2, line 43: Would you consider to add Kaiser et al., 2017 to the list of inversion studies
208 (<https://doi.org/10.5194/acp-2017-1137>)?

209 Done

210

211 Page 2, line 52: To complete the list of HCHO retrievals from LEO it should be added the ones
212 using OMPS measurements
213 (<http://onlinelibrary.wiley.com/doi/10.1002/2015GL063204/abstract>, [https://www.atmos-
214 meas-tech.net/9/2797/2016/](https://www.atmos-meas-tech.net/9/2797/2016/)).

215 Right, done.

216

217 Page 3, line 77: Is there any reference or link available to the S5P HCHO Level 2 Algorithm
218 Theoretical Basis Document v1.0

219 I added the proper reference (De Smedt et al., 2016).

220

221 Page 4, Table 1: What is the meaning of revisit time 24x3 hour. Since h is used in the top line it
222 will be good to make both units consistent (h or hour).

223 24x3h means a revisit time every 3 days, as it was the case for GOME-1 or SCIAMACHY. This is
224 how it appears in the official requirements.

225

226 Page 5, Figure 2: Geolocation and Time information also need to feed the HCHO climatology or
227 TM5 daily forecast.

228 Ok. I have modified the figure.

229

230 Page 6, line 156: "Figure 3 also" would read better if just said "Figure 3 presents"

231 Ok.

232

233 Page 8, Figure 3 caption: Mention that these vertical columns are derived using OMI data. It is
234 said in the text in section 2.2.1 (line 246) but the first time figure 3 is referenced in the text,
235 section 2.2 (line 156) nothing is mention and there may be misunderstandings.

236 Done.

237

238 Page 14, line 337: "average" should be "addition" or "sum".

239 Correct.
240
241 Page 14, line 341: The symbols for solar zenith angle, viewing zenith angle, and relative
242 azimuth angle are not defined in the text. Later on they are defined in table 4 but that only
243 happens in page 15.
244 Corrected.
245
246 Page 16, line 385: "in which a inhomogeneous" should read "in which an inhomogeneous"
247 Corrected.
248
249 Page 19, line 449: Clarify that they are OMI air mass factor for example saying "Yearly averaged
250 OMI air mass factors..."
251 Done
252
253 Page 28, line 642: "Equation (20)" should be "Equation (19)"?
254 right
255
256 Page 29, Table 9. "0.5 and 1.5x10¹⁵ molec.cm⁻²" should be "0.5 to 1.5x10¹⁵ molec.cm⁻²"?
257 Corrected.
258
259 Page 32, Figure 12 caption: Check "Error! Reference source not found"
260 Corrected.
261
262 Page 33, line 774: suggest to change the text between brackets to "(both for ground-based
263 measurements and for satellite columns)"
264 done
265
266 Page 35, line 827: "measurement" should read "measurements"?
267 Corrected.
268
269 Page 37, line 871: Move comma before spectral
270 done
271
272 Page 50, line 1206: Remove sentence "A complete description of the level 2 data..." since it's a repetition
273 of the sentence in line 1203.
274 Done. Thanks
275

276 **Anonymous Referee #2**
277 This work presents an extensive and detailed report of the retrieval algorithm of HCHO for TROPOMI on
278 board of Sentinel-5p. I believe that the manuscript is clearly written and well suits for the AMT so I
279 recommend the publication with a few minor comments for clarification as follows:
280
281 L162: Please enlarge the size of tick and labels in Figures 3, 4.
282 Ok
283
284 L208: "the assumption of a single effective light path" any supporting literature and previous studies
285 would be highly appreciated for this sentence.
286 This directly refers to the key approximations of the DOAS method, described in section 2.2.1 (Platt et
287 al., 1994; Platt and Stutz, 2008; and references therein; Gottwald et al., 2006)
288

289 L237-239: The BrO retrieved from the first fitting is used in the second fitting and it error may affect the
290 retrieval of HCHO in the second fitting but the error analysis associated with this was not included in the
291 manuscript.

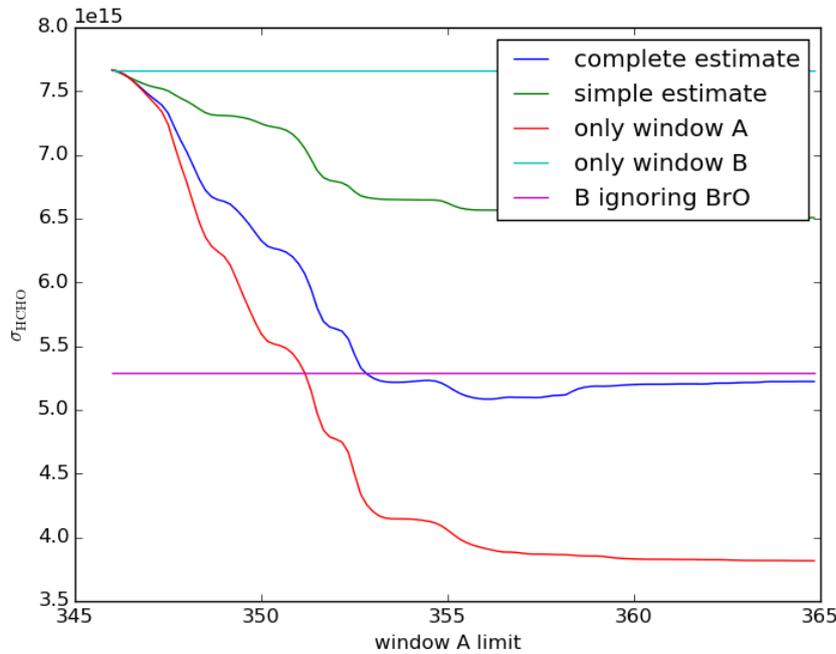
292 We made estimates of full random error calculation, when using coupled analysis intervals. The plot
293 below shows the estimated random error on the HCHO SCD from the least squares estimation, as the
294 limit of the first larger window is increased. Window A is the big window (328.5-...), window B is the
295 small one (328.5-346.0).

296 The red line is the error in the large window (only window A). The light blue is the error when everything
297 is fitted in the small window, without constraints (only window B).

298 The dark blue is the calculated random error when BrO is constrained in window B using the estimate
299 from window A (complete estimate). The error reaches an optimal value when window A goes up to
300 ~356nm, as done in our algorithm. The green "simple estimate" is the result when ignoring the
301 correlations between the BrO estimate and the results in the small window.

302 The violet line (B ignoring BrO) is what our actual version of QDOAS provides in the case where 2
303 coupled fitting intervals are used. This means it just ignores the error from the BrO estimation. So for
304 the settings we currently use, it's quite close to the result from the full calculation.

305 This theoretical estimate confirms the 50% reduction of the random error in the small interval when
306 pre-fitting BrO in the large interval, as observed with real data (figure 1 of this review).



307 **Figure 4: Estimated random error on the HCHO SCD, as the limit of the first window is increased.**
308 **Window A is the large window (328.5-...), window B is the small window (328.5-346.0).**

311 L246: I guess that the retrieved HCHO from the second interval in Fig. 3 is adjusted considering the
312 retrieved BrO from the first interval but is not clearly written in the manuscript.

313 The legend of figure 3 has been modified like this:

314 **Figure 3: Example of regional and monthly averages of the HCHO vertical columns over**
315 **different NMVOC emission regions, derived from OMI observations for the period 2005-2014.**
316 **Results of the retrievals in the two fitting intervals (1:328.5-359 nm and 2: 328.5-346 nm, with**

317 **BrO fitted in interval-1) are shown, as well as the magnitude of the background vertical column**
318 **($N_{v,0}$).**

319 L280, L297-298: Is the sub-interval larger or shorter than a fitting window for each species, for example,
320 HCHO? What if estimated shifts with the shorter subintervals differ within the fitting window, how
321 would you apply this to the calibration?

322 Sub-intervals are typically of the order of 10 nm wide, so shorter than the fitting window, and adjusted
323 to sample the variation of the shift (and therefore the stretch) over a spectral range encompassing the
324 different fitting intervals. To reconstruct the wavelength-dependent shift over the full range, a
325 polynomial is fitted through individual shift values. This ensure optimal wavelength registration of the
326 irradiance spectrum over the full spectral range of interest.

327
328 L299: It appears that the calibration of earth radiance follows the interpolation to the irradiance grid.
329 But the sequence should be reversed I guess, otherwise, a possible shift can interfere interpolated
330 radiance and needs to be corrected before the interpolation to the grids.

331 The final aim of the calibration procedure is to align as accurately as possible the irradiance, radiance
332 and absorption cross-sections. Our approach uses an external wavelength reference (the Kurucz atlas)
333 to align the solar irradiance using solar Fraunhofer lines. This step is very accurate since, in contrast to
334 radiance spectra, the solar irradiance is not perturbed by atmospheric absorption effects and therefore
335 can be very accurately aligned on the Kurucz atlas. Absorption cross-sections are then interpolated on
336 the resulting grid, which ensures very accurate alignment between solar irradiance and cross-sections.
337 The additional shift and stretch between radiance and irradiance spectra is further corrected as part of
338 the DOAS fit itself, again making use of Fraunhofer lines but also accounting for atmospheric absorbers.
339 This latter alignment is extremely accurate since solar and radiance spectra are recorded with the same
340 instrument and atmospheric effects are properly accounted for. We believe that this overall approach is
341 the most stable and accurate to deal with wavelength registration issues.

342
343 L367: Can you take into account the variation of ozone columns with latitude for AMF using the US
344 standard?

345 No, we do not take this into account. But the effect on the HCHO AMF is rather small. See my answer to
346 the same question from first reviewer.

347
348 Line 480-481: Using the measured radiance as reference spectrum instead of irradiance can reduce (or
349 remove) row-dependent offsets. Do we need to remove stripe patterns dependent on the row when the
350 measured radiances are used as reference?

351 This is almost not needed anymore (offsets are reduced from $2e16$ to $1e15$ molec.cm⁻²). We added a
352 figure to illustrate the effect of the difference steps of the background correction (will be figure 8 in the
353 new version).

354
355 Line 482-485, It would be appreciated if you can add some explanations about the cause for the
356 latitudinal dependent offset. For example, Khokhar et al. (2005) suggested that the interference of O3
357 absorption may cause the latitudinal offset. Does this affect HCHO and BrO? Any quantitative
358 information for the latitudinal offset? Does it also change in the two step fitting procedures with the wide
359 and narrow fitting windows, respectively? In addition, if you used the irradiance as reference it may
360 need to account for the latitudinal variation of O3 and BrO absorption in the fitting.

361 We believe that latitudinal offsets are mainly caused by spectral interference with strong ozone
362 absorption, and non-perfect fit for the Ring effect (which is also related to ozone absorption). The
363 magnitude of the offsets at large SZA is larger in the small fitting interval (328.5-346 nm), because ozone
364 absorption has more weight. The magnitude and sign of the offsets vary in space and time, depending
365 on the cause (Ozone or Ring). Absolute values can reach 1 to $2e16$ molec.cm⁻² for SZA larger than 70°.
366 Concerning radiance or irradiance as reference, there is not much difference regarding the latitudinal
367 variations, since radiance as reference are selected only around the equator. In both cases, a correction
368 is needed.

369 As said above, we added a figure. More explanations can also be found in De Smedt et al., 2015.

370
371 Figure 8: It appears that not all M' is linear. So the question is how you include values for the nonlinear
372 M' in the lookup table.

373 We do not consider M' as linear. We have calculated a LUT of M' using adapted grids for albedo, surface
374 elevation, profile height....

375

376 [Table 12: it is not clear what the use of x indicates.](#)

377 We added in the legend:

378 Table 12: Data/Masurement types used for the validation of satellite HCHO columns. The information
379 content of each type of measurement is qualitatively represented by the number of crosses.

380

381 **Algorithm Theoretical Baseline for formaldehyde retrievals**
382 **from S5P TROPOMI and from the QA4ECV project.**

383 Isabelle De Smedt¹, Nicolas Theys¹, Huan Yu¹, Thomas Danckaert¹, Christophe Lerot¹,
384 Steven Compernelle¹, Michel Van Roozendael¹, Andreas Richter², Andreas Hilboll², Enno
385 Peters², Mattia Pederghana³, Diego Loyola³, Steffen Beirle⁴, Thomas Wagner⁴, Henk Eskes⁵,
386 Jos van Geffen⁵, Klaas Folkert Boersma^{5,6}, Peepijn Veeffkind⁵.

387 [1]{Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium}

388 [2]{Institute of Environmental Physics, University of Bremen (IUP-B), Otto-Hahn-Allee 1, 28359 Bremen, Germany}

389 [3]{Institut für Methodik der Fernerkundung (IMF), Deutsches Zentrum für Luft und Raumfahrt (DLR), Oberpfaffenhofen, Germany}

390 [4]{ Max Planck Institute for Chemistry (MPIC), Hahn-Meitner-Weg 1, 55128 Mainz, Germany}

391 [5]{KNMI, De Bilt, The Netherlands}

392 [6]{Wageningen University, Meteorology and Air Quality group, Wageningen, The Netherlands}

393 *Correspondence to:* I. De Smedt (isabelle.desmedt@aeronomie.be)

394 **Abstract:** On board of the Copernicus Sentinel-5 Precursor (S5P) platform, the TROPOspheric Monitoring
395 Instrument (TROPOMI) is a double channel nadir-viewing grating spectrometer measuring solar back-
396 scattered earthshine radiances in the ultraviolet, visible, near-infrared and shortwave infrared with global daily
397 coverage. In the ultraviolet range, its spectral resolution and radiometric performance are equivalent to those
398 of its predecessor OMI, but its horizontal resolution at true nadir is improved by an order of magnitude. This
399 paper introduces the formaldehyde (HCHO) tropospheric vertical column retrieval algorithm implemented in
400 the S5P operational processor, and comprehensively describes its various retrieval steps. Furthermore,
401 algorithmic improvements developed in the framework of the EU FP7-project QA4ECV are described for
402 future updates of the processor. Detailed error estimates are discussed in the light of Copernicus user
403 requirements and needs for validation are highlighted. Finally, verification results based on the application of
404 the algorithm to OMI measurements are presented, demonstrating the performances expected for TROPOMI.

405 **1. Introduction**

406 Long term satellite observations of tropospheric formaldehyde (HCHO) are essential to support air quality and
407 chemistry-climate related studies from the regional to the global scale. Formaldehyde is an intermediate gas in
408 almost all oxidation chains of non-methane volatile organic compounds (NMVOC), leading eventually to CO₂
409 (Seinfeld and Pandis, 2006). NMVOCs are, together with NO_x, CO and CH₄, among the most important
410 precursors of tropospheric ozone. NMVOCs also produce secondary organic aerosols and influence the
411 concentrations of OH, the main tropospheric oxidant (Hartmann et al., 2013). The major HCHO source in the
412 remote atmosphere is CH₄ oxidation. Over the continents, the oxidation of higher NMVOCs emitted from
413 vegetation, fires, traffic and industrial sources results in important and localised enhancements of the HCHO
414 levels (as illustrated in Figure 5, Stavrakou et al., 2009a). With its lifetime of the order of a few hours, HCHO
415 concentrations in the boundary layer Its lifetime being of the order of a few hours, HCHO in the boundary layer
416 can be related to the release of short-lived hydrocarbons, which mostly cannot be observed directly from space.
417 Furthermore, HCHO observations provide information on the chemical oxidation processes in the atmosphere,
418 including CO chemical production from CH₄ and NMVOCs. The seasonal and inter-annual variations of the

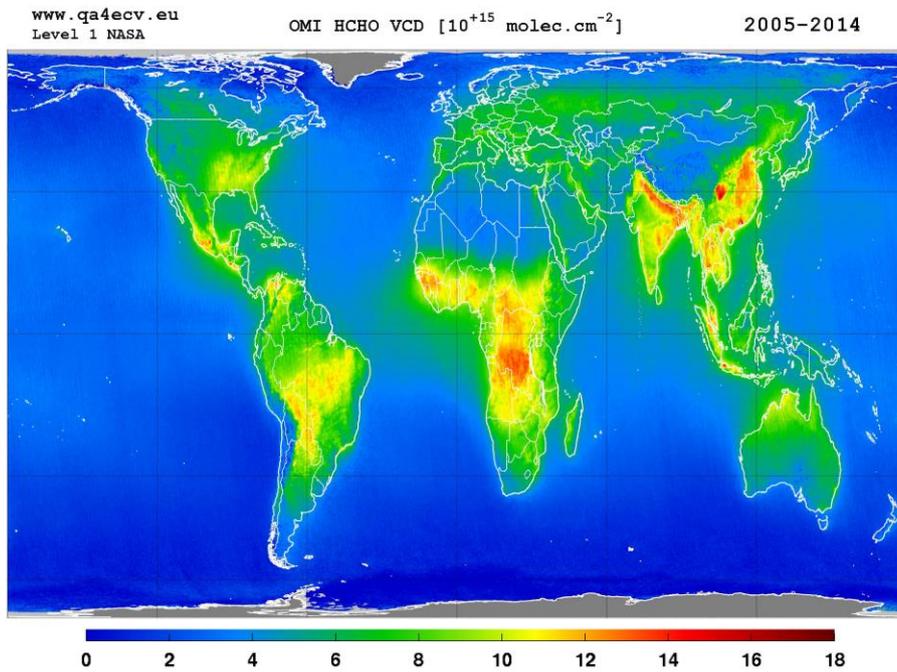
419 formaldehyde distribution are principally related to temperature changes (controlling vegetation emissions)
420 and fire events, but also to changes in anthropogenic activities ([Stavrakou et al., 2009b](#)). For all these reasons,
421 HCHO satellite observations are used in combination with tropospheric chemistry transport models to constrain
422 NMVOC emission inventories in so-called top-down inversion approaches (e.g. Abbot et al., 2003, Palmer et
423 al., 2006; Fu et al., 2007; Millet et al., 2008; Stavrakou et al., 2009a, 2009b, 2012, 2015; Curci et al., 2010;
424 Barkley et al., 2011, 2013; Fortems-Cheiney et al., 2012; Marais et al., 2012; Mahajan et al., 2015; [Kaiser et
425 al., 2017](#)).

426 HCHO tropospheric columns have been successively retrieved from GOME on ERS-2 and from SCIAMACHY
427 on ENVISAT, resulting in a continuous data set covering a period of almost 16 years from 1996 until 2012
428 (Chance et al., 2000; Palmer et al., 2001; Wittrock et al., 2006; Marbach et al., 2009; De Smedt et al., 2008;
429 2010). Started in 2007, the measurements made by the three GOME-2 instruments (EUMETSAT METOP-A,
430 B and C) have the potential to extend by more than a decade the successful time-series of global formaldehyde
431 morning observations (Vrekoussis et al., 2010; De Smedt et al., 2012; Hewson et al., 2012; Hassinen et al.,
432 2016). Since its launch in 2004, OMI on the NASA AURA platform has been providing complementary HCHO
433 measurements in the early afternoon with daily global coverage and a better spatial resolution than current
434 morning sensors (Kurosu et al., 2008; Millet et al., 2008; González Abad et al., 2015; De Smedt et al., 2015).
435 [On the S-NPP spacecraft, OMPS also allows to retrieve HCHO columns since end of 2011 \(Li et al., 2015;
436 González Abad, 2016\)](#). TROPOMI aims to continue this time series of early afternoon observations, with daily
437 global coverage, a spectral resolution and signal-to-noise ratio (SNR) equivalent to OMI, but combined with a
438 spatial resolution improved by an order of magnitude, which potentially offers an unprecedented view of the
439 spatiotemporal variability of NMVOC emissions.

440 To fully exploit the potential of satellite data, applications relying on tropospheric HCHO observations require
441 high quality long-term time series, provided with well characterized errors and averaging kernels, and
442 consistently retrieved from the different sensors. Furthermore, as the HCHO observations are aimed to be used
443 synergistically with other species observations (e.g. with NO₂ for air quality applications), it is essential to
444 homogenize as much as possible the retrieval methods as well as the external databases, in order to minimize
445 systematic biases between the observations. The design of the TROPOMI HCHO prototype algorithm,
446 developed at BIRA-IASB, has been driven by the experience developed with formaldehyde retrievals from the
447 series of precursor missions OMI, GOME(-2) and SCIAMACHY. Furthermore, within the S5P Level 2
448 Working Group project (L2WG), a strong component of verification has been developed involving independent
449 retrieval algorithms for each operational prototype algorithm. For HCHO, the University of Bremen (IUP-UB)
450 has been responsible of the algorithm verification. An extensive comparison of the processing chains of the
451 prototype (the retrieval algorithm presented in this paper) and verification algorithm has been conducted. In
452 parallel, within the EU FP7-project Quality Assurance for Essential Climate Variables (QA4ECV, Lorente et
453 al., 2017), a detailed step by step study has been performed for HCHO and NO₂ DOAS retrievals, including
454 more scientific algorithms (BIRA-IASB, IUP-UB, MPIC, KNMI and WUR), leading to state-of-the art
455 European products (www.qa4ecv.eu). Those iterative processes led to improvements that have been included
456 in the S5P prototype algorithm, or are proposed as options for future improvements of the operational
457 algorithm.

458 This paper gives a thorough description of the TROPOMI HCHO algorithm baseline, as implemented at the
459 German Aerospace Center (DLR) in the S5P operational processor UPAS-2 (Universal Processor for UV/VIS
460 Atmospheric Spectrometers). It reflects the S5P HCHO Level 2 Algorithm Theoretical Basis Document v1.0
461 (De Smedt et al., 2016) and also describes the options to be activated after the S5P launch, as implemented for
462 the QA4ECV OMI HCHO retrieval algorithm (see illustration in Figure 5).

463 In Section 2, we discuss the product requirements and the expected product performance in terms of precision
464 and trueness, and provide a complete description of the retrieval algorithm. In Section 3, the uncertainty of the
465 retrieved columns and the error budget is presented. Results from the algorithm verification exercise are given
466 in Section 4. The possibilities and needs for future validation of the retrieved HCHO data product can be found
467 in Section 5. Conclusions are given in Section 6.



468

469 **Figure 5: 10-years average of HCHO vertical columns retrieved from OMI between 2005 and 2014**
470 (<http://www.qa4ecv.eu/ecv/hcho-p/data>).

