

pdftitle=Comparison of profile retrieval algorithms for MAX-DOAS pdfauthor=T. Vlemmix et al.

title Manuscript prepared for Atmos. Meas. Tech. Discuss.

with version 4.1 of the L<sup>A</sup>T<sub>E</sub>X class copernicus\_discussions.cls.

Date: 20 December 2014

# **MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval approaches**

**Tim Vlemmix<sup>1,2</sup>, Francois Hendrick<sup>2</sup>, Gaia Pinardi<sup>2</sup>, Isabelle De Smedt<sup>2</sup>, Caroline Fayt<sup>2</sup>, Christian Hermans<sup>2</sup>, Ankie Pipers<sup>3</sup>, Pieter Levelt<sup>3,1</sup>, and Michel Van Roozendaal<sup>2</sup>**

<sup>1</sup>Delft University of Technology (TU-Delft), Delft, the Netherlands

<sup>2</sup>Belgian Institute for Space Aeronomy (IASB-BIRA), Brussels, Belgium

<sup>3</sup>Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

Correspondence to: T. Vlemmix (t.vlemmix@tudelft.nl)

## Abstract

abstr A four year data set of MAX-DOAS observations in the Beijing area (2008–2012) is analysed with a focus on  $\text{NO}_2$ , HCHO, and aerosols. Two very different retrieval methods are applied. Method A describes the tropospheric profile with 13 layers and makes use of the optimal estimation method. Method B uses 2–4 parameters to describe the tropospheric profile and an inversion based on a least-squares fit. For each constituent ( $\text{NO}_2$ , HCHO and aerosols) the retrieval outcomes are compared in terms of tropospheric column densities, surface concentrations, and “characteristic profile heights” (i.e. the height below which 75 % of the vertically integrated tropospheric column density resides).

We find best agreement between the two methods for tropospheric  $\text{NO}_2$  column densities, with a standard deviation of relative differences below 10 %, a correlation of 0.99 and a linear regression with a slope of 1.03. For tropospheric HCHO column densities we find a similar slope, but also a systematic bias of almost 10 % which is likely related to differences in profile height. Aerosol optical depths (AODs) retrieved with method B are 20 % high compared to method A. They are more in agreement with AERONET measurements, which are on average only 5 % lower, however with considerable relative differences (standard deviation  $\sim 25$  %). With respect to near surface volume mixing ratios and aerosol extinction we find considerably larger relative differences:  $10 \pm 30$  %,  $-23 \pm 28$  % and  $-8 \pm 33$  % for aerosols, HCHO and  $\text{NO}_2$  respectively. The frequency distributions of these near-surface concentrations show however a quite good agreement, and this indicates that near-surface concentrations derived from MAX-DOAS are certainly useful in a climatological sense. A major difference between the two methods is the dynamic range of retrieved characteristic profile heights which is larger for method B than for method A. This effect is most pronounced for HCHO, where retrieved profile shapes with method A are very close to the a priori, and moderate for  $\text{NO}_2$  and aerosol extinction which on average show quite good agreement for characteristic profile heights below 1.5 km.

One of the main advantages of method A is the stability, even under suboptimal conditions (e.g., in the presence of clouds). Method B is generally more unstable and this explains probably

a substantial part of the quite large relative differences between the two methods. However, despite a relatively low precision for individual profile retrievals it appears as if seasonally averaged profile heights retrieved with method B are less biased towards a priori assumptions than those retrieved with method A. This gives confidence in the result obtained with method B, namely that aerosol extinction profiles tend on average to be higher than NO<sub>2</sub> profiles in spring and summer, whereas they seem on average to be of the same height in winter, a result which is especially relevant in relation to the validation of satellite retrievals.

## 1 Introduction

intro

Multi-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) is a ground-based passive remote sensing technique that is used to detect tropospheric trace gases such as nitrogen dioxide (NO<sub>2</sub>), formaldehyde (HCHO), sulfur dioxide (SO<sub>2</sub>), nitrous acid (HONO), iodine oxide (IO), glyoxal (CHOCHO), bromine oxide (BrO) and aerosols (aerosol extinction) (e.g., Wittrock et al., 2004; Wagner et al., 2004, 2009; Irie et al., 2011; Coburn et al., 2011; Pinardi et al., 2013; Hendrick et al., 2014; Wang et al., 2014). MAX-DOAS instruments take spectral measurements of scattered sunlight in the ultraviolet (UV) and visible (Vis) part of the electromagnetic spectrum. Profile information is obtained from a scan which comprises spectral measurements at different elevation angles but in the same azimuthal direction. The main retrieval products are tropospheric column densities, concentrations near the surface and estimates of the vertical profile shape.

Because of this versatility MAX-DOAS is complementary to ground-based in-situ observations (in a spatial sense) as well as to satellite observations (in a temporal and spatial sense, i.e. the vertical) and it can play an important role in bridging the gap between those techniques (Richter et al., 2013). Knowledge of the relationship between surface concentrations and integrated tropospheric column densities (in urban, suburban and rural regions) is important for the use of satellite observations in studies of air quality (e.g., Boersma et al., 2009; Mendolia et al., 2013).

MAX-DOAS has great potential to be used in regional or global networks similar to the AERONET (sun photometer) and EARLINET (lidar) networks because of its versatility, the relatively low cost per instrument, the fact that a radiometric calibration is not required, and the fact that instruments can operate autonomously. Long-term data sets can be used for e.g. air quality monitoring, validation of chemical transport models, validation of satellite tropospheric column densities retrievals and potentially as input in data assimilation systems for air quality forecasts. With respect to satellite validation it is interesting to note that MAX-DOAS can provide not only tropospheric trace gas column densities for direct comparison, but also profile shape estimates for trace gases and aerosol extinction. These can replace the a priori profile shapes assumed for the satellite retrieval, such that one can assess the impact of the a priori profile shape assumption (both for aerosols and for the trace gas of interest) on the satellite retrieval accuracy (Rodgers and Connor, 2003). Proper knowledge of the accuracy of the profile shape assumptions that are used in the satellite retrieval is crucial for a realistic estimate of the potential biases in the retrieved tropospheric column density.

Mostly in the last decade, much progress has been made with respect to the quantitative interpretation of MAX-DOAS observations (e.g., Wagner et al., 2007; Roscoe et al., 2010), and MAX-DOAS instrumentation or similar (like PANDORA Herman et al., 2009) has been used for a wide range of gases and applications. In comparison to surface concentrations and profile shapes, tropospheric column densities are the most robust retrieval product. Several MAX-DOAS data sets have been used for validation of satellite observations of tropospheric column densities, predominantly for  $\text{NO}_2$  (e.g., Irie et al., 2008b; Halla et al., 2011; Ma et al., 2013; Lin et al., 2014; Kanaya et al., 2014). Near-surface concentrations are generally associated with higher uncertainties (primarily because of the quite limited vertical resolution of MAX-DOAS measurements), but nevertheless some studies have shown promising comparisons compared to independent ground-based in-situ instrumentation, see (e.g., Wagner et al., 2011; Li et al., 2013). Most challenging is the retrieval of vertical tropospheric profiles, and also its validation.

Quite some groups have developed algorithms for the vertical profiles of aerosol extinction and trace gases (e.g., Frieß et al., 2006; Irie et al., 2008a; Clémer et al., 2010; Li et al., 2010; Wagner et al., 2011; Vlemmix et al., 2011; Sinreich et al., 2012). Especially in relation to satel-

lite validation there is a great need for simultaneously measured trace gas and aerosol extinction profiles, and MAX-DOAS is one of the few remote sensing methods which can provide in this need. At the same time it is well known that the MAX-DOAS profiles are only first-order estimates, due to the fact that the information content of MAX-DOAS observations with respect to the vertical distribution of aerosols and trace gases is very limited: the degrees of freedom for signal typically varies from 1 to 3, see Sect. 2.5 and Vlemmix et al. (2011).

Comparatively few studies have been published however which directly address the quality of MAX-DOAS tropospheric profiles obtained from real observations. This is largely due to the fact that suitable long-term (multi-year) data sets which can serve as golden standard in a comparison (e.g., profiles measured with high vertical resolution) do not exist. In turn, the lack of a thorough validation of MAX-DOAS profiles limits their usefulness in validation studies where MAX-DOAS itself would be the reference.

The present study is highly motivated by the need for further assessment of the quality of MAX-DOAS profiles. Our approach is based on three pillars. First, the use of two very different profile retrieval algorithms, both run with various a priori profile shape assumptions. Second, the use of a four year data set covering a wide range of conditions (e.g., pollution levels, seasons, meteorological conditions). Third, analysis of profiles for three different components: formaldehyde,  $\text{NO}_2$  and aerosols (aerosol extinction profile). With this we address in this work the following specific questions: How consistent are the retrievals of individual profiles with different algorithms? How consistent are the retrievals on average? Do the column densities and profiles – on average – show a diurnal and seasonal variation? How strong or weak is the dependence on a priori assumptions? Which atmospheric conditions most critically limit the quality of the profile retrieval? What is the agreement between the profile shapes retrieved for the different constituents?

The two profile retrieval methods that are compared in this study do not retrieve profiles on the same vertical grid. One way to perform a profile comparison would be to interpolate profiles retrieved with both methods to a common vertical grid. A comparison performed in this way would give results for all layers which define the common grid. Such an approach would probably be favorable if the vertical resolution of the measurements (and therefore the DOFS) was

high, but this is not the case for MAX-DOAS measurements, as noted above. Because of the low DOFS (1-3), it was decided to derive from each profile three suitable quantities and to compare the two profile retrieval methods based on those quantities: tropospheric column density, near-surface concentration and “characteristic profile heights”  $H_{75}$ . The latter quantity is defined as the height below which 75 % of the integrated profile resides (75 % of the tropospheric column density). An advantage of this quantity (a scalar) is that it allows a first order description of the profile shape of pollutants which reside primarily in the atmospheric boundary layer.

The paper is structured as follows: In Sect. 2 we describe the data set of MAX-DOAS observations that is used: the instrument characteristics and measurement sites; the settings of the DOAS fitting procedures for the UV and Vis; the two MAX-DOAS profile retrieval algorithms, both of which are run with different “internal” settings to test the dependence on a priori assumptions. The last part of this section describes the criteria that are applied to select data with sufficient quality. Results for selected days and the statistical analysis based on the entire data set are shown and discussed in Sect. 3. Section 4 contains a discussion, and the final conclusions are listed in the last section.

## 2 MAX-DOAS measurements and profile retrieval algorithms

The retrieval of vertical profiles from spectral measurements with MAX-DOAS typically consists of three steps. First, differential slant column densities (of  $O_4$ ,  $NO_2$  and HCHO) are derived by applying the DOAS spectral fitting technique to the measured spectra. Second, differential slant column densities of  $O_4$  are used as input for the aerosol extinction profile retrieval algorithms. Third, differential slant column densities of the trace gas of interest (in this work:  $NO_2$  and HCHO) are used as input for the trace gas profile retrieval algorithm, together with the estimated aerosol extinction profile. In this section each of those steps is described in more detail.

## 2.1 Instrument and measurement site

The MAX-DOAS instrument used in this study has been designed and assembled by the Belgian Institute for Space Aeronomy (BIRA-IASB), see Clémer et al. (2010). It consists essentially of a telescope mounted on a sun-tracker (which can point at any elevation and in any azimuthal direction) combined with two spectrographs: one for the UV (300–390 nm), and one for the Vis (400–720 nm). Although the instrument is also capable of taking direct sun observations, we use here only the scattered sunlight observations taken towards the north. The retrievals described below are based on sequential observations at 2°, 4°, 6°, 8°, 10°, 12°, 15°, 30° and 90° elevation. During the period analysed in this work (2008–2012), the instrument was stationed at two different sites. First it was stationed in the Beijing city centre, at the Institute of Atmospheric Physics (IAP) of the Chinese Academy of Sciences (39.98° N, 116.38° E). From 2010 until present, the instrument was stationed about 55 km away to the east-southeast, at the meteorological observatory in Xianghe (39.75° N, 116.96° E). Compared to Beijing this site has a more suburban character.

## 2.2 DOAS retrieval of differential slant column densities

The DOAS spectral fitting method (Platt and Stutz, 2008) is applied to the spectra measured with the UV and Vis spectrometers. The DOAS analysis is performed with the QDOAS software that has been developed at BIRA (Fayt et al., 2011). Table 1 gives for some relevant parameters the values used in both of the channels. More details of the DOAS settings used can be found in Pinardi et al. (2013) for the UV channel, and Hendrick et al. (2014) for the Vis channel. Note that a scaling factor of 0.8 is applied to the measured differential slant O<sub>4</sub> column densities (see Clémer et al., 2010) in order to obtain sufficient agreement between simulations and measurements. This scaling factor is used for both methods A and B. After the DOAS analysis the differential slant column densities corresponding to each elevation are linearly interpolated in time (with a 20 min sampling), such that as input for the profile retrieval code one “scan” can be provided, as if the measurements were performed at the same time. Since the DOAS analysis is performed with the zenith-noon spectrum as a reference, the (interpolated) zenith differential

slant column densities of a scan is subtracted from all the differential slant column densities. This procedure reduces the sensitivity to trace gases in the stratosphere and upper part of the free troposphere to almost zero.

### 2.3 Method A – algorithm developed at BIRA

5 The first algorithm (method A) has been developed at BIRA Clémer et al. (2010); Hendrick et al. (2014); Wang et al. (2014) and makes use of the optimal estimation method (Rodgers, 2000). Forward simulations of differential slant column densities and weighting functions are performed using the LIDORT radiative transfer model (Spurr, 2008). Trace gas and aerosol extinction profiles are described by partial columns of 13 layers in a fixed altitude grid: the first  
 10 ten layers (below 2 km) each have a vertical extent of 200 m, between 2 and 3 km there are two layers of 0.5 km, and the uppermost layer of the profile goes from 3 to 4 km. An important input parameter for retrieval model A is the a priori profile, which is the initial profile from which the profile retrieval code iteratively searches for a more optimal solution. Retrieved profile shapes can in principle be very different from the a priori, but only if the information content of the  
 15 measurements is sufficiently high (depending on trace gas and measurement conditions). If this is not the case, the retrieved profile shape will be very similar to the a priori. In the original implementation of the retrieval code (Clémer et al., 2010) this a priori profile concentration profile  $n(z)$  was described by an exponential function which is characterized by a specific a priori scale height  $H_s^{\text{prior}}$ :

$$20 \quad n(z) = \frac{N_{\text{prior}}^{\text{V}}}{H_s^{\text{prior}}} \cdot \exp\left(-\frac{z}{H_s^{\text{prior}}}\right). \quad (1)$$

For trace gases, the profile shape is scaled such that the integrated profile corresponds to the first order estimate of the tropospheric trace gas column density ( $N_{\text{prior}}^{\text{V}}$ ), namely the differential slant column density measurement at 30 ° elevation. The corresponding geometrical differential  
 25 air mass factor (see e.g. Brinksma et al., 2008) is equal to one. For aerosols the initial column estimate (the AOD) was set to 0.15 for all retrievals.



