Article

A Proof-of-Concept Algorithm for the Retrieval of Total Column Amount of Trace Gases in a Multi-Dimensional Atmosphere

Adrian Doicu *, Dmitry S. Efremenko [ID] **and Thomas Trautmann**

German Aerospace Center (DLR), Remote Sensing Technology Institute, 82234 Oberpfaffenhofen, Germany;
dmitry.efremenko@dlr.de (D.S.E.); thomas.trautmann@dlr.de (T.T.)
* Correspondence: Adrian.Doicu@dlr.de

Abstract: An algorithm for the retrieval of total column amount of trace gases in a multi-dimensional atmosphere is designed. The algorithm uses (i) certain differential radiance models with internal and external closures as inversion models, (ii) the iteratively regularized Gauss–Newton method as a regularization tool, and (iii) the spherical harmonics discrete ordinate method (SHDOM) as linearized radiative transfer model. For efficiency reasons, SHDOM is equipped with a spectral acceleration approach that combines the correlated k-distribution method with the principal component analysis. The algorithm is used to retrieve the total column amount of nitrogen for two- and three-dimensional cloudy scenes. Although for three-dimensional geometries, the computational time is high, the main concepts of the algorithm are correct and the retrieval results are accurate.

Keywords: NO2 retrieval; SHDOM; multi-dimensional radiative trasnfer

1. Introduction

The retrieval of trace gas products from UV/VIS spectrometers is strongly affected by the presence of clouds. In general, there are three different cloud contributions: (i) the albedo effect, associated with the enhancement of reflectivity for cloudy scenes compared to cloud-free scenes; (ii) the shielding effect when a part of the trace gas column below the cloud is hidden; and (iii) the increase in absorption, related to multiple scattering inside clouds. The albedo and in-cloud absorption effects increase the detectability of trace gases at and above the cloud-top (in case of low clouds), while the shielding effect normally results in an underestimation of the trace gas column [1].

The majority of retrieval algorithms are based on one-dimensional radiative transfer modeling and use the so-called independent pixel approximation (IPA). The IPA for an atmosphere containing partial cloudiness means to compute separately the radiances for completely cloudy and clear skies, and then to express the partially cloudy radiance as a weighted sum of the separate radiances; the weighting factor being provided by the intensity weighted cloud fraction. The clouds within each pixel are assumed to be plane-parallel and homogeneous in both horizontal and vertical directions. The main advantage of the IPA is its computational efficiency, as it requires the solution of only two independent one-dimensional radiative transfer problems. The disadvantage is that for cloudy scenes of small horizontal extent, the errors due to the three-dimensional effects may be very significant [2–4]. This is the case of the new generation of atmospheric composition UV–VIS–NIR sensors, such as Sentinel-5P, Sentinel-4, and Sentinel-5. In [5] the importance of three-dimensional effects on computations of air mass factors has been shown.

To analyze the impact of cloud inhomogeneity on the retrieval, multi-dimensional radiative transfer models are used. In particular, for a specific cloudy scene, a spectral signal is simulated by a multi-dimensional radiative transfer model and then included in a one-dimensional retrieval algorithm to derive for example, the total column amount of a trace gas. By comparing the retrieved total column with the true value, a bias due to cloud effects can be deduced, and possible correction strategies can be explored.

In this paper, we go one step further and present the main concepts of an algorithm for retrieving the total column amount of trace gases in a multi-dimensional atmosphere. To achieve this goal we combine inverse models and regularization tools from inversion theory with a fast linearized version of the spherical harmonic discrete ordinate method [6]. These theoretical results are summarized in Section 2, while in Section 3, the accuracy and efficiency of the algorithm, used to retrieve the total column amount of nitrogen dioxide (NO$_2$) for two- and three-dimensional cloudy scenes, are numerically investigated. We conclude our analysis with some recommendations for operational retrievals.

2. Retrieval Algorithm

For a multi-dimensional geometry, the spectral signal of an instrument that measures the radiance at the top of the atmosphere in direction $\mathbf{\Omega}_m$ at wavelength λ in the spectral interval $[\lambda_{min}, \lambda_{max}]$, is given by

$$\mathcal{I}(\lambda) = \int_{\lambda_{min}-s/2}^{\lambda_{max}+s/2} g(\lambda - \lambda') \mathcal{I}_{FOV}(\lambda') \, d\lambda', \tag{1}$$

where, under the assumption that the distance from the top of the atmosphere to the instrument is large,

$$\mathcal{I}_{FOV}(\lambda') = \int_{S_t} h(\mathbf{r}_t) I(\lambda', \mathbf{r}_t, \mathbf{\Omega}_m) \, dS_t, \tag{2}$$

is the signal integrated over the field of view (FOV) of the instrument at wavelength λ' (see, e.g., in [7] for a general description), $g(\lambda)$ the slit function (defined on all $\mathbb{R}$), s the spectral bandwidth, $h(\mathbf{r}_t)$ the characteristic function that takes the value $1/A_{tm}$ for $\mathbf{r}_t \in S_{tm}$ and 0 otherwise, S_{tm} the footprint of the instrument on the top face S_t, and A_{tm} the area of the instrument footprint. Essentially, Equations (1) and (2) state that the mean radiance across the wavelength and the area S_t at the top of the atmosphere is measured. The top of atmosphere radiance at the exiting point $\mathbf{r}_t$ along the characteristic $\mathbf{\Omega}_m$, $I(\lambda, \mathbf{r}_t, \mathbf{\Omega}_m)$, is computed by using the integral form of the radiative transfer equation

$$\begin{aligned} I(\lambda, \mathbf{r}_t, \mathbf{\Omega}_m) = &\, I(\lambda, \mathbf{r}_b, \mathbf{\Omega}_m) e^{-\int_0^{s_0} \sigma_{ext}(\lambda, \mathbf{r}_b + s\mathbf{\Omega}_m) ds} \\ &+ \int_0^{s_0} \sigma_{ext}(\lambda, \mathbf{r}_b + s\mathbf{\Omega}_m) J(\lambda, \mathbf{r}_b + s\mathbf{\Omega}_m, \mathbf{\Omega}_m) \\ &\times \left[e^{-\int_s^{s_0} \sigma_{ext}(\lambda, \mathbf{r}_b + s'\mathbf{\Omega}_m) ds'} \right] ds, \end{aligned} \tag{3}$$