471 **2. TROPOMI HCHO algorithm**

472 **2.1 Product Requirements**

473 In the UV, the sensitivity to HCHO concentrations in the boundary layer is intrinsically limited from space due
474 to the combined effect of Rayleigh and Mie scattering that limit the fraction of radiation scattered back from
475 low altitudes and reflected from the surface to the satellite. In addition, ozone absorption reduces the number
476 of photons that reach the lowest atmospheric layers. Furthermore, the absorption signatures of HCHO are
477 weaker than those of other UV-Vis absorbers, such as e.g. NO₂. As a result, the retrieval of formaldehyde from
478 space is noise sensitive and error prone. While the precision (or random uncertainty) is mainly driven by the
479 signal to noise ratio of the recorded spectra, the trueness (or systematic uncertainty) is limited by the current
480 knowledge on the external parameters needed in the different retrieval steps.

481 The requirements for HCHO retrievals have been identified as part of the TROPOMI science objectives
482 document (van Weele et al., 2008), the COPERNICUS Sentinels-4/-5 Mission Requirements Document MRD
483 (Langen et al., 2011; 2017), and the S5P Mission Advisory Group report of the review of user requirements
484 for Sentinels-4/-5 (Bovensmann et al., 2011). The requirements for HCHO are summarised in Table 1.
485 Uncertainty requirements include retrieval errors as well as measurement (instrument-related) errors. Absolute
486 requirements (in total column units) relate to background conditions, while percentage values relate to elevated
487 columns.

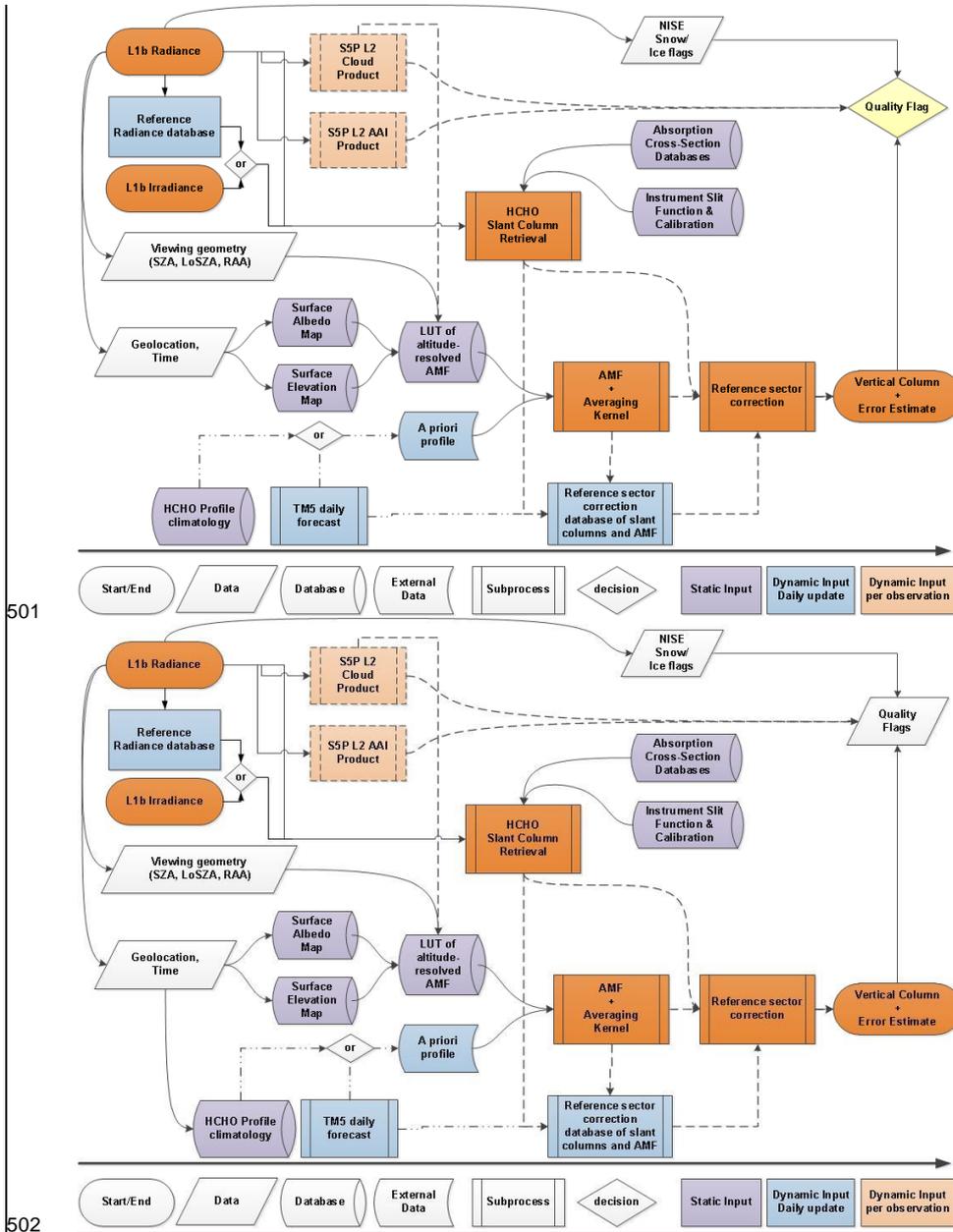
488 Three main COPERNICUS environmental themes have been defined as ozone layer (A), air quality (B), and
489 climate (C) with further division into sub themes. Requirements for HCHO have been specified for a number
490 of these sub themes (B1: Air Quality Protocol Monitoring, B2: Air Quality Near-Real Time, B3: Air Quality
491 Assessment, and C3: Climate Assessment). With respect to air quality protocol monitoring, which is mostly
492 concerned with trend and variability analysis, the requirements are specified for NMVOC emissions on
493 monthly to annual time scales and for larger region/country scale (Bovensmann et al., 2011). In the error
494 analysis section, we discuss these requirements and the expected performances of the HCHO retrieval
495 algorithm.

496 **Table 1: Requirements on HCHO vertical tropospheric column products as derived from the MRD.**
497 **Where numbers are given as "a - b", the first is the target requirement and the second is the threshold**
498 **requirement.**

Horizontal resolution- {km}	Revisit time	Theme	Required uncertainty
5-20 <u>km</u>	0.5-2 <u>hour</u>	B1, B2, B3	30-60% or 1.3×10^{15} molec.cm ⁻² (least stringent)
5-50 <u>km</u>	6 - 24x3 hour	C3	30% or 1.3×10^{15} molec.cm ⁻² (least stringent)

499

500 2.2 Algorithm description



503 Figure 6: Flow Diagram of the L2 HCHO retrieval algorithm implemented in the SSP operational
504 processor.

505 Figure 6 displays a flow diagram of the level-2 (L2) HCHO retrieval algorithm implemented in the S5P
506 operational processor. The baseline operation flow scheme is based on the Differential Optical Absorption
507 Spectroscopy (DOAS) retrieval method (Platt et al., 1994; Platt and Stutz, 2008; and references therein). It is
508 identical in concept to the one of SO₂ (Theys et al., 2017) and very close to the one of NO₂ (van Geffen et al.,
509 2017). The interdependencies with auxiliary data and other L2 retrievals, such as clouds, aerosols or surface
510 reflectance are also represented.

511 Following the diagram in Figure 6, the processing of S5P level-1b (L1b) data proceeds as follows: radiance
512 and irradiance spectra are read from the L1b file, along with geolocation data such as pixel coordinates and
513 observation geometry (sun and viewing angles). The relevant absorption cross section data as well as
514 characteristics of the instrument are used as input for the determination of the HCHO slant columns (N_s). In
515 parallel to the slant column fit, S5P cloud information and absorbing aerosol index (AAI) data are obtained
516 from the operational chain. Alongside, in order to convert the slant column to a vertical column (N_v), an air
517 mass factor (M) that accounts for the average light path through the atmosphere is calculated. For this purpose,
518 several auxiliary data are read from external (operational and static) sources: cloud cover data, topographic
519 information, surface albedo, and the a priori shape of the vertical HCHO profile in the atmosphere. The AMF
520 is computed by combining an a priori formaldehyde vertical profile and altitude-resolved air mass factors
521 extracted from a pre-computed look-up-table (also used as a basis for the error calculation and retrieval
522 characterization module). This look up table has been created using the VLIDORT 2.6 radiative transfer model
523 (Spurr et al., 2008a) at a single wavelength representative for the retrieval interval. It is used to compute the
524 total column averaging kernels (Eskes and Boersma, 2003), which provide essential information on the
525 measurement vertical sensitivity and are required for comparison with other types of data.

526 Background normalization of the slant columns is required in the case of weak absorbers such as formaldehyde.
527 Before converting the slant columns into vertical columns, background values of N_s are normalized to
528 compensate for possible systematic offsets (reference sector correction, see below). The tropospheric vertical
529 column end product results therefore from a differential column to which is added the HCHO background due
530 to methane oxidation, estimated using a tropospheric chemistry transport model.

531 The final tropospheric HCHO vertical column is obtained using the following equation:

$$N_v = \frac{N_s - N_{s,0}}{M} + N_{v,0} \quad (1)$$

532 The main outputs of the algorithm are the slant column density (N_s), the tropospheric vertical column (N_v), the
533 tropospheric air mass factor (M), and the values used for the reference sector correction ($N_{s,0}$ and $N_{v,0}$).

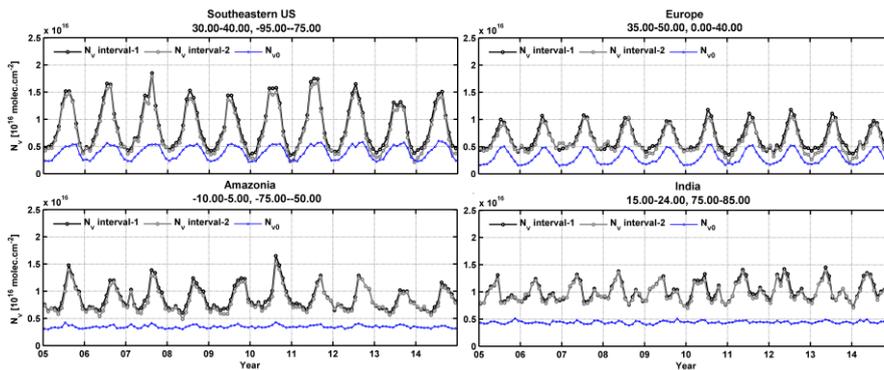
534 Complementary product information includes the clear sky air mass factor, the ~~error-uncertainty~~ on the total
535 column, the averaging kernel, and quality flags. Table 13 in the appendix B gives a non-exhaustive set of data
536 fields that are provided in the level 2 data product. A complete description of the level 2 data format is given
537 in the S5P HCHO Product User Manual (Pedernana et al., 2017).

538 Algorithmic steps are described in more details in the next sections, and settings are summarized in Table 2,
 539 along with algorithmic improvements developed in the framework of the EU FP7-project QA4ECV and
 540 proposed for future TROPOMI processor updates. Figure 7 also presents examples of monthly averaged
 541 HCHO vertical columns over four NMVOC emission regions, along with the background correction values.

542 **Table 2 : Summary of algorithm settings used to retrieve HCHO tropospheric columns from**
 543 **TROPOMI spectra. The last column lists additional features implemented in the QA4ECV HCHO**
 544 **product, which are options for future updates of the SSP Processor.**

Parameter	SSP Operational Algorithm	QA4ECV Algorithm
Slant Columns		
Fitting interval-1	328.5-359 nm	
Fitting interval-2	328.5-346 nm (N_s, BrO fixed by fit in interval-1)	
Absorption cross-sections	HCHO, Meller and Moortgat (2000), 298K NO ₂ , Vandaele et al. (1998), 220K Ozone, Serdyuchenko et al. (2013), 223 + 243K BrO, Fleischmann et al. (2004), 223K O ₂ -O ₂ , Thalman et al. (2013), 293K	
Ring effect	Ring cross-section based on the technique outlined by Chance et al. (1997), defined as I_{rrs}/I_{elas} , where I_{rrs} and I_{elas} are the intensities for inelastic (Rotational Raman Scattering; RRS) and elastic scattering processes.	
Non-linear O3 absorption effect	2 pseudo-cross sections from the Taylor expansion of the ozone slant column into wavelength and the O ₃ vertical optical depth (Puķīte et al., 2010).	
Slit function	One slit function per binned spectrum as a function of wavelength (Pre Flight Model, TROPOMI ISRF Calibration Key Data v1.0.0)	Fit of a prescribed function shape to determine the ISRF during wavelength calibration + online convolution of cross-sections.
Polynomial	5 th order	
Intensity offset correction	Linear offset ($1/I_0$)	
Iterative spike removal	Not activated.	Activated. Tolerance factor 5 (see section 2.2.1)
Reference spectrum I_0	Daily solar irradiance	Daily average of radiances, per row, selected in a remote region.
Air Mass Factors		
Altitude dependent AMFs	VLIDORT , 340 nm, 6-D AMF look-up table	
Treatment of partly cloudy scenes	IPA, no correction for $f_{eff} < 10\%$	
Aerosols	No explicit correction	
A priori profile shapes	TM5-MP 1°x1°, daily forecast (NRT) or reprocessed (Offline)	
Correction of surface pressure	Yes (Equation (10))	
Surface Albedo	OMI-based monthly minimum LER (update of Kleipool et al., 2008)	
Digital elevation map	GMTED2010 (Danielson et al., 2011)	
Cloud product	SSP operational cloud product, treating clouds as Lambertian reflectors (OCRA/ROCINN-CRB, Loyola et al., 2017)	OMI operational cloud algorithm, treating clouds as Lambertian

		reflectors (O ₂ -O ₂ , Veeffkind et al., 2016)
Background Correction		
Correction equation	$N_{v,0} = N_{v,0,CTM}$	$N_{v,0} = \frac{M_0}{M} N_{v,0,CTM}$ (see section 2.2.3)



545

546 **Figure 7: Example of regional and monthly averages of the HCHO vertical columns over different**
547 **NM VOC emission regions, derived from OMI observations for the period 2005-2014. Results of the**
548 **retrievals in the two fitting intervals (-1:328.5-359 nm and -2: 328.5-346 nm, with BrO fitted in**
549 **interval-1) are shown, as well as the magnitude of the background vertical column ($N_{v,0}$).**

550 **2.2.1 Formaldehyde slant column retrieval**

551 The DOAS method relies on the application of Beer-Lambert's law. The backscattered earthshine spectrum as
552 measured by the satellite spectrometer contains the strong solar Fraunhofer lines and additional fainter features
553 due to interactions taking place in the Earth atmosphere during the incoming and outgoing paths of the
554 radiation. The basic idea of the DOAS method is to separate broad and narrowband spectral structures of the
555 absorption spectra in order to isolate the narrow trace gas absorption features. In practice, the application of
556 the DOAS approach to scattered light observations relies on the following key approximations:

- 557
- 558 1. For weak absorbers the exponential function can be linearized and the Lambert-Beer law can be
559 applied to the measured radiance to which a large variety of atmospheric light paths contributes.
 - 560 2. The absorption cross-sections are assumed to be weakly dependent on temperature and
561 independent of pressure. This allows expressing light attenuation in terms of Beer-Lambert's law,
562 and (together with approximation 1) separating spectroscopic retrievals from radiative transfer
563 calculations by introducing the concept of one effective slant column density for the considered
564 wavelength window.
 - 565 3. Broadband variations are approximated by a common low-order polynomial to compensate for
566 the effects of loss and gain from scattering and reflections by clouds/air molecules and/or at the
Earth surface.

567 The DOAS equation is obtained by considering the logarithm of the radiance $I(\lambda)$ and the irradiance $E_0(\lambda)$ (or
 568 another reference radiance selected in a remote sector) and including all broadband variations in a polynomial
 569 function:

$$\ln \frac{I(\lambda)}{E_0(\lambda)} \cong - \sum_j \sigma_j(\lambda) N_{s,j} + \sum_p c_p \lambda^p \quad (2)$$

$$\tau_s^{meas}(\lambda) \cong \tau_s^{diff}(\lambda, N_{s,j}) + \tau_s^{smooth}(\lambda, c_p), \quad (3)$$

570 where the measured optical depth τ_s^{meas} is modelled using a highly structured part τ_s^{diff} and a broadband
 571 variation τ_s^{smooth} .

572 Equation (2) is a linear equation between the logarithm of the measured quantities (I and E_0), the slant column
 573 densities of all relevant absorbers ($N_{s,j}$) and the polynomial coefficients (c_p), at multiple wavelengths. DOAS
 574 retrievals consist in solving an over-determined set of linear equations, which can be done by standard methods
 575 of linear least squares fit (Platt and Stutz, 2008). The fitting process consists in minimizing the chi-square
 576 function, i.e. the weighted sum of squares derived from Equation (3):

$$X^2 = \sum_{i=1}^k \frac{(\tau_s^{meas}(\lambda_i) - \tau_s^{diff}(\lambda_i, N_{s,j}) - \tau_s^{smooth}(\lambda_i, c_p))^2}{\varepsilon_i^2} \quad (4)$$

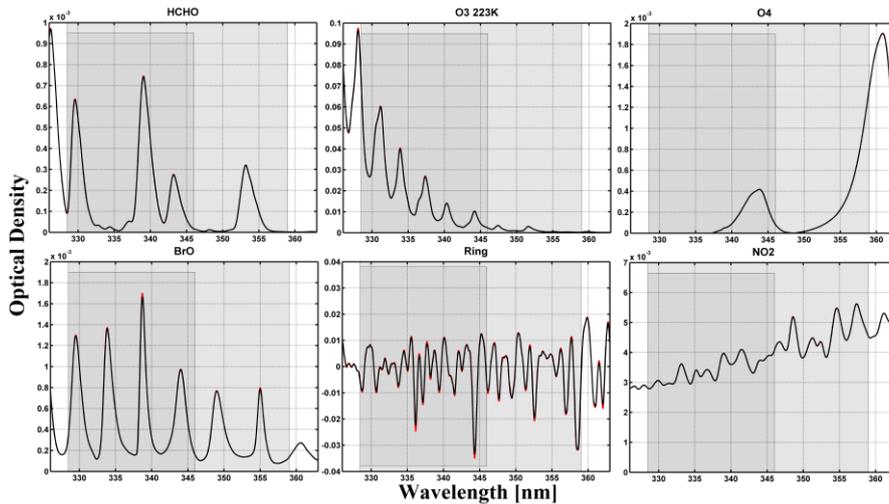
577 where the summation is made over the individual spectral pixels included in the selected wavelength range (k
 578 is the number of spectral pixels in the fitting interval). ε_i is the statistical uncertainty on the measurement at
 579 wavelength λ_i . Weighting the residuals by the instrumental errors ε_i is optional. When no measurement
 580 uncertainties are used (or no error estimates are available), all uncertainties in Equation (4) are set to $\varepsilon_i = 1$,
 581 giving all measurement points equal weight in the fit.

582 In order to optimize the fitting procedure, additional structured spectral effects have to be considered carefully
 583 such as the Ring effect (Grainger and Ring, 1962). Furthermore, the linearity of Equation (3) may be broken
 584 down by instrumental aspects such as small wavelength shifts between I and E_0 .

585 **Fitting intervals, absorption cross-sections and spectral fitting settings**

586 Despite the relatively large abundance of formaldehyde in the atmosphere (of the order of 10^{16} molec.cm⁻²)
 587 and its well-defined absorption bands, the fitting of HCHO slant columns in earthshine radiances is a challenge
 588 because of the low optical density of HCHO compared to other UV-Vis absorbers. The typical HCHO optical
 589 density is one order of magnitude smaller than that of NO₂ and three orders of magnitude smaller than that for
 590 O₃ (see Figure 8). Therefore, the detection of HCHO is limited by the signal to noise ratio of the measured
 591 radiance spectra and by possible spectral interferences and misfits due to other molecules absorbing in the same
 592 fitting interval, mainly ozone, BrO and O₄. In general, the correlation between cross-sections decreases if the
 593 wavelength interval is extended, but the assumption of a single effective light path defined for the entire
 594 wavelength interval may not be fully satisfied, leading to systematic misfit effects that may also introduce

595 biases in the retrieved slant columns. To optimize DOAS retrieval settings, a trade-off has to be found
 596 minimising these effects taking also into consideration the instrumental characteristics. A basic limitation of
 597 the classical DOAS technique is the assumption that the atmosphere is optically thin in the wavelength region
 598 of interest. At shorter wavelengths, the usable spectral range of DOAS is limited by rapidly increasing Rayleigh
 599 scattering and O₃ absorption. The DOAS assumptions start to fail for ozone slant columns larger than 1500 DU
 600 (Van Roozendael et al., 2012). Historically, different wavelength intervals have been selected between 325 and
 601 360 nm for the retrieval of HCHO using previous satellite UV spectrometers (e.g: GOME, Chance et al., 2000;
 602 SCIAMACHY, Wittrock et al., 2006, or GOME-2, Vrekoussis et al., 2010). The TEMIS dataset combines
 603 HCHO observations from GOME, SCIAMACHY, GOME-2 and OMI measurements retrieved in the same
 604 interval (De Smedt et al., 2008; 2012; 2015). The NASA operational and PCA OMI algorithm exploit a larger
 605 interval (Kurosu, 2008; González Abad et al., 2015, Li et al., 2015). The latest QA4ECV product uses the
 606 largest interval, thanks to the good quality of the OMI level 1 spectra. A summary of the different wavelength
 607 intervals is provided in Table 3.