A second important input parameter is the a priori error estimate for each layer. Tests have shown that setting this value high – this would give the algorithm most flexibility to realize diverse profile shapes – leads to frequent retrievals of profile shapes showing oscillations that are not likely to be realistic. For this reason a relatively low value (20 %) of the a priori is chosen, although this limits the potential of the algorithm to deviate significantly from the a priori, see also the discussion in Sect. 4.2.

Tests performed prior to the study presented here, have shown that the interplay between the a priori profile and its error estimate, combined with the fact that the sensitivity of MAX-DOAS decreases with altitude, leads to an undesired effect for relatively high a priori scale heights (> 1.5 km), namely that the retrieved tropospheric trace gas column or AOD in case of aerosol extinction retrieval is systematically too high.

This unwanted mechanism works as follows: for a priori scale heights higher than 1.5 km, the exponentially decreasing a priori profile does not go to (almost) zero in the upper part of the altitude grid (4 km). Because above approximately 1.5 km the information content of the observation is low, the retrieval will have a tendency to stay close to the a priori and not be allowed to go to zero. As a consequence, the retrieved profiles will have a considerable fraction of the partial column above  $\sim 1.5$  km, even when this is not the case in reality. This effect will lift up the mean profile height, and this goes together with a systematic overestimation of the integrated trace gas column (or AOD).

By modifying the definition of the a priori profile shape such that it decreases to zero at the top of the altitude grid, the overestimation of columns and AOD is greatly reduced. The following profile shape definition is forced to low values above 1.5 km and even zero at the top of the altitude grid:

$$n(z) = \left[ \frac{N_{\text{prior}}^V}{H_s^{\text{prior}}} \cdot \exp\left(-\frac{z}{H_s^{\text{prior}}}\right) \right] \cdot (4 - z). \quad (2)$$

Figure 2 shows a priori profile shapes obtained with this definition, for  $H_s^{\text{prior}} = \{0.5, 1.0, 1.5\}$  km. Note that the range in terms of  $H_{75}$  is different:  $\{0.6, 1.0, 1.3\}$  km.

The impact of the a priori profile shape on the retrieved profile can be quite high. For this reason the profile retrieval with method A is performed with three different a priori scale heights ( $H_s^{\text{prior}} = \{0.5, 1.0, 1.5\}$ ), leading to three versions: A1, A2 and A3. The final product that is compared to method B is a composite of the retrievals with these three a priori: for each retrieval quantity (see Sect. 2.6) the mean of the values obtained with A1–A3 is taken as the solution, and the difference between the maximum and minimum as the uncertainty. The reason to follow this approach is that the impact of the a priori is substantial and there is no external information available instead which justifies the choice for one specific a priori. Tests have indicated that errors estimated in this way are in general considerably larger than the smoothing error, a commonly used parameter in the optimal estimation framework to quantify the impact of the a-priori on the retrieval error (Rodgers, 2000). The relative smoothing error per layer is typically less than 20 %, for both the aerosol extinction and the trace gas retrieval. The a-priori based error is about 1.6 times higher in case of the aerosol extinction retrieval and about 4.8 times higher in case of the NO<sub>2</sub> retrieval (both numbers are median values).

## 2.4 Method B – algorithm developed at KNMI

The profile retrieval approach of method B (Vlemmix et al., 2011) is quite different from method A: it makes use of a profile shape parametrisation with just a few (2–4) free parameters; forward simulations are performed by making use of a look-up table which has been created with the Doubling Adding KNMI (DAK) radiative transfer model (De Haan et al., 1987; Stammes et al., 1989); a standard least-squares algorithm is used, without any form of regularization. The main reason to use a low number of free parameters is that the information content of MAX-DOAS observations with respect to the vertical distribution of aerosols and trace gases is quite limited (see Fig. 3). With a suitable choice of free parameters a sufficiently wide range of possible profile shapes can be retrieved, especially in combination with the ensemble approach described below. Compared to the description in Vlemmix et al. (2011) the algorithm has been modified in the following ways: the profile shape parametrization is slightly different, this is described below; the look-up table is compiled to allow for more extreme aerosol optical thicknesses ( $\tau$ ) needed in China with  $\tau = 3.2$  as maximum; the look-up table is expanded

with a UV component (central wavelength: 360 nm); no correction is applied to compensate for the temperature dependence of the differential cross section of NO<sub>2</sub> (similar strategy for method A), a fixed temperature is used (296 K). This will affect the accuracy of both retrieval methods similarly. The up to four free parameters that are used to parametrize the profile (see Fig. 2) are: (i) the tropospheric column density for trace gases and the AOD in case of aerosols; (ii) the top height of the mixing layer; (iii) the “shape parameter”, which determines the linear increase or decrease of the trace gas concentration or aerosol extinction in the mixing layer; (iv) the fraction of the total trace gas column density which resides (uniformly distributed) in the layer starting at the top of the mixing layer up to two kilometers above. The vertical extent of this layer varies with parameter (ii). Parameter (iv) replaces the free tropospheric layer which in the earlier version of the algorithm (Vlemmix et al., 2011) was put at a fixed altitude. Parameter (iii), already tested and introduced as part of a sensitivity study in Vlemmix et al. (2011) is also newly applied here.

An important characteristic of this profile shape parametrization is that with parameter (ii) it can mimic the dynamic behaviour of the cloud free boundary layer, which can be very shallow in the morning (especially after a cold, cloud free night with little wind) and become quite deep during the day, especially in summer. Parameter (iii) is included especially to allow for profile shapes which peak at higher altitudes (e.g., somewhere near the top of the mixing layer). With parameter (iv) elevated trace gas concentrations at higher altitudes can be described. From Vlemmix et al. (2011) it is known that the accuracy of this part of the profile is generally low. For the aerosol extinction profile retrieval, parameter (iv) is not used for practical reasons (computation time). As a consequence it is not possible to perform accurate aerosol extinction profile retrievals under measurement conditions with elevated aerosol layers above the mixing layer (aerosol extinction profiles which peak near the top of the mixing layer can be described with the shape parameter (iii)). Such cases are however indicated by high values of  $\chi^2$  and can therefore be flagged (or excluded), see below. The cost function used for method B is defined

as:

$$\chi^2 = \sum_{i=1}^8 \left( \frac{\Delta N_{\alpha_i}^S - \widehat{\Delta N_{\alpha_i}^S}}{\epsilon_{\alpha_i}} \right)^2, \quad (3)$$

where  $\Delta N_{\alpha_i}^S$  is the measured differential slant column density for elevation  $i$ ,  $\widehat{\Delta N_{\alpha_i}^S}$  is the simulated differential slant column density and  $\epsilon_{\alpha_i}$  is the error estimate for the differential slant column density. Due to the low number of free parameters used in method B (2 to 4), it is more difficult to get optimal agreement between simulations and measurements (i.e., to obtain low residuals) than with method A (13 profile layers). Therefore, and also because there is no a priori to fall back on, the individual retrievals with method B tend to be more unstable with respect to one or more retrieval parameters.

It is important to note that this instability is not necessarily an unwanted effect: it is an expression of the fact that (under some conditions) the MAX-DOAS observations contain very limited information about the profile shape. For such conditions it is desirable to have a good estimate of the uncertainty. This is obtained by making use of an ensemble approach: the retrieval code is run 50 times, each time with slightly different input. The differential slant column density measurements are perturbed by adding Gaussian noise with a standard deviation corresponding to 10 % of the original differential slant column density (obtained with the semi-simultaneous zenith measurement). For each scan an ensemble of solutions is obtained, and for each retrieval quantity the median is taken as the final result. The width of the distribution for each parameter (e.g., described by the end of the first and beginning of the fourth quartile) provides an estimate for the retrieval uncertainty. Note however, that this retrieval uncertainty does not account for the uncertainty with respect to the profile shape parametrization. For this reason the retrieval is run for several profile shape parametrizations at the same time (see Table 2) and a composite retrieval product is constructed a posteriori. A posteriori selection of plausible profile shape parametrizations (among B1–B4) is done by considering the distribution of the reduced  $\chi^2$  ( $\chi_{\nu}^2$ ).

This parameter is defined as:

$$\chi_{\nu}^2 = \frac{\chi^2}{(N - M)}, \quad (4)$$

where  $N$  is the number of observations (differential slant column densities at various elevations) minus the number of model parameters (i.e., 2 to 4). If the median value of the  $\chi_{\nu}^2$  distribution (after 50 runs) for a certain profile shape parametrization is approximately equal to one, then the selected retrieval model is capable of producing simulations that agree with the observations within the estimated measurement error.

After the algorithm is run 50 times for all four models, it is determined which models are included in the a posteriori composite retrieval product, namely all models which have a median  $\chi_{\nu}^2 < 1.5$ . For each model individually the retrieval outcomes for a certain quantity (e.g., surface concentration) is defined as the median value of the distribution (after 50 runs) for that particular quantity. The lower limit of the corresponding uncertainty estimate is defined as the value which marks the transition from the first to the second quartile of the distribution. The upper limit is defined similarly as the value which marks the transition from the third to the fourth quartile of the distribution. This implies that 50 % of the retrievals is within the error bar. The composite product is constructed simply by averaging the medians of the selected models, and the error bars are constructed by averaging the lower limits and upper limits separately. The procedure that is followed here (including all models among B1–B4 which have sufficiently low median of  $\chi_{\nu}^2$ ) yields a more realistic uncertainty estimate than if only the model with lowest median  $\chi_{\nu}^2$  would be used, because it takes into account the uncertainty with respect to the profile shape.

## 2.5 Selection criteria and uncertainty estimates

Comparison of methods A and B is done only for profile pairs which satisfy three criteria: they should pass the quality control criteria of method A, those of method B and they should coincide with AERONET observations. The third criterium provides an indirect way of selecting cloud free periods. MAX-DOAS profiles are only included in the comparison if at least 3 AERONET level 2.0 (cloud screened, quality controlled) measurements are taken within an hour around

the MAX-DOAS measurement. Quality control for method A is based on two quantities: the size of the residual of the profile fit and the degrees of freedom for signal (DOFS, see Rodgers, 2000). The residuals are defined as the sum of squared differences between simulations and measurements, divided by the simulated differential slant column density for an elevation of 30° (this quantity provides a first-order estimate of the tropospheric vertical column density):

$$\delta = \sum_{i=1}^8 \left( \frac{\Delta N_{\alpha_i}^S - \widehat{\Delta N_{\alpha_i}^S}}{\widehat{\Delta N_{\alpha_{i=8}}^S}} \right)^2, \quad (5)$$

Figures 3 and 4 show the histograms of these two parameters before the quality control is applied. These figures illustrate clearly that in general profile retrieval is more challenging for HCHO than for NO<sub>2</sub>: the DOFS for HCHO are often well below 2, whereas for NO<sub>2</sub> the DOFS are often > 2. Also the residuals for HCHO are considerably higher for a considerable fraction of all data (note that Fig. 4 shows the logarithm of the residual). The same is illustrated in Fig. 5: the averaging kernels for HCHO are lower than for NO<sub>2</sub> and are less orthogonal with respect to one another. Profile pairs of A and B are excluded from the comparison if the minimum value of the DOFS is < 1 for one or more of the models A1–A3. Also they are excluded if the maximum residual of A1–A3 is larger than 0.1. Quality control for method B consists of selecting only those profiles where the median value of the reduced  $\chi_\nu^2$  for the profile fit of method B is < 1.5. For aerosol extinction only the median  $\chi_\nu^2$  of the aerosol extinction profile fits is considered, for the trace gas retrieval also the median  $\chi_\nu^2$  of the trace gas profile fits.

The impact of the quality control criteria defined above is discussed in Sect. 4.1.

## 2.6 Retrieval quantities

We compare results mostly based on three quantities: the tropospheric vertical column densities ( $N^V$ ), the concentration ( $n_{\text{surf}}$ ) or volume mixing ratio ( $X_{\text{surf}}$ ) of trace gases near the surface and the characteristic height (or  $H_{75}$ , see Sect. 1) of the retrieved profile. Similar quantities are used in case of aerosols: aerosol optical depth ( $\tau_{\text{aer}}$ ), aerosol extinction near the surface  $e_{\text{surf}}$ , and  $H_{75}$ . A fourth quantity that is used is the a posteriori scale height (not to be confused with

the scale height of the a priori profile of method A ( $H_s^{\text{prior}}$ ), see Sect. 2.3). This a posteriori scale height  $H_s^{\text{post}}$  is a first order profile height estimate derived from column density (or AOD) and surface concentration (or aerosol extinction):

$$H_s^{\text{post}} = \frac{N^V}{n_{\text{surf}}} \quad (6)$$

for the trace gases, and

$$H_s^{\text{post}} = \frac{\tau}{e_{\text{surf}}} \quad (7)$$

for aerosols. The reason to consider this first order profile height estimate in addition to  $H_{75}$  is that, as will be shown in Sect. 3.2, for method A it depends less on the a priori than  $H_{75}$ . This indicates to some extent that the measurements contain information about the profile height that is not extracted in an optimal manner in this particular retrieval set-up.

### 3 Results

#### 3.1 Example day

Figures 6–8 show retrieval results for 19 May 2012. The individual profiles obtained with method A and B (Fig. 6) show good agreement in the sense that in the morning they are all quite low, and in the afternoon they are all quite high. Nevertheless, this example also illustrates that retrieved profile shapes can be very different: not only the four versions of method B show considerable differences but also the two versions of method A, especially those retrieved after 12:00 p.m. (bottom row).

