# Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2

Arno Keppens, Jean-Christophe Lambert, José Granville, Georgina Miles, Richard Siddans, Jacob C. A. van Peet, Ronald J. van Der A, Daan Hubert, Tijl Verhoelst, Andy Delcloo, et al.

Atmos. Meas. Tech., 8, 2093–2120, 2015
www.atmos-meas-tech.net/8/2093/2015/
doi:10.5194/amt-8-2093-2015


Atmospheric Measurement Techniques
Open Access

# Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2

**A. Keppens[1], J.-C. Lambert[1], J. Granville[1], G. Miles[2], R. Siddans[2], J. C. A. van Peet[3], R. J. van der A[3], D. Hubert[1], T. Verhoelst[1], A. Delcloo[4], S. Godin-Beekmann[5], R. Kivi[6], R. Stübi[7], and C. Zehner[8]**

[1]Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[2]Rutherford Appleton Laboratory (RAL), Chilton, Didcot, UK
[3]Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands
[4]Royal Meteorological Institute (KMI-IRM), Brussels, Belgium
[5]Laboratoire Atmosphère, Milieux, Observations Spatiales (LATMOS), Université Pierre et Marie Curie, Université Versailles St-Quentin-en-Yvelines, Centre National de la Recherche Scientifique, Paris, France
[6]Finnish Meteorological Institute (FMI-ARC), Sodankylä, Finland
[7]Federal Office of Meteorology and Climatology (MeteoSwiss), Payerne, Switzerland
[8]European Space Agency (ESA/ESRIN), Frascati, Italy

*Correspondence to:* A. Keppens (arno.keppens@aeronomie.be)

Received: 16 September 2014 – Published in Atmos. Meas. Tech. Discuss.: 21 November 2014
Revised: 17 April 2015 – Accepted: 24 April 2015 – Published: 20 May 2015

**Abstract.** A methodology for the round-robin evaluation and the geophysical validation of ozone profile data retrieved from nadir UV backscatter satellite measurements is detailed and discussed, consisting of data set content studies, information content studies, co-location studies, and comparisons with reference measurements. Within the European Space Agency's Climate Change Initiative on ozone (Ozone_cci project), the proposed round-robin procedure is applied to two nadir ozone profile data sets retrieved at the Royal Netherlands Meteorological Institute (KNMI) and the Rutherford Appleton Laboratory (RAL, United Kingdom), using their respective OPERA v1.26 and RAL v2.1 optimal estimation algorithms, from MetOp-A GOME-2 (i.e. the second generation Global Ozone Monitoring Experiment on the first Meteorological Operational Satellite) measurements taken in 2008. The ground-based comparisons use ozonesonde and lidar profiles as reference data, acquired by the Network for the Detection of Atmospheric Composition Change (NDACC), Southern Hemisphere Additional Ozonesonde programme (SHADOZ), and other stations of the World Meteorological Organisation's Global Atmosphere Watch (WMO GAW). This direct illustration highlights practical issues that inevitably emerge from discrepancies in e.g. profile representation and vertical smoothing, for which different recipes are investigated and discussed. Several approaches for information content quantification, vertical resolution estimation, and reference profile resampling are compared and applied as well. The paper concludes with compliance estimates of the two GOME-2 ozone profile data sets with user requirements from the Global Climate Observing System (GCOS) and from climate modellers.

## 1 Introduction

Atmospheric ozone plays a major role in air quality and the radiation budget of the Earth, and influences climate change through radiative processes in the short wave and long wave spectral domains and through its chemical interaction with other radiatively active trace gases. Climate studies therefore require accurate monitoring of the horizontal and vertical distribution of ozone on the global scale and in the long term (WMO, World Meteorological Organisation, 2010). Global ozone concentration profiles have been retrieved from solar backscatter ultraviolet radiation measurements by nadir viewing satellite spectrometers like the American Solar Backscatter Ultra Violet(/2) series (SBUV(/2)) since the late 1970s and the new genera-

tion European GOME/SCIAMACHY/OMI/GOME-2 series since 1995. Over decades those retrievals have to be frequently assessed and often improved in order to meet climate research user requirements like the Global Climate Observing System (GCOS) targets (WMO, 2010). Both the verification of retrieval algorithm updates and the validation of their outputs are therefore essential parts of the climate monitoring process, to be performed by specialized independent groups.

Several references propose evaluation approaches for nadir ozone profile data products, e.g. Hoogen et al. (1999) and Meijer et al. (2006). They illustrate that the comparison of nadir profiles with reference measurements usually is insufficient to fully appreciate the relative quality of different retrieval products and to verify their compliance with user requirements. The main goal of this publication is to present and rationalise an exhaustive seven-step round-robin (RR) evaluation approach, and to apply this methodology to two nadir ozone profile data sets as an example. This direct implementation highlights practical issues that inevitably emerge from the need for straightforward evaluation of different retrieval products, such as differences in units, in vertical gridding, in vertical smoothing, etc. Such issues are rarely addressed in publications with the level of detail desired by scientists who would like to reproduce the study on their own data sets. Another objective of this paper is therefore to describe the evaluation procedure as a coherent set of generic practices with direct illustration on real data. By means of iterative feedback, it has been possible to incorporate the addressing of such issues into the full RR evaluation methodology, which is reflected in the outline of this paper (Sects. 3–7) and schematized in Appendix A: (1) satellite data collection and post-processing, (2) data set content study, (3) information content study, (4) correlative data selection, (5) co-located data sets study, (6) vertical resampling of reference profiles, and (7) comparative analysis.

Section 2 first introduces the two independent nadir ozone profile retrieval schemes under study in this work: OPERA version 1.26 and RAL version 2.1, developed at the Royal Netherlands Meteorological Institute (KNMI) and at the Rutherford Appleton Laboratory (RAL, United Kingdom), respectively. For this study, both processors retrieved Level-2 (L2) nadir ozone profile data from the same MetOp-A GOME-2 (i.e. the second generation Global Ozone Monitoring Experiment on the first Meteorological Operational Satellite) Level-1 (L1) version 4.0 products of 2008, as delivered operationally by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The respective performance of these two data sets and their compliance with user requirements have been investigated through a round-robin exercise in the context of European Space Agency's Climate Change Initiative (ESA CCI), aiming at the monitoring of essential climate variables (ECVs) from space (ESA, 2013). A major goal of the Ozone CCI subproject is to produce time series of tropospheric and stratospheric ozone distributions from current and historical European, American, and third-party missions that would meet the requirements for reducing the current uncertainty in estimates of global radiative forcing.

It should be emphasized that the focus of this work is on the presentation and application of the RR evaluation methodology, and not on a ranking of the OPERA and RAL retrieval systems. A table summarizing the compliance of each data product with applicable user requirements is nevertheless provided with the conclusions in Sect. 8. The freedom to select data according to the most appropriate evaluation criteria is left to the data user.

## 2 GOME-2 nadir ozone profile retrieval systems

### 2.1 Optimal estimation

The OPERA and RAL retrieval systems are both based on the optimal estimation (OE) method (Rodgers, 2000). This method consists in minimizing the difference between a measured atmospheric spectrum and a spectrum that is simulated by the so-called forward model $F$. The latter includes a radiative transfer model (RTM) which relates the spectral radiance vector $\boldsymbol{y}$ to a vertical atmospheric state vector $\boldsymbol{x}$: $\boldsymbol{y} = F(\boldsymbol{x})$. However, the retrieval is usually performed at higher vertical sampling than the actual amount of independent pieces of information available from the measured irradiance spectrum. Solving for the vertical profile $\boldsymbol{x}$ by inversion therefore generally forms an ill-posed (under-constrained) and consequently unstable problem. The well-established optimal estimation approach outlined by Rodgers (2000) accomplishes such an inversion for weakly nonlinear forward models by linearizing (first order Taylor expansion) the RTM, denoted as $F'$, in search for the minimum of a cost function of the form $\left|\boldsymbol{y} - F'(\boldsymbol{x})\right|$. The OE retrieval scheme includes additional constraints in the form of an extra term in the cost function that accounts for prior information on the profile, its shape, and its allowed covariance. As a result, the retrieved quantity is a mix of information contributed by the measurement and a priori information. The contribution of prior information can be significant where the measurement is weakly or even not sensitive to the atmospheric profile, e.g. in case of vertical fine-scale structures, below optically thick clouds, and at the lower altitudes in twilight conditions. Each retrieved ozone profile $\boldsymbol{x}_\mathrm{r}$ as obtained by the optimal estimation method can thus be regarded at first order, neglecting measurement errors, as a weighted average between prior and measurement information (Rodgers, 2000):

$$\boldsymbol{x}_\mathrm{r} = \mathbf{A}\boldsymbol{x}_\mathrm{t} + (\mathbf{I} - \mathbf{A})\,\boldsymbol{x}_\mathrm{p} \tag{1}$$

with matrix $\mathbf{A}$, the so-called averaging kernel matrix (AKM), constituted by the elements $\mathbf{A}(i,j) = \partial \boldsymbol{x}_\mathrm{r}(i) \,/\, \partial \boldsymbol{x}_\mathrm{t}(j)$. In these equations, $\boldsymbol{x}_\mathrm{t}$ and $\boldsymbol{x}_\mathrm{p}$ represent ozone profile vectors corresponding to the values of the true and prior atmospheric

state, respectively. **I** is the identity matrix equal in size to the AKM.

The averaging kernels reflect the limited sensitivity of the spectral measurement to fine-scale atmospheric structures. They are dependent on the detailed specification of the state vector, the a priori profile, the a priori errors, and the measurement errors, which are all particular to a specific retrieval scheme. For small retrieval errors (from both measurement and RTM) the averaging kernel matrix tends toward the identity matrix, and hence the OE solution becomes less dependent on the a priori profile. For large retrieval errors the averaging kernel elements go to zero, and the solution relies more on the a priori. For the prior errors the situation is reverse. Therefore the settings of the retrieval errors and a priori errors are important in the design of a retrieval system. Although there is a consensus for the former, there is none for the latter. The complete prior covariance matrix (CVM) is hence generally constructed assuming an exponential decrease from its diagonal values (prior variances) using an algorithm-specific correlation length. A more detailed analysis of the averaging kernel matrices is presented in Sect. 5.

### 2.2 KNMI retrieval system

The optimal estimation retrieval system developed by Royal Netherlands Meteorological Institute is called Ozone Profile Retrieval Algorithm (OPERA, version 1.26 for this work) (Mijling et al., 2010; van Peet et al., 2014). OPERA derives the vertical ozone distribution from nadir satellite data between 265 and 330 nm. The multiple-scattered part of a sun-normalized radiance spectrum within this wavelength range is simulated by applying the Linearized Discrete Ordinate Radiative Transfer model (LIDORTA) to an atmospheric state model in six streams (van Oss and Spurr, 2002), while the single-scattered part of the spectrum is simulated by a dedicated single-scatter model. The sphericity of the atmosphere is taken into account using a pseudo-spherical approximation for both the solar direct beam and the line of sight irradiance. The ozone profile elements that are actually retrieved are layer column amounts on a fixed vertical pressure grid.

The LIDORTA model replaces several numerical solvers of its parent LIDORT model by analytical solutions and as a result is a simplified yet faster version of the latter (Spurr et al., 2001). It is a scalar model that does not treat polarization and the vector nature of the radiation field, which leads to errors for the radiance at the top of the atmosphere that can reach 10 % for scattering angles of 90°. A look-up table containing the scalar radiance error for the complete wavelength range under consideration and a wide range of possible atmospheric and viewing conditions enables correcting for this error.

KNMI's atmospheric model handles cloud cover as a Lambertian reflecting layer at the cloud top height for the part of the satellite pixel that is covered with clouds, i.e. for cloud fractions larger than 0.2. For smaller fractions the surface albedo is fitted. Effective cloud fractions and cloud top heights are obtained from the Fast Retrieval Scheme for Cloud Observables (FRESCO) version 6, which extracts cloud information from the oxygen A-band (Wang et al., 2008). Due to its fitting of effective cloud fractions, the presence of aerosols and their effects on the measured radiance spectra are taken into account by the FRESCO retrieval.

OPERA uses the temperature-parameterized ozone cross sections measured by Brion et al. (1998), Daumont et al. (1992), and Malicet et al. (1995). Other trace gases are assumed not to affect the retrieval in the spectral range under consideration. The a priori ozone profile information comes from the global ozone climatology of McPeters, Labow, and Logan (McPeters et al., 2007). Prior covariance information is constructed using the same exponential formulation as described in Hoogen et al. (1999), with a correlation length of 4–5 km.

### 2.3 RAL retrieval system

The optimal estimation retrieval system developed by the Rutherford Appleton Laboratory is a three-step OE scheme for nadir ozone profile retrievals (Munro et al., 1998; Siddans, 2003; Miles et al., 2015), of which version 2.1 was used for this paper. RAL and OPERA use the same Brion-Daumont-Malicet ozone cross-sections.

In the first step, the vertical profile of ozone is retrieved from sun-normalized radiances at selected wavelengths of the ozone Hartley band (GOME-2 Band 1), in the range 265–307 nm. This spectral range primarily contains information on stratospheric ozone. Prior ozone profiles come from the McPeters–Labow–Logan (McPeters et al., 2007) climatology, except in the troposphere where a fixed value of $10^{12}$ molecules $m^{-3}$ is assumed (i.e. 1.5–2 times larger than the climatological values). To avoid too tight an a priori constraint and to avoid spurious effects in the retrieval due to the imperfect sampling of the tropospheric variance by the climatology, the a priori uncertainty is set by default to 100 % for retrieval levels at 0, 6 and 12 km, 30 % at 16 km, 10 % from 20 to 52 km, 50 % at 56 km, and 100 % from 60 to 80 km. The default uncertainty is replaced by the McPeters–Labow–Logan (McPeters et al., 2007) climatological relative variability where the latter exceeds the former. A correlation length of 6 km is applied to construct the covariance matrix. The surface albedo, a scaling factor for the ring effect, and the dark signal are retrieved jointly.

In the second step, the surface albedo for each of the eight Band 2 (B2) ground pixels is retrieved from the sun-normalized radiance spectrum between 335 and 336 nm. Then, in step three, information on lower stratospheric and tropospheric ozone is added by exploiting the temperature dependence of the spectral structure in the ozone Huggins bands. The wavelength range from 323 to 334 nm (GOME-2 B2) is used in conjunction with ECMWF ERA-Interim

(ERA-I) meteorological fields (Dee et al., 2011). Each direct Sun Band 2 spectrum is fitted to a high-resolution (0.01 nm) solar reference spectrum to improve knowledge of wavelength registration and slit function width.

In the third step, the vertical profile of ozone is retrieved in the Huggins band: a third order polynomial is subtracted from the log of the sun-normalized radiance, allowing differential structures to be fitted to a precision of smaller than 0.1 % root-mean-square (vs. order of 1 % in the Hartley band). This removes, to a large degree, independent information on the surface reflectance which modulates the mean layer photon-path profile. It is therefore important to specify an accurate step-two surface albedo as a forward model parameter in this retrieval step. The differential approach leads to improvements in the tropospheric retrieval and results in less stringent requirements on the absolute radiometric accuracy. In this step, the a priori ozone profile and its error are the output of step one, except that a prior correlation length of 8 km is imposed.

The radiative transfer model is derived from GOMETRAN (Rozanov et al., 1997), but the original code has been modified substantially in order to increase its efficiency without losing accuracy. Within the RTM there is no explicit representation of clouds, but their effects are incorporated as part of the Lambertian surface albedo (from step two). Therefore a negative bias in retrieved ozone is to be expected where high or thick cloud is extensive, and there is limited photon penetration (no "ghost column" is added). Methods to improve the characterization of sub-pixel clouds in the GOME-2 field-of-view using VIS-NIR imagery (ATSR and AVHRR) are functional within the RAL scheme but have not been implemented to produce the data for this study.

The linear error analysis is somewhat complicated by the three-step retrieval approach. Particularly as the ozone prior covariance used in step three is not identical to the solution covariance output from step one. This is handled by linearizing each step and propagating the impact of perturbations in parameters affecting the measurements through to the final solution. The estimated standard deviation of the final retrieval is taken to be the square-root of the step-three solution covariance (which includes contributions from other steps through the step-three prior covariance).

### 2.4 Major differences between the OPERA and RAL retrieval systems

A few differences between the OPERA and RAL retrieval settings might influence the round-robin methodology and the interpretation of validation results. The most important difference may be in their respective state vector definitions: OPERA performs a partial ozone column (sub-column) retrieval obtaining 16 layer values in between 17 pressure levels ranging from 0.01 to 1000 hPa, whereas the RAL retrieval results in 17 volume mixing ratio (VMR) or number density (ND) values defined on pressure levels. Sub-column nadir ozone profiles that have been calculated from their VMR profiles are nevertheless also provided by RAL (see Sect. 3.3).

Although both algorithms make use of the same 17-level fixed retrieval grid, OPERA automatically sets the lowest level of the pressure profile to the ECMWF ERA-I surface pressure. The surface pressure is also provided in RAL's L2 output, but nonetheless the lowest level in the vertical profile is kept to 1000 hPa. The RAL forward model uses the actual surface pressure internally, and the 1000 hPa retrieval level is effectively an interpolated one. For the calculation of sub-column properties in this lowest layer, the surface pressure is therefore required. The forward model assumes that between pressure levels the ozone concentration varies linearly in terms of log(VMR) vs. log(pressure). Pressure instead of altitude levels are used by both algorithms because pressure values follow the meridian variation in tropopause height more closely than the geometric elevation does.