608
 609 **Figure 8: Typical optical densities of HCHO, O₃, O₂-O₂, BrO, Ring effect, and NO₂ in the near UV. The**
 610 **slant columns have been taken as 1.3×10^{16} molec.cm⁻² for HCHO, 10^{19} molec.cm⁻² for O₃, 0.4×10^{43}**
 611 **molec.cm⁻⁵ for O₂-O₂, 10^{14} molec.cm⁻² for BrO, and 1×10^{16} molec.cm⁻² for NO₂. High resolution**
 612 **absorption cross-sections of Table 2 have been convolved with the TROPOMI ISFRs v1.0 (row 1 is**
 613 **shown in red and row 225 in black, see also Figure 9). The two fitting intervals (-1 and -2) used to**
 614 **retrieve HCHO slant columns are limited by grey areas.**

615 **Table 3: Wavelength intervals used in previous formaldehyde retrieval studies [nm].**

	GOME	SCIAMACHY	GOME-2	OMI
Chance et al., 2000	337.5-359			
Wittrock et al., 2006		334-348		

Vrekoussis et al., 2010			337-353	
Hewson et al., 2012			328.5-346	
González Abad et al., 2015; Li et al., 2015				328.5-356.5
De Smedt et al., 2008 ; 2012 ; 2015	328.5-346	328.5-346	328.5-346 (BrO in 328.5-359)	328.5-346 (BrO in 328.5-359)
QA4ECV		328.5-359	328.5-359	328.5-359

616 As for the TEMIS OMI HCHO product (De Smedt et al., 2015), the TROPOMI L2 HCHO retrieval algorithm
617 includes a two-step DOAS retrieval approach, based on two wavelength intervals:

- 618 1. 328.5-359 nm: This interval includes six BrO absorption bands and minimizes the correlation with
619 HCHO, allowing a significant reduction of the retrieved slant column noise. Note that this interval
620 includes part of a strong O₄ absorption band around 360 nm, which may introduce geophysical
621 artefacts of HCHO columns over arid soils or high altitude regions.
- 622 2. 328.5-346 nm: in a second step, HCHO columns are retrieved in a shorter interval, but using the BrO
623 slant column values determined in the first step. This approach allows to efficiently de-correlate BrO
624 from HCHO absorption while, at the same time, the O₄-related bias is avoided.

625 The use of a large fitting interval generally allows for a reduction of the noise on the retrieved slant columns.
626 However, a substantial gain can only be obtained if the level 1 spectra are of sufficiently homogeneous quality
627 over the full spectral range. Indeed, experience with past sensors not equipped with polarization scramblers
628 (e.g. GOME(-2) or SCIAMACHY) has shown that this gain can be partly or totally overruled due to the impact
629 of interfering spectral polarization structures (De Smedt et al., 2012; 2015). Assuming spectra free of spectral
630 features, the QA4ECV baseline option using one single large interval (fitting interval-1) will be applicable to
631 TROPOMI. Results of the retrievals from the two intervals applied to OMI are presented in Figure 7. In this
632 case, vertical column differences between the two intervals are generally lower than 10%. They can however
633 reach 20% in winter time.

634 In both intervals, the absorption cross-sections of O₃ at 223K and 243K, NO₂, BrO and O₄ are included in the
635 fit. The correction for the Ring effect, defined as I_{rrs}/I_{elas} where I_{rrs} and I_{elas} are the intensities for inelastic
636 (Rotational Raman Scattering; RRS) and elastic scattering processes, is based on the technique published by
637 Chance et al. (1997). Furthermore, in order to better cope with the strong ozone absorption at wavelengths
638 shorter than 336 nm, the method of Puķīte et al. (2010) is implemented. In this method, the variation of the
639 ozone slant column over the fitting window is taken into account. At the first order, the method consists in
640 adding two cross-sections to the fit: $\lambda\sigma_{O_3}$ and $\sigma_{O_3}^2$ (Puķīte et al., 2010; De Smedt et al.; 2012), using the O₃
641 cross-sections at 223K (close to the temperature at ozone maximum in the tropics). It allows a much better
642 treatment of optically thick ozone absorption in the retrieval and therefore to reduce the systematic
643 underestimation of the HCHO slant columns by 50 to 80%, for SZA from 50° to 70°.

644 To obtain the optical density (Equation (2)), the baseline option is to use the daily solar irradiance. A more
645 advanced option, implemented in QA4ECV, is to use daily averaged radiances, selected for each detector row,
646 in the equatorial Pacific (Lat: [-5° 5°], Long: [180° 240°]). The main advantages of this approach are (1) an
647 important reduction of the fit residuals (by up to 40%) mainly due to the cancellation of O₃ absorption and
648 Ring effect present in both spectra; (2) the fitted slant columns are directly corrected for background offsets
649 present in both spectra; (3) possible row-dependent biases (stripes) are greatly reduced by cancellation of small
650 optical mismatches between radiance and irradiance optical channels directly corrected owing to the use of one
651 reference per detector row; and (4) the sensitivity to instrument degradation affecting radiance measurements
652 is reduced because degradation these effects tend to cancel between the analyzed spectra and the references that
653 are used. It must be noted however that the last three effects can be mitigated when a solar irradiance is used
654 as reference, by means of a post-processing treatment applied as part of the background correction of the slant
655 columns (see section 2.2.3). The option of using an equatorial radiance as reference will be activated in the
656 operational processor after the launch of TROPOMI, during the commissioning phase of the instrument.

657 **Wavelength calibration and convolution to TROPOMI resolution**

658 The quality of the DOAS fit critically depends on the accuracy of the wavelength alignment between the
659 earthshine radiance spectrum, the reference (solar irradiance) spectrum and the absorption cross sections. The
660 wavelength registration of the reference spectrum can be fine-tuned to an accuracy of a few hundredths of a
661 nanometer by means of a calibration procedure making use of the solar Fraunhofer lines. To this end, a
662 reference solar atlas E_s accurate in wavelength to better than 0.01 nm (Chance and Kurucz, 2010) is degraded
663 to the resolution of the instrument, through convolution by the TROPOMI instrumental slit function (see Figure
664 9).

665 Using a non-linear least-squares approach, the shift (Δ_i) between the TROPOMI irradiance and the reference
666 solar atlas is determined in a set of equally spaced sub-intervals covering a spectral range large enough to
667 encompass all relevant fitting intervals. The shift is derived according to the following equation:

$$E_0(\lambda) = E_s(\lambda - \Delta_i) \quad (5)$$

668 where E_s is the reference solar spectrum convolved at the resolution of the TROPOMI instrument and Δ_i is the
669 shift in sub-interval i . A polynomial is fitted through the individual points to reconstruct an accurate wavelength
670 calibration $\Delta(\lambda)$ over the complete analysis interval. Note that this approach allows compensating for stretch
671 and shift errors in the original wavelength assignment. In the case of TROPOMI (or OMI), the procedure is
672 complicated by the fact that such calibrations must be performed and stored for each separate spectral field on
673 the CCD detector array. Indeed due to the imperfect characteristics of the imaging optics, each row of the
674 instrument must be considered as a separate detector for analysis purposes.

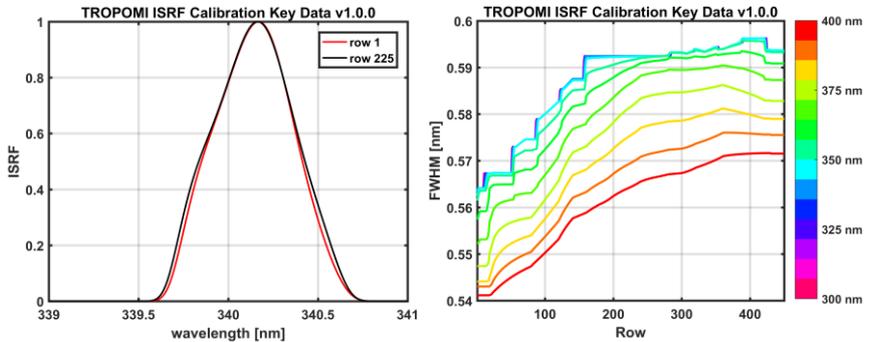
675 In a subsequent step of the processing, the absorption cross-sections of the different trace gases must be
676 convolved with the instrumental slit functions. The baseline approach is to use slit functions determined as part
677 of the TROPOMI key data. Slit functions, or Instrument Spectral Response Functions (ISRF), are delivered for

678 each binned spectrum and as a function of the wavelength as illustrated in Figure 9. Note that an additional
679 feature of the prototype algorithm allows to dynamically fit for an effective slit function of known line shape.
680 This can be used for verification and monitoring purpose during commissioning and later on during the mission.
681 This option is used for the QA4ECV OMI HCHO product.

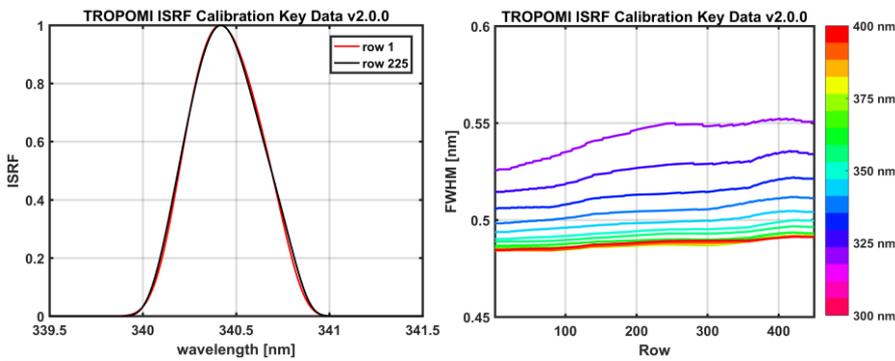
682 More specifically, wavelength calibrations are made for each orbit as follows:

- 683 • The irradiances (one for each binned row of the CCD) are calibrated in wavelength over the 325-360
684 nm wavelength range, using 5 sub-windows.
- 685 • The earthshine radiances are first interpolated on the original L1 irradiance grid. The irradiance
686 calibrated wavelength grid is assigned to those interpolated radiance values.
- 687 • The absorption cross-sections are interpolated (cubic spline interpolation) on the calibrated
688 wavelength grid, prior to the analysis.
- 689 • In the case where averaged radiances are used as reference, an additional step must be performed: the
690 cross-sections are aligned to the reference spectrum by means of shift/stretch values derived from a
691 least-squares fit of the calibrated irradiance towards the averaged reference radiance.
- 692 • During spectral fitting, shift and stretch parameters for the radiance are derived, to align each radiance
693 with cross sections and reference spectrum.

694



695



696 **Figure 9: Right panel: Examples of TROPOMI slit functions around 340 nm, for row 1 and row 225.**
 697 **Left panel: TROPOMI spectral resolution in channel 3, as a function of the row and the wavelength,**
 698 **derived from the instrument key data ISFR v2.0.0.**

699

700 Spike removal algorithm

701 A method to remove individual hot pixels or pixels affected by the South Atlantic Anomaly has been presented
 702 for NO₂ retrievals in Richter et al. (2011). Often only a few individual detector pixels are affected and in these
 703 cases, it is possible to identify and remove the outliers from the fit. However, as the amplitude of the distortion
 704 is usually only of the order of a few percent or less, it cannot always be found in the highly structured spectra
 705 themselves. Higher sensitivity for spikes can be achieved by analysing the residual of the fit where the
 706 contribution of the Fraunhofer lines, scattering, and absorption is already removed. When the residual for a
 707 single pixel exceeds the average residual of all pixels by a chosen threshold ratio (the tolerance factor), the
 708 pixel is excluded from the analysis, in an iterative process. This procedure is repeated until no further outliers
 709 are identified, or until the maximum number of iterations is reached (here fixed to 3). Tests performed with
 710 OMI spectra show that a tolerance factor of 5 improves the HCHO fits. This is especially important to handle
 711 the sensitivity of 2-D detector arrays to high energy particles. However, this improvement of the algorithm has
 712 a non-negligible impact on the time of processing (x 1.8). This option is activated in the QA4ECV algorithm,
 713 and will be activated in the TROPOMI operational algorithm in the next update of the processor.

714 2.2.2 Tropospheric air mass factor

715 In the DOAS approach, an optically thin atmosphere is assumed. The mean optical path of scattered photons
716 can therefore be considered as independent of the wavelength within the relatively small spectral interval
717 selected for the fit. One can therefore define a single effective air mass factor given by the ratio of the slant to
718 the vertical optical depth of a particular absorber j :

$$M_j = \frac{\tau_{s,j}}{\tau_{v,j}}. \quad (6)$$

719 In the troposphere, scattering by air molecules, clouds and aerosols leads to complex light paths and therefore
720 complex altitude-dependent air mass factors. Full multiple scattering calculations are required for the
721 determination of the air mass factors, and the vertical distribution of the absorber has to be assumed *a priori*.
722 For optically thin absorbers, the formulation of Palmer et al. (2001) is conveniently used. It decouples the
723 height-dependent measurement sensitivity from the vertical profile shape of the species of interest, so that the
724 tropospheric AMF (M) can be expressed as the sum average of the altitude dependent air mass factors (m_l)
725 weighted by the partial columns (n_{al}) of the a priori vertical profile in each vertical layer l , from the surface
726 up to the tropopause index (lt):

$$M = \frac{\sum_{l=1}^{lt} m_l(\lambda, \theta_0, \theta, \varphi, A_s, p_s, f_c, A_{cloud}, p_{cloud}) n_{al}(lat, long, time)}{\sum_{l=1}^{lt} n_{al}(lat, long, time)}, \quad (7)$$

727 where A_s is the surface albedo, p_s is the surface pressure, and f_c , A_{cloud} and p_{cloud} are respectively the cloud
728 fraction, cloud albedo and cloud top pressure.

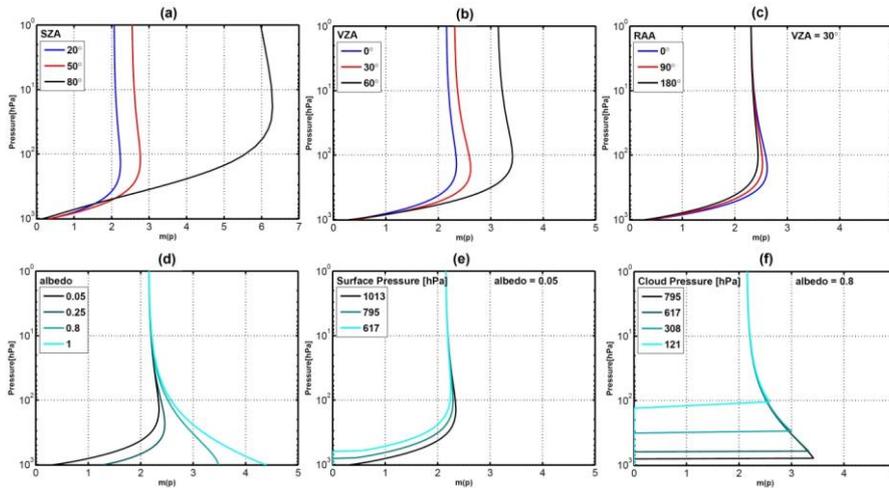
729 The altitude dependent air mass factors represent the sensitivity of the slant column to a change of the partial
730 columns $N_{v,j}$ at a certain level. In a scattering atmosphere, m_l depends on the wavelength, the viewing angles,
731 the surface albedo, and the surface pressure, but not on the partial column amounts or the vertical distribution
732 of the considered absorber (optically thin approximation).

733 LUT of altitude dependent air mass factors

734 Generally speaking, m depends on the wavelength, as scattering and absorption processes vary with
735 wavelength. However, in the case of HCHO, the amplitude of the M variation is found to be small (less than
736 5% for SZA lower than 70°) in the 328.5-346 nm fitting window and a single air mass factor representative for
737 the entire wavelength interval is used at 340 nm (Lorente et al., 2017).

738 Figure 10 illustrates the dependency of m with the observation angles, *i.e.* θ_0 (a), θ (b), and φ (c), and with
739 scene conditions like A_s (d) and p_s , for a weakly (e) or highly reflecting surface (f) (symbols in Table 4). The
740 decrease of sensitivity in the boundary layer is more important for large solar zenith angles and wide
741 instrumental viewing zenith angles. The relative azimuth angle does have relatively less impact on the
742 measurement sensitivity (note however that aerosols and BRDF effects are not included in those simulations).
743 In the UV, surfaces not covered with snow have an albedo lower than 0.1, while snow and clouds generally

744 present larger albedos. For a weakly reflecting surface, the sensitivity decreases near the ground because
 745 photons are mainly scattered, and scattering can take place at varying altitudes. Larger values of the surface
 746 albedo increase the fraction of reflected compared to scattered photons, increasing measurement sensitivity to
 747 tropospheric absorbers near the surface. Over snow or ice also multiple scattering can play an important role
 748 further increasing the sensitivity close to the surface.



749
 750 **Figure 10: Variation of the altitude dependent air mass factor with: (a) solar zenith angle, (b) viewing**
 751 **zenith angle, (c) relative azimuth angle between the sun and the satellite, (d) surface albedo, (e) surface**
 752 **pressure for a weakly reflecting surface, (f) surface pressure for a highly reflecting surface. Unless**
 753 **specified, the parameters chosen for the radiative transfer simulations are: SZA=30°, VZA=0°, RAA=0°, albedo=0.05, surface pressure=1063hPa, $\lambda=340\text{nm}$.**
 754

755 Altitude dependent air mass factors are calculated with the VLIDORT v2.6 radiative transfer model (Spurr,
 756 2008), at 340 nm, using an US standard atmosphere, for a number of representative viewing geometries, surface
 757 albedos and surface pressures (used both for ground and cloud surface pressures), and stored in a look-up table.
 758 Altitude dependent air mass factors are then interpolated within the lookup table for each particular observation
 759 condition and interpolated vertically on the pressure grid of the a priori profile, defined within the TM5-MP
 760 model (Williams et al., 2017). Linear interpolations are performed in $\cos(\theta_0)$, $\cos(\theta)$, relative azimuth angle
 761 and surface albedo, while a nearest neighbour interpolation is performed in surface pressure. The parameter
 762 values chosen for the look-up table are detailed in Table 4. In particular, the grid of surface pressure is very
 763 thin near the ground, in order to minimise interpolation errors caused by the generally low albedo of ground
 764 surfaces. Indeed, as illustrated by Figure 10 (e) and (f), the variation of the altitude dependent air mass factors
 765 is more discontinuous with surface elevation (low reflectivity) than with cloud altitude (high reflectivity).
 766 Furthermore, the LUT and model pressures are scaled to their respective surface pressures, in order to avoid
 767 extrapolations outside the LUT range.

768 **Table 4: Parameters in the altitude dependent air mass factors lookup table**

Parameter name	Nb. of grid points	Grid of values	Symbol
Solar zenith angle [°]	17	0, 10, 20, 30, 40, 45, 50, 55, 60, 65, 70, 72, 74, 76, 78, 80, 85	θ_0
Line of sight zenith angle [°]	10	0, 10, 20, 30, 40, 50, 60, 65, 70, 75	θ
Relative azimuth angle [°]	5	0, 45, 90, 135, 180	φ
Surface albedo	14	0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3 0.4, 0.6, 0.8, 1.0	A_s
Surface pressure [hPa]	17	1063.10, 1037.90, 1013.30, 989.28, 965.83, 920.58, 876.98, 834.99, 795.01, 701.21, 616.60, 540.48, 411.05, 308.00, 226.99, 165.79, 121.11	p_s
Atmospheric pressure [hPa]	64	1056.77, 1044.17, 1031.72, 1019.41, 1007.26, 995.25, 983.38, 971.66, 960.07, 948.62, 937.31, 926.14, 915.09, 904.18, 887.87, 866.35, 845.39, 824.87, 804.88, 785.15, 765.68, 746.70, 728.18, 710.12, 692.31, 674.73, 657.60, 640.90, 624.63, 608.58, 592.75, 577.34, 562.32, 547.70, 522.83, 488.67, 456.36, 425.80, 396.93, 369.66, 343.94, 319.68, 296.84, 275.34, 245.99, 210.49, 179.89, 153.74, 131.40, 104.80, 76.59, 55.98, 40.98, 30.08, 18.73, 8.86, 4.31, 2.18, 1.14, 0.51, 0.14, 0.03, 0.01, 0.001	p_t
Altitude corresponding to the atmospheric pressure, using an US standard atmosphere [km] (for information)	64	-0.35, -0.25, -0.15, -0.05, 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.10, 1.30, 1.50, 1.70, 1.90, 2.10, 2.30, 2.50, 2.70, 2.90, 3.10, 3.30, 3.50, 3.70, 3.90, 4.10, 4.30, 4.50, 4.70, 4.90, 5.25, 5.75, 6.25, 6.75, 7.25, 7.75, 8.25, 8.75, 9.25, 9.75, 10.50, 11.50, 12.50, 13.50, 14.50, 16.00, 18.00, 20.00, 22.00, 24.00, 27.50, 32.50, 37.50, 42.50, 47.50, 55.00, 65.00, 75.00, 85.00, 95.00	z_t

769 **Treatment of partly cloudy scenes**

770 The AMF calculations for TROPOMI will use the cloud fraction (f_c), cloud albedo (A_{cloud}) and cloud pressure
771 (p_{cloud}) from the S5P operational cloud retrieval, treating clouds as Lambertian reflectors (OCRA/ROCINN-
772 CRB, Loyola et al., 2017). The applied cloud correction is based on the independent pixel approximation
773 (Martin et al., 2002 and Boersma et al., 2004), in which an inhomogeneous satellite pixel is considered as a
774 linear combination of two independent homogeneous scenes, one completely clear and the other completely
775 cloudy. The intensity measured by the instrument for the entire scene is decomposed into the contributions
776 from the clear-sky and cloudy fractions. Accordingly, for each vertical layer, the altitude dependent air mass

777 factor of a partly cloudy scene is a combination of two air mass factors, calculated respectively for the cloud-
 778 free and cloudy fractions of the scene:

$$m_l = (1 - w_c)m_{l_clear}(A_s, p_s) + w_cm_{l_cloud}(A_{cloud}, p_{cloud}) \quad (8)$$

779 where m_{l_clear} is the altitude dependent air mass factor for a completely cloud-free pixel, m_{l_cloud} is the altitude
 780 dependent air mass factor for a completely cloudy scene, and the cloud radiance fraction w_c is defined as:

$$w_c = \frac{f_c I_{cloud}(A_{cloud}, p_{cloud})}{(1 - f_c)I_{clear}(A_s, p_s) + f_c I_{cloud}(A_{cloud}, p_{cloud})} \quad (9)$$

781 I_{clear} and I_{cloud} are respectively the radiance intensities for clear-sky and cloudy scenes whose values are
 782 calculated with VLIDORT at 340 nm and stored in look-up tables with the same grids as the altitude dependent
 783 air mass factors. m_{l_clear} and I_{clear} are evaluated for a surface albedo A_s and a surface pressure p_s , while
 784 m_{l_cloud} and I_{cloud} are estimated for a cloud albedo A_{cloud} and at the cloud pressure p_{cloud} . Note that the
 785 variations of the cloud albedo are directly related to the cloud optical thickness. Strictly speaking in a
 786 Lambertian (reflective) cloud model approach, only thick clouds can be represented (one should keep in mind
 787 that still the penetration of photons into the cloud is not covered by the Lambertian model). An effective cloud
 788 fraction corresponding to an effective cloud albedo of 0.8 ($f_{eff} = f_c \frac{A_c}{0.8}$) can be defined, in order to transform
 789 optically thin clouds into equivalent optically thick clouds of reduced horizontal extent. In such altitude
 790 dependent air mass factor calculations, a single cloud top pressure is assumed within a given viewing scene.
 791 For low effective cloud fractions (f_{eff} lower than 10%), the cloud top pressure retrieval is generally highly
 792 unstable and it is therefore reasonable to consider the observation as a clear-sky pixel (i.e. the cloud fraction is
 793 set to 0) in order to avoid unnecessary error propagation through the retrievals. This 10% threshold might be
 794 adjusted according to the quality of the cloud product (Veefkind et al., 2016; Loyola et al., 2017).