Time lines for the different quantities that can be derived from the profile retrieval are shown in Fig. 7 for the same example day. Both for the aerosol extinction retrieval (left column) and the  $\text{NO}_2$  retrieval (right column) the  $\chi^2_V$  are quite low for most of the day, indicating a good quality of fit. On this particular day, the retrievals agree quite well for most quantities, especially for

the column densities (row 1) and surface concentrations (row 3). Agreement is worst for  $H_{75}^{\text{NO}_2}$  and  $H_s^{\text{NO}_2}$  in the afternoon, where method B occasionally shows much higher values. This is a consequence of the fact that the retrieval is not regularized, in combination with relatively low surface concentrations in the afternoon. Because the surface concentration is the denominator in Eq. (6), one can understand that a small change (error) in the surface concentration can lead to a much larger change in  $H_s^{\text{NO}_2}$ . This figure also clearly demonstrates the potential impact of different profile shape assumptions on  $H_{75}^{\text{NO}_2}$ .

Figure 8 shows results for the same day, but this time for the aerosol extinction and HCHO retrieval in the UV. In general there is much more disagreement compared to  $\text{NO}_2$ . There is on this day almost no retrieval where the agreement is good for all quantities at the same time. The agreement between most quantities is especially low between 10:00 a.m. and 4:00 p.m. High values for  $\chi_\nu^2$  in the aerosol extinction retrieval indicate that the retrieval with method B is not successful and therefore this period is flagged with grey bars on top of each figure. Quite remarkable is the disagreement in terms of HCHO column densities in the remaining part of the day (before 10:00 a.m. and after 4:00 p.m.). In the morning of this day the higher column densities (for method B compared to A) seem to go along with higher  $H_{75}^{\text{HCHO}}$ .

## 3.2 Statistical analysis

Results of both retrieval methods are compared for sixteen quantities in terms of correlation, slope and intercept of linear fit, and median, mean and standard deviation of relative differences, see Table 3. The comparison of twelve of these quantities is also shown in Figs. 10, 14, 15 and 21. We will discuss these results separately in terms of tropospheric column densities (AOD for aerosols), profile heights, and surface concentrations (or aerosol extinction).

### 3.2.1 Tropospheric column densities

Figure 9 shows the monthly median values for the column quantities: AOD, tropospheric  $\text{NO}_2$  and HCHO column densities. Note that measurements before 2010 are made in Beijing. From 2010 onwards, the observations are made in Xianghe. A clear seasonal cycle with a winter



minimum of about  $5 \times 10^{15}$  molec cm<sup>-2</sup> and a summer maximum roughly five times as high can be seen for HCHO. Compared to NO<sub>2</sub> and aerosols, the variability per month is quite small. A weaker, but similar seasonal cycle can be seen for aerosols, with typical winter values around 0.2 and a summer median between 0.5 and 1.0. For NO<sub>2</sub> the seasonal cycle of monthly median values is quite weak as well. Winter medians are roughly between  $20 \times 10^{15}$  and  $30 \times 10^{15}$  molec cm<sup>-2</sup>, summer medians between  $10 \times 10^{15}$  and  $20 \times 10^{15}$  molec cm<sup>-2</sup>. Noteworthy is the fact that especially the peak values in winter can be high with values above  $100 \times 10^{15}$ . Peaks in tropospheric NO<sub>2</sub> column densities in midsummer do not exceed  $30 \times 10^{15}$ . Figure 10 and Table 3 show that very good agreement is found for tropospheric NO<sub>2</sub> column densities. The standard deviation of relative differences is however considerable: almost 10 %. The third and fourth column of Fig. 10 show that the relative difference increases with increasing tropospheric column density and with increasing profile height. Also for tropospheric HCHO column densities the agreement is good. Relative differences are however considerably larger than for NO<sub>2</sub>. The dependence of relative differences on the tropospheric column density itself (second row, third column) shows opposite behaviour as for NO<sub>2</sub>, whereas the dependence of relative differences on the profile height shows a similar increase as for NO<sub>2</sub>. Despite the quite high correlation found for the AOD, the agreement between method A and B is moderate, with slopes 1.20 (UV) and 1.39 (Vis) and substantial mean relative differences. Figure 10 (bottom row), shows for the Vis that these differences in AOD are strongly related to the difference in aerosol extinction profile height, but also tend to increase with the AOD itself. The agreement between method B and AERONET is much better, which provides confidence in the AOD retrievals obtained with method B. The frequency distributions of AERONET and AODs retrieved with method B show good agreement and differ with respect to method A in the fact that they include much more cases with AOD between 1.5 and 3.5. Only for the highest 25 % of characteristic profile heights, method B seems to overestimate the AOD systematically by about 20 %. Figure 11 shows for NO<sub>2</sub> and HCHO the relation between AOD (as measured by AERONET) and tropospheric trace gas column densities for different seasons. There are clear seasonal differences with largest differences for NO<sub>2</sub> vs. AOD between summer and winter. The two models show good agreement, with only moderate systematic differences for

HCHO column densities in spring and summer. This is in line with the example day (Fig. 8) which shows considerable differences between tropospheric HCHO column densities retrieved method A and B. Note that on this example day the AOD is high, and the differences in characteristic HCHO profile height are considerable. The quite linear relationship between NO<sub>2</sub> and AOD, and HCHO and AOD illustrate that trace gas emissions are often accompanied by aerosol emissions. From that perspective the flattening of the curves for high AOD's (mainly in summer and autumn) is remarkable. Possibly this is related to aerosols from natural sources (dust) emissions of which do not go along with emissions of trace gases. Another explanation would be that high AOD's cause systematic underestimation of the tropospheric column density, but the flattening of the curves is not seen in winter and spring, even for higher AOD's.

### 3.2.2 Profile heights

Figures 12 and 13 show for methods A and B monthly median values of characteristic heights, with a distinction between retrievals before 10:00 a.m. (red) and retrievals after 12:00 p.m. (blue). For all three species and both retrieval methods, we find higher profiles in the afternoon than in the morning, especially with method B. Only for  $H_{75}^{\text{HCHO}}$  obtained with method A the differences between morning and afternoon are negligible. This is most likely an artefact, which is also seen in Fig. 14, and which is discussed in Sect. 4.2. The morning to afternoon differences found in all other cases are qualitatively in agreement with the expected diurnal variation in the mixing layer height, and provide a first order check to see if the algorithms behave as expected. Variability per month and between months is however much larger with method B. Highest monthly median characteristic profile heights are found with method B for aerosol extinction profiles in summer. This is in agreement with the general expectation that mixing layers are more shallow in winter and grow deeper in summer (see e.g. Luo et al., 2014). That this effect is weaker for NO<sub>2</sub> might be related to the shorter lifetime of NO<sub>2</sub> in summer which limits the effective transport of NO<sub>2</sub> from the surface to the higher parts of the mixing layer, see also Halla et al. (2011) and Mendolia et al. (2013). Figure 14 and Table 3 show that the agreement between the two methods in terms of profile heights is considerably lower than the agreement in terms of column densities, which is to be expected because the information content of MAX-

DOAS with respect to the vertical distribution is limited. The best agreement is found again for  $\text{NO}_2$ , with for the profile heights correlation 0.76, slope 1.44, intercept  $-0.20$  km, and mean relative difference 6.51 %. The standard deviation of relative differences is high: 33.18 %. It can be seen in Fig. 14 that the dynamic range of  $\text{NO}_2$  profile heights found with method A is somewhat lower than with method B. Especially the fraction of profiles with height above 1 km is significantly higher with method B. For the HCHO profile height we have correlation 0.62. This is quite surprising because the dynamic range of profile heights found with method A is very small compared to method B and this also explains the exceptional slope (7.47) and intercept ( $-5.33$  km). Even though no independent data is available, it is quite safe to conclude that this very limited dynamic range is unrealistic, and therefore these HCHO profiles should be used and interpreted with great care. As a result of this effect, it is difficult to judge the quality of the HCHO profile heights obtained with method B. One can see that here the dynamic range is comparable to that of  $\text{NO}_2$  and aerosols, but the mode of the histogram has shifted to higher altitudes compared to  $\text{NO}_2$ . Also for aerosols (Vis) the dynamic range found with method A is limited compared to method B. The correlation (0.62) is somewhat lower than for  $\text{NO}_2$  and HCHO, but the slope and intercept are less extreme than for HCHO (2.82 and  $-1.31$  respectively). Figure 15 demonstrates that in terms of the other estimator of the average profile height ( $H_s^{\text{post}}$ ) the agreement between methods A and B is better. Especially for HCHO, the slope is less extreme for  $H_s^{\text{post}}$  than for  $H_{75}$  which is in line with the higher dynamic range seen for method A in Fig. 15 compared to Fig. 14. We see a similar effect for aerosols. This might indicate a retrieval artefact for method A which causes that information about the profile height that is actually contained in the MAX-DOAS measurements is not efficiently converted into a noticeable effect on  $H_{75}^{\text{NO}_2}$ .

A different view on the quality of the profile height retrieval obtained with both methods is given by Fig. 16. From the first left panel we can conclude that for both methods the internal consistency (UV vs. Vis) of aerosol extinction profile heights below 1.5 km is quite good, especially for method A. Above 1.5 km we have only very few cases with method A, and all of these cases show a strong bias between UV and Vis. For method B the bias appears to be quite constant over the entire range, with UV profiles that are approximately 25 % lower than profiles

in the Vis. The middle panel of Fig. 16 shows a comparison of  $\text{NO}_2$  and aerosol extinction profiles. In contrast to aerosols, we do not expect a strong agreement beforehand. What we hope to see, and this is partially the case, is that the general pattern is similar for both methods. Below 1.5 km the agreement is remarkably good, and this is certainly a confirmation that the results obtained with both methods make some sense. As mentioned before, the limited dynamical range of method A makes it almost impossible to draw conclusions on the reliability of profile heights above 1.5 km found with method B. Nevertheless, a possible explanation for the bias between  $\text{NO}_2$  and aerosol extinction profile heights in this regime is the same as mentioned earlier: higher aerosol extinction profiles occur in summer, but then the lifetime of  $\text{NO}_2$  can be very short, which leads to more shallow  $\text{NO}_2$  profiles. Figure 17 shows for different seasons the characteristic aerosol extinction,  $\text{NO}_2$  and HCHO profile heights as a function of the AOD. For aerosol extinction profile heights, we see a much stronger seasonal cycle with method B than with method A. In principle a seasonal cycle is also expected: higher boundary layers occur in summer, when the thermal convection is strongest. A possible interpretation of the results seen on the top row (decline of  $H_{75}^{\text{aer}}$  with increasing AOD) is that growth of the boundary layer through convection is weakened by the presence of high aerosol loadings (see also Barbaro et al., 2013). Without independent simultaneous observations with other techniques, it can however not be excluded that this effect is related to the measurement technique itself (i.e., a retrieval artefact). Method B shows a weaker seasonal variation in  $\text{NO}_2$  than in aerosol extinction profile heights and highest  $\text{NO}_2$  profiles occur in spring. This might be due to the fact that in spring the  $\text{NO}_2$  lifetime is not as short as in summer (allowing more time for vertical transport), whereas at the same time vertical transport through convection is stronger than in winter. Results for HCHO are more difficult to interpret. Because the lifetime is longer than for  $\text{NO}_2$ , and because formaldehyde sources can be biogenic and anthropogenic (the relative contribution varies by season) the profile shapes can be very different from those of  $\text{NO}_2$ . A quantity that is especially important in the context of satellite validation and satellite retrievals is the relative difference in  $\text{NO}_2$  and aerosol extinction profile height. The impact of the relative characteristic profile heights on the slant column density measurement can be high, and lead to systematic biases if not accounted for in the retrieval. This quantity is shown for both methods as a function

of season in Fig. 18 (also for HCHO). Similar as for the characteristic heights themselves, we see in Fig. 18 a higher dynamic range for method B than for method A. This is partly explained by the lower stability of method B, but also by the ability to retrieve a wider range of profile heights. Both methods detect in spring higher characteristic aerosol than  $\text{NO}_2$  profile heights.

5 In summer method B finds systematically higher values for  $H_{75}^{\text{aer}} - H_{75}^{\text{NO}_2}$  than method A. In winter and autumn, the systematic bias between  $H_{75}^{\text{aer}}$  and  $H_{75}^{\text{NO}_2}$  is smaller. As argued above, results for HCHO are more difficult to interpret (because of the artefact affecting the retrieval with method A). However, based on the results obtained with method B it appears as if aerosol extinction profiles are higher than HCHO profiles in spring and summer, and lower in fall and  
10 winter.

### 3.2.3 Aerosol extinction and trace gas volume mixing ratios near the surface

Seasonal variations of volume mixing ratios and aerosol extinction near the surface are shown in Figs. 19 and 20 for methods A and B respectively. For  $\text{NO}_2$  a systematic difference is seen between morning and afternoon values, and this is clearly related to the dynamics of the mixing layer. For aerosols a similar effect is found. For HCHO however, this contrast is almost  
15 absent. This is related to fact that HCHO profiles shapes retrieved with method A show almost no deviation from the a priori (Sect. 3.2.2). As a consequence, the main driver of the surface concentration is the tropospheric column density of HCHO. This explains why for HCHO retrieval with method A the seasonal variation in volume mixing ratios is so similar to the seasonal variation in column densities. For method B (not shown) the results are quite different in winter  
20 months, when morning values are about three to four times higher than afternoon values of the HCHO volume mixing ratio. In summer months, this effect appears to be less pronounced, unlike for  $\text{NO}_2$ . It is difficult to draw conclusions based on method B only, but this weaker diurnal variation in HCHO surface volume mixing ratios compared to winter could indicate  
25 that in summer local emissions on the surface have a relatively small impact. Based on this data set only, it can however not be excluded that absence of a strong morning to afternoon contrast for HCHO volume mixing ratios in summer is an artefact of the retrieval. Fig. 21 shows the results of the comparison of methods A and B in terms of trace gas volume mixing ratios and

aerosol extinction near the surface (lowest profile layer). In contrast with the results found for the profile heights, the agreement is reasonable, with quite similar histograms for all three constituents. Nevertheless, the systematic relative differences are considerable. For  $\text{NO}_2$  we have a mean relative difference of  $-7.92\%$  with a standard deviation of  $32.97\%$ . For HCHO the relative differences are larger ( $-20.70\%$ ,  $\pm 27.71\%$ ) which is mostly explained by the differences in profile shape, because in terms of column densities the relative difference is smaller and of opposite sign ( $9.43\%$ ). With respect to aerosol extinction near the surface, the agreement between methods A and B is good, with correlation  $0.93$ , slope  $1.33$  and intercept  $-0.08$ . The mean relative difference is however considerable ( $10.27\%$ ) and the the standard deviation of relative differences is high:  $30.43\%$ .