As can be concluded from the two previous sections, the RAL retrieval scheme differs from OPERA in a number of other important respects. The most significant difference is the fit of the Huggins band to a precision of better than 0.1 % (close to the noise level) such that the temperature dependence of the ozone absorption cross-section can be exploited to extract tropospheric information. This is achieved by fitting the differential absorption spectrum (logarithm of the sun-normalized radiance with polynomial subtracted) in the Huggins band rather than the absolute sun-normalized radiance. The latter is necessarily used in the Hartley band fit in order to obtain information at higher altitudes. This distinct treatment of the two spectral ranges leads to the formulation of the retrieval problem in three steps instead of one, as described in Sect. 2.3. An overview of the main OPERA and RAL nadir ozone profile retrieval scheme characteristics and input parameters is provided in Table 1.

## 3 Satellite data collection and post-processing

The round-robin evaluation flowchart for nadir ozone profile retrievals in Appendix A commences with a critical section on data selection and post-processing. It is of major importance that the multiple satellite data sets (see Sects. 2 and 3.1) and reference data sets (see Sect. 6.1) show a maximal agreement in representation (e.g. units and vertical sampling, as discussed in Sects. 3.2 and 3.3) and settings (e.g. flagging and errors) that influence the further evaluation and comparison. Input from other sources that are not directly involved in the validation might thereby be required, for example in order to convert units or vertically extend profiles (cf. vertical smoothing matters in Sect. 6), but should be introduced with great precaution to avoid undesirable corruption of the data to be validated or of the reference data used for validation. The need for mutual flagging of the satellite data is handled in Sect. 4.

**Table 1.** Main OPERA and RAL nadir ozone profile retrieval scheme characteristics and input parameters (for RR algorithms under study, not necessarily corresponding to latest versions).

| Characteristic | OPERA | RAL |
|---|---|---|
| Version number | 1.26 | 2.1 |
| Prior ozone profile source | McPeters et al. (2007) | McPeters et al. (2007) |
| Prior correlation length (km) | 4 to 5 | 6 (step 1) and 8 (step 3) |
| Maximum nr. of iterations | 10 | Unlimited |
| Temperature profile | ECMWF | ECMWF |
| Cloud information source | FRESCO | Albedo/FRESCO |
| Ozone cross-section source(s) | Brion (1998), Daumont (1992), Malicet (1995) | Brion (1998), Daumont (1992), Malicet (1995) |
| Pixel co-adding (B2) | Yes (8 times $80 \times 40\,km^2$ to $160 \times 160\,km^2$ ) | Yes (8 times $80 \times 40\,km^2$ to $160 \times 160\,km^2$ ) |
| Polarization correction | GDP | GDP |
| RTM | LIDORTA | GOMETRAN |
| Retrieval quantity | Partial column (in layers) | VMR and ND (on levels) |
| Profile flagging pass | Convergence = 1, Chi-square criterion van Peet et al. (2014) | Convergence = 1, Chi-square = 1, B1 cost < 200, B2 cost < 120, B1 SCD < 500 DU (January–May only), SZA < 80° |
| Nr. of retrieval steps | 1 | 3 (B1-albedo-B2) |
| Retrieval bands (nm) | 265–303 (B1a) 282–313.5 (B1b) 316.5–330 (B2b) | 265–307 (B1, Hartley) 335–340 (B2, Huggins) 323–334 (B2, Huggins) |
| Vertical profile levels | 17 | 17 |
| Lowest level | ECMWF surface pressure | 1000 hPa, but ECMWF surface pressure used by RTM |
| Aerosol correction | From FRESCO effective cloud fraction | MODTRAN |
| Degradation correction | Yes, by fitting in retrieval | Yes, by prior correction |

### 3.1 Data collection

The methodology for the comparative evaluation of nadir ozone profile data products suggested in this work is hereafter directly applied to data retrieved from measurements by the second generation Global Ozone Monitoring Experiment (GOME-2) aboard MetOp-A (Munro et al., 1998). This platform, launched into an 817 km altitude sun-synchronous polar orbit on 19 October 2006, is the first of three operational meteorological satellites of the EUMETSAT Polar System (EPS) (EUMETSAT, 2013). Based on the heritage of GOME (Burrows et al., 1999), the European atmospheric sounder GOME-2 is a UV-visible spectrometer mounted onto the flight direction side of its host satellite. The instrument scans the atmosphere at nadir across the satellite track within 6 s and with a scan-width of 1920 km, achieving global coverage of the sunlit part of the atmosphere within 1 day. Four spectrometer channels record atmospheric radiance from the UV to the near-infrared (240–790 nm) at 0.24–0.53 nm resolution over ground footprints of about 80 km across track by 40 km along track. Considering the high data availability in that year and the good instrument performance before more severe instrument degradation at later stage, the present study focusses on GOME-2 measurements taken in 2008.

Two types of Level-2 data sets have been provided by the two processing teams from identical GOME-2 L1 data, i.e. version 4.0 radiance data from 2008: global coverage (GC) data sets and station overpass (SO) data sets. Global coverage data are restricted to 3 consecutive days each month (24–26) in order to limit data storage and processing time. Overpass data sets have been provided within a radius of 500 km around all stations archiving ozone profile data to the NDACC Data Host Facility (NDACC, 2013), the SHADOZ (Southern Hemisphere Additional Ozonesonde programme) archive (Thompson et al., 2012), and the World Ozone and Ultraviolet Data Centre (WOUDC, 2013). Note that the co-location criterion with ground-based data is more stringent, with a maximum distance of 200 km (as mentioned in Sect. 6.2).

Although OPERA and RAL L2 retrieval data were provided in HDF5 and net-CDF file formats, respectively, both data sets were converted into homogenized data structures for further processing and analysis (see Sect. 3.3). As KNMI provides nadir ozone profiles, covariance matrices, a priori profiles and averaging kernel matrices for partial columns (in Dobson Unit) only, the latter have been chosen as the nominal quantity for the round-robin exercise. RAL's data delivery approach indeed is significantly different. Its files contain ozone profiles in VMR, number density, and partial column units, with the latter calculated from the former by layer integration over pressure (see next section). RAL also delivers relative covariance matrices, VMR prior profiles, and number density AKMs. Hence, in order to be able to fairly apply the RR evaluation methodology to the same nominal quantity for both retrieval algorithms, RAL covariance matrices, prior

profiles and averaging kernel matrices required unit conversion.

### 3.2 Unit conversion of ozone profile, CVM, and AKM

Performing a proper RR evaluation using identical units for both retrieval scheme outputs is not the only motivation for a profound unit conversion discussion. Co-located ground-based measurements may require a unit conversion as well for the calculation of (relative) ozone profile difference metrics (see Sect. 7). And thirdly, below and in Sect. 5 it is argued that information content studies should be performed from fractional averaging kernel matrices. The creation and conversion of such AKMs therefore deserves proper discussion.

As detailed in Appendix B, all linear unit conversion equations for vertical atmospheric state profiles can be turned into algebraic expressions for vectors and matrices, $\boldsymbol{x}' = \mathbf{M}\boldsymbol{x}$ (Eq. B5), with the size and elements of conversion matrix $\mathbf{M}$ depending on the units of both state vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ involved, as shown in Table B1. The same matrix $\mathbf{M}$ can then be applied for straightforward unit conversion of covariance matrices ($\mathbf{S}' = \mathbf{MSM}^T$ from Eq. B6) and averaging kernel matrices ($\mathbf{A}' = \mathbf{MAM}^{-1}$ from Eq. B7) as well.

When converting averaging kernel matrices between units however, one additionally has to make sure that information content measures, like e.g. vertical sensitivity as the AKM row sums, are independent of those units and always conserved. Information content measures must therefore be determined from fractional (or relative) AKMs only. A relative AKM element $\mathbf{A}_\mathrm{R}(i,j)$ contains the retrieval weight as the relative (unit-less) variation on level $i$ of the retrieved profile $\boldsymbol{x}_\mathrm{r}$ due to a relative (unit-less) unit-perturbation on level $j$ of the true profile $\boldsymbol{x}_\mathrm{t}$:

$$\mathbf{A}_\mathrm{R}(i,j) = \frac{\partial \boldsymbol{x}_\mathrm{r}(i) \,/\, \boldsymbol{x}_\mathrm{r}(i)}{\partial \boldsymbol{x}_\mathrm{t}(j) \,/\, \boldsymbol{x}_\mathrm{t}(j)} = \mathbf{A}(i,j)\,\boldsymbol{x}_\mathrm{r}(i)^{-1}\boldsymbol{x}_\mathrm{t}(j)\,. \tag{2}$$

Switching between a regular AKM form $\mathbf{A}$ and its relative form $\mathbf{A}_\mathrm{R}$ is thus in practice easily accomplished on element level using

$$\mathbf{A}_\mathrm{R}(i,j) = \mathbf{A}(i,j)\,\boldsymbol{x}_\mathrm{r}(i)^{-1}\boldsymbol{x}_\mathrm{r}(j)\,, \tag{3}$$

where the last factor introduces the values of a retrieved profile $\boldsymbol{x}_\mathrm{r}$ as evidently valid estimates for the true atmospheric state, which, based on the last factor in Eq. (2), should actually be used. Analogously, the fractionalization formula for the covariance matrix elements is given by

$$\mathbf{S}_\mathrm{R}(i,j) = \mathbf{S}(i,j)\,\boldsymbol{x}_\mathrm{r}(i)^{-1}\boldsymbol{x}_\mathrm{r}(j)^{-1}. \tag{4}$$

The above Eqs. (3) and (4) implicate that fractional averaging kernel matrices (and covariance matrices) are identical for volume mixing ratio and number density units. Information content measures are therefore automatically conserved. For conversion of VMR or ND fractional covariance or averaging kernel matrices into their sub-column counterparts (defined in between levels), the regular algebraic conversion equations from Appendix B can be applied at the condition of replacing the regular conversion matrix $\mathbf{M}$ by

$$\mathbf{M}_\mathrm{R} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 1 \end{pmatrix} \tag{5}$$

again an $L$ by $N$ matrix, turning the fractional matrix conversion formulas into mere pairwise linear interpolations from the layer edges (levels) to the layer centres, obviously without unit conversion. This $\mathbf{M}_\mathrm{R}$ can therefore be directly obtained by excluding the unit converting quantities $u$ and $\Delta p$ or $\Delta z$ from M3 or M4 in Table B1.

A few important notes should be added to the conversion discussion outlined above.

- Fractional kernel re-normalization using appropriate perturbation function shapes (those of the AKM construction during retrieval) for row integration could additionally be required to ensure vertical sensitivity conservation. It can be easily verified however that, using Eq. (5) as an expression for the relative AKM conversion matrix, sensitivity (next to other information content measures) is automatically conserved if the triangular level perturbation functions of VMR or ND kernels are substituted by rectangular layer perturbation functions for the corresponding partial column AKM (as has been achieved for the RAL data, see vertical sensitivity plots in Fig. 8). Indeed, the sum of two triangular surfaces with base $\Delta z$ or $\Delta p$ and heights $x_i$ and $x_{i+1}$ equals the surface of a rectangle with identical base and with height $(x_i + x_{i+1}) \,/\, 2$.

- Sensitivity is also not necessarily identical for the fractionalized forms of the regular $\mathbf{A}$ and $\mathbf{A}'$ in the absolute kernel conversion formula due to the significantly more complex (and not always unique) pseudo-inverse calculation and multiplication. For information content studies from fractional averaging kernel matrices (see Sect. 5), it is therefore highly recommended first to determine the relative form of the provided AKM and then to apply relative kernel conversion with sensitivity conservation, rather than determining the converted absolute AKM first and fractionalizing this matrix afterwards.

- Vertical regridding has to be additionally accomplished when switching between profile, CVM, or AKM representations with different retrieval levels, especially when interpolations are required. Although this can be accomplished by use of the conversion expressions

Eqs. (B5) to (B7) in Appendix B with the conversion matrix **M** correspondingly adopted, it is much easier to apply this section's methodology first, and to perform a straightforward regridding of the resulting vertical profiles and matrices afterwards (see Sect. 6).

– The above discussion addresses differences in retrieval units and in vertical levels or layers, but not in the vertical coordinate system (i.e. geopotential height, geometric altitude or atmospheric pressure) and its associated units. If data sets are provided in different vertical coordinate systems, without associated transfer function between the two systems being readily available, several conversion options are possible, e.g., using meteorological analyses, a standard atmosphere, or the ideal gas law and hydrostatic equation. Although those data manipulations do not affect the outlined conversion of the vertical atmospheric state profile, covariance matrix, or averaging kernel matrix, they can introduce non-negligible height-allocation errors, and therefore special care is required.

### 3.3 Harmonization of satellite data representations

As both the OPERA and RAL retrieval schemes provide nadir ozone profiles on identical vertical pressure grids, no regridding is required to obtain a harmonized satellite data representation. Unit conversions and mutual flagging (see Sect. 4.2) nevertheless remain compulsory. When producing harmonized data structures from the HDF5 L2 data sets of KNMI, the fractionalized forms of the CVM and AKM are computed from their sub-column counterparts. The prior and retrieved nadir ozone profiles are, as for the retrieved profile, provided in partial column units already. Slant column densities (SCDs), which are required for later analysis, could therefore also be directly determined for all profiles of both algorithms. They have been approximated by dividing the total retrieved ozone column by the solar zenith angle's cosine.

RAL's net-CDF files contain retrieved ozone profiles in VMR, ND, and sub-column units, with the latter calculated from the first on a high resolution pressure grid (41 levels). (Determining sub-column profiles from the provided VMR or number density data by use of matrix M3 or M4 in Table B1, respectively, hence yields slightly different results; see Sects. 6 and 7.) Averaging kernel matrices, provided initially in absolute values, were turned into relative AKMs, and then converted to fractional sub-column kernels. To enable the upcoming vertical averaging kernel smoothing of ground-based reference profiles in ND units (see Sect. 6), the VMR prior profiles have been transformed into the delivered AKM units (ND) by application of M5 from Table B1. RAL's covariance matrices, also provided in ND units, have been added to the data in their fractionalized forms, together with conversions to VMR and sub-column units. As a result, both absolute and relative errors are available in all provided units for both the OPERA and RAL retrieval outcomes.

## 4 Data set content study

The second part of the round-robin flowchart in Appendix A focusses on the data content study for the satellite retrieval algorithms involved. Inspection of the atmospheric state profile's horizontal resolution and distribution in space and time yields important information on where and when the comparative analysis is valid, and whether or not these geographical and temporal ranges are identical (or at least similar) for the data sets under study (see Sect. 4.3). Mutual profile screening should therefore be applied whenever the spatial and temporal sampling agreement between the retrieval outcomes can be improved, or whenever specific retrieval settings can be (further) aligned. This is illustrated for the OPERA and RAL data sets in Sect. 4.2. A more straightforward Level-2 data file inspection is described first.

### 4.1 Data set inspection

The content of the files provided by the retrieval teams has to be examined first. Although OPERA provides significantly more metadata for each nadir ozone profile than the RAL system, about the same (amount of) information has been used for the evaluation processing of all retrievals. Whereas OPERA data for example contain a vertical atmospheric temperature profile, this quantity is not explicitly included in the RAL output. Knowledge of the VMR and number density nadir ozone profiles for RAL nonetheless easily allows calculating the temperature at each retrieval level by use of Eqs. (B1) or (B2). Also thermal tropopause level and tropospheric and stratospheric ozone columns for each profile are provided within the OPERA files. However, applying OPERA's rather harsh tropopause definition, forced to coincide with one of the 17 predefined pressure levels, to RAL's retrievals of the same satellite measurements yields significant differences for the resultant vertically integrated tropospheric columns of up to 50 %, especially around the South Pole. It has therefore been decided not to use the tropopause information delivered by KNMI, but to determine the more classical thermal tropopause heights (WMO, 1957) from co-located ground measurements whenever required in the round-robin analysis (see Sects. 6 and 7).

Histograms and time series plots of sub-columns, slant column densities, vertical ozone columns, surface properties, cloud properties, solar zenith angles (SZAs), and viewing zenith angles (VZAs) have been created and compared for OPERA and RAL retrieval outcomes. Such plots allow rapid examination of quantity distributions and detection of outliers. Except for the tropospheric column concerns discussed in the previous paragraph, no substantial anomalies have been detected.

### 4.2 Impact of data screening

Each nadir ozone profile retrieval team chose its proper quality assessment and quality control (QA/QC) data screening constraints (also see Table 1), nonetheless resulting in fairly comparable relative total numbers of flagged profiles, as presented in Table 2. Additional flagging of NAN values was applied by the evaluation team on the 17 pressure levels and 16 partial ozone columns in each profile.

Identical filtering for all algorithms involved would be optimal for round-robin evaluation. Unfortunately RAL's retrieval-step-specific cost function and slant ozone column constraints impede this approach in this work (see Table 1). It is nevertheless possible to apply RAL's SZA restriction (SZA $< 80°$ as already in the provided data) to the OPERA output as well. This results in an additional almost 6 % of the total amount of OPERA data to be discarded from future analysis (see Table 2). Taking all flagging into account, still a few hundred thousand "good" GOME-2 (2008) profiles remain for each data set.