795 It should be noted that this formulation of the altitude dependent air mass factor for a partly cloudy pixel
 796 implicitly includes a correction for the HCHO column lying below the cloud and therefore not seen by the
 797 satellite, the so-called “ghost column”. Indeed, the total AMF calculation as expressed by (7) and (8) assumes
 798 the same a priori vertical profile in both cloudy and clear parts of the pixel and implies an integration of the
 799 profile from the top of atmosphere to the ground, for each fraction of the scene. The ghost column information
 800 is thus coming from the a priori profiles. For this reason, observations with cloud fractions f_{eff} larger than
 801 30% are assigned with a poor quality flag and have to be used with caution.

802 **Aerosols**

803 The presence of aerosol in the observed scene may affect the quality of the retrieval. No explicit treatment of
804 aerosols (absorbing or not) is foreseen in the operational algorithm as there is no general and easy way to treat
805 the aerosols effect on the retrieval. At computing time, the aerosol parameters (extinction profile, single
806 scattering albedo, ...) are unknown. However, the information on the AAI (Stein Zweers et al., 2016) will be
807 included in the L2 HCHO files as it gives information to the user on the presence of absorbing aerosols and the
808 affected data should be used and interpreted with care.

809 **A priori vertical profile shapes**

810 Formaldehyde concentrations decrease with altitude as a result of the near-surface sources of short-lived
811 NMVOC precursors, the temperature dependence of CH₄ oxidation, and the altitude dependence of photolysis.
812 The profile shape varies according to local NMHC sources, boundary layer depth, photochemical activity, and
813 other factors.

814 To resolve this variability in the TROPOMI near-real time HCHO product, daily forecasts calculated with the
815 TM5-MP chemical transport model (Huijnen et al., 2010, Williams et al., 2017) will be used to specify the
816 vertical profile shape of the HCHO distribution. TM5-MP will also provide a priori profile shapes for the NO₂,
817 SO₂, and CO retrievals. For the QA4ECV OMI products, high-resolution TM5-MP model runs were performed
818 for the period 2004-2016, and the model profiles from this run are used for both HCHO and NO₂ retrievals.

819 TM5-MP is operated with a spatial resolution of 1°x1° in latitude and longitude, and with 34 sigma pressure
820 levels up to 0.1hPa in the vertical direction. TM5-MP uses 3-hourly meteorological fields from the European
821 Centre for Medium Range Weather Forecast (ECMWF) operational model (ERA-Interim reanalysis data for
822 reprocessing, and the operational archive for real time applications and forecasts). These fields include global
823 distributions of wind, temperature, surface pressure, humidity, and (liquid and ice) water content, and
824 precipitation.

825 For the calculation of the HCHO air mass factors, the profiles are linearly interpolated in space and time, at
826 pixel centre and local overpass time, through a model time step of 30 minutes. To reduce the errors associated
827 to topography and the lower spatial resolution of the model compared to the TROPOMI 3.5x7 km² spatial
828 resolution, the a priori profiles need to be rescaled to effective surface elevation of the satellite pixel. Following
829 Zhou et al. (2009) and Boersma et al (2011), the TM5-MP surface pressure is converted by applying the
830 hypsometric equation and the assumption that the temperature changes linearly with height:

$$p_s = p_{s, TM5} \left(\frac{T_{TM5}}{T_{TM5} + \Gamma(z_{TM5} - z_s)} \right)^{-\frac{g}{R\Gamma}} \quad (10)$$

831 Where $p_{s, TM5}$ and T_{TM5} are the TM5-MP surface pressure and temperature, $\Gamma = 0.0065 \text{Km}^{-1}$ the lapse rate,
832 z_{TM5} the TM5-MP terrain height, and z_s surface elevation for the satellite ground pixel from a digital elevation

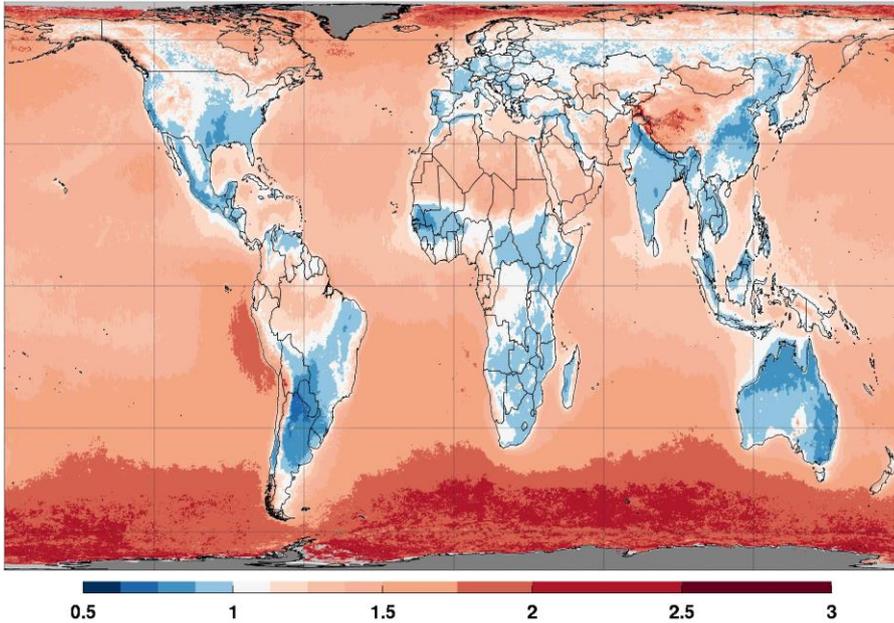
833 map at high resolution. $R=287 J kg^{-1} K^{-1}$ is the gas constant for dry air, and $g = 9.8ms^{-2}$ the gravitational
 834 acceleration.

835 The pressure levels for the a priori HCHO profiles are based on the improved surface pressure level p_s :
 836 $p_l = a_l + b_l p_s$, a_l and b_l being the constants that effectively define the vertical coordinate (Table 13).

837 Yearly averaged **OMI** air mass factors obtained using prior information summarized in Table 5, in particular
 838 TM5-MP HCHO profiles, are presented in Figure 11, in order to give an overview of the tropospheric AMF
 839 values and their global regional variations.

840 **Table 5: Prior information datasets used in the air mass factor calculation in the S5P HCHO**
 841 **operational algorithm and in the QA4ECV OMI algorithm.**

Prior information	Origin of data set	Resolution	Symbol
Surface Albedo	OMI-based monthly minimum LER (update of Kleipool et al., 2008) <u>When available, the TROPOMI-based LER product will be used.</u>	<ul style="list-style-type: none"> month 0.5°x0.5° (lat x long) 342 nm 	A_s
Digital elevation map	GMTED2010 (Danielson et al., 2011)	Average over the ground pixel area.	z_s
Cloud fraction	Operational cloud product based on a Lambertian cloud model (S5P: Loyola et al., 2017; OMI: Veeffkind et al., 2016).	For each ground pixel.	f_c
Cloud pressure			p_{cloud}
Cloud albedo			A_{cloud}
A priori HCHO profiles	Forecast (NRT) or reanalysis from TM5-MP CTM	<ul style="list-style-type: none"> Daily profiles at overpass time 1°x1° (lat x long) 34 sigma pressure levels up to 0.1hPa 	n_a



842

843 **Figure 11: Yearly averaged map of tropospheric air mass factors at 340 nm using the QA4ECV OMI**
 844 **HCHO algorithm. A priori HCHO profiles from high-resolution TM5-MP model runs have been used.**
 845 **The IPA cloud correction is applied for effective cloud fractions f_{eff} larger than 10%. Observations**
 846 **with f_{eff} larger than 30% have been filtered out.**

847 **2.2.3 Across-track and zonal reference sector correction**

848 Residual latitude-dependent biases in the columns, due to unresolved spectral interferences, are known to
 849 remain a limiting factor for the retrieval of weak absorbers such as HCHO. Retrieved HCHO slant columns
 850 can present large offsets depending on minor changes in the fit settings, and on minor instrumental spectral
 851 inaccuracies. Resulting offsets are generally global but also show particular dependencies, mainly with detector
 852 row (across-track) and with latitude (along-track). In the case of a 2D-detector array such as OMI or
 853 TROPOMI, across-track striping can possibly arise, due to imperfect calibration and different dead/hot pixel
 854 masks for the CCD detector regions. Offset corrections are also meant to handle some effects of the time-
 855 dependent degradation of the instrument.

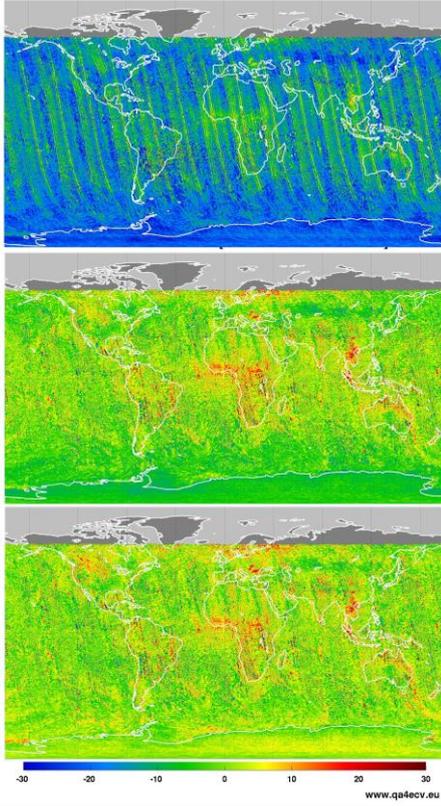
856 A large part of the resulting systematic HCHO slant column uncertainty is reduced by the application of a
 857 background correction, which is based on the assumption that the background HCHO column observed over
 858 remote oceanic regions (Pacific Ocean) is only due to methane oxidation. The natural background level of
 859 HCHO is well estimated from chemistry model simulations of CH₄ oxidation ($N_{v,0,CTM}$). It is ranging from 2
 860 to 4×10^{15} molec.cm⁻², depending on the latitude and the season (De Smedt et al., 2008; 2015; González Abad
 861 et al., 2015).

862 For the HCHO retrieval algorithm, we use a 2-steps normalization of the slant columns (see [Figure 12](#) and
863 [Table 6](#)):

- 864 • Across-track: the mean HCHO slant column is determined for each row in the reference sector around
865 the equator $[-5^\circ 5^\circ]$, $[180^\circ 240^\circ]$. Data selection is based on the slant column errors from the DOAS
866 fit and on the cloud fraction (threshold values are given in [Table 6](#)). Those mean HCHO values are
867 subtracted from all the slant columns of the same day, as a function of the row. The aim is to reduce
868 possible row-dependent offsets. In the case where solar irradiance are used as reference, those offsets
869 can exceed 2×10^{16} molec.cm⁻² (see the first panel of [Figure 12](#)). They are reduced below 10^{15}
870 molec.cm⁻² by this first step, or when row averaged radiances are used as reference, as in the
871 QA4ECV algorithm (middle panel of [Figure 12](#)).
- 872 • Along-track: the latitudinal dependency of the across-track corrected HCHO SCs is modelled by a
873 polynomial fit through their mean values, all rows combined, in 5° latitude bins in the reference sector
874 $[-90^\circ 90^\circ]$, $[180^\circ 240^\circ]$. Again, data selection is based on the slant column errors from the DOAS
875 fit and on the cloud fraction.

876 These two corrections are applied to the global slant columns so that in the reference sector, the mean
877 background corrected slant columns ($\Delta N_s = N_s - N_{s,0}$) are centered around zero ([lower panel of Figure 8](#)).

02 Feb. 2005 OMI HCHO Ns [$\times 10^{15}$ molec.cm⁻²]



878

879 Figure 12: Illustration of the across-track and zonal reference sector correction steps applied to one
880 day of OMI HCHO slant columns (02/02/2005). The upper panel shows the uncorrected slant columns
881 obtained using as DOAS reference spectrum the solar irradiance. The center panel shows the same
882 slant columns after the first across-track correction step or when row averaged radiances selected in
883 the Pacific Ocean are used as reference. The lower panel shows the final background corrected slant
884 columns ΔN_{v} .

885 Table 6: 2-steps normalization of the HCHO vertical columns

Correction	Region	Time frame	Column correction	Observation selection
Across-track	Equatorial Pacific Lat: [-5° 5°], Long: [180° 240°]	NRT: 1-week moving window	$dN_s(\text{row}) = N_s(\text{row}) - \overline{N_{s,0}(\text{row})}$	$\sigma_{N_s} \leq 3\overline{\sigma_{N_s}}$ $f_c \leq 0.4$
Zonal Along-track	Pacific Lat: [-90° 90°], Long: [180° 240°]	Offline: Daily correction	$\Delta N_s(\text{lat}) = dN_s(\text{lat}) - \overline{dN_{s,0}(\text{lat})}$ $\overline{N_{s,0,CTM}(\text{lat})} = \overline{M_0(\text{lat})N_{v,0,CTM}(\text{lat})}$	$\overline{dN_{s,0}(\text{lat})}$ $\leq 5e16$

886 To the corrected slant columns, the background HCHO values from a model have to be added. A latitude-
887 dependent polynomial is fitted daily through 5° latitude bin means of those modelled values in the reference
888 sector. Corresponding values are added to all the columns of the day. Strictly speaking, those background
889 values should be slant columns, derived as the product of air mass factors in the reference sector (M_0) with
890 HCHO vertical columns from the model ($N_{s,0,CTM} = M_0 N_{v,0,CTM}$) (González Abad et al., 2015). However, this
891 option requires the storage of the slant columns, the air mass factors, and their errors, in a separated database
892 (QA4ECV Algorithm and S5P option, see Equation (11)). An approximate solution is to add as background
893 the constant vertical column from the model ($N_{v,0,CTM}$), hence neglecting the variability of the M_0/M ratio. This
894 is the current implementation in the S5P algorithm, which will be updated with equation (11) after launch. For
895 NRT purpose, the evaluation in the reference sector is made using a moving time window of 1 week. For offline
896 processing, the reference sector correction can be refined by using daily evaluations.

$$N_v = \frac{N_s - N_{s,0}}{M} + N_{v,0} = \frac{\Delta N_s}{M} + \frac{M_0}{M} N_{v,0,CTM} = \frac{\Delta N_s + N_{s,0,CTM}}{M} \quad (11)$$

897 Figure 7 presents some examples of monthly and regionally averaged vertical columns, together with the
898 contribution of $N_{v,0}$. It should be realized that this contribution accounts for 20 to 50% of the vertical columns,
899 as expected from the large contribution of methane oxidation to the total HCHO column (Stavrakou et al.,
900 2015).

901 3. Uncertainty analyses

902 3.1 Uncertainty formulation by uncertainty propagation

903 The total uncertainty on the HCHO vertical column is composed of many sources of (random and systematic)
904 errors. In part those are related to the measuring instrument, such as errors due to noise or knowledge of the
905 slit function. In a DOAS-type algorithm, those instrumental errors propagate into the uncertainty of the slant
906 columns. Other types of error can be considered as model errors and are related to the representation of the
907 observation physical properties that are not measured. Examples of model errors are ~~uncertainties-errors~~ on the
908 trace gas absorption cross-sections, the treatment of clouds and ~~uncertainties-errors~~ of the a priori profiles.
909 Model errors can affect the slant columns, the air mass factors or the applied background corrections.

910 A formulation of the uncertainty can be derived analytically by ~~error-uncertainty~~ propagation, starting from the

911 equation of the vertical column (11) which directly results from the different retrieval steps. As the main
 912 algorithm steps are performed independently, they are assumed to be uncorrelated. The total uncertainty on
 913 the tropospheric vertical column can be expressed as (Boersma et al., 2004, De Smedt et al., 2008):

$$\sigma_{N,v}^2 = \left(\frac{\partial N_v}{\partial N_s} \sigma_{N,s}\right)^2 + \left(\frac{\partial N_v}{\partial M} \sigma_M\right)^2 + \left(\frac{\partial N_v}{\partial N_{s,0}} \sigma_{N,s,0}\right)^2 + \left(\frac{\partial N_v}{\partial M_0} \sigma_{M,0}\right)^2 + \left(\frac{\partial N_v}{\partial N_{v,0,CTM}} \sigma_{N,v,0,CTM}\right)^2 \quad (12)$$

$$\sigma_{N,v}^2 = \frac{1}{M^2} \left(\sigma_{N,s}^2 + \frac{(\Delta N_s + M_0 N_{v,0,CTM})^2}{M^2} \sigma_M^2 + \sigma_{N,s,0}^2 + N_{v,0,CTM}^2 \sigma_{M,0}^2 + M_0^2 \sigma_{N,v,0,CTM}^2 \right) \quad (13)$$

914 where $\sigma_{N,s}$, σ_M , $\sigma_{N,s,0}$, $\sigma_{M,0}$ and $\sigma_{N,v,0,CTM}$ are respectively the ~~error~~uncertainties on the slant column, the air
 915 mass factor, and the slant column correction, the air mass factor, and the model vertical column in the reference
 916 sector (indicated by suffix 0). For each of these categories, the following sections provide more details on the
 917 implementation of the uncertainty estimate in the HCHO algorithm. A discussion of the sources of uncertainties
 918 and, where possible, their estimated size are presented, as well as their spatial and temporal patterns.

919 Note that in the current implementation of the operational processor, $M_0 = M$, and the uncertainty formulation
 920 therefore reduces to:

$$\sigma_{N,v}^2 = \frac{1}{M^2} \left(\sigma_{N,s}^2 + \frac{\Delta N_s^2}{M^2} \sigma_M^2 + \sigma_{N,s,0}^2 \right) + \sigma_{N,v,0,CTM}^2 \quad (14)$$

921 Complementing this ~~error~~uncertainty propagation analysis, total column averaging kernels (A) based on the
 922 formulation of Eskes and Boersma (2003) are estimated. Column averaging kernels provide essential
 923 information when comparing measured columns with e.g. model simulations or correlative validation data sets,
 924 because they allow removing the effect of the a-priori HCHO profile shape used in the retrieval (see
 925 APPENDIX C: Averaging Kernel, Boersma et al., 2004; 2016).

926 Section 3.2 presents our current estimates of the precision (random uncertainty) and the trueness (systematic
 927 uncertainty) that can be expected for the TROPOMI HCHO vertical columns. They are discussed along with
 928 the product requirements (Section 2.1).

929 3.1.1 Errors on the slant columns

930 Error sources that contribute to the total uncertainty on the slant column originate both from instrument
 931 characteristics and from ~~error~~uncertainties in the DOAS slant column fitting procedure itself.