## 4 Discussion

In this section we address the main question of this paper: what can be concluded on the quality of aerosol extinction and trace gas profiles retrieved from MAX-DOAS observations. We begin with a discussion of strength and weaknesses of both profile retrieval methods, draw conclusions, and then give recommendations for improvements and use.

### 4.1 Impact of quality control

The ideal selection of high quality data for this comparison study would be based on a validated cloud screening method which performs well under a wide range of aerosol conditions. Such a method was not available when this study was started (in the mean time promising results have been published by Wagner et al., 2013, and Gielen et al., 2014). Therefore a pragmatic approach was chosen, see Sect. 2.5. A disadvantage of this approach is that a high number of retrievals is rejected. For example, there are many cases where the trace gas retrieval is rejected (despite a proper  $\chi_v^2$ ) because the  $\chi_v^2$  in the aerosol extinction retrieval is not sufficiently low. The criterium used might be more appropriate for a quality control intended for profiles – and for that reason it is used in this work – but it is probably too strict for a quality control intended

for column densities only. Several tests have been performed to check the robustness of findings reported in this paper after changing the selection criteria. For example, the criterion on  $\chi^2_{\nu}$  has been relaxed to  $\chi^2_{\nu} < 5$  and the number of AERONET observations in the same hour is lowered from 3 or more to 2 or more. This leads roughly to two times more aerosol extinction profile pairs (see second column of Table 3) and roughly two and a half and three times more profile pairs for HCHO and NO<sub>2</sub> respectively. The impact of these relaxed settings is considerable for the aerosol extinction retrieval (e.g., mean relative difference in  $H_{75}^{\text{aer}}$  increases from 2.74 % to 12.35 %), but quite small for the trace gas retrieval. For example, the mean relative difference in  $H_{75}^{\text{NO}_2}$  increases from 8.71 % to 9.85 %, and the mean relative difference in the volume mixing ratio for NO<sub>2</sub> decreases from -7.92 % to -4.4 %, which is a small change compared to the standard deviation (32.97 %). There are no sign changes for quantities in Table 3 that are significantly different from zero. It should be noted that the results for the aerosol extinction retrieval obtained with these relaxed constraints are clearly considered to be less representative for ideal clear sky conditions. With every set of quality criteria, the results presented here will change slightly (largely due to a different the sampling of the full data set), however the settings used here are considered to be reasonable balance between maintaining sufficient data pairs and rejection of data pairs which are likely to be affected through clouds.

## 4.2 Strength and weaknesses

In Sect. 3.2 it was shown that both methods show good agreement in terms of tropospheric NO<sub>2</sub> and HCHO column densities: the correlation is high, the slopes of linear fit are close to 1 and the intercepts are relatively close to zero. The agreement of characteristic profile heights is reasonable for NO<sub>2</sub> and aerosols, despite clear biases, especially above 1.5 km. The main strength of method A is its robustness (stability). This is a clear advantage especially when differential slant column densities are close to the detection limit, or when the assumptions that are made about fixed parameters (or cloud free conditions) do not hold. In such cases, the retrieval can rely on the a priori. The characteristic profile height retrieval with method B is only stable under cloud free conditions, and if assumptions about fixed parameters are not too far from the truth. The number of profiles which passes the quality control (Sect. 2.5) for

method B is significantly smaller than for method A. A disadvantage of method A is that the combination of a profile parametrization based on 13 layers and a relatively low information content of the MAX-DOAS observations forces one to take measures to stabilize the retrieval. These measures are: (1) a relatively conservative estimate of the a priori error (for each profile layer 20 % of the a priori profile estimate), (2) a profile which decreases to zero rapidly above 1 km. A consequence of this approach is that the absolute values of the a priori error estimate become very low above 1 km. This is believed to be the main reason why it is almost impossible to retrieve profiles with a characteristic height ( $H_{75}$ ) much higher than that of the a priori. In most cases sufficient agreement between observations and simulations can be achieved by modifying the profile shape (compared to the a priori) only below 1.0 km. This explains why for retrieved NO<sub>2</sub> profiles reduction of  $H_{75}$  compared to the a priori is seen much more often than increase. This is not seen for HCHO. For HCHO it appears that the information content is too low to obtain profiles (with method A) which deviate much from the a priori. A strong aspect of method B is that it can realize a high range of quite different profile shapes, with just a few free parameters. It can more easily realize profiles which have a characteristic height ( $H_{75}$ ) well above 1.0 km. In this study, it is however not possible to fully judge the quality of these profiles because these cannot be retrieved with method A. Nevertheless, the monthly averaged morning to afternoon difference in profile height and the seasonal cycle of aerosol extinction profile heights (Fig. 13) correspond to the expected behaviour and this is at least an indirect indication of the quality of the profiles obtained with method B. Independent (e.g., lidar) observations at the same measurement site would be needed to say more about the quality of individual profiles.

This study also makes clear that the main disadvantage of method B is its instability, despite the limited number for free parameters and the ensemble approach. Note however that the retrieval is certainly not always unstable, see for example the retrievals in the Vis on 15 May 2012 (Fig. 7). The advantage of the ensemble approach taken with method B is that most often the instabilities go along with high uncertainty estimates, and this provides a means for additional quality control. Unlikely retrievals with a low uncertainty estimate occur also, but these can



most often be excluded based on high values for  $\chi^2$ , either in the aerosol or in the trace gas part of the retrieval.

### 4.3 Recommendations for algorithm improvements and further validation

Both profile retrieval algorithms have specific strengths and weaknesses, as described above. The challenge for improved algorithms is to combine the stability and precision of method A with the ability of method B to retrieve a high dynamic range of characteristic profile heights. A possible but not so practical solution could be to use method B to obtain an initial estimate of the a priori scaling height for method A and then as a next step to perform a retrieval with method A. This will work however only under strict cloud free conditions because of the limitations for method B. Also it would transfer the impact of instabilities (for individual cases) from method B to method A. An alternative is to use the profile parametrization of method B in the framework of the optimal estimation method. Such a retrieval algorithm could be better capable of retrieving a wide range of profile heights and at the same time be more stable than the present implementation of method B. This would also lead to an algorithm which is considerably faster because there would be no need for an ensemble approach. Improving the stability of the retrieval by making use of a priori data (in combination with the optimal estimation approach) brings a certain risk, which is that systematic biases in the a priori climatology remain present in the a posteriori climatology. An advantage of more simple retrieval schemes (e.g., method B) is that they are predominantly driven by the observations themselves and therefore less prone to inheritance of systematic biases in the a priori, despite a low precision. It is almost impossible to make a choice which combines the best of both worlds: a very stable retrieval (i.e., precision of individual profiles) without introducing systematic biases in a climatological sense. Stability is important for comparison with satellite observations if the number of available cases is very limited, accuracy over a wide range of profile heights is important if MAX-DOAS would be used to provide a climatology of profile heights for better a priori estimates in the satellite retrieval. The recent work by Hartl and Wenig (Hartl and Wenig, 2013) provides indications that the Phillips-Tikhonov regularization method can be used for MAX-DOAS profile retrievals which are more stable and at the same time (potentially) less biased in a climatologi-

cal sense. To our best knowledge, their method has not yet been applied to a long data set of real observations with a similar focus on the ability to retrieve accurate (first order) profile height estimates. The present study has demonstrated the benefit of having a large data set covering a wide range of measurement conditions. Based on a small data set it would have been very difficult to entangle differences in accuracy and precision. More thorough validation requires simultaneous co-located observations with other techniques (lidar, NO<sub>2</sub> sonde). Such validation efforts are especially useful if a sufficiently large data set is available. In the presence of large differences in spatial representativity (this is very different for satellite, MAX-DOAS and in-situ techniques) and a high variability in possible NO<sub>2</sub> and aerosol extinction profile shapes, it is almost impossible to draw conclusions about accuracy of MAX-DOAS profile shapes based on a quite limited number of co-located observations, even if the precision and accuracy of the other techniques are high.

## 5 Summary and conclusions

conclusions

A four year data set of MAX-DOAS observations in the Beijing area is analysed with two different methods for the retrieval of tropospheric NO<sub>2</sub>, HCHO and aerosol extinction profiles. The objective of this study is firstly to assess for each constituent (NO<sub>2</sub>, HCHO, aerosols) and retrieval quantity (AOD or tropospheric column density, characteristic profile height ( $H_{75}$ ), aerosol extinction or surface concentration) the mutual consistency of the retrievals with both methods, and secondly to identify the mechanisms causing the differences. The two profile retrieval methods differ in many respects. Method A uses a profile parametrization with 13 layers (up to 4 km), on-line forward simulations with the LIDORT radiative transfer model and an inversion based on optimal estimation. Method B uses a profile parametrization based on 2 to 4 parameters to describe the profile shape and a look-up table created with the DAK radiative transfer model. The inversion is based on a least-squares minimization and an ensemble approach is used to improve stability of the solutions and to estimate uncertainties. In the following we summarize the results of the comparison, first in a qualitative sense, then quantitatively.

The strength of method A is the stability of the profile shape retrieval, even under cloudy conditions, which is a consequence of the relatively conservative estimate of the uncertainty of the a priori profile. The choice for stability is advantageous for the retrieval of tropospheric column densities and volume mixing ratios near the surface. A negative side effect of this conservative estimate of the uncertainty of the a priori appears to be that the retrieved characteristic profile heights have a relatively small dynamic range. This is most evident for the HCHO profiles retrieved in the UV, but also for aerosol extinction and NO<sub>2</sub> retrieved in the Vis. Method B is generally less stable, and this affects the precision of individual retrievals. The tropospheric column density is least sensitive to instabilities in the profile retrieval, whereas the characteristic profile height and volume mixing ratio near the surface are most sensitive. The most pronounced difference with method A is the higher dynamic range of retrieved profile heights for aerosols, HCHO and NO<sub>2</sub>. Although the higher dynamic range is partly a consequence of the instability of the retrieval (and therefore not necessarily meaningful), diurnal and seasonal patterns that show up after averaging many profiles give some confidence that the retrievals are meaningful. For example, we see low characteristic profile heights in the morning, and higher values in the afternoon, especially for aerosol extinction in summer. This can be related to the periodical cycles of the boundary layer. Also we find in spring and summer lower aerosol extinction profile heights with decreasing aerosol optical thickness. Although it cannot be excluded that this is a retrieval artefact, this might also be real (and therefore add to the credibility of method B), namely that higher aerosol loads reduce the thermal convection in the boundary layer and therefore lead to lower aerosol extinction profile heights. More quantitatively, we find best agreement for the tropospheric NO<sub>2</sub> column densities (correlation 0.99), with almost no systematic bias (slope 1.03, intercept  $-0.56 \times 10^{15}$  molec cm<sup>-2</sup>) and comparatively small relative differences (mean 0.25 % and standard deviation 9.3 %). For formaldehyde column densities we find a high correlation (0.95) and slope close to one (1.02), but also that method B is systematically higher than method A: mean relative difference is 9.4 % and standard deviation of relative differences is 12.1 %. Relative differences in formaldehyde column densities are found to be related to differences in profile height: overestimations of the tropospheric column density (for method B compared to method A) often correspond to overestimations of the characteristic profile height

(for method B compared to method A). Volume mixing ratios near the surface are systematically lower for method B compared to method A: 8 % relative difference for NO<sub>2</sub> and 21 % for HCHO. The differences can again be related to the differences in profile heights between method A and B. The standard deviation of relative differences of surface volume mixing ratios is much higher than for tropospheric column densities: 33 % for NO<sub>2</sub> and 28 % for HCHO. Characteristic profile heights are systematically higher for method B than for method A. The mean relative differences are 6.5 % for NO<sub>2</sub>, 15.7 % for HCHO and 2.7 % for aerosols (Vis). The high standard deviation of relative differences (33 %, 37 % and 44 % for NO<sub>2</sub>, HCHO and aerosols respectively) shows that the precision of characteristic profile heights is low. We find with method B that in spring and summer aerosol extinction profiles are systematically higher than NO<sub>2</sub> profiles. Also we find that in winter and summer mornings HCHO profiles are systematically higher than aerosol extinction profiles, and vice versa in summer afternoons. Note however that these findings are only indicative, because the limitations with method A prevent confirmation of the results obtained with method B. Altogether this study gives some indications about the quality of tropospheric column densities, surface concentrations and profile heights retrieved with MAX-DOAS. Since this study is based solely on MAX-DOAS observations the scope is limited and a more thorough validation is needed. In order to obtain robust validation results which can entangle differences related to accuracy and/or precision for a wide range of pollution and sky conditions, it is recommended to station MAX-DOAS instruments close to continuously monitoring surface in-situ monitors (e.g., for NO<sub>2</sub>), sun photometers and lidars which are sufficiently sensitive to boundary layer aerosols.

*Acknowledgements.* This research was financially supported at IASB-BIRA by the Belgian Federal Science Policy Office, Brussels (AGACC-II project) and the EU 7th Framework Programme projects NORS (contract 284421) and SHIVA (contract 226224). The authors would like to acknowledge R. Spurr for providing the LIDORT package, and P. Stammes and P. Wang for providing DAK. We thank Pucui Wang, Hong-Bin Chen and Philippe Goloub for their effort in establishing and maintaining the Xianghe and Beijing AERONET sites. Furthermore we would like to thank Trissevgeni Stavrou from IASB-BIRA for her contributions to this work.