where $I(\lambda, \mathbf{r}_b, \mathbf{\Omega}_m)$ is the radiance at the surface point $\mathbf{r}_b$, $J(\lambda, \mathbf{r}, \mathbf{\Omega}_m)$ the source function, and $s_0 = |\mathbf{r}_t - \mathbf{r}_b|$ the distance between the entering and exiting points, respectively. Equation (3) can be derived by integrating the radiative transfer equation [8,9]. The first term corresponds to the upwelling radiance at the bottom of the atmosphere and attenuated exponentially as it propagates through the atmosphere. The second term refers to the gain due to sources. The source function J contains only the solar term and the multiple scattering term [10].

In the case of an atmosphere consisting of gas molecules and a cloud, we assume that (i) the optical coefficients of the gas molecules depend on the altitude level and the wavelength, while (ii) the optical coefficients of the cloud depend on the spatial coordinates but not on the wavelength. The second assumption is justified by the fact that a narrow spectral interval is considered in the retrieval. According to the additivity property of optical cross sections [11], in the case of N_g gases, the extinction coefficient is computed as

$$\sigma_{ext}(\lambda, \mathbf{r}) = \sigma_{ext}^{cloud}(\mathbf{r}) + \sigma_{sct}^{mol}(\lambda, z) + \sum_{g=1}^{N_g} \sigma_{abs\,g}^{gas}(\lambda, z), \tag{4}$$

where $\sigma_{\text{ext}}^{\text{cloud}}(\mathbf{r})$ is the extinction coefficient in the cloud (methods for computing $\sigma_{\text{ext}}^{\text{cloud}}(\mathbf{r})$ are summarized in [12]), $\sigma_{\text{sct}}^{\text{mol}}(\lambda, z)$ the molecular scattering coefficient due to Rayleigh scattering [13], $\sigma_{\text{abs}g}^{\text{gas}}(\lambda, z)$ the absorption coefficient of gas g, and (x, y, z) the Cartesian coordinates of point $\mathbf{r}$. Considering a discretization of the atmosphere in N_z levels, i.e., $\{z_j\}_{j=1}^{N_z}$, the absorption coefficient on level z_j is given by

$$\sigma_{\text{abs}g}^{\text{gas}}(\lambda, z_j) = C_{\text{abs}g}(\lambda, z_j) n_g(z_j), \tag{5}$$

where $C_{\text{abs}g}(\lambda, z_j)$ and $n_g(z_j)$ are the absorption cross section and the number density of gas j on level z_j, respectively. Under the assumption that on a layer j bounded below by z_j and above by z_{j+1}, the number density $n_g(z)$ varies linearly with respect to z, the partial and total column of gas g are defined, respectively, by

$$x_{g,j} = \int_{z_j}^{z_{j+1}} n_g(z)\, \mathrm{d}z \approx \frac{\Delta z_j}{2}[n_g(z_j) + n_g(z_{j+1})] \tag{6}$$

and

$$X_g = \int_{z_1}^{z_{N_z}} n_g(z)\, \mathrm{d}z = \sum_{j=1}^{N_z-1} x_{g,j}, \tag{7}$$

where $\Delta z_j = z_{j+1} - z_j$.

The retrieval algorithm is based on the assumption that cloud information are available from co-located imagers, as, for example, the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the Terra and Aqua satellites and the Visible Infrared Imaging Radiometer Suite (VIIRS) on board the Suomi National Polar-Orbiting Partnership spacecraft. The main cloud products delivered by MODIS/VIIRS are the cloud mask (indicator function of the cloud field), cloud optical thickness, cloud-top height (cloud-top pressure), and effective radius of the size distribution. In the MODIS/VIIRS retrieval algorithm, the cloud geometrical thickness is derived under the assumption that the clouds are homogeneous in the vertical direction, or equivalently, that the liquid water content does not change with the altitude. Therefore, it is appropriate to use stochastic cloud models (in particular statistical cloud generators) which operate with the same assumption. For this reason, we will use stochastic cloud models relying on the vertical homogeneity assumption, and in particular, the two-dimensional broken cloud model of Alexandrov et al. [14].

In general, the design of a retrieval algorithm requires the specification of

1. an inverse model,
2. a solution method for solving the inverse problem, and
3. a linearized multi-dimensional radiative transfer model for computing the forward model and the Jacobian matrix at each iteration step.

2.1. Inverse Models

The retrieval of the total columns X_g of N_g gases can be performed by using two differential radiance models (see details in [15–17] and references therein):

1. the Differential Radiance Model with Internal closure (DRMI), in which the measured and simulated differential spectral signals are fitted, and
2. the Differential Radiance Model with External closure (DRME), in which the measured differential spectral signal and a simulated spectral signal with its smooth component extracted are fitted.

Actually, in DRMI we solve the nonlinear equation

$$R_{\text{mes}}^{\delta}(\lambda_k) = R_{\text{sim}}(\lambda_k, \mathbf{X}) + \sum_{j=1}^{N_s} b_j S_j(\lambda_k, \mathbf{X}_a), \quad k = 1, \dots, N_\lambda, \tag{8}$$

where

$$R^\delta_{\text{mes}}(\lambda_k) = \ln \mathcal{I}^\delta_{\text{mes}}(\lambda_k) - P_{\text{mes}}(\lambda_k, \mathbf{c}_{\text{mes}}) \tag{9}$$

and

$$R_{\text{sim}}(\lambda_k, \mathbf{X}) = \ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}) - P_{\text{sim}}(\lambda_k, \mathbf{c}_{\text{sim}}(\mathbf{X})) \tag{10}$$

are the measured and simulated differential spectral signals, respectively, while in DRME we solve the nonlinear equation

$$R^\delta_{\text{mes}}(\lambda_k) = \ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}) + \sum_{j=1}^{N_\text{s}} b_j S_j(\lambda_k, \mathbf{X}_\text{a}) - P(\lambda_k, \mathbf{c}), \quad k = 1, \dots, N_\lambda. \tag{11}$$