932 The retrieval noise for individual observations is limited by the SNR of the spectrometer measurements. A
 933 good estimate of the random variance of the reflectance (which results from the combined noise of radiance

934 and reference spectra) is given by the reduced χ^2 of the fit, which is defined as the sum of squares (4) divided
 935 by the number of degrees of freedom in the fit. The covariance matrix (Σ) of the linear least squares parameter
 936 estimate is then given by:

$$\Sigma = \frac{\chi^2}{(k-n)} (A^T A)^{-1} \quad (15)$$

937 where k is the number of spectral pixels in the fitting interval, n is the number of parameters to fit and the
 938 matrix $A(j \times k)$ is formed by the cross-sections. For each absorber j , the value $\sigma_{N,s,j}$ is usually called the slant
 939 column error (SCE or $\sigma_{N,s,rand}$).

$$\sigma_{N,s,j}^2 = \frac{\chi^2}{(k-n)} (A^T A)^{-1}_{j,j} \quad (16)$$

940 Equation (16) does not take into account systematic errors, that are mainly dominated by slit function and
 941 wavelength calibration uncertainties, absorption cross-section uncertainties, by interferences with other species
 942 (O_3 , BrO or O_4), or by uncorrected stray light effects. The choice of the retrieval interval can have a significant
 943 impact on the retrieved HCHO slant columns. The systematic contributions to the slant column errors are
 944 empirically estimated from sensitivity tests (see Table 7) and can be viewed as part of the structural uncertainty
 945 (Lorente et al., 2017). However, remaining systematic offsets and zonal biases are greatly reduced by the
 946 reference sector correction. All effects summed in quadrature, the various contributions are estimated to
 947 account for an additional systematic ~~error~~ uncertainty of 20% of the background-corrected slant column:

$$\sigma_{N,s,syst} = 0.2\Delta N_s \quad (17)$$

948 The total ~~error~~ uncertainty on slant columns is then:

$$\sigma_{N,s}^2 = \sigma_{N,s,rand}^2 + \sigma_{N,s,syst}^2 \quad (18)$$

949 **Table 7: Summary of the different error sources considered in the HCHO slant column uncertainty**
 950 **budget.**

Error source	Parameter uncertainty	Estimated uncertainty on HCHO SCD	Evaluation method - reference
Measurement noise	S/N=800-1000	1×10^{16} molec.cm ⁻² (random)	Value derived for individual observations by error <u>uncertainty</u> propagation; De Smedt et al., 2015;
HCHO cross-section error	Based on alternative cross-section datasets, offset and polynomial orders.	9%	Mean values derived from sensitivity tests using GOME-2 and OMI data. De Smedt et al., 2008; 2015 Hewson et al., 2013 Pinardi et al., 2013
O ₃ cross-section error		5%	
BrO cross-section error		5%	
NO ₂ cross-section error		3%	
<u>O₃ cross-section error</u>		<u>2%</u>	
Ring correction error		5%	
Choice of offset order		7%	

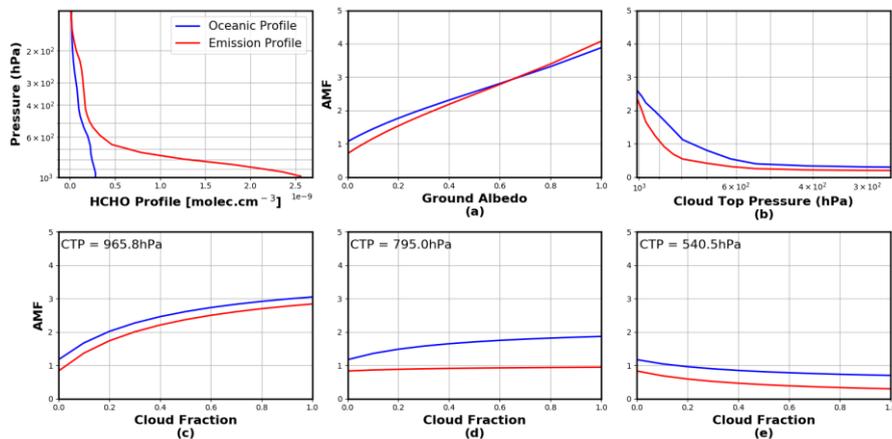
Choice of polynomial order		7%	
Instrumental slit function and wavelength calibration	Based on alternative calibrations	10%	Mean value derived from sensitivity tests using GOME-2 and OMI data.
Choice of wavelength interval	Based on alternative wavelength intervals	10%	Mean value derived from sensitivity tests using GOME-2 and OMI data. Hewson et al., 2013
Temperature dependence of the HCHO XS	0.05%/°K	2%	Mean value derived from sensitivity tests based on Meller and Moorgat (2000)

951 **3.1.2 Errors on air mass factors**

952 The ~~errors~~ **uncertainties** on the air mass factor depend on input parameter uncertainties and on the sensitivity
953 of the air mass factor to each of them. This contribution is broken down into the squared sum (Boersma et al.,
954 2004, De Smedt et al., 2008):

$$\sigma_M^2 = \left(\frac{\partial M}{\partial A_s} \cdot \sigma_{A,s}\right)^2 + \left(\frac{\partial M}{\partial f_c} \cdot \sigma_{f,c}\right)^2 + \left(\frac{\partial M}{\partial p_{cloud}} \cdot \sigma_{p,cloud}\right)^2 + \left(\frac{\partial M}{\partial s} \cdot \sigma_s\right)^2 + (0.2M)^2 \quad (19)$$

955 The contribution of each parameter to the total air mass factor error depends on the observation conditions.
956 The air mass factor sensitivities ($M' = \frac{\partial M}{\partial parameter}$), i.e. the air mass factor derivatives with respect to the
957 different input parameters, can be derived for any particular condition of observation using the altitude-
958 dependent AMF LUT, and using the model profile shapes (see Figure 13). In practice, a LUT of AMF
959 sensitivities has been created using coarser grids than the AMF LUT, and one parameter describing the shape
960 of the profile: the profile height, i.e. the altitude (pressure) below which resides 75% of the integrated HCHO
961 profile. $\frac{\partial M}{\partial s}$ is approached by $\frac{\partial M}{\partial s_h}$ where s_h is half of the profile height. Relatively small variations of this
962 parameter have a strong impact on the total air mass factors, because altitude-resolved air mass factors decrease
963 quickly in the lower troposphere, where the HCHO profiles peak (Figure 10).



964

965 **Figure 13: First panel: TM5-MP HCHO profiles extracted in June over the equatorial Pacific ocean**
 966 **(blue) and over Beijing (red). Those profiles have been used to calculate the tropospheric air mass**
 967 **factors shown in the panels a to e, representing the AMF dependence on (a) the surface albedo, (b) the**
 968 **cloud altitude, (c), (d), (e) the cloud fraction. In all cases, we consider a nadir view and a solar zenith**
 969 **angle of 30°. In (a) the pixel is cloud free, in (b) the albedo is 0.02 and the effective cloud fraction is 0.5,**
 970 **in (c), (d), (e) the ground albedo is 0.02 and the cloud pressure is respectively 966, 795 and 540 hPa.**

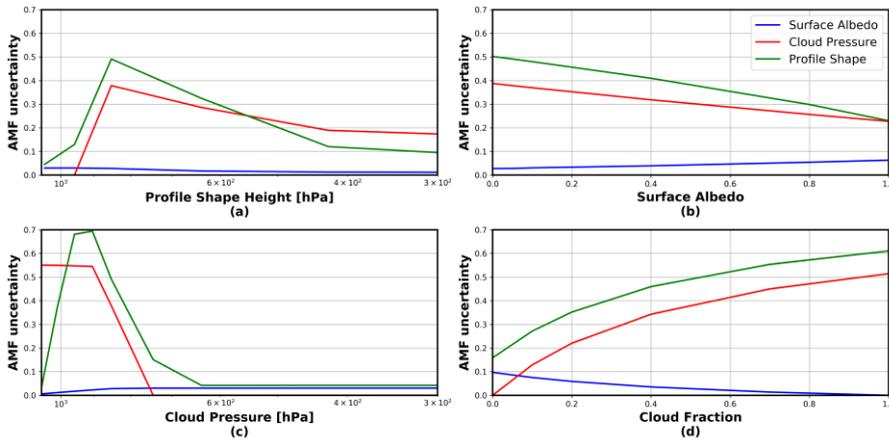
971 The ~~errors-uncertainties~~ $\sigma_{A,s}, \sigma_{f,c}, \sigma_{p,cloud}, \sigma_{s,h}$ are typical uncertainties on the surface albedo, cloud fraction,
 972 cloud top pressure and profile shape, respectively. They are estimated from the literature or derived from
 973 comparisons with independent data (see Table 8). Together with the sensitivity coefficients, these give the first
 974 four contributions on the right of equation (19). The fifth term on the right of equation (19) represents the
 975 uncertainty contribution due to possible errors in the AMF model itself (Lorente et al., 2017). We estimate this
 976 contribution to 20% of the air mass factor (see also section 3.2.2).

977 Estimates of the air mass factor uncertainties and of their impact on the vertical column uncertainties are listed
 978 in Table 8 and represented in Figure 14. They are based on the application of equation (19) to HCHO columns
 979 retrieved from OMI measurements. In expression (19), the impact of possible correlations between
 980 ~~uncertainties-errors~~ on parameters is not considered, like for example the surface albedo and the cloud top
 981 pressure. Note also that errors on the solar angles, the viewing angles and the surface pressure are supposed to
 982 be negligible, which is not totally true in practice, since equation (10) does not yield the true surface pressure
 983 but only a good approximation.

984 **Table 8: Summary of the different error sources considered in the air mass factor uncertainty budget.**

Input parameter error	Symbol	Parameter Uncertainty	Source	Estimated uncertainty on HCHO VCD
Surface albedo	σ_{A_s}	0.02	Kleipool et al., 2008	10-20%
Cloud fraction	$\sigma_{f,c}$	0.05	Veeffkind et al., 2016	05-15%
Cloud height	$\sigma_{p,cloud}$	50hPa		10-20%
Profile shape height	σ_s	75 100hPa	Upper limit of TM5-MP profile height standard deviation.	20-60%
AMF wavelength dependency	Model / Structural uncertainty	20 20%	Lorente et al., 2017	15-35%
LUT interp. errors				
Model atmosphere				
Cloud model/cloud correction/				
No explicit aerosol correction				

985



986

987 **Figure 14: AMF uncertainty related to profile shape, cloud pressure and surface albedo errors, as a**
 988 **function of different observation conditions. In all cases, we consider a nadir viewing and a solar zenith**
 989 **angle of 30°. By default, fixed values have been used. The surface pressure is 1063hPa, the albedo is 0.05,**
 990 **the effective cloud fraction is 0.5, and the profile height and cloud pressure are 795 hPa. AMF uncertainty**
 991 **related to profile shape, cloud pressure and surface albedo errors, as a function of different**
 992 **conditions.**

993

994 **Surface albedo**

995 A reasonable uncertainty on the albedo is 0.02 (Kleipool et al., 2008). This translates to an ~~error~~uncertainty
996 on the air mass factor using the slope of the air mass factor as a function of the albedo and can be evaluated for
997 each satellite pixel (equation (19)). As an illustration, Figure 13 (a) shows the air mass factor dependence on
998 the ground albedo for two typical HCHO profile shapes (in blue: remote profile, in red: emission profile). At
999 340nm, the AMF sensitivity (the slope), is almost constant with albedo, being only slightly higher for low
1000 albedo values. As expected, the AMF sensitivity to albedo is higher for an emission profile peaking near the
1001 surface than for a background profile more spread in altitude. More substantial errors can be introduced if the
1002 real albedo differs considerably from what is expected, for example in the case of the sudden snowfall or ice
1003 cover. Snow/ice cover map will therefore be used for flagging such cases.

1004 **Clouds and aerosols**

1005 An uncertainty on the cloud fraction of 0.05 is considered, while an uncertainty on the cloud top pressure of
1006 50hPa is taken. Figure 13 (b) shows the air mass factor variation with cloud altitude. The AMF is very sensitive
1007 to the cloud top pressure (the slope is steepest) when the cloud is located below or at the level of the
1008 formaldehyde peak. For higher clouds, the sensitivity of the air mass factor to any change in cloud pressure is
1009 very weak. As illustrated in Figure 13 (c), (d) and (e), for which a cloud top pressure of 966, 795 and 540 hPa
1010 is respectively considered, the sensitivity to the cloud fraction is mostly significant when the cloud lies below
1011 the HCHO layer.

1012 The effect of aerosols on the air mass factors are not explicitly considered in the HCHO retrieval algorithm.
1013 To a large extent, however, the effect of the non-absorbing part of the aerosol extinction is implicitly included
1014 in the cloud correction (Boersma et al., 2011). Indeed, in the presence of aerosols, the cloud detection algorithm
1015 is expected to overestimate the cloud fraction. Since non-absorbing aerosols and clouds have similar effects on
1016 the radiation in the UV-visible range, the omission of aerosols is partly compensated by the overestimation of
1017 the cloud fraction, and the resulting error on air mass factor is small, typically below 15% (Millet et al., 2006;
1018 Boersma et al., 2011; Lin et al., 2014; Castellanos et al., 2015; Chimot et al. 2015). In some cases, however,
1019 the effect of clouds and aerosols will be different. For example, when the cloud height is significantly above
1020 the aerosol layer, clouds will have a shielding effect while the aerosol amplifies the signal through multiple
1021 scattering. This will result in an underestimation of the AMF. Absorbing aerosols have also a different effect
1022 on the air mass factors, since they tend to decrease the sensitivity to HCHO concentration. In this case, the
1023 resulting error on the air mass factor can be as high as 30% (Palmer et al., 2001; Martin et al., 2002). This may,
1024 for example, affect significantly the derivation of HCHO columns in regions dominated by biomass burning
1025 as well as over heavily industrialized regions. Shielding and reflecting effect can thus occur, depending on the
1026 observation, decreasing or increasing the sensitivity to trace gas absorption. It has been shown that uncertainties
1027 related to aerosols is reduced by spatiotemporal averaging (Barkley et al., 2012; Lin et al., 2014; Castellanos
1028 et al., 2015; Chimot et al. 2015). Furthermore, the applied cloud filtering effectively removes observations with
1029 the largest aerosol optical depth. In the HCHO product, observations with an elevated absorbing aerosol index
1030 will be flagged, to be used with caution.

1031

1032 **Profile shape**

1033 This contribution to the total AMF error is the largest when considering monthly averaged observations. This
 1034 is supported by validation results using MAX-DOAS profiles measured around Beijing and Wuxi (see De
 1035 Smedt et al. 2015, Wang et al., 2016). Taking into account the averaging kernels allows removing from the
 1036 comparison the error related to the a priori profiles, when validating the results against other modelled or
 1037 measured profiles (see the APPENDIX C: Averaging Kernel).

1038 3.1.3 Errors on the reference sector correction

1039

$$\sigma_{N,v,0}^2 = \frac{1}{M^2} (\sigma_{N,s,0}^2 + N_{v,0,CTM}^2 \sigma_{M,0}^2 + M_0^2 \sigma_{N,v,0,CTM}^2) \quad (20)$$

1040 This ~~error uncertainty~~ includes contributions from the model background vertical column (see the recent study
 1041 of Anderson et al., 2017), from the error on the air mass factor in the reference sector, and from the amplitude
 1042 of the normalization applied to the HCHO columns. As mentioned in 3.1.1, we consider that $\sigma_{N,s,0}$ is taken into
 1043 account in Equation (17). The ~~uncertainty error~~ on the air mass factor in the reference sector $\sigma_{M,0}$ is calculated
 1044 as in Equation (19)(20) and saved during the background correction step. Uncertainty on the model background
 1045 has been estimated as the absolute values of the monthly averaged differences between two different CTM
 1046 simulations in the reference sector: IMAGES (Stavrakou et al., 2009a) and TM5-MP (Huijnen et al., 2010).
 1047 The differences range between 0.5 and 1.5×10^{15} molec.cm⁻².

1048 **Table 9: Estimated errors on the reference sector correction.**

Error source	Uncertainty on HCHO VCD	Evaluation method – reference
Model background	0.5 to and 1.5×10^{15} molec.cm ⁻²	Difference between IMAGES and TM model
Amplitude of the column normalisation ($N_{s,0}$)	0 to 4×10^{15} molec.cm ⁻²	Sensitivity tests using GOME-2 and OMI data.

1049 3.2 HCHO error estimates and product requirements

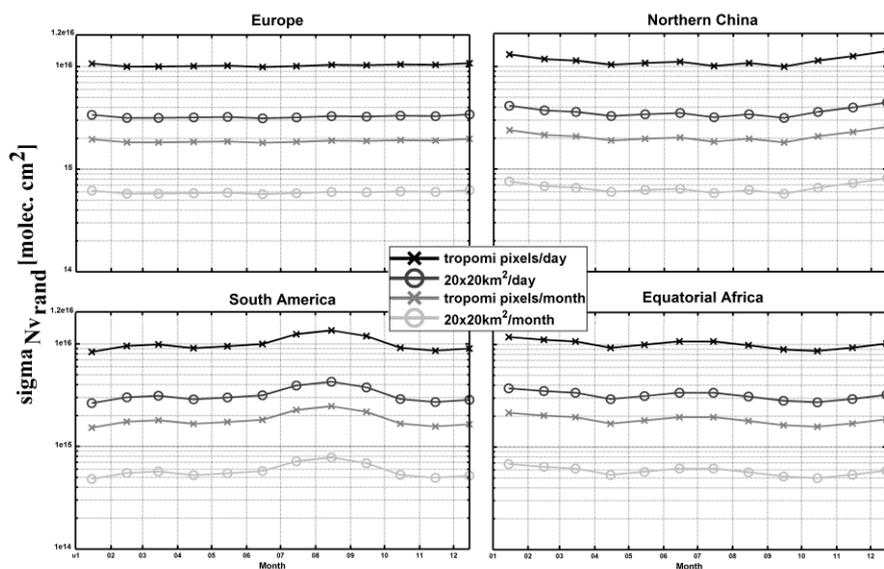
1050 This section presents estimates of the precision (random error) and trueness (systematic error) that can be
 1051 expected for the TROPOMI HCHO vertical columns. These estimates are given in different NMVOC emission
 1052 regions. Precision and trueness of the HCHO product are discussed against the user requirements.

1053 3.2.1 Precision

1054 When considering individual pixels, the total uncertainty is dominated by the random error on the slant
 1055 columns. Our simulations and tests on real satellite measurements show that the precision by which the HCHO
 1056 can be measured is well defined by the instrument signal-to-noise level. For the nominal SNR level (1000), the
 1057 expected precision of single-pixel measurements is equivalent to the precision obtained with OMI HCHO
 1058 retrievals (De Smedt et al., 2015), but with a ground pixel size of about 3.5×7 km², i.e. one order of magnitude
 1059 smaller in surface. Absolute $\sigma_{N,s,rand}$ values typically range between 7 and 12×10^{15} molec.cm⁻² for individual
 1060 pixels, showing an increase as a function of the surface altitude and of the solar zenith angle. Relative values
 1061 range between 100 and 300%, depending on the observation scene. In the case of HCHO retrievals, for
 1062 individual satellite ground pixels, the random ~~uncertainty error~~ on the slant columns is the most important

1063 source of uncertainty on the total vertical column. It can be reduced by averaging the observations, but of
 1064 course at the expense of a loss in time and/or spatial resolution.

1065 The precision of the vertical columns provided in the L2 files corresponds to the precision of the slant column
 1066 divided by the air mass factor
 1067 $\sigma_{N,v,rand} = \frac{\sigma_{N,s,rand}}{M}$ (see Table 13). It is dependent on the air mass factors, and therefore on the observation
 1068 conditions and on the cloud statistics. Figure 15 shows the vertical column precision that is expected for
 1069 TROPOMI, based on OMI observations in 2005. Results are shown in several regions, and at different spatial and
 1070 temporal scales (from individual pixels to monthly averaged column in 20x20km² grids). The product
 1071 requirements for HCHO measurements state a precision of 1.3x10¹⁵ molec.cm⁻². This particular requirement
 1072 cannot be achieved with individual observations at full spatial resolution. However, as represented in Figure
 1073 15, the requirement can be approached using daily observations at the spatial resolution of 20x20km² (close to
 1074 the OMI resolution) or using monthly averaged columns at the TROPOMI resolution. The precision can be
 1075 brought below 1x10¹⁵ molec.cm⁻² if a spatial resolution of 20x20km² is considered for monthly averaged
 1076 columns.



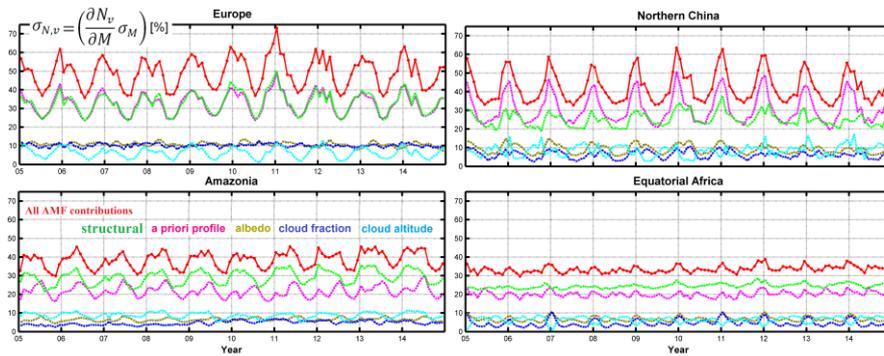
1077
 1078 **Figure 15: Estimated precision on the TROPOMI HCHO columns, in several NMVOC emission**
 1079 **regions, and at different spatial and temporal scales (from individual pixels to monthly averages in**
 1080 **20x20 km² grids). These estimated are based on OMI observations in 2005, using observations with an**
 1081 **effective cloud fraction lower than 40%.**

1082 3.2.2 Trueness

1083 In this section, we present monthly averaged values of the systematic vertical columns uncertainties estimated
 1084 for OMI retrievals between 2005 and 2014. The contribution of the air mass factor uncertainties is the largest

1085 contribution to the vertical column systematic uncertainties (see also Table 10). Figure 16 presents the VCD
 1086 uncertainties due to AMF errors, and the five considered contributions, over Equatorial Africa and Northern
 1087 China, as example of Tropical and mid-latitude sites. The largest contributions are from the a priori profile
 1088 uncertainty and from the structural uncertainty (taken as 20% of the AMF). In the case where the satellite
 1089 averaging kernels are used for comparisons with external HCHO columns, the a priori profile contribution can
 1090 be removed from the comparison uncertainty budget, leading to a total uncertainty in the range of 25% to 50%.
 1091 Table 10 wraps up the estimated relative contributions to the HCHO vertical column uncertainty, in the case
 1092 of monthly averaged columns for typical low and high columns.

1093 Considering these estimates of the HCHO column trueness, the requirements for HCHO product (30%) are
 1094 achievable in regions of high emissions and for certain times of the year. In any case, observations need to be
 1095 averaged to reduce random uncertainties at a level comparable or smaller than systematic uncertainties.



1096
 1097 **Figure 16: Regional and monthly average of the relative systematic vertical column AMF-related**
 1098 **uncertainties in several NMVOC emission regions, for the period 2005-2014. The 5 contributions to the**
 1099 **systematic air mass factor uncertainty are shown: structural (green), a priori profile (pink), albedo**
 1100 **(olive), cloud fraction (blue) and cloud altitude (cyan).**

1101 **Table 10: Estimated HCHO vertical column uncertainty budget for monthly averaged low and**
 1102 **elevated columns (higher than 1×10^{16} molec.cm⁻²). Contributions from the three retrieval steps are**
 1103 **provided, as well as input parameter contributions.**

HCHO vertical error uncertainty	Remote regions / low columns	Elevated column regions / periods
Contribution from systematic slant columns uncertainties	25%	15%
Contribution from air mass factors uncertainties	75%	30%
<ul style="list-style-type: none"> • from a priori profile errors • from model errors • from albedo errors • from cloud top pressure errors • from cloud fraction errors 	<ul style="list-style-type: none"> • 60% • 35% • 20% • 20% • 15% 	<ul style="list-style-type: none"> • 20% • 15% • 10% • 10% • 05%
Contribution from background correction uncertainties	40%	10%
Total	90%	35%
Total without smoothing error	50%	25%

1105 **4. Verification**

1106 In the framework of the TROPOMI L2 WG and QA4ECV projects, extensive comparisons of the prototype
1107 (this paper), the verification (IUP-UB), and alternative scientific algorithms (MPIC, KNMI, WUR) have been
1108 conducted. All follow a common DOAS approach. Prototype and verification algorithms have been applied to
1109 both synthetic and OMI spectra. Here, we present a selection of OMI results. For a complete description of the
1110 verification algorithm as well as results and discussion of the retrievals applied to synthetic spectra, please refer
1111 to the TROPOMI verification report (Richter et al., 2015).