## References

### references

- Barbaro, E., Vilà-Guerau de Arellano, J., Krol, M., and Holtslag, A.: Impacts of aerosol shortwave radiation absorption on the dynamics of an idealized convective atmospheric boundary layer, *Bound.-Lay. Meteorol.*, 148, 31–49, 2013.
- Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J.: Validation of urban NO<sub>2</sub> concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities, *Atmos. Chem. Phys.*, 9, 3867–3879, doi:<http://dx.doi.org/10.5194/acp-9-3867-2009>, 2009.
- Brinksma, E. J., Pinardi, G., Volten, H., Braak, R., Richter, A., Schoenhardt, A., Van Roozendaal, M., Fayt, C., Hermans, C., Dirksen, R. J., Vlemmix, T., Berkhout, A. J. C., Swart, D. P. J., Oetjes, H., Wittrock, F., Wagner, T., Ibrahim, O., de Leeuw, G., Moerman, M., Curier, R. L., Celarier, E. A., Cede, A., Knap, W. H., Veeffkind, J. P., Eskes, H. J., Allaart, M., Rothe, R., PETERS, A. J. M., and Levelt, P. F.: The 2005 and 2006 DANDELIONS NO<sub>2</sub> and aerosol intercomparison campaigns, *J. Geophys. Res.*, 113, D16S46, doi:<http://dx.doi.org/10.1029/2007JD008808>, 2008.
- Clémer, K., Van Roozendaal, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and De Mazière, M.: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, *Atmos. Meas. Tech.*, 3, 863–878, doi:<http://dx.doi.org/10.5194/amt-3-863-2010>, 2010.
- Coburn, S., Dix, B., Sinreich, R., and Volkamer, R.: The CU ground MAX-DOAS instrument: characterization of RMS noise limitations and first measurements near Pensacola, FL of BrO, IO, and CHOCHO, *Atmos. Meas. Tech.*, 4, 2421–2439, doi:<http://dx.doi.org/10.5194/amt-4-2421-2011>, 2011.
- De Haan, J. F., Bosma, P. B., and Hovenier, J. W.: The adding method for multiple scattering calculations of polarized light, *Astron. Astrophys.*, 183, 371–393, 1987.
- Fayt, C., De Smedt, I., Letocart, V., Merlaud, A., Pinardi, G., and Van Roozendaal, M.: QDOAS Software user manual, available at: <http://uv-vis.aeronomie.be/software/QDOAS/index.php> (last access: 30 June 2013), 2011.
- Gielen, C., Van Roozendaal, M., Hendrick, F., Pinardi, G., Vlemmix, T., De Bock, V., De Backer, H., Fayt, C., Hermans, C., Gillotay, D., and Wang, P.: A simple and versatile

- cloud-screening method for MAX-DOAS retrievals, *Atmos. Meas. Tech. Discuss.*, 7, 5883–5920, doi:<http://dx.doi.org/10.5194/amtd-7-5883-2014>, 2014.
- Halla, J. D., Wagner, T., Beirle, S., Brook, J. R., Hayden, K. L., O'Brien, J. M., Ng, A., Majonis, D., Wenig, M. O., and McLaren, R.: Determination of tropospheric vertical columns of NO<sub>2</sub> and aerosol optical properties in a rural setting using MAX-DOAS, *Atmos. Chem. Phys.*, 11, 12475–12498, doi:<http://dx.doi.org/10.5194/acp-11-12475-2011>, 2011.
- Hartl, A. and Wenig, M. O.: Regularisation model study for the least-squares retrieval of aerosol extinction time series from UV/Vis MAX-DOAS observations for a ground layer profile parameterisation, *Atmos. Meas. Tech.*, 6, 1959–1980, doi:<http://dx.doi.org/10.5194/amt-6-1959-2013>, 2013.
- Hendrick, F., Müller, J.-F., Clémer, K., Wang, P., De Mazière, M., Fayt, C., Gielen, C., Hermans, C., Ma, J. Z., Pinardi, G., Stavrou, T., Vlemmix, T., and Van Roozendael, M.: Four years of ground-based MAX-DOAS observations of HONO and NO<sub>2</sub> in the Beijing area, *Atmos. Chem. Phys.*, 14, 765–781, doi:<http://dx.doi.org/10.5194/acp-14-765-2014>, 2014.
- Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N.: NO<sub>2</sub> column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, *J. Geophys. Res.*, 114, D13307, doi:<http://dx.doi.org/10.1029/2009JD011848>, 2009.
- Irie, H., Kanaya, Y., Akimoto, H., Iwabuchi, H., Shimizu, A., and Aoki, K.: First retrieval of tropospheric aerosol profiles using MAX-DOAS and comparison with lidar and sky radiometer measurements, *Atmos. Chem. Phys.*, 8, 341–350, doi:<http://dx.doi.org/10.5194/acp-8-341-2008>, 2008a.
- Irie, H., Kanaya, Y., Akimoto, H., Tanimoto, H., Wang, Z., Gleason, J. F., and Bucsel, E. J.: Validation of OMI tropospheric NO<sub>2</sub> column data using MAX-DOAS measurements deep inside the North China Plain in June 2006: Mount Tai Experiment 2006, *Atmos. Chem. Phys.*, 8, 6577–6586, doi:<http://dx.doi.org/10.5194/acp-8-6577-2008>, 2008b.
- Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendael, M.: Eight-component retrievals from ground-based MAX-DOAS observations, *Atmos. Meas. Tech.*, 4, 1027–1044, doi:<http://dx.doi.org/10.5194/amt-4-1027-2011>, 2011.
- Kanaya, Y., Irie, H., Takashima, H., Iwabuchi, H., Akimoto, H., Sudo, K., Gu, M., Chong, J., Kim, Y. J., Lee, H., Li, A., Si, F., Xu, J., Xie, P.-H., Liu, W.-Q., Dzhola, A., Postylakov, O., Ivanov, V., Grechko, E., Terpigova, S., and Panchenko, M.: Long-term MAX-DOAS network observations of

NO<sub>2</sub> in Russia and Asia (MADRAS) during the period 2007–2012: instrumentation, elucidation of climatology, and comparisons with OMI satellite observations and global model simulations, *Atmos. Chem. Phys.*, 14, 7909–7927, doi:<http://dx.doi.org/10.5194/acp-14-7909-2014>, 2014.

- 5 Li, X., Brauers, T., Shao, M., Garland, R. M., Wagner, T., Deutschmann, T., and Wahner, A.: MAX-DOAS measurements in southern China: retrieval of aerosol extinctions and validation using ground-based in-situ data, *Atmos. Chem. Phys.*, 10, 2079–2089, doi:<http://dx.doi.org/10.5194/acp-10-2079-2010>, 2010.
- 10 Li, X., Brauers, T., Hofzumahaus, A., Lu, K., Li, Y. P., Shao, M., Wagner, T., and Wahner, A.: MAX-DOAS measurements of NO<sub>2</sub>, HCHO and CHOCHO at a rural site in Southern China, *Atmos. Chem. Phys.*, 13, 2133–2151, doi:<http://dx.doi.org/10.5194/acp-13-2133-2013>, 2013.
- 15 Lin, J.-T., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendael, M., Clémer, K., and Irie, H.: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, *Atmos. Chem. Phys.*, 14, 1441–1461, doi:<http://dx.doi.org/10.5194/acp-14-1441-2014>, 2014.
- 20 Luo, T., Yuan, R., and Wang, Z.: Lidar-based remote sensing of atmospheric boundary layer height over land and ocean, *Atmos. Meas. Tech.*, 7, 173–182, doi:<http://dx.doi.org/10.5194/amt-7-173-2014>, 2014.
- Ma, J. Z., Beirle, S., Jin, J. L., Shaiganfar, R., Yan, P., and Wagner, T.: Tropospheric NO<sub>2</sub> vertical column densities over Beijing: results of the first three years of ground-based MAX-DOAS measurements (2008–2011) and satellite validation, *Atmos. Chem. Phys.*, 13, 1547–1567, doi:<http://dx.doi.org/10.5194/acp-13-1547-2013>, 2013.
- 25 Frieß, U., Monks, P. S., Remedios, J. J., Rozanov, A., Sinreich, R., Wagner, T., , and Platt, U.: MAX-DOAS O<sub>4</sub> measurements: a new technique to derive information on atmospheric aerosols: 2. Modeling studies, *J. Geophys. Res.*, 111, D14203, doi:<http://dx.doi.org/10.1029/2005JD006618>, 2006.
- 30 Mendolia, D., D’Souza, R. J. C., Evans, G. J., and Brook, J.: Comparison of tropospheric NO<sub>2</sub> vertical columns in an urban environment using satellite, multi-axis differential optical absorption spectroscopy, and in situ measurements, *Atmos. Meas. Tech.*, 6, 2907–2924, doi:<http://dx.doi.org/10.5194/amt-6-2907-2013>, 2013.

- Pinardi, G., Van Roozendael, M., Abuhassan, N., Adams, C., Cede, A., Clémer, K., Fayt, C., Frieß, U., Gil, M., Herman, J., Hermans, C., Hendrick, F., Irie, H., Merlaud, A., Navarro Comas, M., Peters, E., Piter, A. J. M., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vrekoussis, M., Wagner, T., Wittrock, F., and Yilmaz, S.: MAX-DOAS formaldehyde slant column measurements during CINDI: intercomparison and analysis improvement, *Atmos. Meas. Tech.*, 6, 167–185, doi:<http://dx.doi.org/10.5194/amt-6-167-2013>, 2013.
- Platt, U. and Stutz, J.: *Differential Optical Absorption Spectroscopy*, Springer-Verlag, Berlin, Heidelberg, 2008.
- Richter, A., Weber, M., Burrows, J., Lambert, J.-C., and van Gijsel, A.: Validation strategy for satellite observations of tropospheric reactive gases, *Ann. Geophys.-Italy*, 56, doi:<http://dx.doi.org/10.4401/ag-633510.4401/ag-6335>, 2013.
- Rodgers, C. D.: *Inverse Methods For Atmospheric Sounding: Theory and Practice*, Ser. Atmos. Oceanic Planet. Phys., World Scientific Publishing, 2000.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, 4116, doi:<http://dx.doi.org/10.1029/2002JD002299>, 2003.
- Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Friess, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Piter, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO<sub>2</sub> and O<sub>4</sub> by MAX-DOAS and zenith-sky UV and visible spectrometers, *Atmos. Meas. Tech.*, 3, 1629–1646, doi:<http://dx.doi.org/10.5194/amt-3-1629-2010>, 2010.
- Sinreich, R., Merten, A., Molina, L., and Volkamer, R.: Parameterizing radiative transfer to convert MAX-DOAS dSCDs into near-surface box-averaged mixing ratios, *Atmos. Meas. Tech.*, 6, 1521–1532, doi:<http://dx.doi.org/10.5194/amt-6-1521-2013>, 2013.
- Spurr, R.: LIDORT and VLIDORT: linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: *Light Scattering Reviews*, 3, edited by: Kokhanovsky, A., 2008.
- Stammes, P., de Haan, J. F., and Hovenier, J. W.: The polarized internal radiation field of a planetary atmosphere, *Astron. Astrophys.*, 225, 239–259, 1989.



- Vlemmix, T., Pipers, A. J. M., Berkhout, A. J. C., Gast, L. F. L., Wang, P., and Levelt, P. F.: Ability of the MAX-DOAS method to derive profile information for NO<sub>2</sub>: can the boundary layer and free troposphere be separated?, *Atmos. Meas. Tech.*, 4, 2659–2684, doi:<http://dx.doi.org/10.5194/amt-4-2659-2011>, 2011.
- 5 Wagner, T., Burrows, J. P., Deutschmann, T., Dix, B., von Friedeburg, C., Frieß, U., Hendrick, F., Heue, K.-P., Irie, H., Iwabuchi, H., Kanaya, Y., Keller, J., McLinden, C. A., Oetjen, H., Palazzi, E., Petritoli, A., Platt, U., Postyyakov, O., Pukite, J., Richter, A., van Roozendael, M., Rozanov, A., Rozanov, V., Sinreich, R., Sanghavi, S., and Wittrock, F.: Comparison of box-air-mass-factors and radiances for Multiple-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) geometries calculated from different UV/visible radiative transfer models, *Atmos. Chem. Phys.*, 7, 1809–1833, doi:<http://dx.doi.org/10.5194/acp-7-1809-2007>, 2007.
- 10 Wagner, T., Dix, B., von Friedeburg, C., Frieß, U., Sanghavi, S., Sinreich, R., and Platt, U.: MAX-DOAS O<sub>4</sub> measurements: a new technique to derive information on atmospheric aerosols – principles and information content, *J. Geophys. Res.*, 109, D22205, doi:<http://dx.doi.org/10.1029/2004JD004904>, 2004.
- 15 Wagner, T., Deutschmann, T., and Platt, U.: Determination of aerosol properties from MAX-DOAS observations of the Ring effect, *Atmos. Meas. Tech.*, 2, 495–512, doi:<http://dx.doi.org/10.5194/amt-2-495-2009>, 2009.
- 20 Wagner, T., Beirle, S., Brauers, T., Deutschmann, T., Frieß, U., Hak, C., Halla, J. D., Heue, K. P., Junkermann, W., Li, X., Platt, U., and Pundt-Gruber, I.: Inversion of tropospheric profiles of aerosol extinction and HCHO and NO<sub>2</sub> mixing ratios from MAX-DOAS observations in Milano during the summer of 2003 and comparison with independent data sets, *Atmos. Meas. Tech.*, 4, 2685–2715, doi:<http://dx.doi.org/10.5194/amt-4-2685-2011>, 2011.
- 25 Wagner, T., Apituley, A., Beirle, S., Dörner, S., Friess, U., Remmers, J., and Shaiganfar, R.: Cloud detection and classification based on MAX-DOAS observations, *Atmos. Meas. Tech.*, 7, 1289–1320, doi:<http://dx.doi.org/10.5194/amt-7-1289-2014>, 2014.
- Wang, T., Hendrick, F., Wang, P., Tang, G., Clémer, K., Yu, H., Fayt, C., Hermans, C., Gielen, C., Pinardi, G., Theys, N., Brenot, H., and Van Roozendael, M.: Evaluation of tropospheric SO<sub>2</sub> retrieved from MAX-DOAS measurements in Xianghe, China, *Atmos. Chem. Phys. Discuss.*, 14, 6501–6536, doi:<http://dx.doi.org/10.5194/acpd-14-6501-2014>, 2014.
- 30 Wittrock, F., Oetjen, H., Richter, A., Fietkau, S., Medeke, T., Rozanov, A., and Burrows, J. P.: MAX-DOAS measurements of atmospheric trace gases in Ny-Ålesund – Radiative transfer stud-

ies and their application, *Atmos. Chem. Phys.*, 4, 955–966, doi:<http://dx.doi.org/10.5194/acp-4-955-2004>, 2004.