### 4.3 Geographical and temporal coverage

Except for the notable absence of RAL profiles around the South-Atlantic Anomaly (SAA) due to the very tight quality control of Band 1 spectra, the geographical coverage of both the OPERA and RAL GC data sets is rather dense and well distributed, as displayed in Fig. 1. SAA profiles are only partially filtered out by KNMI. Filtering moreover reduces the initial higher number of RAL profiles at high latitudes to become comparable with the number of OPERA profiles above 30° N.

Figure 2, which depicts the daily relative amount of flagged GOME-2 2008 profiles (for GC and SO data combined), reveals another major ozone profile distribution difference between the OPERA and RAL algorithms. Due to its effective January–May B1 SCD filtering (see Table 1), RAL provides a noticeably increased number (20–40 %) of "bad" (thus screened) profiles during this period of the year. From June 2008 onwards, RAL's average daily filtering is reduced to about 5 %, which is lower than OPERA's. The temporal profile distribution of OPERA is more homogeneous around the 10–20 % level, but displays several outliers of $\sim 50$ % in February–March and September–October. For RAL, additional spikes appear in September and December.

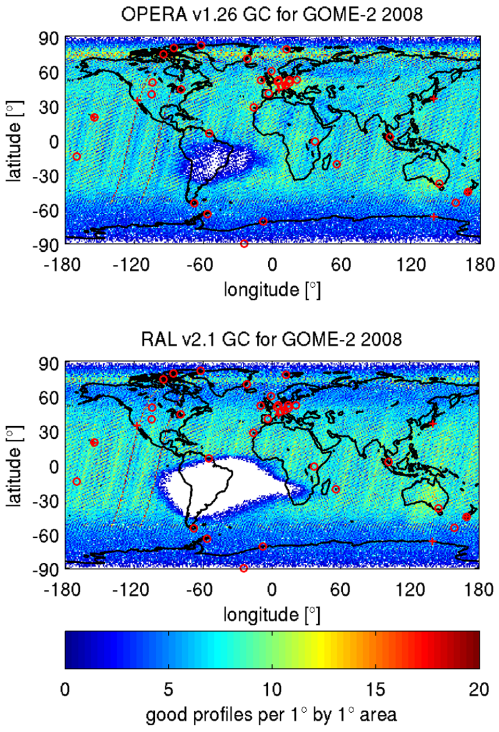



**Figure 1.** Spatial distribution of "good" (not screened) GOME-2 profiles (2008 only) in the OPERA v1.26 and RAL v2.1 global coverage (GC) data sets. Contributing ozonesonde and lidar stations are indicated by red circles and crosses, respectively.

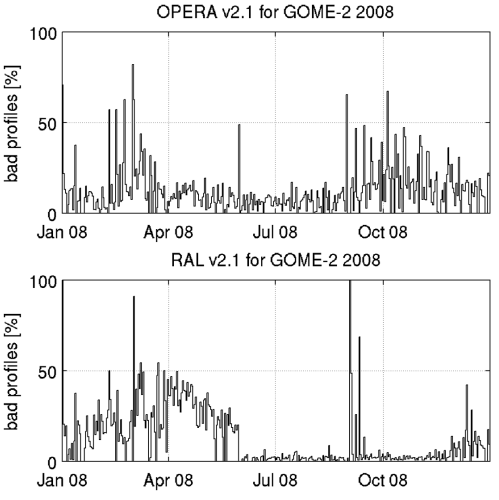



**Figure 2.** Daily relative amount of "bad" (filtered) GOME-2 profiles (2008 only) in the OPERA v1.26 and RAL v2.1 global coverage and station overpass data sets combined.

## 5 Information content study

As aforementioned, retrieved quantities are mixes of a priori constraints and of information that is contributed by the satellite measurement. Each row of the averaging kernel matrix shows, for a given altitude, how the retrieval system either smoothes or amplifies departures of the true profile from the prior profile. A study of algebraic properties of this averaging kernel matrix helps in understanding how the system captures actual atmospheric signals, and is therefore included as a third segment in the round-robin evaluation scheme in Appendix A. Starting from the provided AKMs, one can derive and study several diagnostic parameters and quality indicators of the retrieval, such as the degrees of freedom of the sig-

**Table 2.** Numbers of profiles and screening (profiles omitted from future analysis) percentages for both GOME-2 (2008) nadir RR data sets.

| | OPERA v1.26 | | | | RAL v2.1 | | | |
|---|---|---|---|---|---|---|---|---|
| | Total | Flagged (%) | SZA > 80° (%) | Good | Total | Flagged (%) | SZA > 80° (%) | Good |
| Global coverage | 442 335 | 5.7 | 5.7 | 391 626 | 387 452 | 7.7 | 0.0 | 357 661 |
| Station overpass | 405 363 | 4.6 | 5.5 | 364 714 | 236 817 | 13.8 | 0.0 | 201 143 |

nal (DFS), measurement quality quantifier (MQQ), eigenvectors and eigenvalues, vertical sensitivity, vertical resolution, and centroid offset. Latitudinal and temporal dependences of information content metrics and vertical sensitivity are of particular interest, and their variation is directly related to changes in the solar zenith angle and slant ozone column.

Through straightforward data analysis it can be easily demonstrated that the typical retrieval quality quantifiers discussed in this section (from GC data only) mostly depend on the units when derived from absolute averaging kernel matrices, sometimes called integrating kernels (Bhartia et al., 2013). As these measures however should be unit-independent, fractional AKMs have to be considered instead (although for the small ozone concentrations below the tropopause it can sometimes be insightful to look at figures of merit derived from absolute kernels as well) (Meijer et al., 2006). E.g. vertical resolution measures and their indicative altitudes can only be interpreted correctly if relative perturbations of similar magnitude can be expected at all the altitudes to which a particular retrieved value is sensitive. Exemplary fractional averaging kernel matrices for both OPERA and RAL (from the same satellite measurement) are shown in Fig. 3 (see Sect. 5.3 for centroid definition).

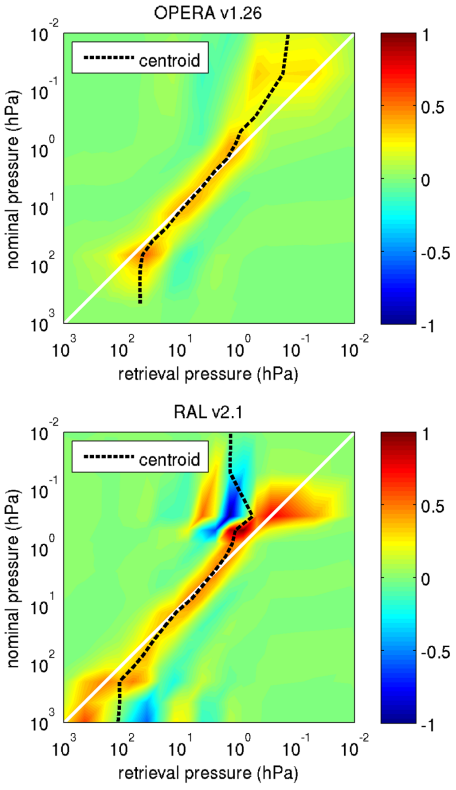



**Figure 3.** Illustrative relative averaging kernel matrix representations with indication of centroid position (see Sect. 5.3) for OPERA v1.26 and RAL v2.1 retrievals from the first GOME-2 2008 measurement over Uccle (Belgium, January 2, 10:06 a.m.).

### 5.1 Information content metrics

Rodgers (Rodgers, 2000, p. 33–34) provides a somewhat intuitive definition for information content: "The information content of a measurement can be defined qualitatively as [the base two logarithm of] the factor by which knowledge of a quantity is improved by making the measurement" or, in other words, as the base two logarithm of the signal-to-noise ratio (SNR), which "corresponds to the number of digits of a binary number required to identify the state." In practice, several closely related but nevertheless different information content measures exist, requiring different algebraic approaches (Rodgers, 2000; Ceccherini et al., 2012):

- signal information load, which is directly determined from the measured irradiance spectrum and therefore retrieval-independent;
- Shannon information content, representing the entropy reduction of the probability distribution function due to the measurement, thus being equal in signal space and in atmospheric state space;
- degrees of freedom of the signal (DFS), representing the number of independent quantities measured, which largely depends on the retrieval process;
- fisher information content, which is calculated from the Fisher information matrix (FIM);
- measurement quality quantifier (MQQ) and derived measures (in particular the grid normalized relative MQQ), determined from the FIM and quantifying the quality of the measurements with respect to the retrieved parameters independently of the retrieval constraints.

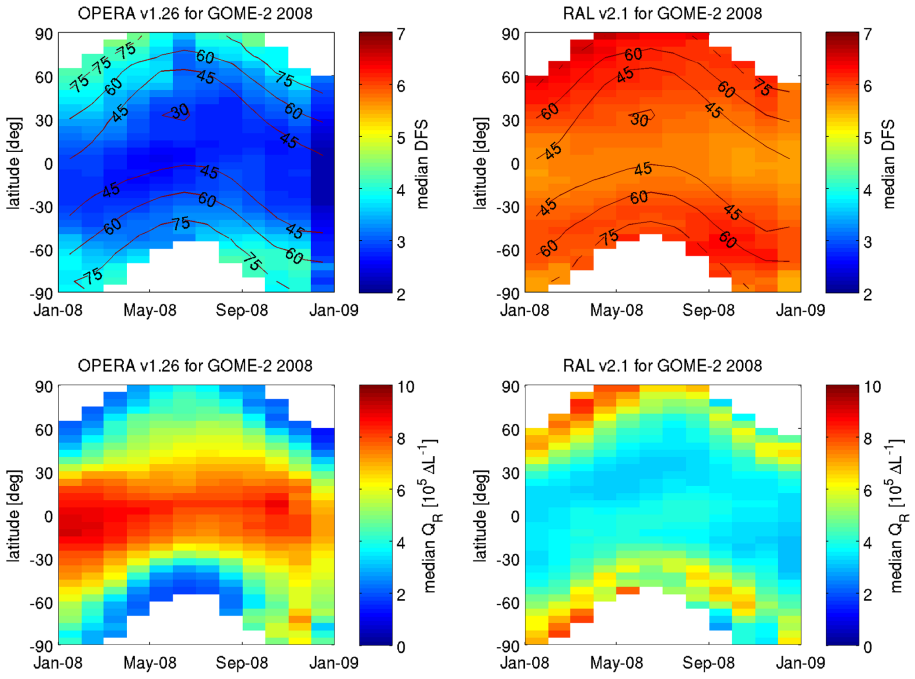



**Figure 4.** Latitude-time cross-section (5° latitude bands, monthly) of median information content measures applied to OPERA and RAL fractional averaging kernel matrices: DFS (upper line) and $Q_R$ (lower line). Solar zenith angle contour plots (values in degrees) have been added to the DFS graphs. The layer thickness metric $\Delta L$ is calculated as log(pressure) difference.

In this work we limit ourselves to discussion and analysis of DFS and grid-normalized relative MQQ (noted $Q_R$). Both quantities are independent of the retrieval units and vertical retrieval grid, but can nevertheless be considered complementary. Whereas the DFS is a non-linear measure (loosely related to the Shannon information content) of the number of independent layers of information that can be determined, the grid-normalized relative MQQ is purely additive and provides an alternative characterization of the (non-additive) signal information load (Ceccherini et al., 2012).

Despite their complementarity, DFS and $Q_R$ can be determined very similarly, i.e. as a vertically integrated density distribution of the (fractional) AKM diagonal elements and fractional Fisher information matrix ($\mathbf{F}_R$) diagonal elements, respectively. The mathematical analogy between both quantities is outlined in Table 3 from top to bottom. In practice, the DFS equals the trace of the AKM and is therefore automatically independent of the retrieval units, vertical retrieval grid, and prior nadir ozone profile. The fractional FIM on the other hand is calculated as $\mathbf{F}_R = \mathbf{S}_R^{-1}\mathbf{A}_R$ for optimal estimation retrievals, yielding unit-independent MQQ values that differ from those that are calculated from the absolute $\mathbf{F} = \mathbf{S}^{-1}\mathbf{A}$ (Ceccherini et al., 2012).

Vertically resolved information (density) quantifiers like the layer DFS density $\boldsymbol{d}(i)$ and the relative information distribution $\boldsymbol{f}_R(i)$ (see Table 3) are not discussed here, although their analysis and complementarity can add to the understanding of atmospheric state measurement and retrieval processes and their vertical dependences; e.g. see Ceccherini et al. (2013). Within the presented round-robin type retrieval quality assessment however, the respective total amounts of information and the vertical sensitivity distribution (see Sect. 5.2) have – for reasons that are further elucidated – been selected to be of major focus.

The DFS difference between OPERA and RAL is remarkable, with RAL's being typically about three points higher, as can be seen in Fig. 4. This difference partially reflects the signal degradation correction on the prior by RAL, which was not implemented for OPERA. The reduced OPERA DFS is also due to its stronger a priori constraint (APC) in the retrieval, mainly up to the middle stratosphere. OPERA's APC moreover has greater seasonal variability than RAL's, leading to a somewhat stronger seasonal DFS variation. Especially for the more fluctuating OPERA kernels, the meridian and temporal DFS dependences can be related to the solar zenith angle (and slant ozone column) behaviour (also in Fig. 4). The effect is smaller for the RAL retrievals, where seasonal DFS variations of only about 0.5 on 5.5 to 6.5 can be observed. The 17 vertical retrieval levels nevertheless oversample the vertical resolution expected on the basis of the number of AKMs' degrees of freedom for both algorithms.

The grid-normalized relative MQQ discrepancy between OPERA and RAL may seem just as important as their DFS difference. Although $Q_R$ ranges between 3 and 8 in units $10^5$ per log(hPa) for both retrieval schemes, OPERA and RAL results almost act oppositely in their temporal-meridian

**Table 3.** Degrees of freedom of the signal (DFS) and relative measurement quality quantifier (MQQ), together with related quantities, with $\mathbf{A}_\mathrm{R}$ the fractional AKM, $\mathbf{F}_\mathrm{R}$ the fractional FIM, and $\Delta L$ the layer thickness in km or log pressure. Units are provided between brackets.

| Quantifier | DFS $= \sum_i \mathbf{A}_\mathrm{R}(i,i)$ [/] | Rel. MQQ $= \sum_i \mathbf{F}_\mathrm{R}(i,i)$ [/] |
|---|---|---|
| Quantifier distribution (or vertical components) | $\mathbf{A}_\mathrm{R}(i,i)$ [/] | $\mathbf{F}_\mathrm{R}(i,i)$ [/] |
| Quantifier density distribution (or vertical density components, which are grid independent) | $\boldsymbol{d}(i) = \frac{\mathbf{A}_\mathrm{R}(i,i)}{\Delta L_i}$ [$\Delta L^{-1}$] | $\boldsymbol{f}_\mathrm{R}(i) = \frac{\mathbf{F}_\mathrm{R}(i,i)}{\Delta L_i^2}$ [$\Delta L^{-2}$] "relative information distribution" |
| Vertically integrated quantifier density distribution (grid independent) | $\sum_i \boldsymbol{d}(i)\,\Delta L_i =$ DFS [/] | $Q_\mathrm{R} = \sum_i \boldsymbol{f}_\mathrm{R}(i)\,\Delta L_i$ [$\Delta L^{-1}$] "grid-normalized relative MQQ" |

behaviours (see Fig. 4). This effect may be understood as a strong SZA dependence of the grid-normalized relative MQQ with rather prominent threshold at about 60°. OPERA attains significantly more information for lower solar zenith angles, whereas RAL performs better for higher SZA. As the SZA dependence of the DFS is much smaller than that of the MQQ, it has to be concluded that the OPERA and RAL covariances (CVM values) reflect the SZA-dependent information content variation of the retrieval process more strongly than the respective averaging kernel matrices.

Information metrics studies often include AKM eigenvectors and eigenvalues discussions (Rodgers, 2000). They can however be considered to echo the vertical sensitivity distribution (see next section) and DFS behaviour (for individual degrees of freedom, as the DFS equals the eigenvalue sum), respectively, and are as such not included here.

### 5.2 Vertical sensitivity and its variations

The areas of the fractional AKM rows (i.e. AKM row sums in practice) are a unit-independent normalized measure for how sensitive the retrieval at a certain height is to all heights (levels) of the true atmospheric state. The vertical sensitivity therefore also "in general can be thought of as a rough measure of the fraction of the retrieval that comes from the data, rather than from the a priori" (Rodgers, 2000, p. 47) and, within each averaging kernel, ideally peaks at the nominal retrieval altitude (see next section). Yet the vertical sensitivity is also defined by the physics in the retrieval's radiative transfer function. As a result, nadir ozone profiles from UV spectrometry for example characteristically show less sensitivity in the UT/LS (upper troposhere to lower stratosphere) region right up to the underside of the stratospheric ozone peak. It is therefore no surprise that the two round-robin retrieval algorithms under study show certain common sensitivity behaviour with height.

Figure 5 displays the OPERA and RAL vertical sensitivities, averaged over the complete global coverage data sets, and represented as a function of altitude (vertical scale in pressure unit), time (top), latitude (middle), and ozone slant column density (bottom). Between about 5 and 50 km in altitude (900 and 1 hPa), sensitivity values go up to 1 or even 2 (over-sensitivity), while outside of this altitude range the sensitivity drops down to vanish in the lowest layer and in the mesosphere (0–0.5 as under-sensitivity). Looking simultaneously at the middle and bottom graphs of Fig. 5, one concludes that the main driver of sensitivity changes is its strong dependence on the ozone SCD. The temporal and meridian evolutions of sensitivity reflect changes in the SCD with solar zenith angle, latitude, and time.