1112 **4.1 Harmonized DOAS fit settings using OMI test data**

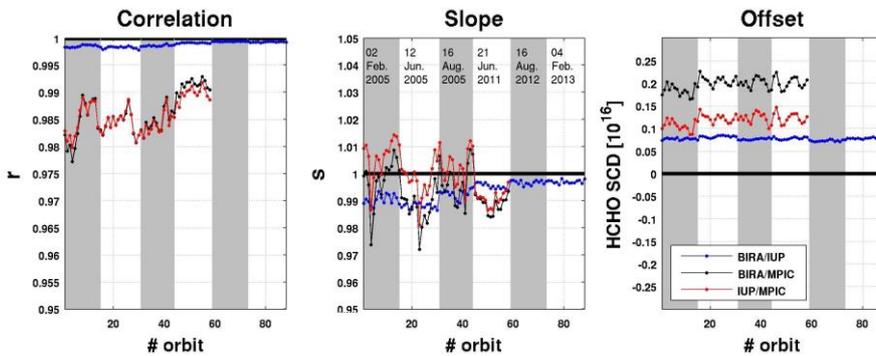
1113 For this exercise, a common set of DOAS fit parameters has been agreed upon. The goal of the intercomparison
1114 of harmonized fit settings was to ensure that the software implementation of the different algorithms behaves
1115 as expected in a large range of realistic measurement scenarios. Another objective was to gain knowledge on
1116 the level of agreement/disagreement of results from different groups when using the same settings, as well as
1117 on the main drivers for differences. Common and simple fit parameters based on the operational and
1118 verification algorithm were selected. They are summarized in Table 11.

1119 **Table 11: Common DOAS fit settings for HCHO using OMI data.**

Parameter	Values
Fitting interval-1	328.5-359 nm
Calibration	1 interval (328-359 nm), using the SAO 2010 solar atlas (Chance and Kurucz, 2010).
Molecular species	HCHO, NO ₂ , Ozone, BrO, O ₂ -O ₂ : same cross-sections as in Table 4
Ring effect	Ring cross-section based on the technique outlined by Chance et al. (1997)
Slit function	One slit function per binned spectrum as a function of wavelength (60 OMI ISRF, Dirksen et al., 2006).
Polynomial	5 th order
Intensity offset correction	Linear offset (1/I ₀)
Reference spectrum I ₀	Daily solar irradiance

1120
1121 The intercomparison of results using common settings allowed to identify and fix several issues in the different
1122 codes leading to an overall consolidation of the algorithms. It has been found that minor changes in the fit
1123 settings may lead to large offsets ($\pm 10 \times 10^{15}$ molec.cm⁻²) in the HCHO SCDs. However, an excellent level of
1124 agreement ($\pm 2 \times 10^{15}$ molec.cm⁻²) between the different retrieval codes was obtained after several iterations of
1125 the common settings. The main sources of discrepancies were found to be related to (1) the solar I₀ correction
1126 applied on the O₃ cross-sections, (2) the intensity offset correction, (3) the details of the wavelength calibration
1127 of the radiance and irradiance spectra, and (4) the OMI slit functions and their implementation in the
1128 convolution tools (Boersma et al., 2015).

1129 An overview of the final SCD comparison is shown on Figure 17 for six test days at the beginning and the end
 1130 of the OMI time series, and for a particular OMI orbit on the left panel of Figure 18. The correlation coefficient,
 1131 slope and offset of linear regression fits performed on each comparison orbit are displayed. The correlation of
 1132 slant columns from BIRA and IUP-UB is extremely high in most cases. It is > 0.998 for all orbits. The slope
 1133 of the regression line between BIRA and IUP-UB results is close to 1.0. There is a constant offset of less than
 1134 1×10^{15} molec.cm⁻². The comparison between MPIC results and the two other algorithms gives somehow lower
 1135 correlations, but still larger than 0.98 from the beginning to the end of the OMI lifetime. Final deviations on
 1136 OMI HCHO SCD when using common settings were found to be of maximum $\pm 2\%$ (slope) and 2.5×10^{15}
 1137 molec.cm⁻². When relating the remaining differences in retrieved SCDs using common settings to the slant
 1138 column errors from the DOAS fit ($\sigma_{N,S,rand}$), it can be concluded that the differences between the results are
 1139 significantly smaller than the uncertainties (from 10 to 20% of $\sigma_{N,S,rand}$). Moreover, remaining offsets in SCDs
 1140 are further reduced by the background correction procedure.
 1141



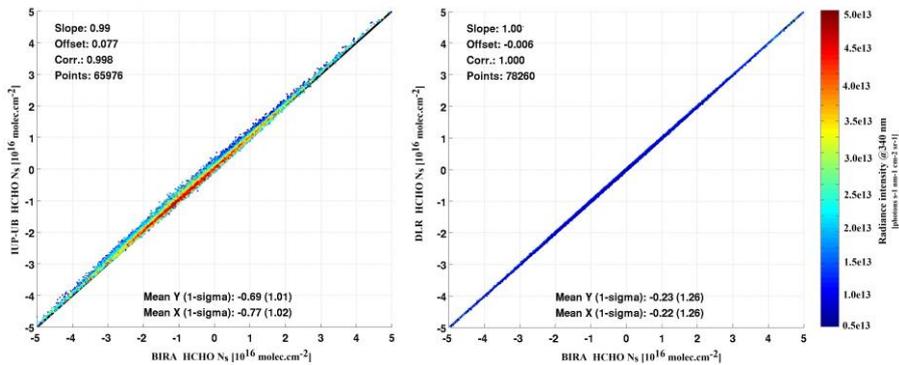
1142

1143 **Figure 17: Correlation (left), slope (middle) and offset (right) from a linear regression performed for**
 1144 **the common fit settings (see Table 11) for each orbit of OMI test days. A correlation plot for an**
 1145 **example orbit is provided in the left panel of Figure 18.** ~~Figure 13 Error! Reference source not found.~~

1146 4.2 Verification of the operational implementation

1147 A similar intercomparison exercise was performed with the operational algorithm UPAS, developed at DLR,
 1148 but using the exact settings of the prototype algorithm as detailed in Table 2. An example of resulting
 1149 correlation fit is shown in the right panel of Figure 18 for the same OMI orbit as for the comparison with the
 1150 IUP-UB results. The level of agreement between the prototype and operational results is found to be almost
 1151 perfect (correlation coefficient of 1, slope of 1.003 and offset of less than 0.2×10^{15} molec.cm⁻²), and very
 1152 satisfactory considering the sensitivity on small implementation changes.

1153



1154

1155 **Figure 18: Correlation plots of HCHO slant columns retrieved with the BIRA prototype algorithm and**
 1156 **(left) the IUP-UB verification algorithm, (right) the operational processor, for OMI orbit number 2339**
 1157 **on 02/02/2005, including all pixels with $SZA < 80^\circ$.**

1158

1159 5. Validation

1160 Independent validation activities are proposed and planned by the S5P Validation Team (Fehr, 2016) and within
 1161 the ESA S5P Mission Performance Center (MPC). The backbone of the formaldehyde validation is the MAX-
 1162 DOAS and FTIR networks operated as part of the Network for the Detection of Atmospheric Composition
 1163 Change (NDACC, www.ndsc.ncep.noaa.gov/) complemented by PANDONIA (pandonia.net/) and national
 1164 activities. In addition, model datasets will be used for validation as well as independent satellite retrievals.
 1165 Finally, airborne campaigns are planned to support the formaldehyde and other trace gases validation.

1166 5.1 Requirements for validation

1167 To validate the TROPOMI formaldehyde data products, comparisons with independent sources of HCHO
 1168 measurements are required. This includes comparisons with ground-based measurements, aircraft observations
 1169 and satellite data sets from independent sensors and algorithms. Moreover, not only information on the total
 1170 (tropospheric) HCHO column is needed but also information on its vertical distribution, especially in the lowest
 1171 three kilometres where the bulk of formaldehyde generally resides. In this altitude range, the a-priori vertical
 1172 profile shapes have the largest systematic impact on the satellite column errors. HCHO and aerosol profile
 1173 measurements are therefore needed.

1174 The diversity of the NMVOC species, lifetimes and sources (biogenic, biomass burning or anthropogenic) calls
 1175 for validation data in a large range of locations worldwide (tropical, temperate and boreal forests, urban and
 1176 sub-urban areas). Continuous measurements are needed to obtain good statistics ([as well both](#) for ground-based
 1177 measurements [and as](#) for satellite columns) and to capture the seasonal variations. Validation and assessment
 1178 of consistency with historical satellite datasets require additional information on the HCHO diurnal variation,
 1179 which depends on the precursor emissions and on the local chemical regime.

1180 The main emphasis is on quality assessment of retrieved HCHO column amounts on a global scale and over
 1181 long time periods. The validation exercise will establish whether HCHO data quality meets the requirements
 1182 of geophysical research applications like long term trend monitoring on the global scale, NMVOC source
 1183 inversion, and research on the budget of tropospheric ozone. In addition, the validation will investigate the
 1184 consistency between TROPOMI HCHO data and HCHO data records from other satellites.

1185 **5.2 Reference measurement techniques**

1186 Table 12 summarizes the type of data and measurements that can be used for the validation of the TROPOMI
 1187 HCHO columns. The advantages and limitations of each technique are discussed. It should be noted that, unlike
 1188 tropospheric O₃ or NO₂, the stratospheric contribution to the total HCHO column can be largely neglected
 1189 which simplifies the interpretation of both satellite and ground-based measurements.

1190 **Table 12: Data/Masurement types used for the validation of satellite HCHO columns. The**
 1191 **information content of each type of measurement is qualitatively represented by the number of crosses.**

Type of measurement	Sensitivity in the boundary layer	Vertical profile information	Diurnal variation	Seasonal Variation	Total column	Earth coverage
MAX-DOAS	xxx	xx (3)	xxx	xxx	xx	xx
FTIR	x	-	xxx	xx	xxx	x
Direct Sun	xxx	-	xxx	xxx	xxx	x
In situ (1)	xx	-	xxx	xxx	-	xx
Aircraft (2)	xx	xxx	x	-	xx (4)	x
Satellite instruments	x	-	x	xxx	xx	xxx (5)

- 1192
 1193 (1) Surface measurements that could be combined with regional modelling.
 1194 (2) Including ultra-light and unmanned airborne vehicles.
 1195 (3) Up to 2-3 km.
 1196 (4) Profiles generally need to be extrapolated.
 1197 (5) Different daily coverage and spatial resolutions.

1198 The Multi-axis DOAS (MAX-DOAS) measurement technique has been developed to retrieve stratospheric and
 1199 tropospheric trace gas total columns and profiles. The most recent generation of MAX-DOAS instruments
 1200 allows for measurement of aerosols and a number of tropospheric pollutants, such as NO₂, HCHO, SO₂, O₄
 1201 and CHOCHO (e.g. Irie et al., 2011). With the development of operational networks such as Pandonia
 1202 (<http://pandonia.net/>), it is anticipated that many more MAX-DOAS instruments will become available in the
 1203 near future to extend validation activities in other areas where HCHO emissions are significant. The locations
 1204 where HCHO measurements are required are reviewed in the next section. Previous comparisons between
 1205 GOME-2 and OMI HCHO monthly averaged columns with MAX-DOAS measurements recorded by BIRA-
 1206 IASB in the Beijing city centre and in the sub-urban site of Xianghe showed that the systematic differences
 1207 between the satellite and ground-based HCHO columns (about 20 to 40%) are almost completely explained

1208 when taking into account the vertical averaging kernels of the satellite observations (De Smedt et al., 2015,
1209 Wang et al., 2017), showing the importance of validating the a priori profiles as well.

1210 HCHO columns can also be retrieved from the ground using FTIR spectrometers. In contrast to MAXDOAS
1211 systems which essentially probe the first two kilometres of the atmosphere, FTIR instruments display a strong
1212 sensitivity higher up in the free troposphere and are thus complementary to MAXDOAS (Vigouroux et al.,
1213 2009). The deployment of FTIR instruments of relevance for HCHO is mostly taking place within the NDACC
1214 network. Within the project NIDFORVal (SPP Nitrogen Dioxide and Formaldehyde Validation using NDACC
1215 and complementary FTIR and UVVis networks), the number of FTIR stations providing HCHO time-series
1216 has been raised from only 4 (Vigouroux et al., 2009; Jones et al., 2009; Viatte et al., 2014; Franco et al., 2015)
1217 to 21. These stations are covering a wide range of HCHO concentrations, from clean Arctic or oceanic sites to
1218 sub-urban and urban polluted sites, as well as sites with large biogenic emissions such as Porto Velho (Brazil)
1219 or Wollongong (Australia).

1220 Although ground-based remote-sensing DOAS and FTIR instruments are naturally best suited for the validation
1221 of column measurements from space, in-situ instruments can also bring useful information. This type of
1222 instrument can only validate surface HCHO concentrations, and therefore additional information on the vertical
1223 profile (e.g. from regional modelling) is required to make the link with the satellite retrieved column. However,
1224 in-situ instruments (where available) have the advantage to be continuously operated for pollution monitoring
1225 in populated areas, allowing for extended and long term comparisons with satellite data (see e.g. Dufour et al.,
1226 2009). Although more expensive and with a limited time and space coverage, aircraft campaigns provide
1227 unique information on the HCHO vertical distributions (Zhu et al., 2017).

1228 **5.3 Deployment of validation sites**

1229 Sites operating correlative measurements should preferably be deployed at locations where significant
1230 NMVOC sources exist. This includes:

- 1231 • Tropical forests (Amazonian forest, Africa, Indonesia): The largest HCHO columns worldwide are
1232 observed over these remote areas that are difficult to access. Biogenic and biomass burning emissions are
1233 mixed. A complete year is needed to discriminate the various effects on the HCHO retrieval. Clouds tend
1234 to have more systematic effects in tropical regions. Aircraft measurements are needed over biomass
1235 burning areas.
- 1236 • Temperate forests (South-Eastern US, China, Eastern Europe): In summer time, HCHO columns are
1237 dominated by biogenic emissions. Those locations are useful to validate particular a-priori assumptions
1238 such as model isoprene chemistry and OH oxidation scheme. Measurements are mostly needed from April
1239 to September.
- 1240 • Urban and sub-urban areas (Asian cities, California, European cities): Anthropogenic NMVOCs are more
1241 diverse, and have a weaker contribution to the total HCHO column than biogenic NMVOCs. This type of
1242 signal is therefore more difficult to validate. Continuous observations at mid-latitudes over a full year are
1243 needed, to improve statistics.

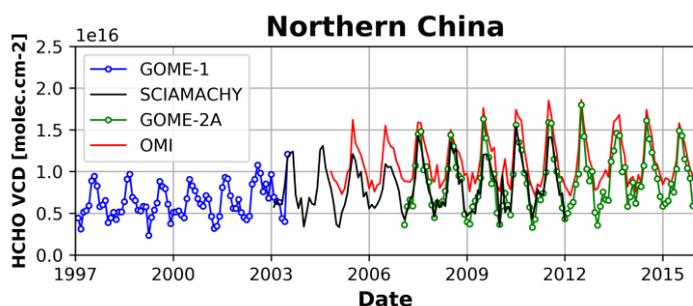
1244 For adequate validation, the long-term monitoring should be complemented by dedicated campaigns. Ideally
1245 such campaigns should be organised in appropriate locations such as e.g. South-Eastern US, Alabama where
1246 biogenic NMVOCs and biogenic aerosols are emitted in large quantities during summer time, and should
1247 include both aircraft and ground-based components.

1248 5.4 Satellite-satellite intercomparisons

1249 Satellite-satellite intercomparisons of HCHO columns are generally more straightforward than validation using
1250 ground-based correlative measurements. Such comparisons are evaluated in a meaningful statistical sense
1251 focusing on global patterns and regional averages, seasonality, scatter of values and consistency between
1252 results and reported uncertainties. When intercomparing satellite measurements, special care has to be drawn
1253 to:

- 1254 • differences in spatial resolutions, resulting in possible offsets between satellite observations (van
1255 der A et al., 2008; De Smedt et al., 2010; Hilboll et al., 2013),
- 1256 • differences in overpass times, that holds valuable geophysical information about diurnal cycles
1257 in emissions and chemistry (De Smedt et al., 2015; Stavrakou et al., 2015)
- 1258 • differences in a priori assumptions.
- 1259 • differences in the cloud algorithms and cloud correction schemes.

1260 Assessing the consistency between successive satellite sensors is essential to allow for scientific studies making
1261 use of the combination of several sensors. For example trends in NMVOC emissions have been successfully
1262 derived from GOME(-2), SCIAMACHY, and OMI measurements (Figure 19). It is anticipated that TROPOMI,
1263 the next GOME-2 instruments, OMPS, GEMS, TEMPO and the future Sentinel-4 and -5, will allow to extend
1264 these time series.



1265

1266 **Figure 19: HCHO columns over Northern China as observed with GOME (in blue), SCIAMACHY (in**
1267 **black), GOME-2 (in green), and OMI (in red) (De Smedt et al., 2008; 2010; 2015).**

1268 **6. Conclusions**

1269 The retrieval algorithm for the TROPOMI formaldehyde product generation is based on the heritage from
1270 algorithms successfully developed for the GOME, SCIAMACHY, GOME-2 and OMI sensors. A double-
1271 interval fitting approach is implemented, following an algorithm baseline demonstrated on the GOME-2 and
1272 OMI sensors. The HCHO retrieval algorithm also includes a post-processing across-track reference sector
1273 correction to minimize OMI-type striping effects, if any. Additional features for future processor updates
1274 include the use of daily earthshine radiance as reference selected in the remote Pacific₂ spectral₇ outlier
1275 screening during the fitting procedure (spike removal algorithm), and a more accurate background correction
1276 scheme.

1277 A detailed uncertainty budget is provided for every satellite observation. The precision of the HCHO
1278 tropospheric column is expected to come close to the COPERNICUS product requirements in regions of high
1279 emissions and, at mid-latitude, for summer (high sun) conditions. The trueness of the vertical columns is also
1280 expected to be improved, owing to the use of daily forecasts for the estimation of HCHO vertical profile shapes,
1281 that will be provided by a new version of the TM5-MP model, running at the spatial resolution of 1x1 degree
1282 in latitude and longitude.

1283 The validation of satellite retrievals in the lower troposphere is known to be challenging. Ground-based
1284 measurements, where available, often sample the atmosphere at different spatial and temporal scales than the
1285 satellite measurements, which leads to ambiguous comparisons. Additional correlative measurements are
1286 needed over a variety of regions, in particular in the Tropics and at the sub-urban level in mid-latitudes. These
1287 aspects are covered by a number of projects developed in the framework of the TROPOMI validation plan
1288 (Fehr, 2016).

1289

1290 **Acknowledgements**

1291 The TROPOMI HCHO algorithmic developments have been supported by the ESA Sentinel-5 Precursor Level-
1292 2 Development project, as well as by the Belgian PRODEX (TRACE-S5P project). Multi-sensor HCHO
1293 developments have been funded by the EU FP7 QA4ECV project (grant no. 607405), in close cooperation with
1294 KNMI, University of Bremen, MPIC-Mainz and WUR.

1295 **7. References**

1296 Abbot, D. S., Palmer, P. I., Martin, R. V., Chance, K. V., Jacob, D. J. and Guenther, A.: Seasonal and
1297 interannual variability of North American isoprene emissions as determined by formaldehyde column
1298 measurements from space, *Geophys. Res. Lett.*, 30(17), 1886, 2003.

1299 [Anderson, D. C., Nicely, J. M., Wolfe, G. M., Hanisco, T. F., Salawitch, R. J., Canty, T. P., Dickerson, R. R.,](#)
1300 [Apel, E. C., Baidar, S., Bannan, T. J., Blake, N. J., Chen, D., Dix, B., Fernandez, R. P., Hall, S. R., Hornbrook,](#)
1301 [R. S., Gregory Huey, L., Josse, B., Jöckel, P., Kinnison, D. E., Koenig, T. K., Le Breton, M., Marécal, V.,](#)
1302 [Morgenstern, O., Oman, L. D., Pan, L. L., Percival, C., Plummer, D., Revell, L. E., Rozanov, E., Saiz-Lopez,](#)
1303 [A., Stenke, A., Sudo, K., Tilmes, S., Ullmann, K., Volkamer, R., Weinheimer, A. J. and Zeng, G.:](#)
1304 [Formaldehyde in the Tropical Western Pacific: Chemical Sources and Sinks, Convective Transport, and](#)
1305 [Representation in CAM-Chem and the CCMI Models, *J. Geophys. Res. Atmos.*, 122\(20\), 11,201-11,226,](#)
1306 [doi:10.1002/2016JD026121, 2017.](#)

1307 Barkley, M. P., Palmer, P. I., Ganzeveld, L., Arneth, A., Hagberg, D., Karl, T., Guenther, A., Paulot, F.,
1308 Wennberg, P. O., Mao, J., Kurosu, T. P., et al.: Can a state of the art chemistry transport model simulate
1309 Amazonian tropospheric chemistry?, *J. Geophys. Res.*, 116(D16), D16302, doi:10.1029/2011JD015893, 2011.

1310 Barkley, M. P., Kurosu, T. P., Chance, K., Smedt, I. De, Van Roozendaal, M., Arneth, A., Hagberg, D.,
1311 Guenther, A. and De Smedt, I.: Assessing sources of uncertainty in formaldehyde air mass factors over tropical
1312 South America: Implications for top-down isoprene emission estimates, *J. Geophys. Res.*, 117(D13), D13304,
1313 doi:10.1029/2011JD016827, 2012.

1314 Barkley, M. P., De Smedt, I., Van Roozendaal, M., Kurosu, T. P., Chance, K. V., Arneth, A., Hagberg, D.,
1315 Guenther, A. B., Paulot, F., Marais, E. A., others, et al.: Top-down isoprene emissions over tropical South
1316 America inferred from SCIAMACHY and OMI formaldehyde columns, *J. Geophys. Res. Atmos.*, 118(12),
1317 n/a–n/a, doi:10.1002/jgrd.50552, 2013.

1318 Boersma, K. F., Eskes, H. J. and Brinksma, E. J.: Error analysis for tropospheric NO₂ retrieval from space, *J.*
1319 *Geophys. Res.*, 109(D4), doi:10.1029/2003JD003962, 2004.

1320 Boersma, K.F., Lorente, A., Muller, J. and the QA4ECV consortium: Recommendations (scientific) on best
1321 practices for retrievals for Land and Atmosphere ECVs, QA4ECV D4.2, v0.8,
1322 <http://www.qa4ecv.eu/sites/default/files/D4.2.pdf>, 2015.