**Table 1.** DOAS settings used for the UV and Vis. For more details, see Pinardi et al. (2013) for the UV and Hendrick et al. (2014) for the Vis.  
table

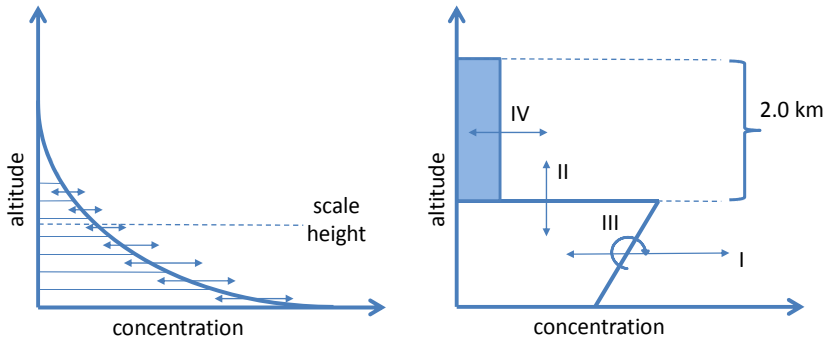
|                       | UV  | Vis  |
|-----------------------|---|--|
| wavelength range (nm) | 336.5–359   | 425–490  |
| cross sections        | HCHO, O <sub>4</sub> , O <sub>3</sub> , NO <sub>2</sub> , BrO, Ring | NO <sub>2</sub> , O <sub>4</sub> , O <sub>3</sub> , H <sub>2</sub> O, Ring |
| polynomial            | 3rd order   | 3rd order  |

**Table 2.** Retrieval with method B is performed for different combinations of free parameters which describe the profile shape. See also Fig. 1.

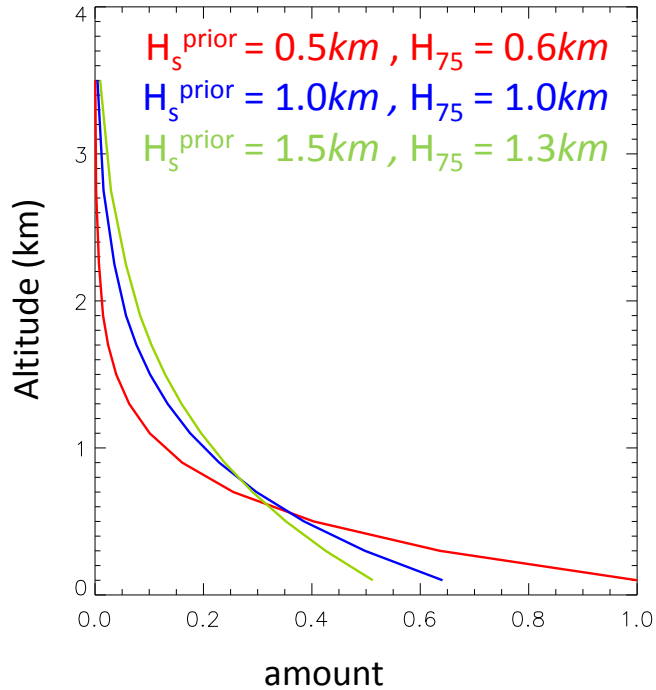
| Profile<br>Parameterization | Free Parameters<br>included |
|-----------------------------|-----------------------------|
| B1                          | I, II                       |
| B2                          | I, II, III                  |
| B3                          | I, II, IV                   |
| B4                          | I, II, III, IV              |

**Table 3.** Statistical comparison of methods A and B. The last two columns refer to median, mean and standard deviation (std.) of percentual relative differences (RD). The linear fit results are defined for method A on the abscissa. Relative differences have a sign determined by B–A. Note that the intercepts have unit km for  $H_{75}$  and  $H_s^{\text{post}}$ , unit  $10^{15}\text{molec cm}^{-2}$  for the tropospheric column densities, unit ppb for the volume mixing ratio (vmr) near the surface, and unit  $\text{km}^{-1}$  for the aerosol extinction near the surface.

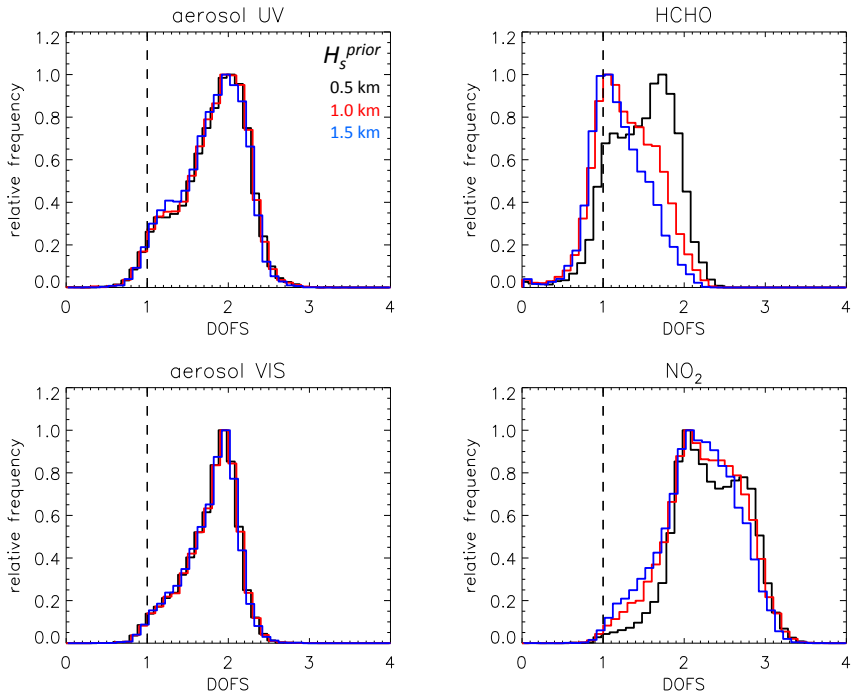
| quantity                | $N$  | corr. | slope | interc. | median (mean) of RD (%) | std. of RD (%) |
|-------------------------|------|-------|-------|---------|-------------------------|----------------|
| UV, aerosols            |      |       |       |         |                         |                |
| AOD                     | 2734 | 0.92  | 1.20  | −0.02   | 17.16 (22.81)           | 24.92          |
| $H_{75}$                | 2723 | 0.77  | 4.08  | −2.42   | 7.55 (6.69)             | 47.17          |
| $H_s^{\text{post}}$     | 2735 | 0.88  | 1.50  | −0.37   | 10.92 (10.25)           | 25.91          |
| extinction near surface | 2734 | 0.93  | 1.12  | 0.01    | 11.44 (14.93)           | 20.07          |
| UV, HCHO                |      |       |       |         |                         |                |
| trop. column density    | 2509 | 0.95  | 1.02  | 0.67    | 9.64 (9.43)             | 12.14          |
| $H_{75}$                | 2498 | 0.67  | 7.47  | −5.33   | 22.34 (15.65)           | 36.93          |
| $H_s^{\text{post}}$     | 2498 | 0.77  | 3.01  | −1.47   | 41.96 (37.87)           | 35.57          |
| vmr near surface        | 2504 | 0.80  | 0.95  | −0.32   | −23.03 (−20.70)         | 27.71          |
| Vis, aerosols           |      |       |       |         |                         |                |
| AOD                     | 4001 | 0.91  | 1.39  | −0.05   | 18.26 (23.07)           | 23.54          |
| $H_{75}$                | 3936 | 0.62  | 2.82  | −1.31   | 5.67 (2.74)             | 43.63          |
| $H_s^{\text{post}}$     | 3821 | 0.63  | 1.20  | 0.11    | 12.65 (14.40)           | 36.64          |
| extinction near surface | 3907 | 0.93  | 1.33  | −0.08   | 11.24 (10.27)           | 30.43          |
| Vis, NO <sub>2</sub>    |      |       |       |         |                         |                |
| trop. column density    | 3360 | 0.99  | 1.03  | −0.56   | 0.51 (0.25)             | 9.26           |
| $H_{75}$                | 3298 | 0.76  | 1.44  | −0.20   | 8.71 (6.51)             | 33.18          |
| $H_s^{\text{post}}$     | 3309 | 0.80  | 1.79  | −0.35   | 18.71 (17.30)           | 38.03          |
| vmr near surface        | 3313 | 0.76  | 1.16  | −2.25   | −11.82 (−7.92)          | 32.97          |



**Fig. 1.** Schematic of profile parametrizations for methods A (left) and B (right). Method A uses 13 layers (not drawn) between 0 and 4 km. The number of free variables for method B varies, see Table 2. figure

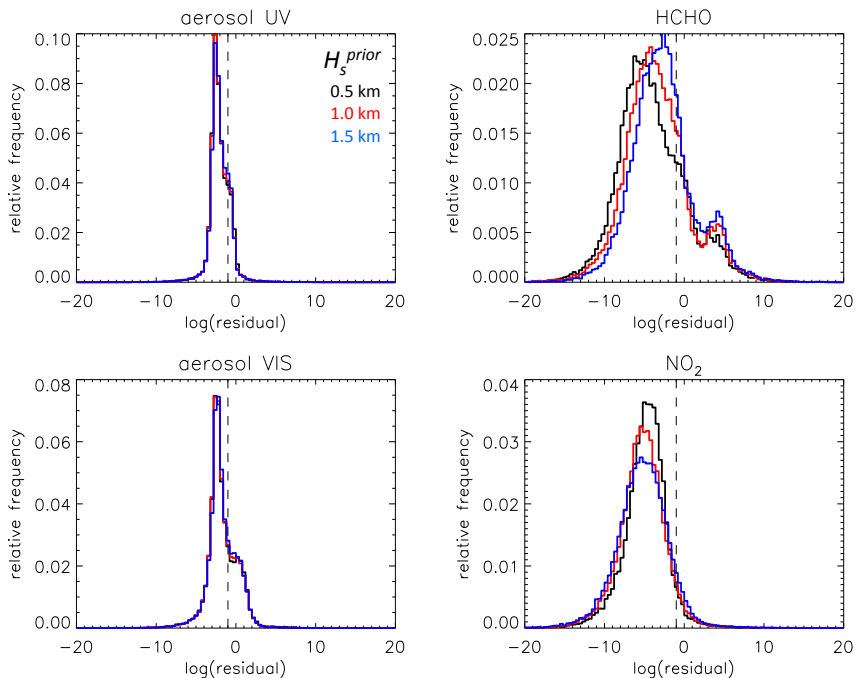


**Fig. 2.** Method A is run with three different a priori profile shapes (see Eq. 2), each with a different characteristic profile height ( $H_{75}$ ).

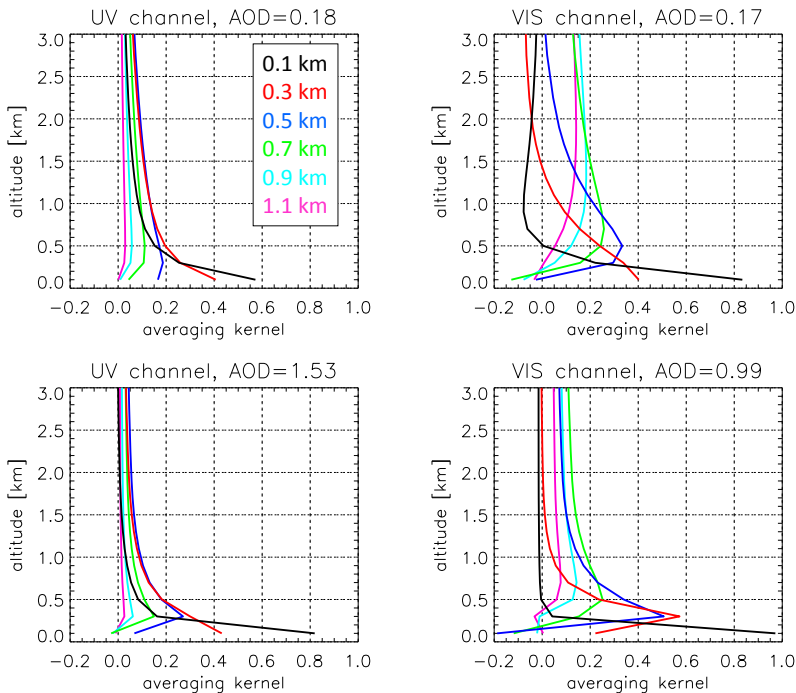


**Fig. 3.** Histograms showing the degrees of freedom for signal (DOFS) for the profile fits obtained with methods A1–A3, based on all MAX-DOAS scans analyzed for the Xianghe station. The upper row shows results for the UV, the bottom row shows results for the Vis. The dashed line indicates the threshold that is used for the quality control: retrievals with  $\text{DOFS} < 1$  are excluded from the comparison.