In Equations (8)–(11), $\mathbf{X} = [X_1, \dots, X_{N_g}]$ is the vector comprising the total columns of trace gases; $\mathbf{X}_\text{a}$ an a priori estimate of $\mathbf{X}$; $\mathcal{I}^\delta_{\text{mes}}(\lambda_k)$ the (noisy) spectral signal measured by the instrument at wavelength λ_k, where $k = 1, \dots N_\lambda$, N_λ the number of measurement wavelengths; $\mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X})$ the simulated spectral signal computed by a multi-dimensional radiative transfer model; $S_j(\lambda_k, \mathbf{X}_\text{a})$ with $j = 1, \dots, N_\text{s}$, the correction spectra including, for example, the polarization correction spectrum and the Ring spectrum, N_s the number of correction spectra; and b_j with $j = 1, \dots, N_\text{s}$, the (wavelength independent) amplitudes of the correction spectra. The polynomials $P_{\text{mes}}(\lambda, \mathbf{c}_{\text{mes}})$, $P_{\text{sim}}(\lambda, \mathbf{c}_{\text{sim}}(\mathbf{X}))$, and $P(\lambda, \mathbf{c})$ with degree $N_\text{p} - 1$ account for the low-order spectral structure due to the scattering by clouds and aerosols, while the differential spectral signals R_{sim} and R^δ_{mes} are mostly due to the gaseous absorption processes (see Ref. [18] for details). Specifically, in DRMI, the coefficients

$$\mathbf{c}_{\text{mes}} = [c_{\text{mes}0}, \dots, c_{\text{mes}N_\text{p}-1}] \text{ and } \mathbf{c}_{\text{sim}}(\mathbf{X}) = [c_{\text{sim}0}(\mathbf{X}), \dots, c_{\text{sim}N_\text{p}-1}(\mathbf{X})]$$

of the smoothing polynomials $P_{\text{mes}}(\lambda, \mathbf{c}_{\text{mes}})$ and $P_{\text{sim}}(\lambda, \mathbf{c}_{\text{sim}}(\mathbf{X}))$, respectively, are computed as the solutions of the least-squares problems

$$\mathbf{c}_{\text{mes}} = \arg\min_{\mathbf{c}} \sum_{k=1}^{N_\lambda} \left[\ln \mathcal{I}^\delta_{\text{mes}}(\lambda_k) - P_{\text{mes}}(\lambda_k, \mathbf{c})\right]^2 \tag{12}$$

and

$$\mathbf{c}_{\text{sim}}(\mathbf{X}) = \arg\min_{\mathbf{c}} \sum_{k=1}^{N_\lambda} \left[\ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}) - P_{\text{sim}}(\lambda_k, \mathbf{c})\right]^2, \tag{13}$$

respectively. Thus, these coefficients, which are uniquely determined by $\ln \mathcal{I}^\delta_{\text{mes}}(\lambda_k)$ and $\ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X})$, are not a part of the retrieval, while in DRMI, the coefficients $\mathbf{c} = [c_0, \dots, c_{N_\text{p}-1}]$ of the smoothing polynomial $P(\lambda, \mathbf{c})$ are included in the retrieval.

Similar inversion models can be used for tropospheric column retrieval. These models together with the underlying retrieval algorithms are described in Appendix A.

To speed up the calculation, two approximate inverse models can be designed.

1. In DRME, we assume the linearized model [19]

$$\ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}) = \ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}_\text{a}) + \sum_{g=1}^{N_g} (X_g - X_{\text{a}g}) \frac{\partial \ln \mathcal{I}_{\text{sim}}}{\partial X_g}(\lambda_k, \mathbf{X}_\text{a}) + \delta R_\text{p}(\lambda_k), \tag{14}$$

where the second term in the right-hand side of the equation describes the variations of $\ln \mathcal{I}_{\text{sim}}$ with respect to the total column amounts of trace gases X_g with $g = 1, \dots, N_g$, and the third term $\delta R_\text{p}(\lambda_k)$ comprises the contributions of all other atmospheric parameters. Approximating $\delta R_\text{p}(\lambda_k)$ by a smoothing polynomial, i.e., $\delta R_\text{p}(\lambda_k) \approx P_\text{p}(\lambda_k, \mathbf{c}_\text{p})$, and putting $P(\lambda_k, \mathbf{c}) - P_\text{p}(\lambda_k, \mathbf{c}_\text{p}) \rightarrow P(\lambda_k, \mathbf{c})$, we are led to the linear equation

$$R_{\text{mes}}^{\delta}(\lambda_k) = \ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}_a) + \sum_{g=1}^{N_g} (X_g - X_{ag}) \frac{\partial \ln \mathcal{I}_{\text{sim}}}{\partial X_g}(\lambda_k, \mathbf{X}_a)$$

$$+ \sum_{j=1}^{N_s} b_j S_j(\lambda_k, \mathbf{X}_a) - P(\lambda_k, \mathbf{c}), \quad k = 1, \ldots, N_\lambda, \tag{15}$$

which is equivalent with the Differential Optical Absorption Spectroscopy (DOAS) equation [18]. This equivalence is proved in Appendix B.