1323 Boersma, K. F., Vinken, G. C. M., and Eskes, H. J.: Representativeness errors in comparing chemistry transport
1324 and chemistry climate models with satellite UV–Vis tropospheric column retrievals, *Geosci. Model Dev.*, 9,
1325 875–898, <https://doi.org/10.5194/gmd-9-875-2016>, 2016.

1326 Bovensmann, H., Peuch, V.-H., van Weele, M., Erbertseder, T., and Veihelmann, B.: Report Of The Review
1327 Of User Requirements For Sentinels-4/-5, ESA, EO-SMA-/1507/JL, issue: 2.1, 2011.

1328 Brion, J., et al.: Absorption spectra measurements for the ozone molecule in the 350-830 nm region, *J. Atmos.*
1329 *Chem.*, 30, 291-299, 1998.

1330 Castellanos, P., Boersma, K. F., Torres, O., and de Haan, J. F.: OMI tropospheric NO₂ air mass factors over
1331 South America: effects of biomass burning aerosols, *Atmos. Meas. Tech.*, 8, 3831-3849, doi:10.5194/amt-8-
1332 3831-2015, 2015.

1333 Chance, K. and R. J. Spurr: Ring effect studies: Rayleigh scattering including molecular parameters for
1334 rotational Raman scattering, and the Fraunhofer spectrum, *Applied Optics*, 36, 5224-5230, 1997.

1335 Chance, K. V., Palmer, P. I., Martin, R. V., Spurr, R. J. D., Kurosu, T. P. and Jacob, D. J.: Satellite observations
1336 of formaldehyde over North America from GOME, *Geophysical Research Letters*, 27(21), 3461-3464,
1337 doi:10.1029/2000GL011857, 2000.

1338 Chance, K. and Kurucz, R. L.: An improved high-resolution solar reference spectrum for earth's atmosphere
1339 measurements in the ultraviolet, visible, and near infrared, *J. Quant. Spectrosc. Radiat. Transf.*, 111(9), 1289-
1340 1295, 2010.

1341 Chimot, J., Vlemmix, T., Veefkind, J. P., de Haan, J. F. and Levelt, P. F.: Impact of aerosols on the OMI
1342 tropospheric NO₂ retrievals over industrialized regions: how accurate is the aerosol correction of cloud-free
1343 scenes via a simple cloud model?, *Atmos. Meas. Tech. Discuss.*, 8(8), 8385-8437, doi:10.5194/amt-d-8-8385-
1344 2015, 2015.

1345 Clémer, K., Van Roozendaal, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and
1346 De Mazière, M.: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS
1347 measurements in Beijing, *Atmos. Meas. Tech.*, 3, 863-878, 2010.

1348 Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K. and Visconti, G.: Estimating European volatile organic
1349 compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument,
1350 *Atmos. Chem. Phys.*, 10(23), 11501-11517, 2010.

1351 Danckaert, T., Fayt, C., Van Roozendaal, M., De Smedt, I., Letocart, V., Merlaud, A., Pinardi, G: Qdoas
1352 Software User Manual, Version 2.1, [http://uv-](http://uv-vis.aeronomie.be/software/QDOAS/QDOAS_manual_2.1_201212.pdf)
1353 [vis.aeronomie.be/software/QDOAS/QDOAS_manual_2.1_201212.pdf](http://uv-vis.aeronomie.be/software/QDOAS/QDOAS_manual_2.1_201212.pdf), 2012.

1354 Danielson, J.J., and Gesch, D.B.: Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S.
1355 Geological Survey Open-File Report 2011-1073, 26 p, 2011.

1356 Daumont, M., Brion, J., Charbonnier, J., and Malicet, J.: Ozone UV spectroscopy, I: Absorption cross-sections
1357 at room temperature, *J. Atmos. Chem.*, 15, 145-155, 1992.

1358 De Smedt, I., Müller, J.-F., Stavrou, T., van der A, R., Eskes, H. and Van Roozendael, M.: Twelve years of
1359 global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, *Atmos.*
1360 *Chem. Phys.*, 8(16), 4947-4963, 2008.

1361 De Smedt, I., Stavrou, T., Müller, J. F., van Der A, R. J. and Van Roozendael, M.: Trend detection in satellite
1362 observations of formaldehyde tropospheric columns, *Geophys. Res. Lett.*, 37(18), L18808,
1363 doi:10.1029/2010GL044245, 2010.

1364 De Smedt, I.: Long-Term Global Observations of Tropospheric Formaldehyde Retrieved from Spaceborne
1365 Nadir UV Sensors, Ph.D. thesis, Universite Libre De Bruxelles, Laboratoire de Chimie Quantique et
1366 Photophysique, Faculté de Sciences Appliquées, 2011.

1367 De Smedt, I., Van Roozendael, M., Stavrou, T., Müller, J.-F., Lerot, C., Theys, N., Valks, P., Hao, N., and
1368 van der A, R.: Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A
1369 addressing noise reduction and instrumental degradation issues, *Atmos. Meas. Tech. Discuss.*, 5, 5571-5616,
1370 doi:10.5194/amtd-5-5571-2012, Special Issue: GOME-2: calibration, algorithms, data products and validation,
1371 2012.

1372 De Smedt, I., Stavrou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen,
1373 C., Vigouroux, C., Hermans, C., et al.: Diurnal, seasonal and long-term variations of global formaldehyde
1374 columns inferred from combined OMI and GOME-2 observations, *Atmos. Chem. Phys. Discuss.*, 15(8),
1375 12241–12300, doi:10.5194/acpd-15-12241-2015, 2015.

1376 [De Smedt, I., Theys, N., van Gent, J., Danckaert, T., Yu, H. and Van Roozendael, M.: S5P/TROPOMI HCHO](http://www.tropomi.eu/sites/default/files/files/S5P-BIRA-L2-ATBD-HCHO_400F_TROPOMI_v1p0p0-20160205.pdf)
1377 [ATBD, S5P-BIRA-L2-400F-ATBD, v1.0.0, 2016-02-19, Level-2 Algorithm Developments for Sentinel-5](http://www.tropomi.eu/sites/default/files/files/S5P-BIRA-L2-ATBD-HCHO_400F_TROPOMI_v1p0p0-20160205.pdf)
1378 [Precursor., \[http://www.tropomi.eu/sites/default/files/files/S5P-BIRA-L2-ATBD-\]\(http://www.tropomi.eu/sites/default/files/files/S5P-BIRA-L2-ATBD-HCHO_400F_TROPOMI_v1p0p0-20160205.pdf\)](http://www.tropomi.eu/sites/default/files/files/S5P-BIRA-L2-ATBD-HCHO_400F_TROPOMI_v1p0p0-20160205.pdf)
1379 [HCHO 400F TROPOMI v1p0p0-20160205.pdf](http://www.tropomi.eu/sites/default/files/files/S5P-BIRA-L2-ATBD-HCHO_400F_TROPOMI_v1p0p0-20160205.pdf), 2016.

1380 Dirksen, R., Dobber, M., Voors, R., and Levelt, P.: Pre-launch characterization of the Ozone Monitoring
1381 Instrument transfer function in the spectral domain, *Appl. Opt.*, 45(17), 3972-3981, 2006.

1382 Dufour, G., F. Wittrock, M. Camredon, M. Beekmann, A. Richter, B. Aumont, and J. P. Burrows,
1383 SCIAMACHY formaldehyde observations: constraint for isoprene emission estimates over Europe?, *Atmos.*
1384 *Chem. Phys.*, 9(5), 1647-1664, 2009.

1385 Eskes, H. J. and K. F. Boersma, Averaging kernels for DOAS total-column satellite retrievals, *Atmos. Chem.*
1386 *Phys.*, 3, 1285-1291, 2003.

1387 Fayt, C. and M. Van Roozendael: Windoas 2.1, Software User Manual, BIRA-IASB, 2001.

1388 Fehr, T.: Sentinel-5 Precursor Scientific Validation Implementation Plan, EOP-SM/2993/TF-tf, 1.0,
1389 <http://doi.org/10.5281/zenodo.165739>, 2016.

1390 Fleischmann, O. C., et al.: New ultraviolet absorption cross-sections of BrO at atmospheric temperatures
1391 measured by time-windowing Fourier transform spectroscopy, *J. Photochem. Photobiol. A*, 168, 117–132,
1392 2004.

1393 Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Saunois, M., Szopa, S., Cressot, C., Kurosu, T. P.,
1394 Chance, K. and Fried, A.: The formaldehyde budget as seen by a global-scale multi-constraint and multi-
1395 species inversion system, *Atmos. Chem. Phys. Discuss.*, 12(3), 6909-6955, doi:10.5194/acpd-12-6909-2012,
1396 2012.

1397 Franco, B., Hendrick, F., Van Roozendaal, M., Müller, J.-F., Stavrakou, T., Marais, E. A., Bovy, B., Bader,
1398 W., Fayt, C., Hermans, C., Lejeune, B., Pinardi, G., Servais, C., and Mahieu, E.: Retrievals of formaldehyde
1399 from ground-based FTIR and MAX-DOAS observations at the Jungfraujoch station and comparisons with
1400 GEOS-Chem and IMAGES model simulations, *Atmos. Meas. Tech.*, 8, 1733-1756,
1401 <https://doi.org/10.5194/amt-8-1733-2015>, 2015.

1402 Fu, T.-M., Jacob, D. J., Palmer, P. I., Chance, K. V., Wang, Y. X., Barletta, B., Blake, D. R., Stanton, J. C. and
1403 Pilling, M. J.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions
1404 in east and south Asia and implications for ozone, *J. Geophys. Res.*, 112(D6), D06312, 2007.

1405 González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T. P. and Suleiman, R.: Updated Smithsonian
1406 Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval, *Atmos. Meas.*
1407 *Tech.*, 8(1), 19–32, doi:10.5194/amt-8-19-2015, 2015.

1408 [González Abad, G., Vasilkov, A., Sefstor, C., Liu, X., and Chance, K.: Smithsonian Astrophysical Observatory](https://doi.org/10.5194/amt-9-2797-2016)
1409 [Ozone Mapping and Profiler Suite \(SAO OMPS\) formaldehyde retrieval, *Atmos. Meas. Tech.*, 9, 2797-2812,](https://doi.org/10.5194/amt-9-2797-2016)
1410 [https://doi.org/10.5194/amt-9-2797-2016, 2016.](https://doi.org/10.5194/amt-9-2797-2016)

1411 Gonzi, S., Palmer, P. I., Barkley, M. P., De Smedt, I. and Van Roozendaal, M.: Biomass burning emission
1412 estimates inferred from satellite column measurements of HCHO: Sensitivity to co-emitted aerosol and
1413 injection height, *Geophys. Res. Lett.*, 38(14), L14807, doi:10.1029/2011GL047890, 2011.

1414 Gottwald, M., Bovensmann, H. et al.: SCIAMACHY, Monitoring the Changing Earth's Atmosphere, DLR,
1415 Institut für Methodik der Fernerkundung (IMF), 2006.

1416 Grainger, J. F. and J. Ring: Anomalous Fraunhofer line profiles, *Nature*, 193, 762, 1962.

1417 Greenblatt, G. D., Orlando, J. J., Burkholder, J. B., and Ravishankara, A. R.: Absorption measurements of
1418 oxygen between 330 and 1140 nm, *J. Geophys. Res.*, 95(D11), 18 577–18 582, doi:10.1029/90JD01375, 1990.

1419 [Hartmann, D. L., Klein Tank, A.M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y.,](#)
1420 [Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P.W., Wild, M., and](#)
1421 [Zhai, P. M.: Observations: Atmosphere and Surface, in: *Climate Change 2013: The Physical Science Basis.*](#)

1422 [Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate](#)
1423 [Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A.,](#)
1424 [Xia, Y., Bex, V., and Midgley P. M., Cambridge University Press, Cambridge, United Kingdom and New](#)
1425 [York, NY, USA, 2013.](#)

1426 Hassinen, S., Balis, D., Bauer, H., Begoin, M., Delcloo, A., Eleftheratos, K., Gimeno Garcia, S., Granville, J.,
1427 Grossi, M., Hao, N., Hedelt, P., Hendrick, F., Hess, M., Heue, K.-P., Hovila, J., Jönch-Sørensen, H., Kalakoski,
1428 N., Kauppi, A., Kiemle, S., Kins, L., Koukoulis, M. E., Kujanpää, J., Lambert, J.-C., Lang, R., Lerot, C., Loyola,
1429 D., Pedergnana, M., Pinardi, G., Romahn, F., Van Roozendaal, M., Lutz, R., De Smedt, I., Stammes, P.,
1430 Steinbrecht, W., Tamminen, J., Theys, N., Tilstra, L. G., Tuinder, O. N. E., Valks, P., Zerefos, C., Zimmer, W.
1431 and Zyrichidou, I.: Overview of the O3M SAF GOME-2 operational atmospheric composition and UV
1432 radiation data products and data availability, *Atmos. Meas. Tech.*, 9(2), 383–407, doi:10.5194/amt-9-383-2016,
1433 2016.

1434 Heckel, A., Kim, S.-W., Frost, G. J., Richter, A., Trainer, M. and Burrows, J. P.: Influence of low spatial
1435 resolution a priori data on tropospheric NO₂ satellite retrievals, *Atmos. Meas. Tech.*, 4(9), 1805–1820,
1436 doi:10.5194/amt-4-1805-2011, 2011.

1437 Hewson, W., Bösch, H., Barkley, M. P. and De Smedt, I.: Characterisation of GOME-2 formaldehyde retrieval
1438 sensitivity, *Atmospheric Measurement Techniques*, 6(2), 371–386, doi:10.5194/amt-6-371-2013, 2013.

1439 Hilboll, A., Richter, A. and Burrows, J. P.: Long-term changes of tropospheric NO₂ over megacities derived
1440 from multiple satellite instruments, *Atmospheric Chemistry and Physics*, 13(8), 4145–4169, doi:10.5194/acp-
1441 13-4145-2013, 2013.

1442 Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S.,
1443 Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade,
1444 F., Scheele, R., Nédélec, P., and Pätz, H.-W., The global chemistry transport model tm5: description and
1445 evaluation of the tropospheric chemistry version 3.0., *Geoscientific Model Development*, 3(2):445-473, 2010.

1446 Jones, N. B., Riedel, K., Allan, W., Wood, S., Palmer, P. I., Chance, K., and Notholt, J.: Long-term tropospheric
1447 formaldehyde concentrations deduced from ground-based fourier transform solar infrared measurements,
1448 *Atmos. Chem. Phys.*, 9, 7131-7142, <https://doi.org/10.5194/acp-9-7131-2009>, 2009.

1449 [Kaiser, J., Jacob, D. J., Zhu, L., Travis, K. R., Fisher, J. A., González Abad, G., Zhang, L., Zhang, X., Fried,](#)
1450 [A., Crounse, J. D., St. Clair, J. M., and Wisthaler, A.: High-resolution inversion of OMI formaldehyde columns](#)
1451 [to quantify isoprene emission on ecosystem-relevant scales: application to the Southeast US, *Atmos. Chem.*](#)
1452 [Phys. Discuss., <https://doi.org/10.5194/acp-2017-1137>, in review, 2017](#)

1453 Kleipool, Q. L., Dobber, M. R., de Haan, J. F. and Levelt, P. F.: Earth surface reflectance climatology from 3
1454 years of OMI data, *J. Geophys. Res.*, 113(D18), D18308, doi:10.1029/2008JD010290, 2008.

1455 Koelemeijer, R. B. A., Stammes, P., Hovenier, J. W. and de Haan, J. F.: A fast method for retrieval of cloud
1456 parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment, *J. Geophys.*
1457 *Res.*, 106(D4), 3475-3490, doi:10.1029/2000JD900657, 2001.

1458 Khokhar, M. F.: Spatio-Temporal Analyses of Formaldehyde over Pakistan by Using SCIAMACHY and
1459 GOME-2 Observations, *Aerosol Air Qual. Res.*, 1–14, doi:10.4209/aaqr.2014.12.0339, 2015.

1460 Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener,
1461 F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and
1462 applications., *Atmos. Chem. Phys.*, 5(2):417-432, 2005.

1463 Kurosu, T. P., OMHCHO README FILE,
1464 [http://www.cfa.harvard.edu/tkurosu/SatelliteInstruments/OMI/PGEReleases/READMEs/OMHCHO_READ](http://www.cfa.harvard.edu/tkurosu/SatelliteInstruments/OMI/PGEReleases/READMEs/OMHCHO_README.pdf)
1465 [ME.pdf](http://www.cfa.harvard.edu/tkurosu/SatelliteInstruments/OMI/PGEReleases/READMEs/OMHCHO_README.pdf), last access: 14/08/2012, 2008.

1466 Langen, J., Meijer, Y., Brinksma, E., Veihelmann, B., and Ingmann, P.: GMES Sentinels 4 and 5 Mission
1467 Requirements Document (MRD), ESA, EO-SMA-/1507/JL, issue: 3, 2011.

1468 Langen, J., Meijer, Y., Brinksma, E., Veihelmann, B., and Ingmann, P.: Copernicus Sentinels 4 and 5 Mission
1469 Requirements Traceability Document (MRTD), ESA, EO-SMA-/1507/JL, issue: 2, 2017.

1470 Leitao, J., Richter, A., Vrekoussis, M., Kokhanovsky, A., Zhang, Q.J., Beekmann, M., and Burrows, J. P.: On
1471 the improvement of NO₂ satellite retrievals – aerosol impact on the air mass factors, *Atmos. Meas. Tech.*, 3,
1472 475–493, doi:10.5194/amt-3-475-2010,2010.

1473 Leue, C.: Detektion der troposphärischen NO₂ Daten anhand von GOME. Ph.D. thesis, Univ. Heidelberg,
1474 Heidelberg, Germany, 1999.

1475 Li, C., Joiner, J., Krotkov, N. A. and Dunlap, L.: A New Method for Global Retrievals of HCHO Total Columns
1476 from the Suomi National Polar-orbiting Partnership Ozone Monitoring and Profiler Suite, *Geophys. Res. Lett.*,
1477 doi:10.1002/2015GL063204, 2015.

1478 Lin, J. T., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendaal, M.,
1479 Clemer, K. and Irie, H.: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument:
1480 Effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, *Atmos. Chem.*
1481 *Phys.*, 14(3), 1441–1461, doi:10.5194/acp-14-1441-2014, 2014.

1482 Lorente, A., Boersma, K. F., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M.,
1483 De Smedt, I., Van Roozendaal, M., Wang, Y., Wagner, T., Beirle, S., Lin, J. T., Krotkov, N., Stammes, P.,
1484 Wang, P., Eskes, H. J., and Krol, M.: Structural uncertainty in air mass factor calculation for NO₂ and HCHO
1485 satellite retrievals, *Atmos. Meas. Tech. Discuss.*, doi:10.5194/amt-2016-306, in review, 2016.

1486 Loyola, D. G., Gimeno García, S., Lutz, R., Romahn, F., Spurr, R. J. D., Pedernana, M., Doicu, A., and
1487 Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor,
1488 Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2017-128>, in review, 2017.

1489 Mahajan, A. S., De Smedt, I., Biswas, M. S., Ghude, S., Fadnavis, S., Roy, C. and van Roozendael, M.: Inter-
1490 annual variations in satellite observations of nitrogen dioxide and formaldehyde over India, Atmos. Environ.,
1491 116, 194–201, doi:10.1016/j.atmosenv.2015.06.004, 2015.

1492 Malicet, C., Daumont, D., Charbonnier, J., Parisse, C., Chakir, A., and Brion, J.: Ozone UV spectroscopy, II:
1493 Absorption cross-sections and temperature dependence, J. Atmos. Chem., 21, 263–273, 1995.

1494 Marais, E. A., Jacob, D. J., Kurosu, T. P., Chance, K., Murphy, J. G., Reeves, C., Mills, G., Casadio, S., Millet,
1495 D. B., Barkley, M. P., Paulot, F., et al.: Isoprene emissions in Africa inferred from OMI observations of
1496 formaldehyde columns, Atmos. Chem. Phys. Discuss., 12(3), 7475–7520, doi:10.5194/acpd-12-7475-2012,
1497 2012.

1498 Marbach, T., Beirle, S., Platt, U., Hoor, P., Wittrock, F., Richter, A., Vrekoussis, M., Grzegorski, M., Burrows,
1499 J. P. and Wagner, T.: Satellite measurements of formaldehyde linked to shipping emissions, Atmos. Chem.
1500 Phys., 9(21), 2009.

1501 Martin, R. V., Chance, K. V., Jacob, D. J., Kurosu, T. P., Spurr, R. J. D., Bucsele, E. J., Gleason, J., Palmer, P.
1502 I., Bey, I., Fiore, A. M., Li, Q., et al.: An improved retrieval of tropospheric nitrogen dioxide from GOME, J.
1503 Geophys. Res., 107(D20), doi:10.1029/2001JD001027, 2002.

1504 Meller, R., and Moortgat, G. K.: Temperature dependence of the absorption cross section of HCHO between
1505 223 and 323K in the wavelength range 225–375 nm, J. Geophys. Res., 105(D6), 7089–7102,
1506 doi:10.1029/1999JD901074, 2000.

1507 Millet, D. B., Jacob, D. J., Boersma, K. F., Fu, T.-M., Kurosu, T. P., Chance, K. V., Heald, C. L. and Guenther,
1508 A.: Spatial distribution of isoprene emissions from North America derived from formaldehyde column
1509 measurements by the OMI satellite sensor, Journal of Geophysical Research, 113(D2), 1–18,
1510 doi:10.1029/2007JD008950, 2008.

1511 Palmer, P. I., Jacob, D. J., Chance, K. V., Martin, R. V., D, R. J., Kurosu, T. P., Bey, I., Yantosca, R. and Fiore,
1512 A.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde
1513 retrievals from the Global Ozone Monitoring Experiment, Journal of Geophysical Research, 106(D13), 14539–
1514 14550, doi:10.1029/2000JD900772, 2001.

1515 Palmer, P. I., Abbot, D. S., Fu, T.-M., Jacob, D. J., Chance, K. V., Kurosu, T. P., Guenther, A., Wiedinmyer,
1516 C., Stanton, J. C., Pilling, M. J., Pressley, S. N., et al.: Quantifying the seasonal and interannual variability of
1517 North American isoprene emissions using satellite observations of the formaldehyde column, Journal of
1518 Geophysical Research, 111(D12), 1–14, doi:10.1029/2005JD006689, 2006.