The OPERA and RAL vertical sensitivities however follow clearly distinctive temporal-meridian patterns. RAL shows a much more constant behaviour with clear and increased sensitivity edges in altitude, while the OPERA sensitivity varies more strongly and does not always show sharp edges, especially near the upper boundary (around 1 hPa). This behaviour can be directly related to the DFS and MQQ dependences on SZA and SCD as discussed in the previous section and are confirmed by the slant ozone column density behaviour of the sensitivity profiles (again see Fig. 5): especially for OPERA the sensitivity profile differences between low slant ozone columns and low latitudes on the one side, and between high slant columns and high latitudes on the other hand are very clear. Going from the former to the latter, the over-sensitivity at low altitudes seems to be redistributed, resulting in a unit-sensitivity region reaching higher stratospheric regions (mainly at the poles). RAL's over-sensitivities likewise slightly decrease towards increasing slant column densities, however without the partially compensating effect seen in the OPERA data and therefore resulting in a more reduced overall sensitivity.

### 5.3 Vertical resolution and height registration

User requirements in the form of upper limits for the vertical resolution and the altitude registration uncertainty usually apply for nadir ozone profile retrievals (van der A, 2011). In the literature, several methods have been proposed to esti-

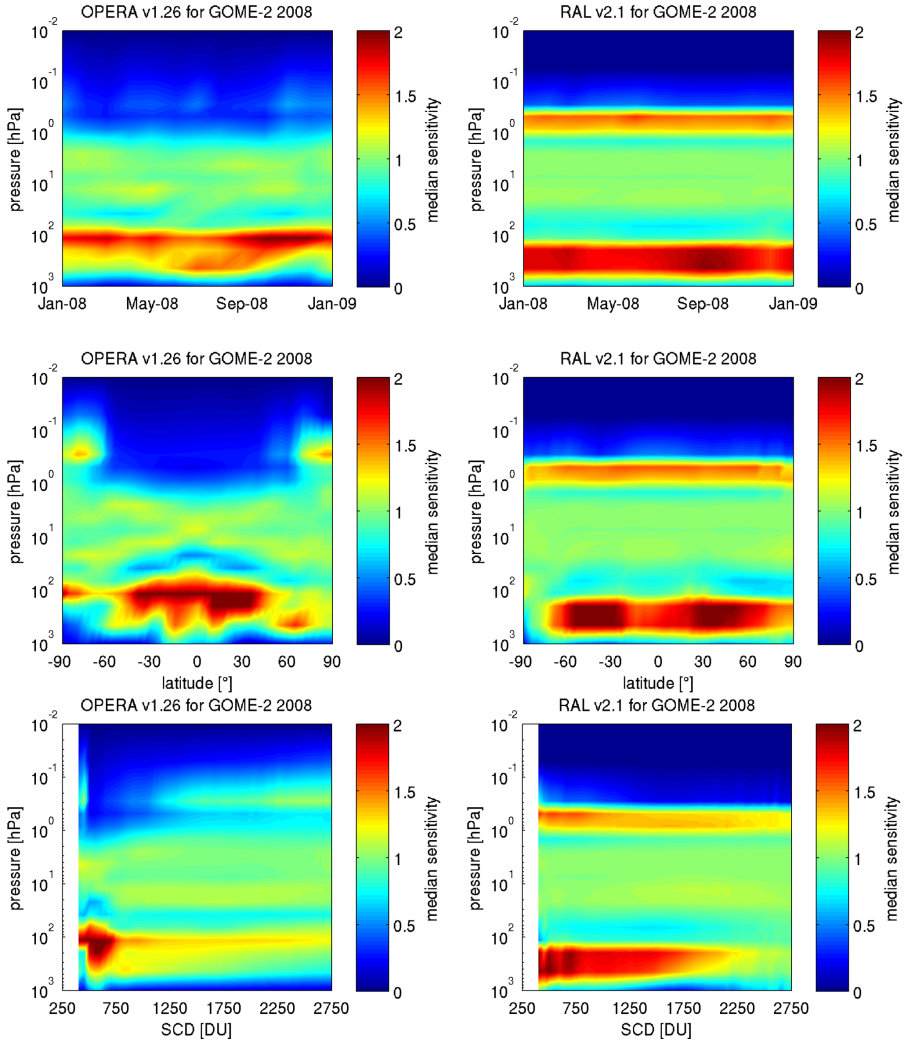



**Figure 5.** OPERA v1.26 (left) and RAL v2.1 (right) GOME-2 (2008) median sensitivity profiles as a function of time (top), latitude (middle), and slant column density (SCD, bottom) for the global coverage data sets.

mate the height-dependent (hence vectorial) vertical resolution of a nadir ozone profile retrieval value from the associated vertical averaging kernels (Rogers, 2000):

- Purser–Huang data density reciprocal;
- sinusoidal disturbance response;
- direct FWHM ($\boldsymbol{W}_{\mathrm{D}}$);
- Gaussian FWHM ($\boldsymbol{W}_{\mathrm{G}}$);
- second moment about the nominal altitude;
- second moment about the kernel mean's altitude;
- Backus–Gilbert spread around the nominal altitude ($\boldsymbol{s}_{\mathrm{BG}}$);
- Backus–Gilbert spread around the centroid altitude or resolving length ($\boldsymbol{r}_{\mathrm{BG}}$).

The Purser and Huang (1993) data density reciprocal equals the range of the height covered divided by the number of independent quantities measured or $\boldsymbol{d}(i)^{-1}$ (see Table 3). This value thus represents the altitude range (in km) covering 1 degree of freedom. The response of the retrieval to sine wave perturbations in the state on the other hand is easy to graph but hard to interpret quantitatively (Rogers, 2000) and therefore not further considered in this work. The next six estimates in the list above are more direct vertical resolution measures in terms of the width of the averaging kernel or point-spread function. Second moment definitions however are "unsatisfactory, as there are regions where the integrand is negative, so the square root cannot be taken." (Rodgers, 2000, p. 61), mostly resulting in discontinuous vertical resolutions. Very straightforward full width at half maximum (FWHM) values can be determined from the averaging kernels directly or after Gaussian fitting. Finally, Backus and

Gilbert (1970) provided a method for determining the vertical resolution as a kernel-weighed and normalized square distance. The Backus–Gilbert spread $\boldsymbol{s}_{\mathrm{BG}}(i)$ about the nominal altitude $z_i$ of level $i$ is given by (in discretized form)

$$\boldsymbol{s}_{\mathrm{BG}}(i) = 12\frac{\sum_j \left(z_j - z_i\right)^2 \mathbf{A}_{\mathrm{R}}^2(i,j)\,\Delta z_j}{\left(\sum_j \mathbf{A}_{\mathrm{R}}(i,j)\,\Delta z_j\right)^2}. \quad (6)$$

This expression however does not take into account the fact that some kernels have dislocated centroids, meaning that the retrieval's sensitivity barycentre is located far away from the nominal altitude. Therefore the Backus–Gilbert spread about the centroid or "resolving length" has been defined. With $z_i$ replaced by $\boldsymbol{c}(i)$ in Eq. (6), one obtains:

$$\boldsymbol{r}_{\mathrm{BG}}(i) = 12\frac{\sum_j \left(z_j - \boldsymbol{c}(i)\right)^2 \mathbf{A}_{\mathrm{R}}^2(i,j)\,\Delta z_j}{\left(\sum_j \mathbf{A}_{\mathrm{R}}(i,j)\,\Delta z_j\right)^2}, \quad (7)$$

where the centroid altitude $\boldsymbol{c}(i)$ of level $i$ is calculated as the altitude-weighed mean of the squared vertical averaging kernel, taking the (again discretized) form (Rodgers, 2000):

$$\boldsymbol{c}(i) = \frac{\sum_j z_j \mathbf{A}_{\mathrm{R}}^2(i,j)\,\Delta z_j}{\sum_j \mathbf{A}_{\mathrm{R}}^2(i,j)\,\Delta z_j}. \quad (8)$$

Being in possession of an expression for determining the barycentre of the retrieved information, it seems valuable to include the centroid offset $\boldsymbol{c}(i) - z_i$ in the retrieval study as well. The centroid offset quantifies the displacement of the centroid from the nominal altitude and as such provides an estimate of the uncertainty on the retrieval height registration (also see dashed lines in Fig. 3). Note that the vertical resolution estimated by the data density reciprocal does not at all account for non-vanishing centroid offsets, nor for the actual averaging kernel shape, while the FWHM definitions automatically yield vertical resolutions from kernel values that are by rough approximation located symmetrically around the retrieval centroid.

Altitude cross-sections of time-averaged OPERA and RAL centroid offsets are represented on the right in Fig. 6. Except for a 5 km difference below about 15 km in altitude (roughly at the tropopause), median relative AK centroid offsets are very similar for OPERA and RAL. Between approximately 15 km and 50 to 60 km (100 to 0.2 hPa) in altitude (pressure), the centroid of each retrieval coincides with the nominal altitude within a few kilometres, thus well below the values of the resolving length (see below). Above and below this altitude range, thus at altitudes of poor sensitivity, deviations of typically up to 10 or 20 km appear, which are values of the order of the resolving length. Reduced offsets at low altitudes occur for RAL profiles that correspond to small solar zenith angles (i.e. with the Sun close to zenith). More generally, annual changes in the slant column density in polar areas, and thus in sensitivity, drive a similar annual change in the centroid offset at low altitudes.

Median vertical resolutions calculated by use of the five selected approaches and applied to both OPERA and RAL data are also shown in Fig. 6. Despite the rudimentary definition, the data density reciprocal sometimes agrees rather well with the four other vertical resolution estimators that take account of kernel shape, mainly for OPERA. RAL $\boldsymbol{d}(i)^{-1}$ values often deviate from other vertical resolution estimators and are significantly smaller than the OPERA values, especially at low altitudes. This effect is due to RAL's larger averaging kernel elements near the earth's surface, as a result from its somewhat smaller retrieval barycentre offsets.

OPERA's median direct FWHM resolution estimate is considerably different. A forthright assessment of the averaging kernel shape around the centroid clearly results in the lowest vertical resolution values. The FWHM estimate obtained from a Gaussian fit to the AK deviates from the direct FWHM. Gaussian fitting widens the kernel shape at high altitudes (above 60 km), while being undefined (set to zero) below the tropopause due to incompleteness (i.e. too large a fraction of the Gaussian function is located below the ground level).

The effect of the nominal vs. centroid reference altitude in the Backus–Gilbert spread definition becomes very clear in Fig. 6. OPERA's resolving length ($\boldsymbol{r}_{\mathrm{BG}}$) for example is much closer to its direct FWHM counterpart ($\boldsymbol{W}_{\mathrm{D}}$): in the mesosphere and the lower troposphere, where ozone values are small, the spread increases to unrealistic values of over 40 km due to kernel fluctuations (see example in Fig. 3) far from the diagonal that are not picked up by the FWHM definitions. The Backus–Gilbert spread about the centroid is nevertheless considered to be the optimal vertical resolution estimate, as (i) it properly takes into account both the negative and positive contributions in the averaging kernel, (ii) it gives values for all shapes of kernels, and (iii) it handles most effectively the kernels with a dislocated retrieval barycentre (Meijer et al., 2006).

Not shown in Fig. 6 is that in polar areas an annual variation appears clearly from the ground up to the lower stratosphere and also above the middle stratosphere, with maxima in winter and minima in summer. This behaviour correlates directly with the annual variation of the slant column density (highest in winter and lowest in summer) and thus of the sensitivity: the poorer the sensitivity, the larger the resolving length, or the more the retrieval at a certain level is influenced by ozone concentrations at adjacent levels. The resolving length of both OPERA and RAL for example reaches its minimum of about 10–20 km in the UT/LS and middle atmosphere (15–60 km), although this region is considerably smaller for RAL.

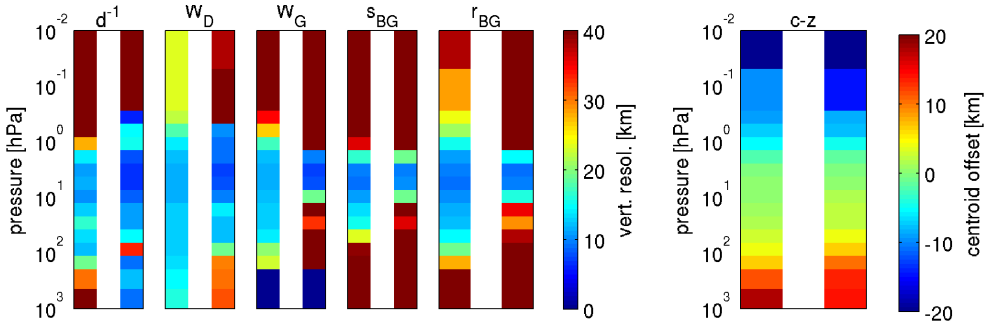



**Figure 6.** Median vertical resolution estimates (abbreviated as vert. resol.) and centroid offset of GOME-2 (2008) data retrieved by OPERA v1.26 (each left colour profile) and RAL v2.1 (each right colour profile). Note that the centroid offset plots and vertical resolution plots have different colour scales, although their absolute ranges are identical (40 km). $\boldsymbol{d}^{-1}$, $\boldsymbol{W}_{\mathrm{D}}$, $\boldsymbol{W}_{\mathrm{G}}$, $\boldsymbol{s}_{\mathrm{BG}}$, $\boldsymbol{r}_{\mathrm{BG}}$, and $c - z$ denote the data density reciprocal, direct and Gaussian FWHM, Backus–Gilbert spread around the nominal and centroid altitude, and centroid offset, respectively.

## 6 Comparison pre-processing

### 6.1 Correlative measurement selection

Ground-based data records obtained by ozonesondes and stratospheric ozone lidars are used as a transfer standard against which the OPERA and RAL GOME-2 retrievals are compared. As described in Sect. 3, the selection and post-processing of reference data is included in the first segment of the evaluation scheme (Appendix A). Like for the satellite data, prior to searching for co-locations, data screening has been applied to the ground-based correlative measurements, both on the entire profiles and on individual levels. The recommendations of the ground data providers to discard unreliable measurements are followed. Next to that, measurements with unphysical (e.g. negative or NAN) pressure, temperature, or ozone readings are rejected automatically. Ozonesonde measurements at pressures below 5 hPa (beyond 30–33 km) and lidar measurements outside of the 15–47 km vertical range are rejected as well.

#### 6.1.1 Ozonesondes

In situ measurements of ozone are carried out regularly by ozonesondes on-board small meteorological balloons launched at numerous sites around the world. They measure the vertical profile of ozone partial pressure with order of 10 m vertical sampling (100–150 m actual vertical resolution) from the ground up to the burst point of the balloon, usually between 30 and 33 km. An interfaced radiosonde provides the pressure, temperature and GPS data necessary to geolocate each measurement and to convert ozone partial pressure into other units. Normalization factors are not applied.

Different types of ozonesondes were developed over the years. Those still in use today are based on the electrochemical reaction of ozone with a potassium-iodide sensing solution. Laboratory tests and field campaigns indicate that between the tropopause and about 28 km altitude all sonde types produce consistent results when the standard operating procedures are followed (Smit and ASOPOS panel, 2011). The estimated bias is smaller than ±5 %, and the precision remains within the order of 3 % (Smit et al., 2007; Deshler et al., 2008). Above 28 km the bias increases for all sonde types. Below the tropopause, due to lower ozone concentrations, the precision decreases slightly to 3–5 %, depending on the sonde type. The tropospheric bias also becomes larger, between ±5 and ±7 %. Other factors besides ozonesonde type influence the data quality as well. A detailed overview can be found in (Smit and ASOPOS panel, 2011).

This work uses ozonesonde data acquired at stations of the Network for the Detection of Atmospheric Composition Change (NDACC), Southern Hemisphere Additional Ozonesonde programme (SHADOZ), and other sites of WMO's Global Atmosphere Watch. The ozonesonde stations considered in this work are marked by red circles in Fig. 1.

#### 6.1.2 Lidars

A differential absorption lidar (DIAL) operates mostly during clear-sky nights, simultaneously emitting two pulsed laser beams at wavelengths with a different ozone absorption cross-section. The backscattered signal is integrated over a few hours to retrieve the vertical distribution of ozone (Mégie et al., 1977). A stratospheric ozone lidar system emits beams at 308 nm for the absorbed wavelength and 353 or 355 nm for the reference one, which makes it sensitive from the tropopause up to about 45–50 km altitude with a vertical resolution that declines with altitude from 0.3 to 3–5 km (Godin et al., 1999). The profiles are reported as ozone number densities vs. geometric altitude.

The DIAL technique is in principle self-calibrating since the ozone profile is retrieved directly from the returned signals without introducing instrumental constants. However, interference by aerosols, signal induced noise in photomultipliers and saturation of the data acquisition system can degrade the quality of the measurements. Unreliable mea-

surements can be discarded based on the reported precision, which decreases with altitude. The estimated bias and precision are about ±2 % between 20 and 35 km and increase to ±10 % outside this altitude range where the signal-to-noise ratio is smaller (Keckhut et al., 2004). The consistency between six lidars in the NDACC network was recently studied by (Nair et al., 2012) using various satellite data sets. They concluded that the different lidar records agree within ±5 % of the space-based observations over the range of 20–40 km.