2. We adopt an approximate inverse model in which we compute the spectral signal by an one-dimensional radiative transfer model, and (similar to the Ring and polarization correction spectra) introduce a correction spectrum $S_{3D}(\lambda_k, \mathbf{X}_a)$ that accounts on three-dimensional effects. To define the correction spectrum, we take into account that, for example, in DRME, we can write

$$\ln \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}) - P(\lambda_k, \mathbf{c})$$
$$= \ln \mathcal{I}_{\text{sim}}^{1D}(\lambda_k, \mathbf{X}) + \ln[1 + \delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X})] - P(\lambda_k, \mathbf{c})$$
$$\approx \ln \mathcal{I}_{\text{sim}}^{1D}(\lambda_k, \mathbf{X}) + \delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}) - P(\lambda_k, \mathbf{c})$$
$$= \ln \mathcal{I}_{\text{sim}}^{1D}(\lambda_k, \mathbf{X}) + [\delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}) - P_{3D}(\lambda_k, \mathbf{c}_{3D}(\mathbf{X}))] - P_{1D}(\lambda_k, \mathbf{c}_{1D}), \tag{16}$$

where $\mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X})$ and $\mathcal{I}_{\text{sim}}^{1D}(\lambda_k, \mathbf{X})$, are the spectral signals computed by a three- and an one-dimensional radiative transfer model, respectively, and

$$\delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}) = \frac{\mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}) - \mathcal{I}_{\text{sim}}^{1D}(\lambda_k, \mathbf{X})}{\mathcal{I}_{\text{sim}}^{1D}(\lambda_k, \mathbf{X})}, \tag{17}$$

is a relative correction spectrum. In deriving Equation (16), we used the first-order Taylor approximation $\ln(1 + \delta \mathcal{I}_{\text{sim}}^{3D}) \approx \delta \mathcal{I}_{\text{sim}}^{3D}$ for $\delta \mathcal{I}_{\text{sim}}^{3D} \ll 1$, and the decomposition $P(\lambda_k, \mathbf{c}) = P_{3D}(\lambda_k, \mathbf{c}_{3D}(\mathbf{X}) + P_{1D}(\lambda_k, \mathbf{c}_{1D})$, in which, the coefficients $\mathbf{c}_{3D}(\mathbf{X})$ are uniquely determined by $\delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X})$ (see below), while the coefficients $\mathbf{c}_{1D}$ are unknown (as the coefficients $\mathbf{c}$ are). In view of Equation (16), we assume that in the retrieval, the differential spectral term

$$\delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}) - P_{3D}(\lambda_k, \mathbf{c}_{3D}(\mathbf{X}))$$

can be described by $b_{3D} S_{3D}(\lambda_k, \mathbf{X}_a)$, where

$$S_{3D}(\lambda_k, \mathbf{X}_a) = \delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}_a) - P_{3D}(\lambda_k, \mathbf{c}_{3D}(\mathbf{X}_a)) \tag{18}$$

is the (differential) correction spectrum due to three-dimensional effects, $P_{3D}(\lambda_k, \mathbf{c}_{3D}(\mathbf{X}_a))$ the associated smoothing polynomial, and b_{3D} the amplitude of the correction spectrum (to be included in the retrieval). From Equation (18), we see that the correction spectrum $S_{3D}(\lambda_k, \mathbf{X}_a)$ reproduces the differential structure of the relative correction spectrum $\delta \mathcal{I}_{\text{sim}}^{3D}(\lambda_k, \mathbf{X}_a)$.

2.2. Solution Method

The nonlinear Equations (8) and (11) can be written in compact form as

$$\mathbf{y}^{\delta} = \mathbf{F}(\mathbf{x}), \tag{19}$$

where $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is the forward model, $\mathbf{x} \in \mathbb{R}^n$ the state vector, and $\mathbf{y}^{\delta} \in \mathbb{R}^m$ the noisy data vector. In DRMI and DRME, the state vectors are $\mathbf{x} = [\mathbf{X}, \mathbf{b}]^T$ and $\mathbf{x} = [\mathbf{X}, \mathbf{b}, \mathbf{c}]^T$, respectively, where $\mathbf{b} = [b_1, \ldots, b_{N_s}]$ is the vector comprising the amplitudes of the correction spectra b_j, $j = 1, \ldots, N_s$. Furthermore, the number of data points is $m = N_\lambda$, and the

measured differential spectral signals affected by noise $R_{mes}^{\delta}(\lambda_k)$ for $k = 1, \ldots N_{\lambda}$ are the components of the noisy data vector $\mathbf{y}^{\delta}$.

Because the nonlinear Equation (19) is usually ill-posed, a regularization method, as, for example, the iteratively regularized Gauss–Newton method, is used to compute a solution with physical meaning [15,20,21]. In this approach, at the iteration step k, we consider a linearization of $\mathbf{F}(\mathbf{x})$ around the current iterate $\mathbf{x}_k^{\delta}$ and solve the linearized equation by means of Tikhonov regularization with the penalty term $\alpha_k ||\mathbf{L}(\mathbf{x} - \mathbf{x}_a)||^2$, where α_k is the regularization parameter at the iteration step k, $\mathbf{L}$ the regularization matrix, and $\mathbf{x}_a$ the a priori state vector, the best beforehand estimate of the solution. Thus, the new iterate minimizing the function

$$\mathcal{F}_{lk}(\mathbf{x}) = \left\| \mathbf{y}_k^{\delta} - \mathbf{K}_k(\mathbf{x} - \mathbf{x}_a) \right\|^2 + \alpha_k \left\| \mathbf{L}(\mathbf{x} - \mathbf{x}_a) \right\|^2, \tag{20}$$

is given by

$$\mathbf{x}_{k+1}^{\delta} = \mathbf{x}_a + \mathbf{K}_k^{\dagger} \mathbf{y}_k^{\delta}, \tag{21}$$

where

$$\mathbf{K}_k^{\dagger} = (\mathbf{K}_k^T \mathbf{K}_k + \alpha_k \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}_k^T \tag{22}$$

is the regularized generalized inverse, $\mathbf{K}_k = \mathbf{K}(\mathbf{x}_k^{\delta}) = \partial \mathbf{F}(\mathbf{x}_k^{\delta})/\partial \mathbf{x}$ the Jacobian matrix evaluated at $\mathbf{x}_k^{\delta}$, and

$$\mathbf{y}_k^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_k^{\delta}) + \mathbf{K}_k(\mathbf{x}_k^{\delta} - \mathbf{x}_a) \tag{23}$$

the linearized data vector at the iteration step k. The following peculiarities of the iteratively regularized Gauss–Newton method deserve to be mentioned [15].

1. In contrast to the method of Tikhonov regularization [22], the regularization parameter is not constant during the iterative process. Instead, the regularization parameters α_k are the terms of a decreasing (geometric) sequence, i.e., $\alpha_k = q\alpha_{k-1}$ with $q < 1$. In this way, the amount of regularization is gradually decreased during the iterative process.