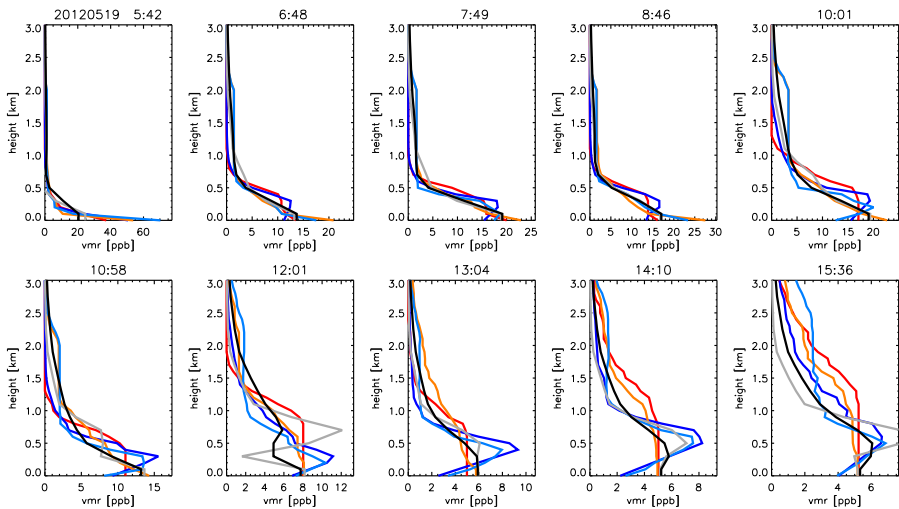




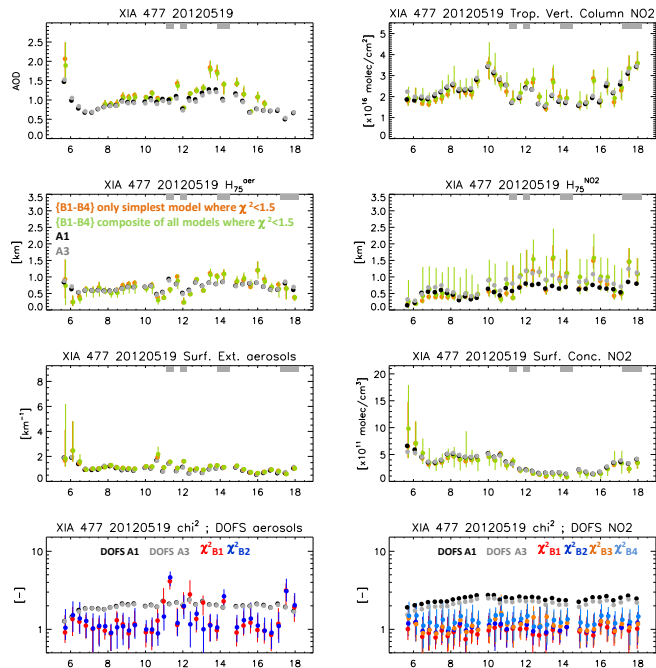
**Fig. 4.** Histograms of residuals of the profile fits obtained with methods A1–A3, based on all MAX-DOAS scans analyzed for the Xianghe station. The upper row shows results for the UV, the bottom row shows results for the Vis. The dashed line indicates the threshold (0.1, or  $-1$  on a logarithmic scale) that is used for the quality control: retrievals with residuals above this threshold are excluded from the comparison.



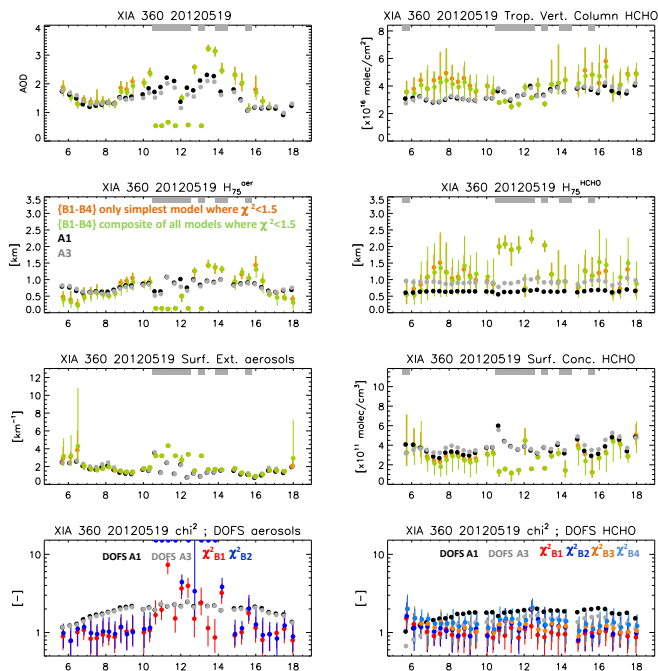
**Fig. 5.** Examples of averaging kernels for retrievals performed with method A. Both for the UV (left column) and Vis (right column). The upper row shows averaging kernels for low AOD (representative for the winter season), the bottom row for high AOD (representative for the summer season).



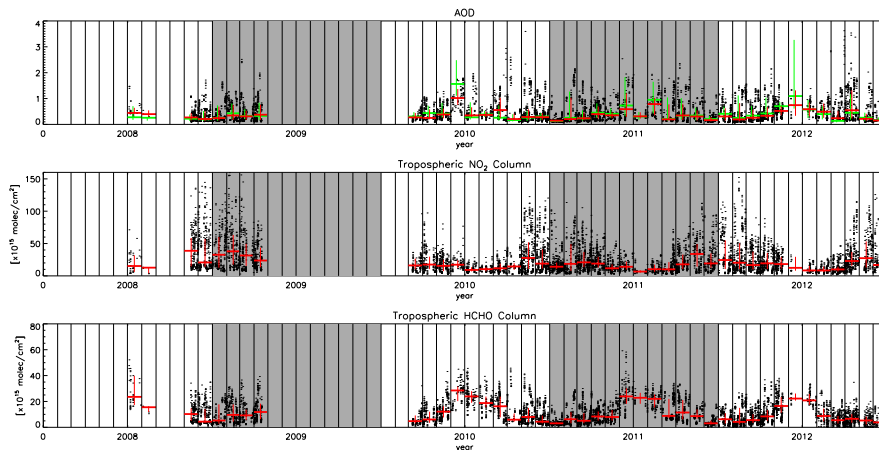
**Fig. 6.** Examples of  $\text{NO}_2$  profiles retrieved with method A (gray ( $H_s = 0.5$  km) and black ( $H_s = 1.5$  km)) and method B (parametrizations B1–B4 (see Fig. 1 and Table 2) shown in red, blue, orange, light blue respectively).



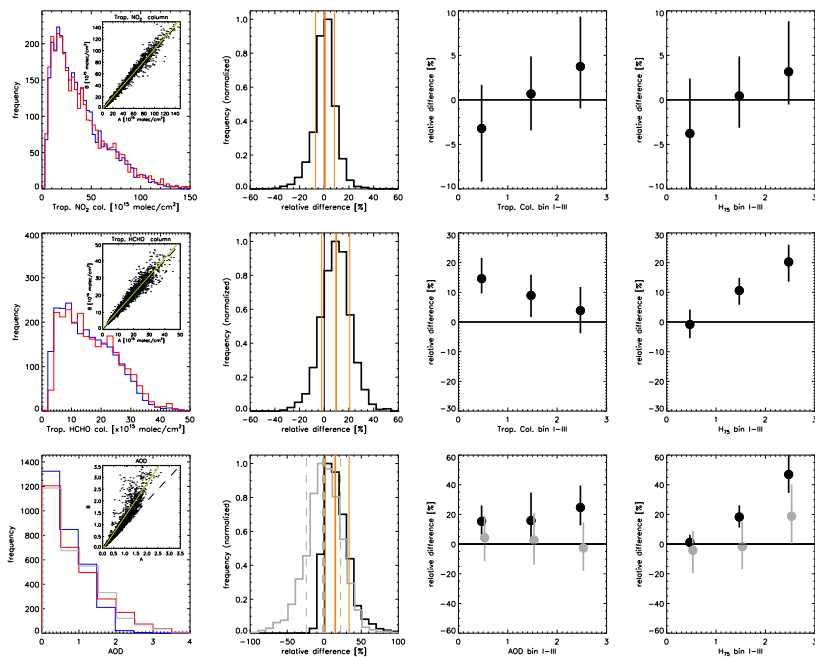
**Fig. 7.** Example of quantities derived from the profile retrieval in the Vis (Xianghe, 19 May 2012). The left column of the figure shows the results of the aerosol retrieval, the right column of the figure shows results for the NO<sub>2</sub> retrieval. Grey horizontal bars above each figure indicate periods that are flagged because of high values of  $\chi^2$  in the aerosol or in the trace gas retrieval.



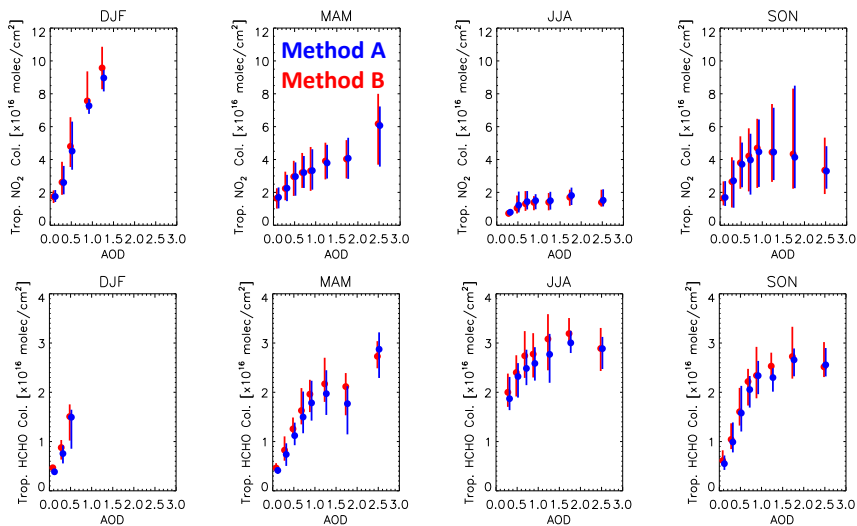
**Fig. 8.** Example of quantities derived from the profile retrieval in the UV (Xianghe, 19 May 2012). The left column of the figure shows the results of the aerosol retrieval, the right column of the figure shows results for the HCHO retrieval. Grey horizontal bars above each figure indicate periods that are flagged because of high values of  $\chi^2$  in the aerosol or in the trace gas retrieval.



**Fig. 9.** Time series of individual data points and monthly medians of AOD (Vis), and tropospheric column densities of  $\text{NO}_2$  and HCHO, obtained with method A. The black dots refer to individual profiles and the red lines refer to monthly medians. AOD from AERONET is shown in green (upper row). In 2008/09 the instrument was installed in Beijing, from 2010 onwards it was stationed in Xianghe.

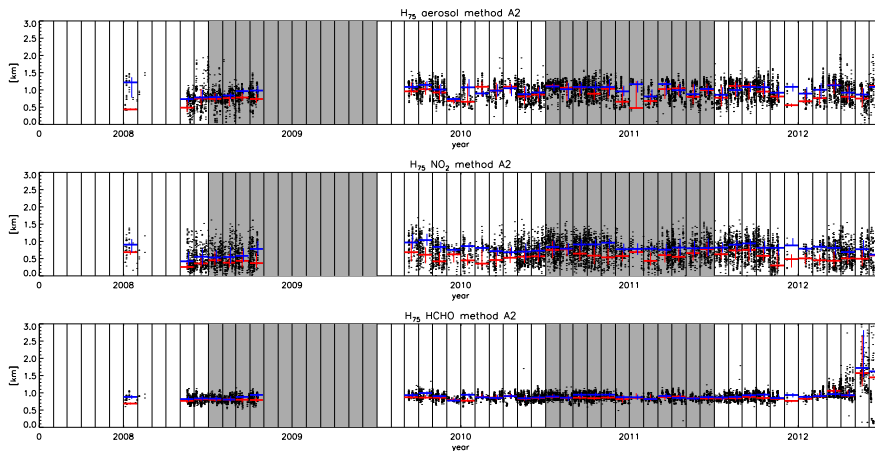


**Fig. 10.** Statistics of column density retrievals. The three rows refer to tropospheric column densities of  $\text{NO}_2$ , HCHO and aerosols respectively. Left column: frequency distributions obtained with method A (blue) and B (red). Second column: frequency distribution of relative differences (B minus A). Lines in orange indicate the quartiles. Column 3: relative difference sorted as a function of the tropospheric column density (AOD for row 3) where the three bins refer to the lowest 25 %, middle 50 % and highest 25 % respectively. Columns 4: similar to column 3, but here sorted as a function of the profile height ( $H_{75}$ ). The grey line on the bottom row refers to AOD's from AERONET. The relative differences indicated in grey (bottom row, columns 2–4) refer to method B minus AERONET.

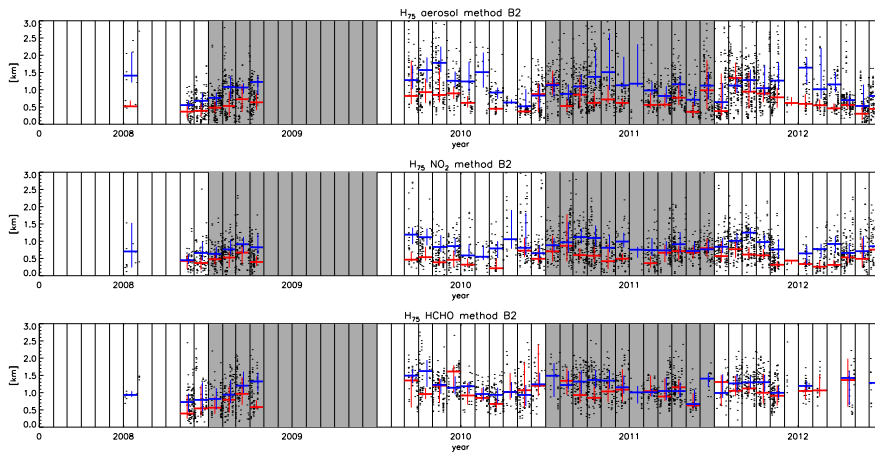


**Fig. 11.** Median values of the tropospheric  $\text{NO}_2$  column (upper row) and HCHO column (bottom row) for a range of AOD bins, specified per season.