This work uses data from all stratospheric ozone lidars that have been operational in the NDACC network during the year 2008. The selected lidar stations sample latitudes from 80.0° N–67° S, but most sites are located in the Northern Hemisphere (see red crosses in Fig. 1 for those considered here).

### 6.2 Co-location settings

The setting of coincidence criteria and the subsequent extraction of co-located ozone profiles from both the satellite and reference data sets is included in the fourth part of the validation scheme (see Appendix A). From the screened GOME-2 station overpass data sets and the approved NDACC (ozonesonde and lidar) and WOUDC (ozonesonde only) ground-based measurement data, several co-located subsets have been extracted. Only co-locations with a maximal spatial distance $\Delta r$ of 200 km and a maximal time difference $\Delta t$ of 2 h for ozonesondes and 12 h for lidars were allowed. The spatial co-location window is thus chosen to be of the order of the horizontal resolution of the co-added satellite pixels, while the limit temporal differences are a compromise between a sufficient overlap of the sampled air masses and a sufficiently large sample for statistical analysis (see next section). When multiple satellite profiles co-locate with one unique ground measurement, only the closest GOME-2 measurement is kept, being mostly the same for OPERA and RAL. The ground measurement is typically located within the selected satellite pixel as a result. The closest (space-time) distance $\Delta s$ is thereby calculated by incorporating an estimated average air mass drift at 100 km h$^{-1}$: $\Delta s^2 = \Delta r^2 + (100 \times \Delta t)^2$.

### 6.3 Co-located data sets study

The correlative measurement selection from the previous section must be justified by a study of the co-located data sets, as outlined in the fifth section of the round-robin validation flowchart (Appendix A): based on (i) the spatial and temporal scale of the atmospheric ozone variation and on (ii) the geographical and temporal sampling statistics of the satellite-reference coincidences, the co-location criteria should be iteratively adjusted.

The latitude-time sampling of the closest satellite–ground co-locations (after filtering) resulting from the constraints outlined above is displayed in Fig. 7 for the two GOME-2 retrieval schemes under study. As only stations are considered for which both retrieval schemes provided co-located data, their coincidence sampling is very similar, and RAL's lack of atmospheric monitoring around the SAA will not affect the comparison statistics in Sect. 7. The north–south latitude range is sufficiently covered throughout the year 2008, local polar winters apart (where GOME-2 does not measure). This is mainly due to the high number of balloon-launching stations (32). Appropriate flagging unfortunately reduces the number of lidar stations to seven, with five in the Northern and two in the Southern Hemisphere.

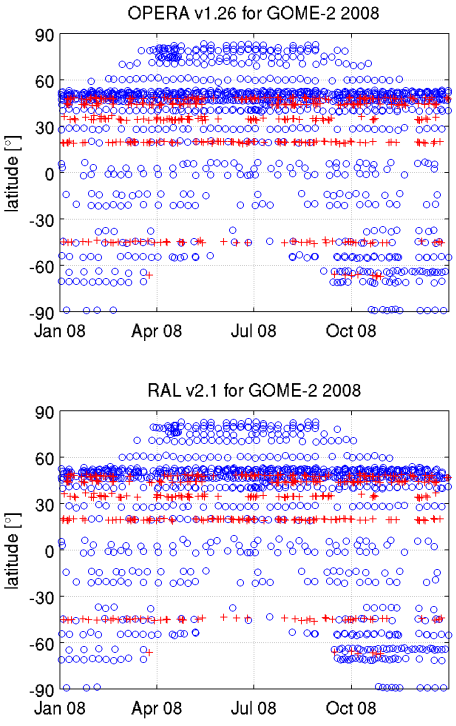



**Figure 7.** Latitude-time sampling of closest co-locations between OPERA v1.26 and RAL v2.1 nadir ozone profile retrievals of GOME-2 (2008) observations on the one hand, and ground-based ozonesonde (blue circles, 200 km and 2 h distance at maximum) and lidar (red crosses, 200 km and 12 h distance at maximum) measurements on the other hand. Latitude values correspond to the satellite position during measurement. Ground measurements are obtained from the NDACC, SHADOZ, and WOUDC databases.

The average number of GOME-2 2008 co-locations per ozonesonde station is 32, with a maximum of 147 at Payerne (for OPERA) and a minimum of 9 at Höhenpeißenberg (for RAL). For the lidar stations the number of co-locations per station is somewhat more uniform, with an average of 56, a maximum of 90 at Mauna Loa (for OPERA), and a minimum of 10 at Dumont d'Urville (for RAL). For both ozonesonde and lidar co-locations the mean spatial distance equals about 70 km, with all values ranging between 2.5 and 199.0 km. Measurement time differences vary from a few seconds (simultaneously in practice) up to 2 h for balloon

measurements, and from 7 to 12 h for night-time lidar registrations, with mean temporal distance values of 55 min and 10.5 h, respectively. Taking into account the typical scales of ozone variation, these numbers justify the comparative validation from (relative) difference metrics in Sect. 7.

### 6.4 Vertical resampling of reference profiles

The satellite nadir ozone profiles and co-located ozonesonde and lidar measurements, next to unit differences, exhibit very different vertical resolutions and vertical smoothing characteristics. These differences are tackled in the sixth part of the round-robin approach outlined here (see Appendix A). Two complementary methods are applied to adapt the high vertical sampling of the ground-based data (and associated uncertainties) to the lower vertical resolution of the satellite data. The first method used is the classical one described by Rodgers and Connor (Rodgers, 2000), which consists in smoothing the high resolution profile (ozonesonde or lidar) with the averaging kernels and a priori profile associated with the satellite retrieval to be validated. The second method used is the independent pseudo-inverse regridding technique described by Calisesi et al. (2005), which avoids the corruption of the correlative measurements by the averaging kernels and prior profile associated with the retrieval to be validated.

Both for regridded and kernel smoothed comparisons, the ground-based ozone profile measurements were first vertically extended (both upwards and downwards) using the McPeters and Labow climatology of 2011 (McPeters and Labow, 2012), resampled to a 100 m vertical grid by Gaussian smoothing, and converted to ozone partial column units (DU) by vertical integration. Ozonesonde measurement profiles are typically provided in partial pressure, easily converted into ND units ($cm^{-3}$) and VMR units (ppmv) thanks to the on-board pressure and temperature measurements. Lidar data on the other hand are mostly given in number density and in general do not contain temperature profiles for a beforehand ND to VMR conversion by use of Eq. (B1). The latter has therefore been accomplished by consistently applying pressure and temperature fields that were extracted from the latest ERA-Interim reanalysis (Dee et al., 2011).

Diverse unit conversion and kernel smoothing approaches can (and sometimes must) be applied (see 12 boxes in section six of the flowchart in Appendix A): ground-based ozone profiles in VMR or number density units require conversion to sub-columns for comparison with OPERA, while for RAL a more direct comparison in VMR or ND is possible as well. Unit conversion of ground-based reference data anyhow optimally (in order to minimize data manipulation) takes place before kernel smoothing for OPERA (pre-conversion) and after kernel smoothing for RAL (post-conversion) because of the partial column and number density units of the initial OPERA and RAL averaging kernel matrices, respectively. Finally, two different AK smoothing methods are common practice but yield somewhat dissimilar results (see Sect. 7). The first contains a direct multiplication of the AKM with a regridded ground-based ozone profile (coarse method), while the second is based on a multiplication of a row-interpolated AKM with the full ground profile (fine method), which maximally exploits the high-resolution reference measurement without adding information to the retrieval data (Ridolfi et al., 2006). Both procedures will be applied and compared in this work, resulting in nine evaluation approaches to be studied in total (corresponding to the nine boxes with black text in the sixth part of Fig. A1), as denoted by Roman numerals.

- *Comparison of OPERA (I) and RAL (II) nadir ozone sub-column profiles with regridded and unit converted ground-based data.* Three different methods can be applied for fine grid VMR or ND to coarse grid partial column conversion of a ground profile: the two most general approaches consist of a unit conversion by use of Eq. (B5) (with conversion matrix M3 or M4 in Table B1), preceded (a) or followed (b) by a pseudo-inverse-regridding (Calisesi et al., 2005) that conserves the profile integral (total column). Remark that partial column *densities* ($\boldsymbol{x}^{\mathrm{PC}}(i) \,/\, \Delta z_i$) have to be considered for sub-column regridding. An additional method (c), that can only be used for conversions to sub-columns and is applied here, consists of a fine resolution conversion of the ground data, followed by a more straightforward summation of the resulting sub-columns over the satellite retrieval pressure levels.

- *Comparison of RAL nadir ozone number density profiles with pseudo-inverse-regridded ground-based profiles (III).* This approach is added to study the effect of unit conversions on the comparisons.

- *Comparison of OPERA partial column profiles with pre-converted and kernel smoothed ground-based data.* For the coarse kernel smoothing, basically obtained by approaching the true atmospheric state $\boldsymbol{x}_{\mathrm{t}}$ by the ground-based ozone profile measurement $\boldsymbol{x}_{\mathrm{g}}$ in Eq. (1) (Connor et al., 1991), from VMR we have (IV)

$$\widetilde{\boldsymbol{x}}_{\mathrm{g}}^{\mathrm{PC}} = \mathbf{A}^{\mathrm{PC}} \left( \mathbf{M}_{\mathrm{c}} \overline{\boldsymbol{x}}_{\mathrm{g}}^{\mathrm{VMR}} \right) + \left( \mathbf{I} - \mathbf{A}^{\mathrm{PC}} \right) \boldsymbol{x}_{\mathrm{p}}^{\mathrm{PC}}, \tag{9}$$

  while the fine kernel smoothing method from the same units takes the form (V):

$$\widetilde{\boldsymbol{x}}_{\mathrm{g}}^{\mathrm{PC}} = \mathbf{A}^{*\mathrm{PC}} \left( \mathbf{M}_{\mathrm{f}} \boldsymbol{x}_{\mathrm{g}}^{\mathrm{VMR}} \right) + \left( \mathbf{I} - \mathbf{A}^{\mathrm{PC}} \right) \boldsymbol{x}_{\mathrm{p}}^{\mathrm{PC}}. \tag{10}$$

  Note that conversion matrix $\mathbf{M}$ has different sizes for the coarse and fine smoothing methods, i.e. coarse by coarse (subscript c) and fine by fine (subscript f), respectively. In the above equations, a bar indicates the representation of ground data on the coarse satellite grid (with $\overline{\boldsymbol{x}}_{\mathrm{g}}^{\mathrm{VMR}}$ in Eq. (9) obtained by pseudo-inverse-regridding), while an upper tilde ($\sim$) indicates that the

profile has been averaging kernel smoothed as well. $\mathbf{A}^*$ is used for a row-interpolated AKM with dimensions coarse grid by fine grid. Going from $\mathbf{A}$ to $\mathbf{A}^*$, the AKM's vertical sensitivity is conserved by imposing the following row-renormalization constraint, with $\Delta L$ representing the layer thickness in km or log pressure:

$$\mathbf{A}_{(\mathrm{c}\times\mathrm{c})}\begin{pmatrix}\Delta L_1\\ \vdots\\ \Delta L_\mathrm{c}\end{pmatrix}=\mathbf{A}^*_{(\mathrm{c}\times\mathrm{f})}\begin{pmatrix}\Delta L_1\\ \vdots\\ \Delta L_\mathrm{f}\end{pmatrix}. \quad (11)$$

– *Comparison of RAL number density profiles with kernel smoothed ground-based data.* The coarse smoothing in number densities is calculated as (VI)

$$\widetilde{\boldsymbol{x}}_\mathrm{g}^\mathrm{ND}=\mathbf{A}^\mathrm{ND}\overline{\boldsymbol{x}}_\mathrm{g}^\mathrm{ND}+\left(\mathbf{I}-\mathbf{A}^\mathrm{ND}\right)\boldsymbol{x}_\mathrm{p}^\mathrm{ND}, \quad (12)$$

while the fine kernel smoothing becomes (VII)

$$\widetilde{\boldsymbol{x}}_\mathrm{g}^\mathrm{ND}=\mathbf{A}^{*\mathrm{ND}}\boldsymbol{x}_\mathrm{g}^\mathrm{ND}+\left(\mathbf{I}-\mathbf{A}^\mathrm{ND}\right)\boldsymbol{x}_\mathrm{p}^\mathrm{ND}. \quad (13)$$

– *Comparison of RAL partial column profiles with kernel smoothed and post-converted ground-based data.* The kernel smoothing steps then are identical to those in the previous case, but both the coarse (VIII) and fine (IX) methods are now followed by a post-conversion of the form:

$$\widetilde{\boldsymbol{x}}_\mathrm{g}^\mathrm{PC}=\mathbf{M}_\mathrm{c}\widetilde{\boldsymbol{x}}_\mathrm{g}^\mathrm{ND}. \quad (14)$$

The sub-column nadir ozone profiles that are provided in the RAL data sets are however calculated from their ND retrievals quite differently (see Sect. 3.3). RAL's VMR integration over pressure differences on a 41 level grid after conversion and interpolation of the kernel smoothed number density profile yields a slightly different result from Eq. (14) and is therefore also considered in the result discussion.

It should be noted that the kernel smoothing outlined in the above equations can also be expressed differently. For the fine-gridded kernel smoothing approaches for example, one can write (ignoring unit conversions):

$$\widetilde{\boldsymbol{x}}_\mathrm{g}=\boldsymbol{x}_\mathrm{p}+\mathbf{A}^*\left(\boldsymbol{x}_\mathrm{g}-\boldsymbol{x}_\mathrm{p}^*\right). \quad (15)$$

Despite the mathematical identity of both forms, it should be clear that for the latter an additional interpolation of the prior satellite profile from the coarse to the fine grid is required, denoted as $\boldsymbol{x}_\mathrm{p}^*$. As it is however preferential to minimize data manipulation, the outlined Eqs. (9) to (14) contain the privileged fine-gridded approach.

Analogous to the OPERA and RAL GOME-2 data, the bottom ozonesonde and lidar partial ozone columns in the comparison data have always been attributed to the full fixed lowest retrieval layer, i.e. roughly from 0–6 km in altitude, although they represent the integrated amount of ozone from the surface upwards. Finally, information on the altitudes of the thermal tropopause and the ozone maximum, as calculated from their closest ground-based co-locations (WMO, 1957), has been added to all comparisons.

## 7 Comparisons with reference data

The median and the 68 % interpercentile (IP 68) of the difference in ozone have been assessed as statistical estimators of the bias and spread between the satellite and reference data, instead of the usual mean and $1\sigma$ standard deviation. In case of a normal distribution of the ozone differences, median and IP 68 are equivalent to the mean and standard deviation, respectively, but they offer the advantage to be much less sensitive to occasional outliers. The calculation of these bias and spread estimators is detailed in Sect. 7.1. Comparison results are reported immediately afterwards, first in the form of global statistics on the bias and spread of ozone differences (Sect. 7.2), and second as a function of different parameters of importance (Sect. 7.3): time, latitude, ozone slant column density, and cloud fractional cover.

### 7.1 Comparison statistics

For each pair of coincident profiles, absolute and relative (in %) difference profiles can be directly calculated as $\boldsymbol{\Delta x}=\boldsymbol{x}_\mathrm{r}-\overline{\boldsymbol{x}}_\mathrm{g}$ and $\boldsymbol{\Delta x}_\mathrm{R}=100\left(\boldsymbol{x}_\mathrm{r}\,/\,\overline{\boldsymbol{x}}_\mathrm{g}-1\right)$, respectively. The absolute bias profile $\boldsymbol{b}$ for a selection of closest co-locations is then determined as the median (50 % quantile $Q$) of the set of its absolute differences, i.e. $Q_{50}\left(\{\boldsymbol{\Delta x}\}\right)$, and provides a vertical profile of the accuracy estimate for the satellite measurement and retrieval combined. The random uncertainty on all absolute bias levels or layers is assessed by the spread on the absolute difference profiles within the same set $\{\boldsymbol{\Delta x}\}$, calculated as their 68 % interpercentile, i.e. their 16 and 84 % interquantile distance:

$$\boldsymbol{s}=\frac{1}{2}\left[Q_{84}\left(\{\boldsymbol{\Delta x}\}\right)-Q_{16}\left(\{\boldsymbol{\Delta x}\}\right)\right]. \quad (16)$$

Similar expressions hold for the relative comparison bias $\boldsymbol{b}_\mathrm{R}$ and spread $\boldsymbol{s}_\mathrm{R}$ profiles, obtained when replacing $\{\boldsymbol{\Delta x}\}$ by $\{\boldsymbol{\Delta x}_\mathrm{R}\}$.

Absolute and relative random satellite errors $\sqrt{\mathbf{S}\left(i,i\right)}$ and $\sqrt{\mathbf{S}_\mathrm{R}\left(i,i\right)}$ have been calculated for comparison with their respective median differences, in order to compare the bias with the random uncertainty of the retrieved nadir ozone profiles. Next to that, combined satellite and ground-based uncertainty profiles – which are also regridded using the method of Calisesi et al. (2005) – are assessed in relation to the spread. This combined uncertainty is given by

$\sqrt{\mathbf{S}(i,i)+\overline{\boldsymbol{e}}(i)^2}$ at level or layer $i$, with $\overline{\boldsymbol{e}}$ the regridded ground-based error (identical for relative quantities with subscript $R$). For both retrieval algorithms, the relative satellite errors amount to about 1 % at the ozone maximum and up to about 15 % (OPERA) and 30 % (RAL) in the troposphere and middle atmosphere.