2. For iterative regularization methods, the number of iteration steps k plays the role of the regularization parameter, and the iterative process is stopped after an appropriate number of steps $k^{\star}$ in order to avoid an uncontrolled expansion of the errors in the data. The stopping rule used in this study is the discrepancy principle [23], according to which, the iterative process is terminated after $k^{\star}$ steps such that

$$\left\| \mathbf{r}_{k^{\star}}^{\delta} \right\| \leq \tau \Delta < \left\| \mathbf{r}_k^{\delta} \right\|, \ 0 \leq k < k^{\star}, \tag{24}$$

where $\mathbf{r}_k^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_k^{\delta})$ is the residual vector at $\mathbf{x}_k^{\delta}$, $\tau > 1$ a control parameter, and Δ the noise level (an upper bound for the noise in the data). Because in practice the noise level cannot be a priori estimated, we adopt a practical approach based on the observation that the residual $||\mathbf{r}_k^{\delta}||$ decreases during the iterative process and attains a plateau at approximately Δ. Thus, if the nonlinear residuals $||\mathbf{r}_k^{\delta}||$ converge to $||\mathbf{r}_{\infty}^{\delta}||$ within a prescribed tolerance, we use the estimate $\Delta = ||\mathbf{r}_{\infty}^{\delta}||$.

3. The regularization matrix is chosen as a diagonal matrix, that is, the penalty term $\alpha_k ||\mathbf{L}(\mathbf{x} - \mathbf{x}_a)||^2$ is taken as

$$\alpha_k \left[\sum_{g=1}^{N_g} \frac{w_g}{X_{ag}} (X_g - X_{ag})^2 + \sum_{j=1}^{N_s} \frac{w_{bj}}{b_{aj}} (b_j - b_{aj})^2 \right] \tag{25}$$

for DRMI and

$$\alpha_k \left[\sum_{g=1}^{N_g} \frac{w_g}{X_{ag}} (X_g - X_{ag})^2 + \sum_{j=1}^{N_s} \frac{w_{bj}}{b_{aj}} (b_j - b_{aj})^2 + \sum_{p=0}^{N_p-1} \frac{w_{cp}}{c_{ap}} (c_p - c_{ap})^2 \right] \tag{26}$$

for DRME. Here, the scalars $\{w_g\}_{g=1}^{N_g}$, $\{w_{bj}\}_{j=1}^{N_s}$, and $\{w_{cp}\}_{p=0}^{N_p-1}$ give the weight of each component of the state vector into the regularization matrix.

As compared to the method of Tikhonov regularization, the iteratively regularized Gauss–Newton method is less sensitive to overestimations of the regularization parameter, but requires more iteration steps to achieve convergence.

We conclude this section by mentioning that in the case of DRME, one iteration step of the iteratively regularized Gauss–Newton method corresponds to the linear problem described by Equations (11) and (14).

2.3. Linearized Radiative Transfer Model

In order to solve the nonlinear Equation (19) by means of the iteratively Gauss–Newton method, the forward model $\mathbf{F}(\mathbf{x})$ and the Jacobian matrix $\mathbf{K}(\mathbf{x})$ have to be calculated at each iteration step. From Equation (3) we infer that the computation of $\mathbf{F}(\mathbf{x})$ requires the computation of the radiance $I(\lambda, \mathbf{r}_b, \mathbf{\Omega})$ and the source function $J(\lambda, \mathbf{r}, \mathbf{\Omega})$ at all surface and grid points, respectively, while the computation of $\mathbf{K}(\mathbf{x})$ requires the computation of the partial derivatives of the extinction field $\partial \sigma_{\text{ext}}(\cdot, \mathbf{r})/\partial X_g$, the surface radiance $\partial I(\cdot, \mathbf{r}_b, \cdot)/\partial X_g$, and the source function $\partial J(\cdot, \mathbf{r}, \cdot)/\partial X_g$.

For derivative calculations, we employ the same assumption as in the standard DOAS model, that is, the (discrete) number density profile $\{n_{g,j}\}_{j=1}^{N_z}$, where $n_{g,j} = n_g(z_j)$, is assumed to be a scaled version of an a priori profile $\{n_{ag,j}\}_{j=1}^{N_z}$. Consequently, if s_g is the scale factor of gas g, i.e., $n_{g,j} = s_g n_{ag,j}$, for all $j = 1, \ldots, N_z$, the following equalities,

$$s_g = \frac{n_{g,j}}{n_{ag,j}} = \frac{\sigma_{\text{abs}g,j}^{\text{gas}}(\lambda)}{\sigma_{\text{abs}ag,j}^{\text{gas}}(\lambda)} = \frac{x_{g,j}}{x_{ag,j}} = \frac{X_g}{X_{ag}} \tag{27}$$

with $\sigma_{\text{abs}g,j}^{\text{gas}}(\lambda) = \sigma_{\text{abs}g}^{\text{gas}}(\lambda, z_j)$, are valid. Thus, for any value of the total column X_g, the absorption coefficient is updated as

$$\sigma_{\text{abs}g,j}^{\text{gas}}(\lambda) = \frac{\sigma_{\text{abs}ag,j}^{\text{gas}}(\lambda)}{X_{ag}} X_g, \tag{28}$$

and the corresponding top of atmosphere radiance $I(\lambda, \mathbf{r}_t, \mathbf{\Omega}_m)$ is determined by means of a radiative transfer calculation. In general, for computing the derivative of a spectral quantity $F(\lambda)$ with respect to the total column X_g, we regard $F(\lambda)$ as a function of $\sigma_{\text{abs}g,j}^{\text{gas}}(\lambda)$, i.e., $F = F(\sigma_{\text{abs}g,1}^{\text{gas}}, \ldots, \sigma_{\text{abs}g,N_z}^{\text{gas}})$, and assume that the partial derivatives $\partial F/\partial \sigma_{\text{abs}g,j}^{\text{gas}}$, $j = 1, \ldots, N_z$, are known. Taking into account the one-to-one correspondence (28), we are led to the computational formula

$$\frac{\partial F}{\partial X_g}(\lambda) = \sum_{j=1}^{N_z} \frac{\partial F}{\partial \sigma_{\text{abs}g,j}^{\text{gas}}}(\lambda) \frac{\partial \sigma_{\text{abs}g,j}^{\text{gas}}}{\partial X_g}(\lambda) = \sum_{j=1}^{N_z} \frac{\sigma_{\text{abs}ag,j}^{\text{gas}}(\lambda)}{X_{ag}} \frac{\partial F}{\partial \sigma_{\text{abs}g,j}^{\text{gas}}}(\lambda). \tag{29}$$