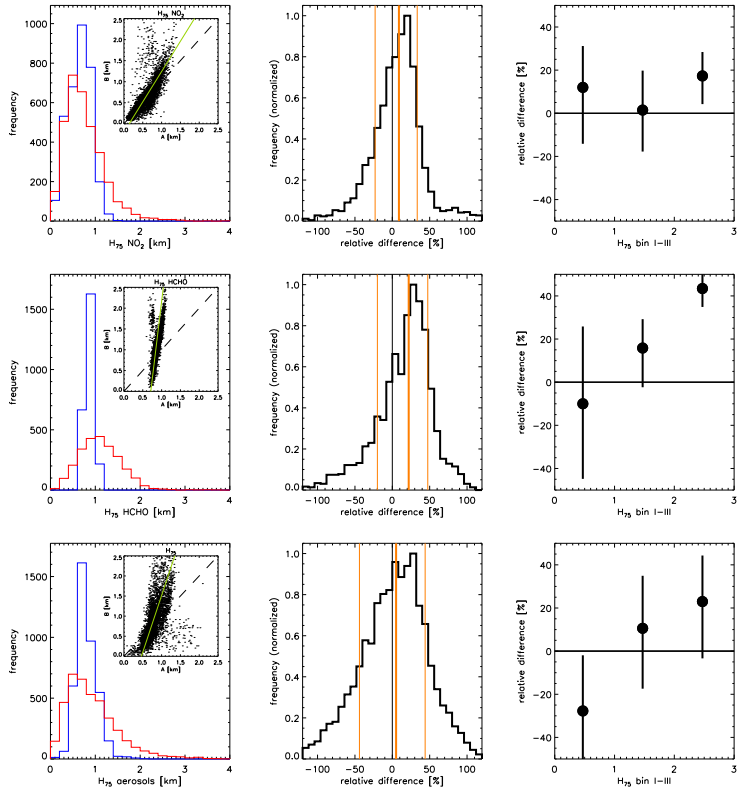




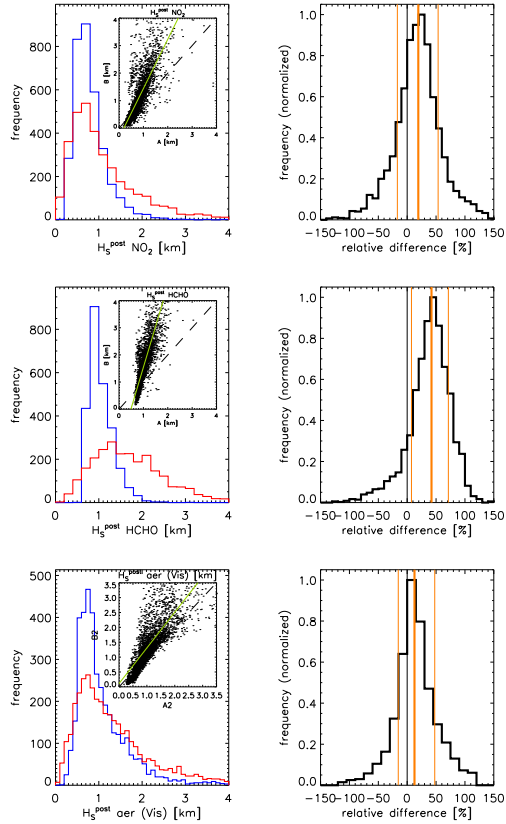
**Fig. 12.** Time series of individual data points and monthly medians of profile heights ( $H_{75}$ ) for aerosol (Vis), NO<sub>2</sub> and HCHO obtained with method A. The black dots refer to individual profiles. The red and blue lines refer to monthly medians for the morning (all observations before 10:00 a.m.) and afternoon respectively (all observations after 12:00 p.m.).



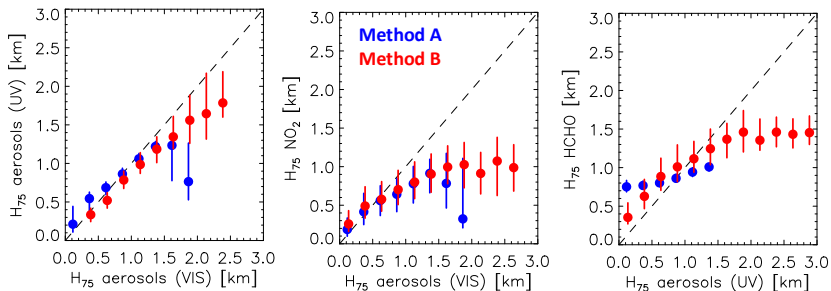
**Fig. 13.** Similar to Fig. 12, but now for method B.



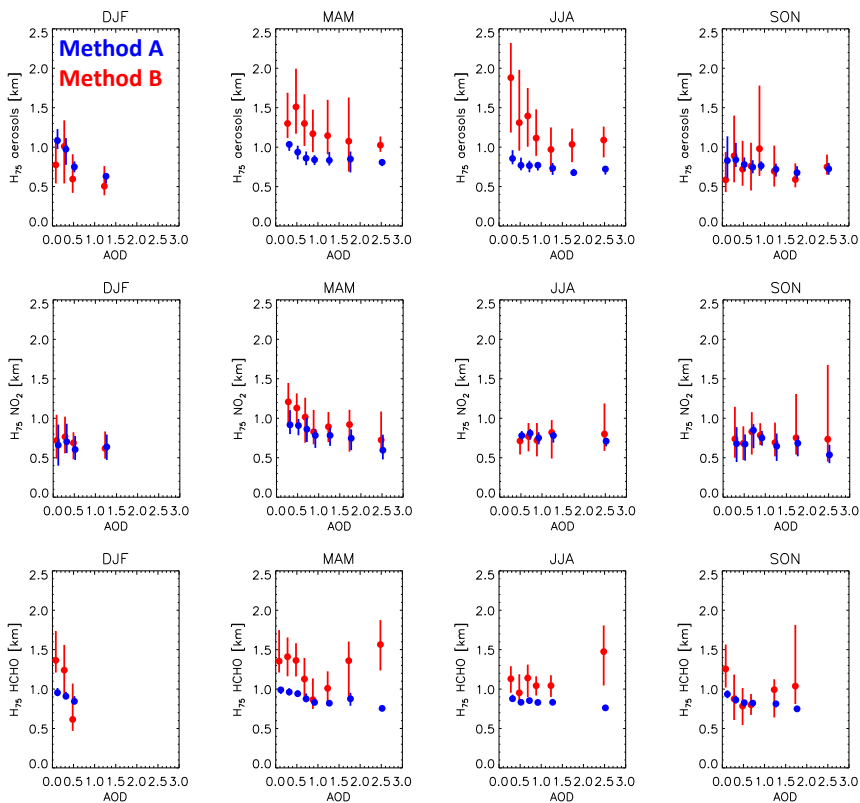
**Fig. 14.** Statistics of characteristic profile height retrievals. The three rows refer to profile heights ( $H_{75}$ ) of  $\text{NO}_2$ , HCHO and aerosols respectively. Left column: frequency distributions obtained with method A (blue) and B (red). Second column: frequency distribution of relative differences (B minus A). Lines in orange indicate the quartiles. Column 3: relative difference sorted as a function of  $H_{75}$  where the three bins refer to the lowest 25 %, middle 50 % and highest 25 % respectively.



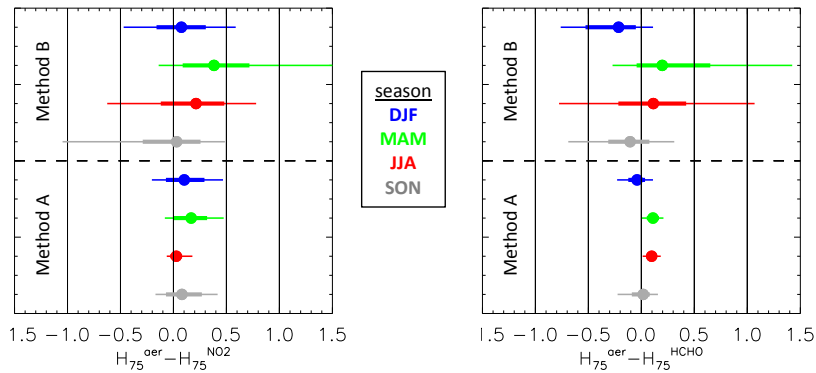
**Fig. 15.** Similar to Fig. 14, but here for the a posteriori scale height ( $H_s^{\text{post}}$ , Eqs. 6 and 7). The three rows refer  $\text{NO}_2$ , HCHO and aerosols respectively. Left column: frequency distributions obtained with method A (blue) and B (red). Second column: frequency distribution of relative differences (B minus A). Lines in orange indicate the quartiles.



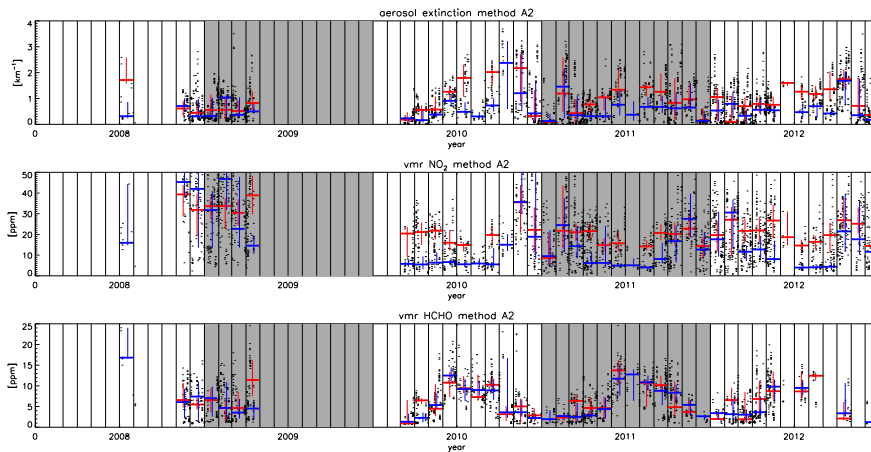
**Fig. 16.** Three binned scatterplots: aerosol profile heights retrieved in the Vis, vs. profile heights retrieved in the UV (left), aerosol profile height retrieved in the Vis vs. NO<sub>2</sub> profile heights (middle), aerosol profile heights retrieved in the UV vs. HCHO profile heights (right). Note that models A and B have very different frequency distributions of characteristic profile heights for the three constituents, see Fig. 14.



**Fig. 17.** Median values of the characteristic profile height ( $H_{75}$ ) for aerosols (upper row), NO<sub>2</sub> (middle row) and HCHO column (bottom row) for a range of AOD bins, specified per season.

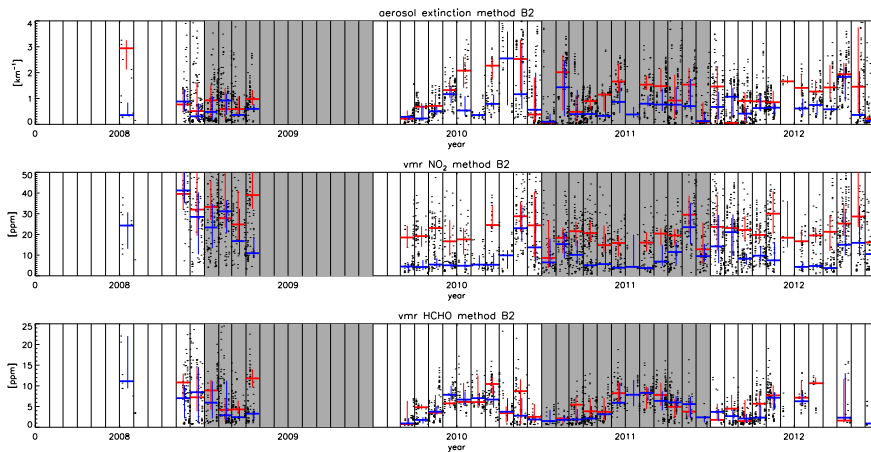


**Fig. 18.** Boxplots of difference between aerosol and trace gas profile heights (NO<sub>2</sub> left, HCHO right), specified per season and retrieval method.

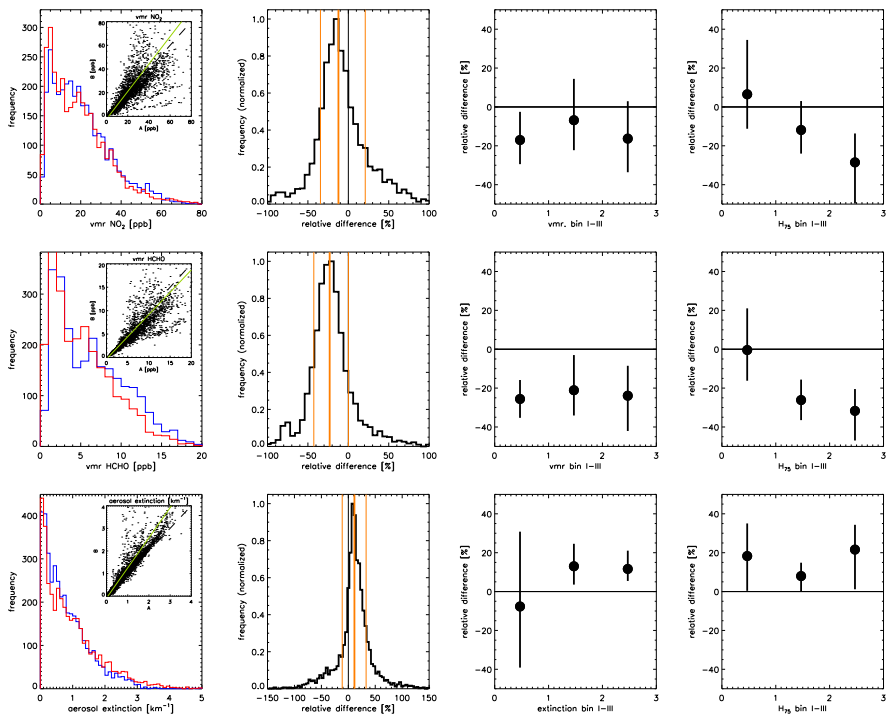


**Fig. 19.** Time series of individual data points and monthly medians of near-surface aerosol extinction (upper row) and volume mixing ratios for NO<sub>2</sub> and HCHO (bottom row), retrieved with method A. Red refers to morning observations, blue to afternoon observations.





**Fig. 20.** Time series of individual data points and monthly medians of near-surface aerosol extinction (upper row) and volume mixing ratios for NO<sub>2</sub> and HCHO (bottom row), retrieved with method B. Red refers to morning observations, blue to afternoon observations.



**Fig. 21.** Statistics of near-surface concentration retrievals. The three rows refer to volume mixing ratios of NO<sub>2</sub> (row 1) and HCHO (row 2) and aerosol extinction (row 3). Left column: frequency distributions obtained with method A (blue) and B (red). Second column: frequency distribution of relative differences (B minus A). Lines in orange indicate the quartiles. Column 3: relative differences sorted as a function of the volume mixing ratio (rows 1 and 2) and aerosol extinction (row 3). Column 4: relative difference sorted as a function of  $H_{75}$  where the three bins refer to the lowest 25 %, middle 50 % and highest 25 % respectively.