### 7.2 Full comparison results

Absolute and relative difference statistics for all GOME-2 2008 closest ozonesonde and lidar co-locations and all outlined evaluation approaches (as discussed in Sect. 6) are collected in Fig. 8. Results are provided for (from top to bottom) OPERA v1.26 in partial columns, RAL v2.1 in partial columns, RAL v2.1 in partial columns determined using RAL's high-resolution VMR integration method, and RAL v2.1 number densities. Within each subplot, comparisons with regridded, coarse resolution AK smoothed, and fine resolution AK smoothed ground-based data are shown (in green, blue, and red, respectively). Satellite random uncertainties and satellite–ground combined uncertainties have been included for direct bias and spread appraisal, respectively. Median vertical sensitivities are also added to each plot.

From these full range comparison results, the following conclusions can be drawn, allocated to a number of previously specified RR evaluation concerns (cf. user requirement document and the flowchart in Appendix A):

- *On the consistency of ground-based instruments.* Except for some significant (5–10 %) deviations at the lidars' lower edges, the ozonesonde and lidar bias and spread results agree very well in the vertical overlap region (about 10–100 hPa) and are hence regarded to provide valid reference ozone profile values.

- *On the ground-based vertical resampling approaches.* Regridding and smoothing methodologies typically yield similar results in the middle atmosphere (MA) where only lidar statistics are available. Only in relative units does the fine resolution AK smoothing seem to result in increasingly more negative bias values, though for ever smaller ozone concentrations, with increasing altitude than the other methods. Below about 20 hPa clear differences between ground-based profile regridding on the one hand and its kernel smoothing on the other hand begin to appear and increase towards the ground surface. In the troposphere and UT/LS, median differences get opposite signs for both the OPERA and RAL retrieval results. Here, comparison spreads are up to the order of 10 % higher for the regridded ground-based data in the same vertical region. A forced reduction of the vertical smoothing difference between ground-based and satellite measurements, by application of the averaging kernels to the former, thus has a substantial impact on their resultant difference figures. Small differences can nevertheless be observed between coarse and fine resolution smoothing approaches as well. Again in the troposphere and around the UT/LS, the partial column bias of the fine resolution smoothing often seems to be slightly shifted ($\sim 3$ %) towards more positive values with respect to the coarse smoothing. This might be due to approximation errors during kernel interpolation and summation of the fine-resolution profile over a larger number of ozone layers or an overestimation of the ozone concentrations during regridding of the ground-based profile for the coarse method (or both).

- *On the comparison consistency for different units.* All three comparison approaches applied to the RAL scheme data return qualitatively similar results, as should be expected. Except for discrepancies at the highest and lowest comparison layers, the high-resolution VMR integration bias is very close (deviations of few percentages only) to the regular partial column results (its surface layer drops out because of the additional integration). Both sub-column spread values are nearly indistinguishable. Larger dissimilarities can be perceived with respect to the number density median and interpercentile differences, predominantly around and below the tropopause. The notable fluctuation of the ND bias there is lacking in partial column comparisons, although AK smoothed median differences show much better agreements. Number density spreads reach higher values in the same altitude range.

- *On the RR validation of the retrieval schemes, taking vertical sensitivity into account.* Focussing on the averaging kernel smoothed regular partial column comparisons (blue and red lines in first two rows of Fig. 8), it becomes clear that the OPERA and RAL retrieval schemes demonstrate roughly analogous vertical comparison behaviour. Negative biases of $-10$ % to up to $-20$ % (0 to $-2$ DU) are found in the MA, while below 20 hPa the bias becomes positive and remains quite stable between 5 and 10 % (1 to 3 DU). An important distinction can however be pointed to. The RAL bias is typically lower than the OPERA bias in the lower stratosphere, whereas the opposite is true below the tropopause (see discussion of difference dependences and detailed numbers in Table 4). The lowest retrieved OPERA ozone layer however has nearly zero sensitivity, thus representing the a priori climatology rather than the atmospheric state under observation. This, together with the climatology being derived from ozonesonde measurements, explains both the markedly low bias and spread for the OPERA surface layer. Apart from that, OPERA and RAL 68 % interpercentiles agree very well. No other sensitivity dead zones are detected, but both algorithms show over-sensitivities (i.e. excessively attributed during retrieval) around the UT/LS, partially

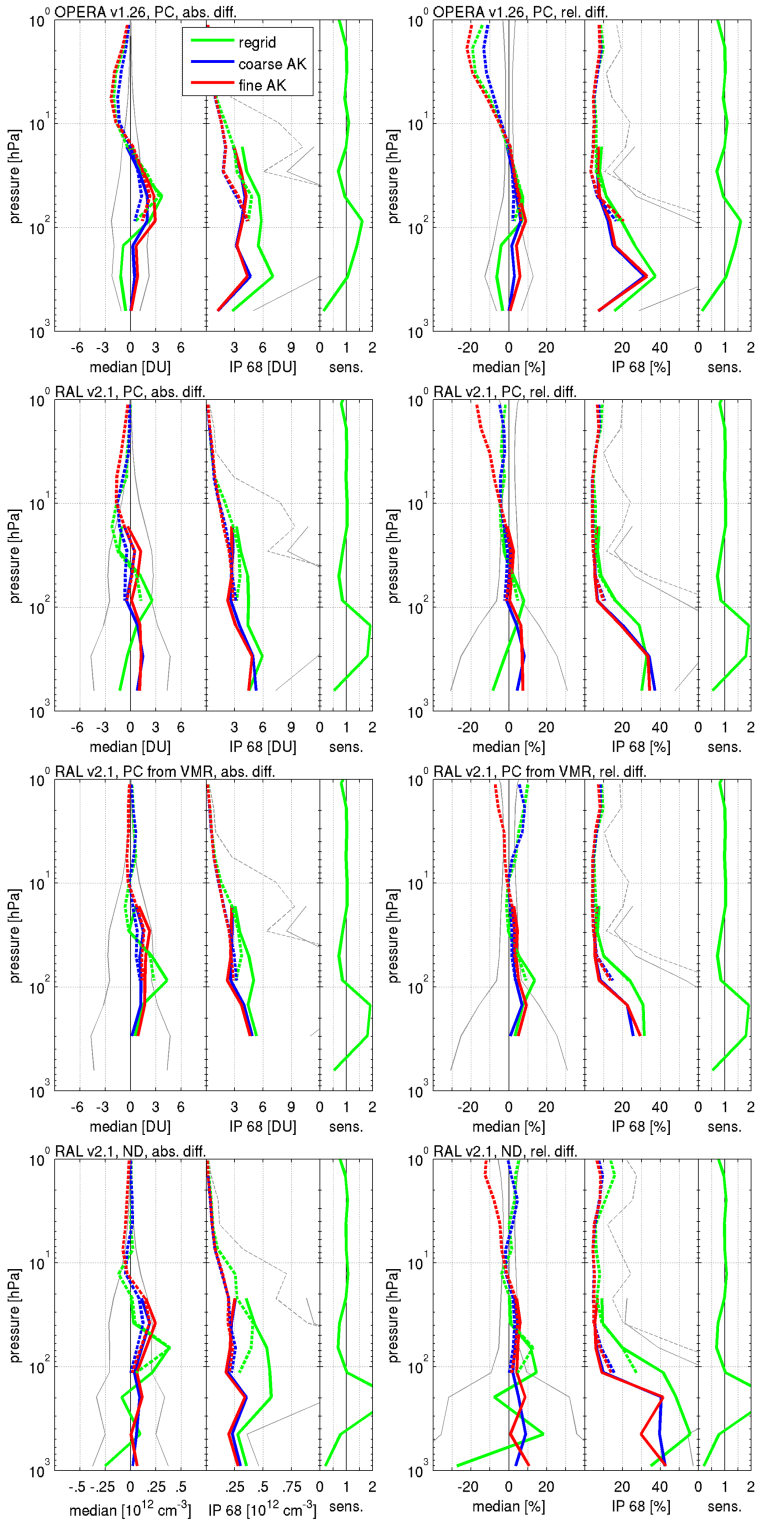



**Figure 8.** Ground-based comparison results for GOME-2 2008 (from top to bottom) OPERA v1.26 in partial columns (PC), RAL v2.1 in partial columns, RAL v2.1 in partial columns determined using RAL's high-resolution VMR integration method, and RAL v2.1 number densities (ND). Absolute and relative difference statistics are grouped in the left and right column, respectively. Full lines are for ozonesonde differences, dashed lines for lidar comparisons, with green, blue, and red colours for regridded, coarse resolution AK smoothed, and fine resolution AK smoothed ground-based data, respectively. Light grey lines indicate satellite and combined uncertainties for bias and spread evaluation (see main text). The median sensitivity is also added to each plot.

compensating the vertical sensitivity dip right above (also see Sect. 5.2).

– *On the difference statistics in relation to the satellite and ground-based random uncertainties.* For both the OPERA and RAL retrieval processes, the resulting satellite and combined satellite–ground random uncertainties equal at least double the comparison biases and spreads, respectively, below the UT/LS. This discrepancy indicates an overestimation of the random satellite uncertainty by both retrieval teams. The OPERA and RAL comparison spreads remain below the combined random uncertainty by roughly the same relative amount throughout the vertical comparison range (1–1000 hPa). Their biases however become comparable with or exceed the satellite random uncertainty above the 100 hPa level, with this effect being more pronounced for OPERA than for RAL. To be significantly detectable above the random noise level in a single measurement however, the bias should exceed the combined satellite–ground random uncertainty, which is hardly the case somewhere here. The total satellite measurement and retrieval uncertainty is an unknown number because of precision ignorance, but can be estimated to range in between the combined (quadratic sum) bias and satellite random uncertainty and the combined bias and comparison spread (although the latter contains error contributions that are not part of the satellite observation, like co-location mismatch). Total uncertainty numbers for both retrieval schemes under study are compared with the user requirements in Table 4. The only ascertained compliance occurs for RAL around the UT/LS.

### 7.3 Dependences of the bias

Figure 9 contains qualitative relative median bias plots showing its dependences on time, latitude, slant column density, and cloud fraction for the OPERA and RAL GOME-2 2008 closest ozonesonde and lidar co-locations. The regular fine resolution averaging kernel smoothed ground-based profiles have hereby been used as reference profiles, as they result from a minimal reference profile manipulation. Regularities that can be forthrightly observed from these graphs are summarized in Table 4.

Major findings are (1) the increased bias above the tropopause for OPERA and below the tropopause for RAL (also see previous section) and (2) the appearance of strong (up to 30 %) biases near the tropics for both algorithms. The first can be most clearly seen from the temporal and cloud fraction (CF) dependences, which are rather uniform. Augmented biases mainly occur near the tropopause during Northern Hemisphere winters and for cloud fractions below 0.2. The second finding can be directly related to the high biases arising for minimal slant column densities, although RAL also shows a remarkable inverse bias around the tropopause for SCD below 500 DU. Positive and negative RAL biases moreover persist towards high slant column densities below the tropopause. As already remarked from Fig. 8, both the OPERA and RAL median relative differences decrease down to about −20 % when going towards the 1 hPa level.

The slight bias complementarity with respect to the tropopause between the OPERA and RAL retrieval schemes might be partially understood by reconsidering the grid-normalized relative MQQ discussed in Sect. 5. In combination with the vertical sensitivity profiles, this metric revealed a clear SZA or SCD dependence of the amount of information contained within the respective retrieval outcomes of the algorithms under study. RAL's increased amount of information contained in the retrievals for large solar zenith angles (i.e. above 60° corresponding to a less penetrating sideways solar irradiation) nicely addresses the stratosphere, but is accompanied by a reduced assessment of the troposphere. OPERA on the other hand is related to higher amounts of information for smaller solar zenith angles (below 60°) which cause a deeper vertical atmospheric penetration of the solar radiation. Note however that this reasoning cannot be claimed generally valid, as OPERA's (like RAL's) reduced sensitivity near the earth's surface (resulting in a comparison with the prior's climatology) also has to be taken into account.

## 8 Conclusions

This work thoroughly discusses a methodology for the round-robin evaluation and the geophysical validation of nadir ozone profile retrievals, as summarized in the flowchart in Appendix A, and applies the proposed best-practice to a pair of optimal estimation algorithms run on MetOp-A GOME-2 level-1 version 4.0 radiance measurements of 2008. The quality assessment combines data set content studies, information content studies, and comparisons with ground-based reference measurements. Key results are summarized in Table 4, which also reproduces data quality criteria established by the Climate Research Group in ESA's Ozone_cci User Requirement Document (van der A, 2011) and Product Validation Plan (Lambert et al., 2012). These documents include directives on observation frequency (3 day coverage), horizontal (20–200 km) and vertical (3–10 km, depending on altitude) resolution, total uncertainty (8–20 %, depending on altitude and time frame), and stability (1–3 % per decade). Other evaluation parameters that appeared valuable throughout the retrieval outputs' quality control have been added. The practical outlining of unit conversion options for atmospheric state profiles and the corresponding fractional averaging kernel matrices (and covariance matrices) has also been an important part of the satellite data set pre-processing in this work.

**Table 4.** Overview of nadir ozone profile round-robin evaluation criteria and their corresponding OPERA and RAL outcome for GOME-2 (2008). Where possible, results have been differentiated between troposphere (TS), UT/LS, and middle atmosphere (MA).

| Criterion | User requirements | OPERA v1.26 | RAL v2.1 |
|---|---|---|---|
| Horizontal resol. | 20 to 200 km | 160 by 160 km$^2$ | 160 by 160 km$^2$ |
| Filtering | – | ∼ 10 % | ∼ 10 % |
| Spatial sampling | – | Full | SAA missing |
| Observation frequency (revisit time) | 3 days | ∼ 3 days | ∼ 3 days |
| DFS | – | 3 to 4 | 5.5 to 6.5 |
| Vertical resol. (resolving length estimate) | TS: 6 km to tropo-column<br>UT/LS: 3 to 6 km<br>MA: 3 to 10 km | TS: > 40 km<br>UT/LS: 10 to 20 km<br>MA: 20 to 40 km | TS: > 40 km<br>UT/LS: 10 to 20 km<br>MA: > 40 km |
| Height registration offset | – | TS: 10 to 20 km<br>UT/LS: negligible<br>MA: −20 to 0 km | TS: 10 to 15 km<br>UT/LS: negligible<br>MA: −20 to 0 km |
| Closest co-locations (2008) | – | 1464 | 1254 |
| Accuracy (bias) | – | TS: ∼ 5 % (1 DU)<br>UT/LS: −5 to 10 % (−1 to 3 DU)<br>MA: −20 to −5 % (−2 to −1 DU) | TS: 7 to 8 % (1 DU)<br>UT/LS: 1 to 4 % (1 DU)<br>MA: −15 to 0 % (−2 to 0 DU) |
| Temporal dependence of bias | – | Increased bias mainly between tropopause and ozone maximum for Northern Hemisphere summer, negative MA bias throughout | Increased bias around and below tropopause for Northern Hemisphere winter and Antarctic spring, negative MA bias throughout |
| Meridian dependence of bias | – | High: Negative bias around and above south pole tropopause, positive around south pole ozone maximum<br>Mid: Small mainly positive biases<br>Low: Increased bias up to ozone maximum, related to bias for small SCD | High: Negative bias around and below tropopause, positive around south pole ozone maximum<br>Mid: Negative bias<br>Low: Increased bias below tropopause, related to bias for small SCD |
| SCD dependence of bias | – | Clearly related to meridian dependence | Clearly related to meridian dependence |
| Cloud fraction dependence of bias | – | Slightly increased cloud free bias around tropopause | Increased cloud free bias below tropopause |
| Comparison spread | – | TS: 10 to 30 % (1 to 4 DU)<br>UT/LS: ∼ 5 % (3 DU)<br>MA: 5 to 10 % (0 to 3 DU) | TS: 30 to 35 % (4 to 5 DU)<br>UT/LS: ∼ 5 % (3 DU)<br>MA: 5 to 10 % (0 to 3 DU) |
| Satellite random uncertainty | – | TS: 6 to 12 %<br>UT/LS: 2 to 5 %<br>MA: 2 to 3 % | TS: 10 to 30 %<br>UT/LS: ∼ 5 %<br>MA: 3 to 5 % |
| Total uncertainty | TS: 10 %<br>UT/LS: 8 %<br>MA: 8 % | TS: 8 to 30 %<br>UT/LS: 5 to 11 %<br>MA: 5 to 22 % | TS: 12 to 36 %<br>UT/LS: 5 to 6 %<br>MA: 3 to 18 % |

The most remarkable RR evaluation findings are RAL's missing SAA sampling, the significantly higher (about 3 points) RAL DFS in combination with the fairly complementary grid-normalized MQQ values (with respect to a roughly 60° SZA threshold value), and the somewhat related complementarity of the median comparison differences. OPERA typically results in a few percentages lower bias below the UT/LS (omitting its surface layer with nearly zero sensitivity), while RAL performs correspondingly better above (order of 1 DU). Both algorithms show augmented biases around the tropics, an overall decreasing negative bias towards the middle atmosphere, and comparable spreads (slightly smaller for RAL).

The statistics in Table 4 indicate that CCI user requirements are met for the horizontal resolution (order of 180 km), revisit time (order of 3 days), and RAL's total uncertainty around the UT/LS and partially above (below 8 %) only. OPERA's total uncertainty nevertheless is similar in the same vertical region. The total uncertainty is thereby estimated between the combined bias and satellite random uncertainty on the one hand and the combined bias and comparison spread on the other hand, although the latter contains error contributions that are not part of the satellite observation. The vertical resolution figures, estimated by the Backus–Gilbert spread about the centroid, do not meet the user requirements either. Minima of about 10 km are obtained for both algorithms where maximum values of 6 km are actually required.