1519 Pedergnana, M., Loyola, D., Apituley, A., Sneep, M., Veeffkind, J. P.: Sentinel-5 precursor/TROPOMI Level
1520 2 Product User Manual Formaldehyde HCHO, S5P-L2-DLR-PUM-400F, 0.11.4,
1521 [http://www.tropomi.eu/sites/default/files/files/S5P-L2-DLR-PUM-400F-](http://www.tropomi.eu/sites/default/files/files/S5P-L2-DLR-PUM-400F-Product_User_Manual_for_the_Sentinel_5_precursor_Formaldehyde_HCHO-00.11.04-20170601_signed.pdf)
1522 [Product_User_Manual_for_the_Sentinel_5_precursor_Formaldehyde_HCHO-00.11.04-](http://www.tropomi.eu/sites/default/files/files/S5P-L2-DLR-PUM-400F-Product_User_Manual_for_the_Sentinel_5_precursor_Formaldehyde_HCHO-00.11.04-20170601_signed.pdf)
1523 [20170601_signed.pdf](http://www.tropomi.eu/sites/default/files/files/S5P-L2-DLR-PUM-400F-Product_User_Manual_for_the_Sentinel_5_precursor_Formaldehyde_HCHO-00.11.04-20170601_signed.pdf), 2017.

1524 Pinardi, G., Van Roozendael, M., Abuhassan, N., Adams, C., Cede, a., Clémer, K., Fayt, C., Frieß, U., Gil, M.,
1525 Herman, J., Hermans, C., et al.: MAX-DOAS formaldehyde slant column measurements during CINDI:
1526 intercomparison and analysis improvement, *Atmospheric Measurement Techniques*, 6(1), 167–185,
1527 doi:10.5194/amt-6-167-2013, 2013.

1528 Platt, U.: Differential optical absorption spectroscopy (DOAS), in *Air Monitoring by Spectroscopic*
1529 *Techniques*, M.W. Sigrist ed., Chemical Analysis Series, Wiley, New York, 127, 27-84, 1994.

1530 Platt, U and Stutz, J.: *Differential Optical Absorption Spectroscopy: Principles and Applications (Physics of*
1531 *Earth and Space Environments)*, Springer-Verlag, Berlin, Heidelberg, ISBN 978-3540211938, 2008.

1532 Puķīte, J., Kühn, S., Deutschmann, T., Platt, U., and Wagner, T.: Extending differential optical absorption
1533 spectroscopy for limb measurements in the UV, *Atmos. Meas. Tech.*, 3, 631-653, 2010.

1534 Richter, A., Begoin, M., Hilboll, A. and Burrows, J. P.: An improved NO₂ retrieval for the GOME-2 satellite
1535 instrument, *Atmos. Meas. Tech.*, 4(6), 213-246, doi:10.5194/amt-4-1147-2011, 2011.

1536 Richter, A. and S5-P verification teams: S5P/TROPOMI Science Verification Report, S5P-IUP-L2-ScVR-RP,
1537 v2.1, 2015-12-22, in *Level-2 Algorithm Developments for Sentinel-5 Precursor.*, 2015.

1538 Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding, Theory and Practice*, World Scientific Publishing,
1539 Singapore-New-Jersey-London-Hong Kong, 2000.

1540 Rodgers, C. D., and B. J. Connor: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108,
1541 doi:10.1029/2002JD002299, 2003.

1542 [Seinfeld, J. H. and S. N. Pandis, Atmospheric Chemistry and Physics: From air pollution to climate change,](#)
1543 [second edition, John Wiley and Sons, New-York, 2006.](#)

1544 Serdyuchenko, A., Gorshchev, V., Weber, M., Chehade, W., and Burrows, J. P.: High spectral resolution ozone
1545 absorption cross-sections – Part 2: Temperature dependence, *Atmos. Meas. Tech.*, 7, 625-636,
1546 doi:10.5194/amt-7-625-2014, 2014.

1547 Spurr, R. J. D.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate
1548 radiative transfer models for use in remote sensing retrieval problems, in *Light Scattering Reviews*, edited by
1549 A. Kokhanovsky, pp. 229–271, Berlin, 2008a.

1550 Spurr, R. J. D., J. de Haan, R. van Oss, and A. Vasilkov, Discrete ordinate radiative transfer in a stratified
1551 medium with first-order rotational Raman scattering, *J.Q.S.R.T* 109, Iss. 3, 404425, 2008b.

1552 Stavrakou, T., Müller, J. F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Giglio, L. and Guenther,
1553 A.: Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns
1554 through 2003–2006, *Atmos. Chem. Phys.*, 9(3), 1037-1060, 2009a.

1555 Stavrakou, T., Smedt, I. D., Roozendael, M. V., Vrekoussis, M., Wittrock, F., Burrows, J., Building, M., Lane,
1556 B., Gifford, C. and Kingdom, U.: The continental source of glyoxal estimated by the synergistic use of
1557 spaceborne measurements and inverse modelling, 2009b.

1558 Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Van Roozendael, M., Guenther, a., Wild, M. and Xia,
1559 X.: Isoprene emissions over Asia 1979–2012: impact of climate and land-use changes, *Atmos. Chem. Phys.*,
1560 14(9), 4587–4605, doi:10.5194/acp-14-4587-2014, 2014.

1561 Stavrakou, T., Müller, J., Bauwens, M., Smedt, I. De and Roozendael, M. Van: How consistent are top-down
1562 hydrocarbon emissions based on formaldehyde observations from GOME-2 and OMI ?, *Atmos. Chem. Phys.*
1563 *Discuss.*, 12007–12067, doi:10.5194/acpd-15-12007-2015, 2015.

1564 Stein Zweers et al., TROPOMI ATBD of the UV aerosol index, S5P-KNMI-L2-0008-RP, 1.0,
1565 [http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0008-RP-TROPOMI_ATBD_UVAI-v1p0p0-](http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0008-RP-TROPOMI_ATBD_UVAI-v1p0p0-20160203.pdf)
1566 [20160203.pdf](http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0008-RP-TROPOMI_ATBD_UVAI-v1p0p0-20160203.pdf), 2016

1567 Tanskanen, A. Lambertian Surface Albedo Climatology at 360 nm from TOMS Data Using Moving Time-
1568 Window Technique. In: Proceedings of the XX Quadrennial Ozone Symposium, 1-8 June 2004, Kos, Greece.

1569 Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O₂-O₂ collision pairs
1570 between 340 and 630 nm and at atmospherically relevant pressure., *Phys. Chem. Chem. Phys.*, 15(37), 15371–
1571 81, doi:10.1039/c3cp50968k, 2013.

1572 Theys, N., De Smedt, I., Yu, H., Danckaert, T., van Gent, J., Hörmann, C., Wagner, T., Hedelt, P., Bauer, H.,
1573 Romahn, F., Pedernana, M., Loyola, D. and Van Roozendael, M.: Sulfur dioxide retrievals from TROPOMI
1574 onboard Sentinel-5 Precursor: algorithm theoretical basis, *Atmos. Meas. Tech.*, 10(January), 119–153,
1575 doi:10.5194/amt-10-119-2017, 2017.

1576 U.S. Standard Atmosphere, U.S. Government Printing Office, Washington, D.C., 1976.

1577 Vandaele A.C., C. Hermans, P.C. Simon, M. Carleer, R. Colin, S. Fally, M.F. Mérianne, A. Jenouvrier, and B.
1578 Coquart, Measurements of the NO₂ absorption cross-section from 42000 cm⁻¹ to 10000 cm⁻¹ (238-1000 nm)
1579 at 220 K and 294 K, *J.Q.S.R.T.*, 59, 171-184, 1998.

1580 van der A, R.J., H.J. Eskes, K.F. Boersma, T.P. van Noije, et al., Trends, seasonal variability and dominant
1581 NO_x source derived from a ten year record of NO₂ measured from space, *J. Geophys. Res.*, 113, D04302, doi:
1582 10.1029/2007JD009021, 2008.

1583 van Geffen, J.H.G.M., K.F. Boersma, H.J. Eskes, J.D. Maasackers and J.P. Veefkind, TROPOMI ATBD of
1584 the total and tropospheric NO₂ data products, S5P-KNMI-L2-0005-RP, 1.1.0,
1585 [http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0005-RP-](http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0005-RP-TROPOMI_ATBD_NO2_data_products-v1p1p0-20170816_signed.pdf)
1586 [TROPOMI_ATBD_NO2_data_products-v1p1p0-20170816_signed.pdf](http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0005-RP-TROPOMI_ATBD_NO2_data_products-v1p1p0-20170816_signed.pdf), 2017.

1587 Van Roozendael, M., V. Soebijanta, C. Fayt, and J.-C. Lambert: Investigation of DOAS Issues Affecting the
1588 Accuracy of the GDP Version 3.0 Total Ozone Product, in ERS-2 GOME GDP 3.0 Implementation and Delta
1589 Validation, ERSE-DTEX-EOAD-TN-02-0006, ESA/ESRIN, Frascati, Italy, Chap.6, pp.97-129, 2002.

1590 Van Roozendael, M., Spurr, R., Loyola, D., Lerot, C., Balis, D., Lambert, J.-C., Zimmer, W., Van Gent, J.,
1591 Van Geffen, J., Koukouli, M., Granville, J., Doicu, A., Fayt, C. & Zehner, C.: Sixteen Years Of GOME/ERS-
1592 2 Total Ozone Data: The New Direct-Fitting Gome Data Processor (Gdp) Version 5 - Algorithm Description,
1593 *J.geophys. Res.*, 117, D03305, Doi: 10.1029/2011jd016471, 2012.

1594 van Weele, M., Levelt, P., Aben, I., Veefkind, P., Dobber, M., Eskes, H., Houweling, S., Landgraf, J.,
1595 Noordhoek, R.: Science Requirements Document for TROPOMI. Volume 1, KNMI & SRON, RS-
1596 TROPOMI-KNMI-017, issue: 2.0, 2008.

1597 Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J.
1598 F., Kleipool, Q., van Weele, M., et al.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for
1599 global observations of the atmospheric composition for climate, air quality and ozone layer applications,
1600 *Remote Sensing of Environment*, 120(0), 70-83, 2012.

1601 Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F.: Improvements to the OMI O₂-O₂ operational
1602 cloud algorithm and comparisons with ground-based radar-lidar observations, *Atmos. Meas. Tech.*, 9, 6035-
1603 6049, <https://doi.org/10.5194/amt-9-6035-2016>, 2016.

1604 Viatte, C., Strong, K., Walker, K. A., and Drummond, J. R.: Five years of CO, HCN, C₂H₆, C₂H₂, CH₃OH,
1605 HCOOH and H₂CO total columns measured in the Canadian high Arctic, *Atmos. Meas. Tech.*, 7, 1547-1570,
1606 <https://doi.org/10.5194/amt-7-1547-2014>, 2014.

1607 Vigouroux, C., F. Hendrick, T. Stavrakou, B. Dils, I. De Smedt, C. Hermans, A. Merlaud, F. Scolas, C. Senten,
1608 G. Vanhaelewyn, S. Fally, M. Carleer, J.-M. Metzger, J.-F. Müller, M. Van Roozendael, and M. De Mazière,
1609 Ground-based FTIR and MAX-DOAS observations of formaldehyde at Réunion Island and comparisons with
1610 satellite and model data, *Atmos. Chem. Phys.*, 9, 9523-9544, doi:10.5194/acp-9-9523-2009.

1611 Vountas, M., Rozanov, V. V. and Burrows, J. P.: Ring effect: impact of rotational Raman scattering on radiative
1612 transfer in earth's atmosphere, *J. of Quant. Spec. and Rad. Trans.*, 60(6), 943-961, 1998.

1613 Vrekoussis, M., Wittrock, F., Richter, A. and Burrows, J. P.: GOME-2 observations of oxygenated VOCs:
1614 what can we learn from the ratio glyoxal to formaldehyde on a global scale?, *Atmos. Chem. Phys.*, 10(21),
1615 10145-10160, 2010.

1616 Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-
1617 MP for optimized satellite retrievals: description and validation, *Geosci. Model Dev.*, 10, 721-750,
1618 doi:10.5194/gmd-10-721-2017, 2017.

1619 Wittrock, F., Richter, A., Oetjen, H., Burrows, J. P., Kanakidou, M., Myriokefalitakis, S., Volkamer, R., Beirle,
1620 S., Platt, U. and Wagner, T.: Simultaneous global observations of glyoxal and formaldehyde from space,
1621 *Geophysical Research Letters*, 33(16), 1-5, doi:10.1029/2006GL026310, 2006.

1622 Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio,
1623 M. P., De Smedt, I., Gonzalez Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J. W., Hanisco, T. F.,
1624 Richter, D., Scarino, A. J., Walega, J., Weibring, P. and Wolfe, G. M.: Observing atmospheric formaldehyde
1625 (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A,
1626 GOME2B, OMPS) with SEAC4RS aircraft observations over the Southeast US, *Atmos. Chem. Phys.*, 0, 1–24,
1627 doi:10.5194/acp-2016-162, 2016.

1628 Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P.: An improved tropospheric NO₂ retrieval
1629 for OMI observations in the vicinity of mountainous terrain, *Atmos. Meas. Tech.*, 2, 401-416, doi:10.5194/amt-
1630 2-401-2009, 2009.

1631 **APPENDIX A: Acronyms and abbreviations**

A	Averaging Kernel
AMF	Air mass factor
AOD	Aerosol optical depth
AAI	Aerosol absorbing index
ATBD	Algorithm Theoretical Basis Document
BIRA-IASB	Royal Belgian Institute for Space Aeronomy
BrO	Bromine Monoxide
BRDF	Bidirectional reflectance distribution function
CH ₄	Methane
CO	Carbon Monoxide
CAPACITY	Composition of the Atmosphere: Progress to Applications in the user Community
CCD	Charged Coupled Device
CF	Climate and Forecast metadata conventions
CRB	Clouds as Reflecting Boundaries
CTM	Chemical Transport Model
DOAS	Differential optical absorption spectroscopy
DU	Dobson Unit (1 DU = 2.6867x10 ¹⁶ molecules cm ⁻²)
ECMWF	European Centre for Medium Range Weather Forecast
ESA	European Space Agency
FWHM	Full Width Half Maximum
GMES	Global Monitoring for Environment and Security
GOME	Global Ozone Monitoring Experiment
HCHO	Formaldehyde (or H ₂ CO)
IPA	Independent Pixel Approximation
IR	Infrared
ISRF	Instrument Spectral Response Function
L2	Level-2
L2WG	Level-2 Working Group
LER	Lambertian Equivalent Reflector
VLIDORT	Vector Linearized Discrete Ordinate Radiative Transfer
LOS	Line-of-sight angle
LS	Lower stratosphere
LUT	Look-up table
MAX-DOAS	Multi-axis DOAS
MPC	Mission Performance Center
NDACC	Network for the Detection of Atmospheric Composition Change
NMVO	Non-Methane Volatile Organic Compound
NO ₂	Nitrogen Dioxide

NRT	Near-real time
OCRA	Optical Cloud Recognition Algorithm
OD	Optical Depth
O ₃	Ozone
OMI	Ozone Monitoring Instrument
OMPS	Ozone Mapping Profiler Suite
(P)BL	Planetary Boundary Layer
PCA	Principal Component Analysis
QA4ECV	Quality Assurance For Essential Climate Variables
RAA	Relative Azimuth Angle
ROCINN	Retrieval Of Cloud Information using Neural Networks
RRS	Rotational Raman Scattering
RTM	Radiative transfer model
S5P	Sentinel-5 Precursor
S5	Sentinel 5
SAA	Solar Azimuth Angle
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric ChartographY
SC(D)	Slant column density
SCDE	Slant column density error
SNR	Signal-to-noise ratio
SO ₂	Sulfur dioxide
SOW	Statement Of Work
SWIR	Short-wave infrared
SZA	Solar zenith angle
TM 4/5	Data assimilation / chemistry transport model (version 4 or 5)
TROPOMI	Tropospheric Monitoring Instrument
UPAS	Universal Processor for UV/VIS Atmospheric Spectrometers
UV	Ultraviolet
UVN	Ultraviolet/Visible/Near-infrared
VAA	Viewing Azimuth Angle
VZA	Viewing Zenith Angle
VC(D)	Vertical column density

1632 **APPENDIX B: High level L2 HCHO data product description**

1633 In addition to the main product results, such as HCHO slant column, tropospheric vertical column and air mass
 1634 factor, the level 2 data files contain a number of additional ancillary parameters and diagnostic information. A
 1635 selection of important parameters is given in Table 13. [A complete description of the level 2 data format is](#)
 1636 [given in the Product User Manual \(Pedernana et al., 2017\).](#)

1637 **Table 13: Selective list of output fields in the TROPOMI HCHO product. Scanline and ground_pixel**
 1638 **are respectively the number of pixels in an orbit along track and across track. Layer is the number of**
 1639 **vertical levels in the averaging kernels and the a-priori profiles.**

Symbol	Unit*	Variable name	Number of entries
N_v	mol.m ⁻²	formaldehyde_tropospheric_vertical_column	scanline x ground_pixel
N_s	mol.m ⁻²	fitted_slant_columns	scanline x ground_pixel x number_of_slant_columns
$N_s - N_{s,0}$	mol.m ⁻²	formaldehyde_slant_column_corrected	scanline x ground_pixel
$N_{v,0}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_correction	scanline x ground_pixel
M	n.u.	formaldehyde_tropospheric_air_mass_factor	scanline x ground_pixel
M_{clear}	n.u.	formaldehyde_clear_air_mass_factor	scanline x ground_pixel
f_c	n.u.	cloud_fraction_crb	scanline x ground_pixel
w_c	n.u.	cloud_fraction_intensity_weighted	scanline x ground_pixel
p_{cloud}	Pa	cloud_pressure_crb	scanline x ground_pixel
A_{cloud}	n.u.	cloud_albedo_crb	scanline x ground_pixel
A_s	n.u.	surface_albedo	scanline x ground_pixel
z_s	m	surface_altitude	scanline x ground_pixel
$\sigma_{N,v,rand}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_precision	scanline x ground_pixel
$\sigma_{N,v,syst}$	mol.m ⁻²	formaldehyde_tropospheric_vertical_column_trueness	scanline x ground_pixel
$\sigma_{N,s,rand}$	mol.m ⁻²	fitted_slant_columns_precision	scanline x ground_pixel x number_of_slant_columns
$\sigma_{M,rand}$	n.u.	formaldehyde_tropospheric_air_mass_factor_precision	scanline x ground_pixel
$\sigma_{N,s,0}$	mol.m ⁻²	formaldehyde_slant_column_corrected_trueness	scanline x ground_pixel
A	n.u.	averaging_kernel	layer x scanline x ground_pixel
n_a	vmr	formaldehyde_profile_apriori	layer x scanline x ground_pixel
p_s	Pa	surface_pressure	scanline x ground_pixel
a_l	Pa	tm5_constant_a	layer
b_l	n.u.	tm5_constant_b	layer

Symbol	Unit*	Variable name	Number of entries
$N_{s,l}$	mol.m ⁻²	fitted_slant_columns_win1	scanline x ground_pixel x number_of_slant_columns_win1
$\sigma_{N,s,l,rand}$	mol.m ⁻²	fitted_slant_columns_precision_win1	scanline x ground_pixel x number_of_slant_columns_win1

1640 * multiplication factor to convert mol.m⁻² to molec.cm⁻²: 6.022x10¹⁹

1641 **APPENDIX C: Averaging Kernel**

1642 Retrieved satellite quantities always represent a weighted average over all parts of the atmosphere that
 1643 contribute to the signal observed by the satellite instrument. The DOAS total column retrieval is implicitly
 1644 dependant on the a priori trace gas profile n_a . Radiative transfer calculations account for the sensitivity of the
 1645 measurement to the HCHO concentrations at all altitudes and these sensitivities are weighted with the assumed
 1646 a priori profile shape to produce the vertical column. The averaging kernel (A) is proportional to the
 1647 measurement sensitivity profile, and provides the relation between the retrieved column N_v and the true tracer
 1648 profile x (Rodgers, 2000; Rodgers and Connor, 2002):

$$N_v - N_{v,a} = A. (x^{pc} - n_a^{pc}) \quad (21)$$

1649 where the profiles are expressed in partial columns (pc). For total column observations of optically thin
 1650 absorbers DOAS averaging kernels are calculated as follows (Eskes and Boersma, 2003): $A(p) = \frac{m(p)}{M}$, where
 1651 $m(p)$ is the altitude-resolved air mass factor and M is the tropospheric air mass factor. The air mass factor and
 1652 therefore the retrieved vertical column, depends on the a priori profile shape, in contrast to the altitude-resolved
 1653 air mass factor which describes the sensitivity of the slant column to changes in trace gas concentrations at a
 1654 given altitude and does not depends on the a priori profile in an optically thin atmosphere. From the definition
 1655 of A , we have $N_{v,a} = A. n_a^{pc}$ and Equation (21) simplifies to:

$$N_v = A. x^{pc} \quad (22)$$

1656 The averaging kernel varies with the observation conditions. In the HCHO retrieval product, A is provided
 1657 together with the error budget for each individual pixel. The provided HCHO vertical columns can be used in
 1658 two ways, each with its own associated error (Boersma et al., 2004):

- 1659 1. For independent study and/or comparison with other independent measurements of total column amounts.
 1660 In this case, the total error related to the column consists of slant column measurement errors, reference
 1661 sector correction errors, and air mass factor errors. The latter consists of errors related to uncertainties in
 1662 the assumed profile n_a and errors related to the m parameters.
- 1663 2. For comparisons with chemistry transport models or validation with independent profile measurements,
 1664 if the averaging kernel information is used, the a priori profile shape error no longer contributes to the
 1665 total error. Indeed, the relative difference between the retrieved column N_v and an independent profile x_i
 1666 is:

$$\delta = \frac{N_v - A. x_i^{pc}}{N_v} \quad (23)$$

1667

1668 The total AMF M cancels since it appears as the denominator of both N_v and A . Because only the total
1669 AMF depends on the a priori tracer profile n_a , the comparison using the averaging kernel is not influenced
1670 by the chosen a priori profile shape. The a priori profile error does not influence the comparison, but of
1671 course, it still does influence the error on the retrieved vertical column.