Thus, the radiative transfer model should be linearized with respect to the absorption coefficient, that is, the linearized radiative transfer model should deliver the partial derivatives $\partial F/\partial \sigma_{\text{abs}g,j}^{\text{gas}}$ for $j = 1, \ldots, N_z$ and $g = 1, \ldots, N_g$.

The multi-dimensional radiative transfer model used in our retrieval algorithm is the spherical harmonic discrete ordinate method (SHDOM) [6,24]. Linearizations of SHDOM by means of a forward and a forward-adjoint approach have been discussed in [25]. These can be summarized as follows.

1. The linearized forward approach relies on an analytical computation of the derivatives. The method is accurate and has the advantage that no assumptions rather than those of the forward model have to be made. However, the method is time consuming and memory demanding when the number of parameters to be retrieved is large. The

reason is that not only the source function has to be stored as a spherical harmonic series at each grid point, but also its derivatives with respect to the atmospheric parameters of interest.

2. The linearized forward-adjoint approach relies on the application of the adjoint radiative transfer theory. The method requires less storage for derivatives calculation, is much faster, but relatively less accurate. The main reason for this lower accuracy is that different interpolation schemes are used for radiance and derivative calculations.

Because in our application the number of retrieved quantities is relatively small, we decided in the favor of the linearized forward approach. Specifically, the partial derivatives are computed in an analytical manner by using the chain rule in a sort of backward procedure starting with the output radiance and following the chain of dependencies back to the inputs to the model.

In [26], a simplified approach, leading to a considerable reduction of the computation time, was proposed. The idea is to neglect in Equation (3) the contributions of the partial derivatives of the surface radiance $\partial I(\cdot, \mathbf{r}_\mathrm{b}, \cdot)/\partial X_g$ and the source function $\partial J(\cdot, \mathbf{r}, \cdot)/\partial X_g$. In other words, when computing the partial derivative of the top of atmosphere radiance $\partial I(\cdot, \mathbf{r}_\mathrm{t}, \cdot)/\partial X_g$, only the contributions of the partial derivatives of the extinction field $\partial \sigma_\mathrm{ext}(\cdot, \mathbf{r})/\partial X_g$ are taken into account.

To speed up the computations, the linearized SHDOM is equipped with a spectral acceleration approach that combines the correlated k-distribution method with dimensionality reduction techniques. This approach was applied for computing the spectral signal in [27], and can be extended to derivative calculations as follows.

1. *Correlated k-distribution method.* Consider a discretization of the spectral interval $[\lambda_\mathrm{min} - s/2, \lambda_\mathrm{max} + s/2]$ into a set $\{\overline{\lambda}_k\}_{k=1}^{\overline{W}}$ of $\overline{W}$ equally spaced wavelengths with the discretization step $\Delta\lambda$, and assume that the transmission within a spectral interval depends only on the distribution of the gas absorption coefficient $\sigma_\mathrm{abs}^\mathrm{gas}(\lambda)$ within the spectral interval [28]. Letting $F = F(\sigma_{\mathrm{abs}k}^\mathrm{gas})$ be the cumulative density function of $\sigma_\mathrm{abs}^\mathrm{gas}(\lambda)$ in the spectral interval $[\overline{\lambda}_k - \Delta\lambda/2, \overline{\lambda}_k + \Delta\lambda/2]$, $\sigma_{\mathrm{abs}k}^\mathrm{gas}(F)$ the inverse distribution function, and $\{F_l, \varpi_l\}_{l=1}^{N_\mathrm{q}}$ a set of N_q quadrature points and weights in the interval $[0, 1]$, the spectral signal (1) and its partial derivative with respect to the total column are computed as

$$\mathcal{I}(\lambda) = \sum_{w=1}^{W} \omega_w g(\lambda - \lambda_w) \mathcal{I}_\mathrm{FOV}(\sigma_\mathrm{abs}^\mathrm{gas}(\lambda_w)), \tag{30}$$

$$\frac{\partial \mathcal{I}}{\partial X_g}(\lambda) = \sum_{w=1}^{W} \omega_w g(\lambda - \lambda_w) \frac{\partial \mathcal{I}_\mathrm{FOV}}{\partial X_g}(\sigma_\mathrm{abs}^\mathrm{gas}(\lambda_w)), \tag{31}$$

where $\lambda_w = \overline{\lambda}_k$, $\omega_w = \Delta\lambda\,\varpi_l$, and $\sigma_\mathrm{abs}^\mathrm{gas}(\lambda_w) = \sigma_{\mathrm{abs}k}^\mathrm{gas}(F_l)$ for $w = l + (k-1)N_\mathrm{q}$, $k = 1, \ldots, \overline{W}$, $l = 1, \ldots, N_\mathrm{q}$, and $W = \overline{W}N_\mathrm{q}$. Thus, in the framework of the correlated k-distribution method, W monochromatic radiative transfer calculations are required for computing $\mathcal{I}_\mathrm{FOV}(\lambda_w)$ and $\partial \mathcal{I}_\mathrm{FOV}(\lambda_w)/\partial X_g$, and so, $\mathcal{I}(\lambda)$ and $\partial \mathcal{I}(\lambda)/\partial X_g$. A further acceleration can be achieved when $\mathcal{I}_\mathrm{FOV}(\lambda_w)$ and $\partial \mathcal{I}_\mathrm{FOV}(\lambda_w)/\partial X_g$ are computed by using dimensionality reduction techniques, as for example, the principal component analysis.