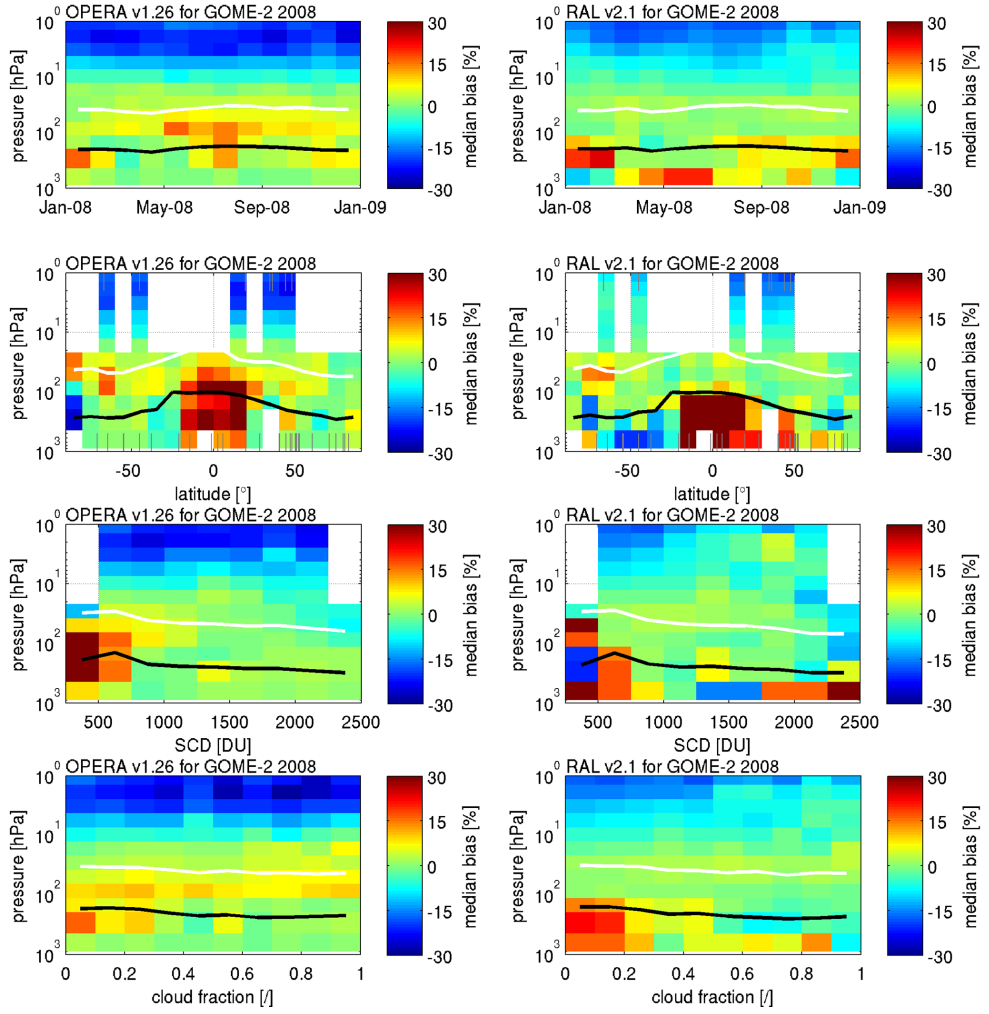



**Figure 9.** Temporal, meridian, SCD, and cloud fraction dependences of the median relative bias of OPERA v1.26 (left) and RAL v2.1 (right) GOME-2 2008 nadir ozone profile retrievals with respect to ground-based ozonesonde and lidar measurements (averaging-kernel-based smoothing at fine resolution). Black and white lines represent the tropopause and ozone maximum, respectively.

Below the UT/LS and in the middle atmosphere, where the height registration uncertainty (centroid offset) significantly deviates from the nominal altitude, unphysical resolution estimates even occur. Satellite measurement stability has not been assessed as a retrieval quality indicator in this RR evaluation study because of the focus on 1 year of satellite data. This will be the subject of future work. Moreover, the current RAL scheme does not use measurements between 307 and 322 nm (largely due to instrumental issues encountered during development), while this omitted range is included in OPERA and includes potentially valuable information on the ozone profile. Work is therefore ongoing within the Ozone CCI to assess whether the approaches used by the two schemes could be combined to allow this potential to be fulfilled in practice.

## Appendix A

The nadir ozone profile retrieval round-robin evaluation flowchart proposal that has also been applied for this work is displayed in Fig. A1.

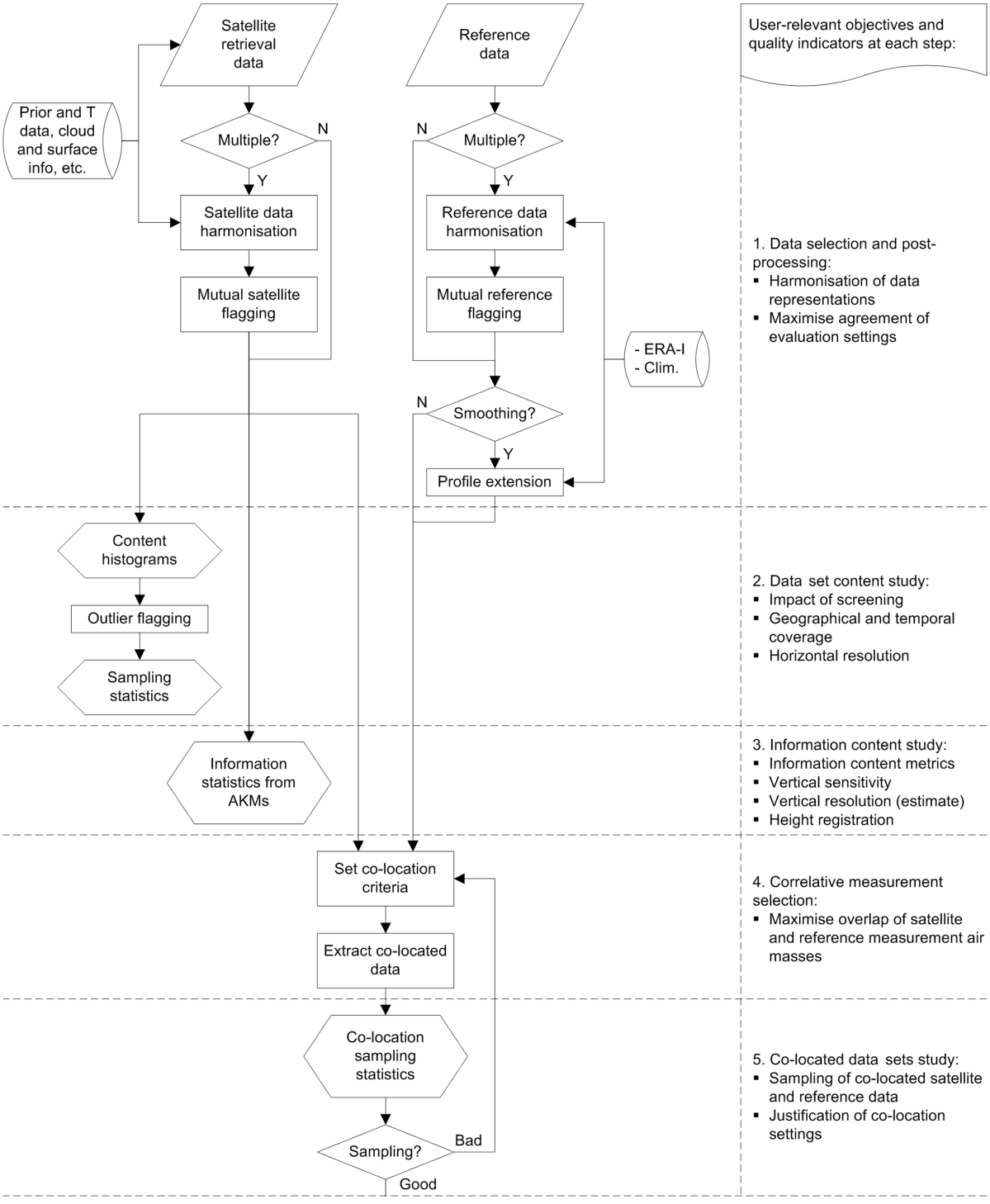
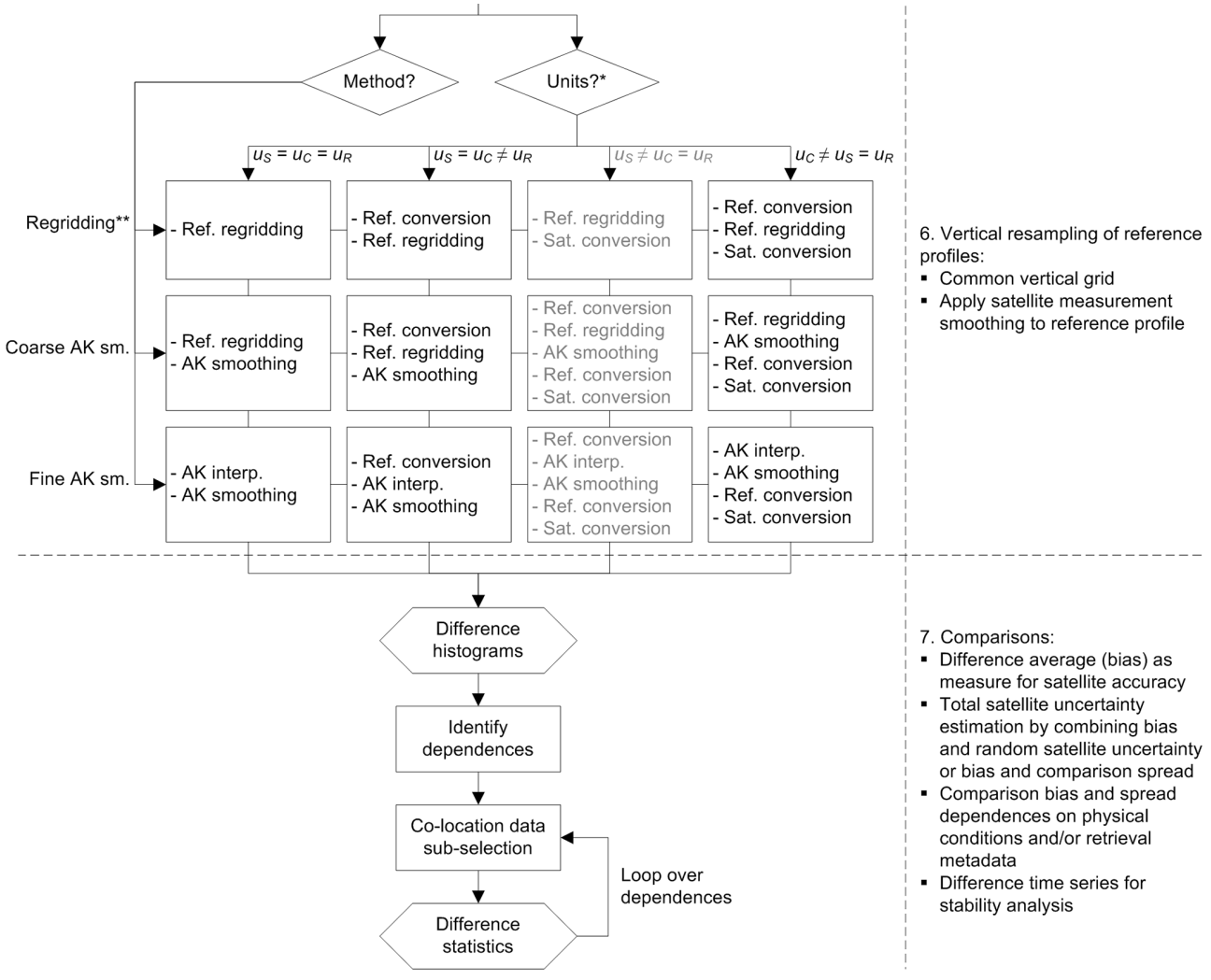



**Figure A1.**




* $u_S$, $u_C$, and $u_R$ for satellite, comparison, and reference units, respectively.
** Regridding only, or regridding including function smoothing (Gaussian, triangular, others). Regridding possibly by summation for sub-columns.

**Figure A1.** Nadir ozone profile retrieval round-robin evaluation flowchart. The unit option marked in grey is not of use in this publication.

## Appendix B

For a conversion of a number density (ND) ozone value $\boldsymbol{x}^{\mathrm{ND}}(i)$ to a volume mixing ratio (VMR) ozone value $\boldsymbol{x}^{\mathrm{VMR}}(i)$ at level $i$, the following equation can be applied:

$$\boldsymbol{x}^{\mathrm{VMR}}(i) = \frac{k_{\mathrm{B}} T(i)}{\boldsymbol{p}(i)} \boldsymbol{x}^{\mathrm{ND}}(i), \tag{B1}$$

which is related to the ideal gas law, with $k_{\mathrm{B}}$ the Boltzmann constant and with $T(i)$ and $\boldsymbol{p}(i)$ the atmospheric temperature and pressure at the same level $i$, respectively. Simple inversion yields the following:

$$\boldsymbol{x}^{\mathrm{ND}}(i) = \frac{\boldsymbol{p}(i)}{k_{\mathrm{B}} T}(i) \boldsymbol{x}^{\mathrm{VMR}}(i) \tag{B2}$$

as a level expression for VMR to ND conversion. Things get more complicated however when converting VMR or ND values to partial column (PC) values. This transformation consists of a conversion of $N$ level values to $L = N - 1$ layer values (with here $N = 17$). Conversion of sub-column profiles in layers to VMR or ND profiles on levels on the other hand are not considered because of this being an under-constraint problem. A conversion to VMR or ND profiles defined on the layer centres (in between levels) is nevertheless possible (Doicu and von Bargen, 2006), but not handled in this work.

The sub-column $\boldsymbol{x}^{\mathrm{PC}}(k)$ of layer $k$ can be approximated by discrete integration of the adjacent number densities (at levels $i$ and $i+1$) over half the layer thickness (Eq. B3a), or, equivalently, by discrete integration of the number density at the layer centre (obtained by linear interpolation) over the full layer thickness (Eq. B3b):

$$\boldsymbol{x}^{\mathrm{PC}}(k) = \boldsymbol{x}^{\mathrm{ND}}(i) \frac{\Delta z_k}{2} u + \boldsymbol{x}^{\mathrm{ND}}(i+1) \frac{\Delta z_k}{2} u \tag{B3a}$$

$$\boldsymbol{x}^{\mathrm{PC}}(k) = \left[ \frac{\boldsymbol{x}^{\mathrm{ND}}(i) + \boldsymbol{x}^{\mathrm{ND}}(i+1)}{2} \right] \Delta z_k u. \tag{B3b}$$

For the conversion of VMR values to partial columns, a similar reasoning is valid, now based on the ideal gas law for pressure differences over layers (Ziemke et al., 2001):

$$\boldsymbol{x}^{\mathrm{PC}}(k) = \boldsymbol{x}^{\mathrm{VMR}}(i) \frac{\Delta p_k}{2} u + \boldsymbol{x}^{\mathrm{VMR}}(i+1) \frac{\Delta p_k}{2} u \tag{B4a}$$

$$\boldsymbol{x}^{\mathrm{PC}}(k) = \left[ \frac{\boldsymbol{x}^{\mathrm{VMR}}(i) + \boldsymbol{x}^{\mathrm{VMR}}(i+1)}{2} \right] \Delta p_k u. \tag{B4b}$$

$u$ represents a conversion constant depending on the units to be converted. $\Delta z_k$ equals the thickness of layer $k$ (in km), while $\Delta p_k$ equals the pressure difference over the same layer (in hPa). The lowest pressure (or altitude) level of the profile is thereby always set to the surface pressure (or altitude), in order to properly integrate over the layer thickness that best approximates the satellite measurement penetration for a given surface geometry. (Note however that for convenient plotting this variable height is attributed to a fixed pressure value on the vertical axis.) This setting is automatically included for the pressure profiles provided by OPERA, but has to be introduced by the user for the RAL profiles (see Sect. 2.4).

For immediate unit conversion of a full ozone profile, the above linear equations can be turned into algebraic expressions for vectors and matrices:

$$\boldsymbol{x}' = \mathbf{M}\boldsymbol{x} \tag{B5}$$

with the size and elements of conversion matrix $\mathbf{M}$ depending on the units of both state vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ involved, as shown in the second column of Table B1 (M1 to M4). For conversion of VMR or ND units to partial columns, $\mathbf{M}$ includes the pairwise linear interpolation from the layer edges (levels) to the layer centres. Except for M3 and M4, the conversion matrix has to be recalculated for each profile pair.

A second set of conversion matrix definitions (M5 to M8) has been added in the third column of Table B1. The possibility exists that two profiles in different retrieval units are provided, whereas the prior profile, CVM, or AKM for one of those units is unknown. The unit conversion matrix $\mathbf{M}$ can than directly be constructed from taking ratios of the known state vector values. This approach may even be the only possible when e.g. vertical temperature or pressure information is missing.

The conversion matrix $\mathbf{M}$ can be applied for unit conversion of covariance matrices (going from $\mathbf{S}$ to $\mathbf{S}'$) and averaging kernel matrices (going from $\mathbf{A}$ to $\mathbf{A}'$) as well:

$$\mathbf{S}' = \mathbf{M}\mathbf{S}\mathbf{M}^T \tag{B6}$$

and

$$\mathbf{A}' = \mathbf{M}\mathbf{A}\mathbf{M}^{-1}. \tag{B7}$$

The latter expression is easily attained by e.g. combining Eqs. (1) (in main text) and (B5) in two ways and equating the results Eqs. (B8a) and (B8b) (van Peet and van der A, 2013):

$$\boldsymbol{x}'_{\mathrm{r}} = \boldsymbol{x}'_{\mathrm{p}} + \mathbf{A}'\left(\boldsymbol{x}'_{\mathrm{t}} - \boldsymbol{x}'_{\mathrm{p}}\right) = \mathbf{M}\boldsymbol{x}_{\mathrm{p}} + \mathbf{A}' M\left(\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_{\mathrm{p}}\right) \tag{B8a}$$

$$\boldsymbol{x}'_{\mathrm{r}} = \mathbf{M}\left[\boldsymbol{x}_{\mathrm{p}} + \mathbf{A}\left(\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_{\mathrm{p}}\right)\right] = \mathbf{M}\boldsymbol{x}_{\mathrm{p}} + \mathbf{M}\mathbf{A}\left(\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_{\mathrm{p}}\right). \tag{B8b}$$

Given the possible forms of $\mathbf{M}$, its inverse $\mathbf{M}^{-1}$ in Eq. (B7) can be both the regular inverse or the right-hand pseudo-inverse. Possible unphysical values (e.g. zeroes or divides thereby) must be screened from further processing.

**Table B1.** Conversion matrix definitions for various unit conversions (rows) and for two possibilities regarding conversion factor construction (columns). The number of rows or columns always equals the number of retrieval levels $N$ or the number of intermediate layers $L = N - 1$. The abbreviations VMR, ND, and PC stand for volume mixing ratio, number density and partial column, respectively. $u$ represents a conversion constant depending on the units to be converted.

| | Conversion from pressure and temperature profiles or vertical grid definitions | Conversion from available ozone profile vectors |
|---|---|---|
| VMR to ND<br>Size: $N \times N$ | $u \begin{pmatrix} T_1 / p_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & T_N / p_N \end{pmatrix}$ (M1) | $\begin{pmatrix} x_1^{\mathrm{ND}} / x_1^{\mathrm{VMR}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_N^{\mathrm{ND}} / x_N^{\mathrm{VMR}} \end{pmatrix}$ (M5) |
| ND to VMR<br>Size: $N \times N$ | $u \begin{pmatrix} p_1 / T_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_N / T_N \end{pmatrix}$ (M2) | $\begin{pmatrix} x_1^{\mathrm{VMR}} / x_1^{\mathrm{ND}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_N^{\mathrm{VMR}} / x_N^{\mathrm{ND}} \end{pmatrix}$ (M6) |
| VMR to PC<br>Size: $L \times N$ | $\frac{u}{2} \begin{pmatrix} \Delta p_1 & \Delta p_1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & \Delta p_L & \Delta p_L \end{pmatrix}$ (M3) | $\begin{pmatrix} c_1 & c_1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & c_L & c_L \end{pmatrix}$ (M7) with<br>$c_k = x_k^{\mathrm{PC}} \left( x_i^{\mathrm{VMR}} + x_{i+1}^{\mathrm{VMR}} \right)^{-1}$ for $k = i$ |
| ND to PC<br>Size: $L \times N$ | $\frac{u}{2} \begin{pmatrix} \Delta z_1 & \Delta z_1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & \Delta z_L & \Delta z_L \end{pmatrix}$ (M4) | $\begin{pmatrix} c_1 & c_1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & c_L & c_L \end{pmatrix}$ (M8) with<br>$c_k = x_k^{\mathrm{PC}} \left( x_i^{\mathrm{ND}} + x_{i+1}^{\mathrm{ND}} \right)^{-1}$ for $k = i$ |

*Acknowledgements.* The reported work was funded by ESA via the CCI – ECV Ozone project, with support from the Belgian Federal Science Policy Office (BELSPO) and ProDEx via project A3C. The ground-based ozonesonde and lidar data used in this publication were obtained as part of WMO's Global Atmospheric Watch (GAW) programme, including the Network for the Detection of Atmospheric Composition Change (NDACC) and NASA's Southern Hemisphere Additional Ozonesonde programme (SHADOZ), and are publicly available via the NDACC Data Host Facility, the SHADOZ archive, and World Ozone and Ultraviolet Data Center (WOUDC) (see http://www.ndacc.org/, http://croc.gsfc.nasa.gov/shadoz/, and http://www.woudc.org/, respectively). We warmly thank several members of the NDACC ozonesonde and lidar working groups for fruitful discussions. Lidar operation is funded through national collaborators and we are grateful to the following institutes and their co-workers who contributed to generating these data: CNRS and CNES (Dumont d'Urville station and Observatoire Haute Provence, PI is S. Godin-Beekmann), DWD (Höhenpeißenberg station, PI is H. Claude), RIVM and NIWA (Lauder station, PI is D. P. J. Swart), NASA/JPL (Mauna Loa Observatory and Table Mountain Facility, PIs are I. S. McDermid and T. Leblanc), and NIES (Tsukuba station, PI is H. Nakane).

Edited by: P. Stammes

## References

Backus, G. E. and Gilbert, F.: Uniqueness in the Inversion of inaccurate Gross Earth Data, Philos. T. R. Soc. A, 266, 123–192, 1970.

Bhartia, P. K., McPeters, R. D., Flynn, L. E., Taylor, S., Kramarova, N. A., Frith, S., Fisher, B., and DeLand, M.: Solar Backscatter UV (SBUV) total ozone and profile algorithm, Atmos. Meas. Tech., 6, 2533–2548, doi:10.5194/amt-6-2533-2013, 2013.

Brion, J., Chakir, A., Charbonnier, J., Daumont, D., Parisse, C., and Malicet, J.: Absorption spectra measurements for the ozone molecule in the 350–830 nm region, J. Atmos. Chem., 30, 291–299, 1998.

Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., Debeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., Eisinger, M., and Perner, D.: The Global Ozone Monitoring Experiment (GOME): mission concept and first scientific results, J. Atmos. Sci., 56, 151–175, 1999.

Calisesi, Y., Soebijanta, V. T., and van Oss, R. F.: Regridding of remote soundings: Formulation and application to ozone profile comparison, J. Geophys. Res.-Atmos., 110, D23306, doi:10.1029/2005JD006122, 2005.

Ceccherini, S., Carli, B., and Raspollini, P.: Quality quantifier of indirect measurements, Opt. Exp., 20, 5151–5167, 2012.

Ceccherini, S., Carli, B., and Raspollini, P.: Quality of MIPAS operational products, J. Quant. Spectrosc. Ra., 121, 45–55, doi:10.1016/j.jqsrt.2013.01.021, 2013.

Connor, B. J., Parrish, A., and Tsou, J.-J.: Detection of stratospheric ozone trends by ground-based microwave observations, Proc. SPIE Int. Soc. Opt. Eng., 1491, 218–230, 1991.

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

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Deshler, T., Mercer, J. L., Herman, G. J. S., Stübi, R., Levrat, G., Johnson, B. J., Oltmans, S. J., Kivi, R., Thompson, A. M., Witte, J., Davies, J., Schmidlin, F. J., Brothers, G., and Sasaki, T.: Atmospheric comparison of electrochemical cell ozonesondes from different manufacturers, and with different cathode solution strengths: The Balloon Experiment on Standards for Ozonesondes, J. Geophys. Res., 113, 1–17, doi:10.1029/2007JD008975, 2008.

Doicu, A. and von Bargen, A.: SCIAMACHY 1b to 2 Off-line Processing Instructions for the Usage Of the Level 2 Product Limb MDS, SCIAMACHY technical note, ENV-TN-DLR-SCIA-0077, ESA Document Library, available at: https://earth.esa.int/web/guest/document-library (last access: November 2014), 1, 2006.

ESA: Ozone Climate Change Initiative, available at: http://www.esa-ozone-cci.org (last access: September 2014), 2013.

EUMETSAT: An Introduction to the EUMETSAT Polar System, available at: http://www.eumetsat.int/eps_webcast (last access: September 2014), 2013.

Godin, S., Carswell, A. I., Donovan, D. P., Claude, H., Steinbrecht, W., McDermid, I. S., McGee, T. J., Gross, M. R., Nakane, H., Swart, D. P. J., Bergwerff, H. B., Uchino, O., von der Gathen, P., and Neuber, R.: Ozone differential absorption lidar algorithm intercomparison, Appl. Opt., 38, 6225–6236, doi:10.1364/AO.38.006225, 1999.

Hoogen, R., Rozanov, V. V., and Burrows, J. P.: Ozone profiles from GOME satellite data: algorithm description and first validation, J. Geophys. Res., 104, 8263–8280, doi:10.1029/1998JD100093, 1999.

Keckhut, P., McDermid, S., Swart, D., McGee, T., Godin-Beekmann, S., Adriani, A., Barnes, J., Baray, J.-L., Bencherif, H., Claude, H., di Sarra, A., Fiocco, G., Hansen, G., Hauchecorne, A., Leblanc, T., Lee, C., Pal, S., Megie, G., Nakane, H., Neuber, R., Steinbrecht, W., and Thayer, J.: Review of ozone and temperature lidar validations performed within the 1385 framework of the Network for the Detection of Stratospheric Change, J. Environ. Monit., 6, 721–733, doi:10.1039/b404256e, 2004.

Lambert, J.-C., Balis, D., De Backer, H., Granville, J., Hubert, D., Kivi, R., Koukouli, M., and Stübi, R.: Ozone-cci Product Validation Plan (PVP), ESA project deliverable, Ozone_cci_PVP_1.0, Ozone_cci website, 2012.

Malicet, J., Daumont, D., Charbonnier, J., Parisse, C., Chakir, A., and Brion, J.: Ozone UV spectroscopy. II - Absorption cross-

sections and temperature dependence, J. Atmos. Chem., 21, 263–273, 1995.

McPeters, R. D. and Labow, G. J.: Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, J. Geophys. Res., 117, D10303, doi:10.1029/2011JD017006, 2012.

McPeters, R. D., Labow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, J. Geophys. Res.-Atmos., 112, D05308, doi:10.1029/2005JD006823, 2007.

Mégie, G., Allain, J. Y., Chanin, M. L., and Blamont, J. E.: Vertical profile of stratospheric ozone by lidar sounding from the ground, Nature, 270, 329–331, doi:10.1038/270329a0, 1977.

Meijer, Y. J., Swart, D. P. J., Baier, F., Bhartia, P. K., Bodeker, G. E., Casadio, S., Chance, K., Del Frate, F., Erbertseder, T., Felder, M. D., Flynn, L. E., Godin-Beekmann, S., Hansen, G., Hasekamp, O. P., Kaifel, A., Kelder, H. M., Kerridge, B. J., Lambert, J.-C., Landgraf, J., Latter, B., Liu, X., McDermid, I. S., Pachepsky, Y., Rozanov, V., Siddans, R., Tellmann, S., van der A, R. J., van Oss, R. F., Weber, M., and Zehner, C.: Evaluation of Global Ozone Monitoring Experiment (GOME) ozone profiles from nine different algorithms, J. Geophys. Res.-Atmos., 111, 21306, doi:10.1029/2005JD006778, 2006.

Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G., and Richards, N. A. D.: Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmos. Meas. Tech., 8, 385–398, doi:10.5194/amt-8-385-2015, 2015.

Mijling, B., Tuinder, O. N. E., van Oss, R. F., and van der A, R. J.: Improving ozone profile retrieval from spaceborne UV backscatter spectrometers using convergence behaviour diagnostics, Atmos. Meas. Tech., 3, 1555–1568, doi:10.5194/amt-3-1555-2010, 2010.

Munro, R., Siddans, R., Reburn, W. J., and Kerridge, B. J.: Direct measurement of tropospheric ozone distributions from space, Nature, 392, 168–171, doi:10.1038/32392, 1998.

Nair, P. J., Godin-Beekmann, S., Froidevaux, L., Flynn, L. E., Zawodny, J. M., Russell III, J. M., Pazmiño, A., Ancellet, G., Steinbrecht, W., Claude, H., Leblanc, T., McDermid, S., van Gijsel, J. A. E., Johnson, B., Thomas, A., Hubert, D., Lambert, J.-C., Nakane, H., and Swart, D. P. J.: Relative drifts and stability of satellite and ground-based stratospheric ozone profiles at NDACC lidar stations, Atmos. Meas. Tech., 5, 1301–1318, doi:10.5194/amt-5-1301-2012, 2012.

NDACC: Network for the Detection of Atmospheric Composition Change (NDACC), NOAA National Weather Service, available at: http://www.ndsc.ncep.noaa.gov (last access: September 2014), 2013.

Purser, R. J. and Huang, H.-L.: Estimating Effective Data Density in a Satellite Retrieval or an Objective Analysis, J. Appl. Meteor., 32, 1092–1107, 1993.

Ridolfi, M., Ceccherini, S., and Carli, B.: Optimal interpolation method for intercomparison of atmospheric measurements, Opt. Lett., 31, 7, 855–857, 2006.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, World Scientific Publishing, Singapore, 238 pp., 2000.

Rozanov, V. V., Diebel, D., Spurr, R. J. D., and Burrows, J. P.: GOMETRAN: A radiative transfer model for the satellite project GOME – the plane-parallel version, J. Geophys. Res., 102, 16683–16695, 1997.

Siddans, R.: Height Resolved Ozone Retrievals from GOME, PhD Thesis, University of Reading, 2003.

Smit, H. and ASOPOS panel: Quality Assurance and Quality Control for Ozonesonde Measurements in GAW, WMO Global Atmosphere Watch report series 201, World Meteorological Organization, available at: http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html (last access: September 2014), 2011.

Smit, H., Straeter, W., Johnson, B., Oltmans, S., Davies, J., Tarasick, D., Hoegger, B., Stübi, R., Schmidlin, F., Northam, T., Thompson, A., Witte, J.-C., Boyd, I., and Posny, F.: Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Juelich Ozone Sonde Intercomparison Experiment (JOSIE) 1550, J. Geophys. Res., 112, 1–18, doi:10.1029/2006JD007308, 2007.

Spurr, R. J. D., Kurosu, T. P., and Chance, K. V.: A linearized discrete ordinate radiative transfer model for atmospheric remote sensing retrieval, J. Quant. Spectrosc. Ra., 68, 689–735, 2001.

Thompson, A. M., Miller, S. K., Tilmes, S., Kollonige, D. W., Witte, J. C., Oltmans, S. J., Johnson, B. J., Fujiwara, M., Schmidlin, F. J., Coetzee, G. J. R., Komala, N., Maata, M., bt Mohamad, M., Nguyo, J., Mutai, C., Ogino, S.-Y., Raimundo Da Silva, F., Paes Leme, N. M., Posny, F., Scheele, R., Selkirk, H. B., Shiotani, M., Stübi, R., Levrat, G., Calpini, B., Thouret, V., Tsuruta, H., Valverde Canossa, J., Vömel, H., Yonemura, S., Andrés Diaz, J., Tan Thanh, N. T., and Thuy Ha, H. T.: Southern Hemisphere Additional Ozonesondes (SHADOZ) ozone climatology (2005–2009): Tropospheric and tropical tropopause layer (TTL) profiles with comparisons to OMI-based ozone products, J. Geophys. Res., 117, D23301, doi:10.1029/2011JD016911, 2012.

van der A, R. J.: Ozone-cci User Requirement Document (URD), ESA project deliverable, Ozone_cci_URD_2.1, Ozone_cci website, 2011.

van Oss, R. F. and Spurr, R. J. D.: Fast and accurate 4 and 6 stream linearized discrete ordinate radiative transfer models for ozone profile retrieval, J. Quant. Spectrosc. Ra., 75, 177–220, doi:10.1016/S0022-4073(01)00246-1, 2002.

van Peet, J. C. A. and van der A, R. J.: Converting the averaging kernel and covariance matrix, Internal Note (version 1.0), 2013.

van Peet, J. C. A., van der A, R. J., Tuinder, O. N. E., Wolfram, E., Salvador, J., Levelt, P. F., and Kelder, H. M.: Ozone Profil E Retrieval Algorithm (OPERA) for nadir-looking satellite instruments in the UV-VIS, Atmos. Meas. Tech., 7, 859–876, doi:10.5194/amt-7-859-2014, 2014.

Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendael, M.: FRESCO+: an improved $O_2$ A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565–6576, doi:10.5194/acp-8-6565-2008, 2008.

WMO: Meteorology – a three-dimensional science, WMO Bull., 230, 6, 134–138, 1957.

WMO: Implementation plan for the global observing system for climate in support of the UNFCCC (2010 update), Report nr. GCOS-138, Geneva, Switzerland, 2010.

WOUDC: World Ozone and Ultraviolet Radiation Data Centre, Environment Canada, available at: http://www.woudc.org (last access: September 2014), 2013.

Ziemke, J. R., Chandra, S., and Bhartia, P. K.: "Cloud slicing": A new technique to derive upper tropospheric ozone from satellite measurements, J. Geophys. Res., 106, 9853–9867, 2001.