2. *Principal component analysis.* At wavelength λ_w, the integrated signal $\mathcal{I}_\mathrm{FOV}(\lambda_w)$ is related to the integrated signal calculated by a simplified (approximate) radiative transfer model $\mathcal{I}_\mathrm{FOV}^\mathrm{s}(\lambda_w)$ through the relation

$$\ln \frac{\mathcal{I}_\mathrm{FOV}(\lambda_w)}{\mathcal{I}_\mathrm{FOV}^\mathrm{s}(\lambda_w)} = f(\lambda_w), \tag{32}$$

i.e.,

$$\mathcal{I}_\mathrm{FOV}(\lambda_w) = \mathcal{I}_\mathrm{FOV}^\mathrm{s}(\lambda_w) \mathrm{e}^{f(\lambda_w)}. \tag{33}$$

The correction factor $f(\lambda_w)$ in Equation (32) is actually the quantity that is calculated by means of the principal component analysis. To summarize this approach, we assume that for each wavelength λ_w, the spectral variability of the optical parameters can be described by a vector $\mathbf{x}_w \in \mathbb{R}^N$, defined by

$$\mathbf{x}_w^T = \left[\ln \sigma_{\mathrm{abs}1}^{\mathrm{gas}}(\lambda_w), ..., \ln \sigma_{\mathrm{abs}L}^{\mathrm{gas}}(\lambda_w), \ln \sigma_{\mathrm{sct}1}^{\mathrm{mol}}(\lambda_w), ..., \ln \sigma_{\mathrm{sct}L}^{\mathrm{mol}}(\lambda_w), \ln F_0(\lambda_w) \right], \quad (34)$$

where $\sigma_{\mathrm{abs}k}^{\mathrm{gas}}$ and $\sigma_{\mathrm{sct}k}^{\mathrm{mol}}$ are the optical coefficients in the kth level, $N = 2N_z + 1$, and N_z is the number of altitude levels. Denoting by $\bar{\mathbf{x}} = (1/W) \sum_{w=1}^{W} \mathbf{x}_w$ the sample mean of the data, the aim is to find an M-dimensional subspace ($M < N$) spanned by a set of linear independent vectors $\{\mathbf{a}_k\}_{k=1}^{M}$, such that the centered data $\mathbf{x}_w - \bar{\mathbf{x}}$ belong to this subspace, i.e., $\mathbf{x}_w \approx \bar{\mathbf{x}} + \sum_{k=1}^{M} y_{wk}\mathbf{a}_k = \bar{\mathbf{x}} + \mathbf{A}\mathbf{y}_w$ yielding $\mathbf{y}_w = \mathbf{A}^\dagger(\mathbf{x}_w - \bar{\mathbf{x}})$ for $w = 1, \ldots, W$, where $\mathbf{A} = [\mathbf{a}_k]_{k=1}^{M} \in \mathbb{R}^{N \times M}$, $\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \in \mathbb{R}^{M \times N}$ is the pseudoinverse of $\mathbf{A}$, and y_{wk} is the kth component of the vector of parameters $\mathbf{y}_w \in \mathbb{R}^M$. In the principal component analysis, the original N-dimensional data $\mathbf{X} = [\mathbf{x}_w]_{w=1}^{W} \in \mathbb{R}^{N \times W}$ are projected on the M-dimensional subspace spanned by the dominant singular vectors of the data covariance matrix $\mathbf{C}_\mathbf{x} = (1/W)\mathbf{X}\mathbf{X}^T \in \mathbb{R}^{N \times N}$, that is, with σ_k and $\mathbf{u}_k$ being the kth singular value and vector of the matrix $\mathbf{C}_\mathbf{x}$, we choose $\mathbf{A} = \mathbf{U}_M\mathbf{\Sigma}_M^{1/2}$ and $\mathbf{A}^\dagger = \mathbf{\Sigma}_M^{-1/2}\mathbf{U}_M^T$, where $\mathbf{U}_M = [\mathbf{u}_k]_{k=1}^{M} \in \mathbb{R}^{N \times M}$ and $\mathbf{\Sigma}_M = \mathrm{diag}[\sigma_k]_{k=1}^{M} \in \mathbb{R}^{M \times M}$. Furthermore, approximating the correction factor $f(\mathbf{x}_w)$ by a second-order Taylor expansion around $\bar{\mathbf{x}}$, and the gradient and the Hessian of f by central differences, we are led to the computational formula

$$f(\mathbf{x}_w) \approx f(\bar{\mathbf{x}}) + \frac{1}{2}\sum_{k=1}^{M}[f(\bar{\mathbf{x}} + \mathbf{a}_k) - f(\bar{\mathbf{x}} - \mathbf{a}_k)]y_{wk}$$

$$+ \frac{1}{2}\sum_{k=1}^{M}[f(\bar{\mathbf{x}} + \mathbf{a}_k) - 2f(\bar{\mathbf{x}}) + f(\bar{\mathbf{x}} - \mathbf{a}_k)]y_{wk}^2. \quad (35)$$

To compute the derivative of the integrated signal with respect to the total column X_g, we may use the principal component analysis to calculate the derivative correction factor $f_{X_g}(\lambda_w)$, defined by [29]

$$f_{X_g}(\lambda_w) = \ln\left[\frac{\partial \mathcal{I}_{\mathrm{FOV}}}{\partial X_g}(\lambda_w) \Big/ \frac{\partial \mathcal{I}_{\mathrm{FOV}}^{\mathrm{s}}}{\partial X_g}(\lambda_w)\right], \quad (36)$$

in which case, we have

$$\frac{\partial \mathcal{I}_{\mathrm{FOV}}}{\partial X_g}(\lambda_w) = \frac{\partial \mathcal{I}_{\mathrm{FOV}}^{\mathrm{s}}}{\partial X_g}(\lambda_w)\mathrm{e}^{f_{X_g}(\lambda_w)}. \quad (37)$$

Alternatively, taking the derivative of Equation (33) with respect to X_g, i.e.,

$$\frac{\partial \mathcal{I}_{\mathrm{FOV}}}{\partial X_g}(\lambda_w) = \frac{\partial \mathcal{I}_{\mathrm{FOV}}^{\mathrm{s}}}{\partial X_g}(\lambda_w)\mathrm{e}^{f(\lambda_w)} + \mathcal{I}_{\mathrm{FOV}}(\lambda_w)\frac{\mathrm{d}f}{\mathrm{d}X_g}(\lambda_w), \quad (38)$$

we may use the principal component analysis to calculate the derivative correction factor $f_{X_g}(\lambda_w)$, defined by

$$f_{X_g}(\lambda_w) = \frac{\mathrm{d}f}{\mathrm{d}X_g}(\lambda_w) = \frac{1}{\mathcal{I}_{\mathrm{FOV}}(\lambda_w)}\frac{\partial \mathcal{I}_{\mathrm{FOV}}}{\partial X_g}(\lambda_w) - \frac{1}{\mathcal{I}_{\mathrm{FOV}}^{\mathrm{s}}(\lambda_w)}\frac{\partial \mathcal{I}_{\mathrm{FOV}}^{\mathrm{s}}}{\partial X_g}(\lambda_w). \quad (39)$$

3. Numerical Simulations

In this section, we analyze the accuracy and efficiency of the algorithms for retrieving total column amount of NO_2 under cloudy conditions. The cloudy scenes considered in

our simulations are similar to those described in [27]. Specifically, the geometrical and optical parameters are chosen as follows.

1. The domain of analysis is a rectangular prism of lengths $L_x = L_y = 15$ km and $L_z = 50$ km. The discretization steps along the horizontal directions are $\Delta x = \Delta y = 0.5$ km. Along the vertical direction, the atmosphere between 0 and 50 km is discretized with a step of 0.5 km between 0 and 3 km, 0.1 km between 3 and 4 km, 0.5 km between 4 and 10 km, 1.0 km between 10 and 14 km, 2 km between 14 and 30 km, and 5 km between 30 and 50 km.

2. A homogeneous cloud is placed between 3 and 4 km. The cloud extinction field is given by $\sigma_{ext}^{cloud}(x, y) = \sigma_0 f(x, y)$, where $\sigma_0 = 6\,km^{-1}$ and $f(x, y)$ is the indicator function (note that $f(x, y)$ takes the values 1 and 0 inside and outside the cloud, respectively). The cloud phase function is a Henyey–Greenstein phase function [30] with the asymmetry parameter $g = 0.8$, and the cloud single-scattering albedo is $\omega_{cloud} = 0.99$. Eight cloudy scenes are generated by a two-dimensional broken cloud model [14] with a cloud fraction of about 0.4. The extinction field $\sigma_{ext}^{cloud}(x, y)$ is smoothed at the boundary of a cloudy region in order to avoid abrupt changes in the horizontal plane. The indicator functions corresponding to the eight cloudy scenes are illustrated in Figure 1a–h. For two-dimensional geometries, slices at $y = 7$ km are selected from the cloud fields 1, 2, 3, 5, 6, and 8. The corresponding indicator functions are shown in Figure 2. Note that the slices corresponding to the cloudy scenes 4 and 7 are similar to the slice corresponding to the cloudy scene 2, and are therefore omitted.

3. The number of discrete zenith and azimuth angles are $N_\mu = 16$ and $N_\varphi = 2N_\mu$, respectively, the solar and instrument zenith angles are $\theta_0 = 30°$ and $\theta_m^0 = 45°$, respectively, and the relative azimuth angle is $\Delta\varphi = 0$.

4. A Lambertian reflecting surface with the surface albedo $A = 0.2$ is considered.

5. The footprint of the detector is a square of length $2a = 10\Delta x$ centered at $x_0 = y_0 = L_x/2$, and $z_0 = L_z$, and a wavelength-dependent slit function corresponding to the TROPOMI instrument is assumed.

6. In addition to the scattering and absorption by the cloud, molecular Rayleigh scattering and the absorption by NO_2, ozone (O_3), oxygen dimer (O_4), and water vapor (H_2O) are considered. The measurement spectral grid roughly resembles the TROPOMI's spectral resolution and consists of 119 spectral points between 425 nm and 450 nm.

The process of generating synthetic measurement spectra is organized as follows.

1. For a clean scenario, we use the a priori partial column profile of NO_2 [17] illustrated in Figure 3 to generate the true (exact) partial column profile. In this regard, denoting the a priori partial columns of gas g by $x_{ag,j}$, we choose the true partial columns as $x_{g,j}^\dagger = s_g x_{ag,j}$, $j = 1, \ldots, N_z$ with $s_{NO_2} = 2.0$ and $s_{O_3} = s_{O_4} = s_{H_2O} = 1.2$. The true total column $X_g^\dagger$ of gas g is then computed as $X_g^\dagger = \sum_{j=1}^{N_z} x_{g,j}^\dagger = s_g X_{ag}$; thus, $X_{NO_2}^\dagger = 2X_{aNO_2}$.

2. For $\mathbf{X}^\dagger = [X_{NO_2}^\dagger, X_{O_3}^\dagger, X_{O_4}^\dagger, X_{H_2O}^\dagger]$, we generate the simulated spectral signal $\mathcal{I}_{sim}(\lambda_k, \mathbf{X}^\dagger)$ by means of SHDOM.

3. For cubic smoothing polynomials, i.e., $N_p = 4$, we determine the coefficients $\mathbf{c}_{sim}(\mathbf{X}^\dagger)$ of the polynomial $P_{sim}(\lambda, \mathbf{c}_{sim}(\mathbf{X}^\dagger))$ as the solutions of the least-squares problem (13).

4. We compute the noisy spectral signal as $\mathcal{I}_{sim}^\delta(\lambda_k, \mathbf{X}^\dagger) = \mathcal{I}_{sim}(\lambda_k, \mathbf{X}^\dagger) + \delta_k$, where the measurement errors δ_k are assumed to be independent Gaussian random variables with zero mean and standard deviation $\sigma_k = \mathcal{I}_{sim}(\lambda_k, \mathbf{X}^\dagger)/SNR$, where SNR is the signal-to-noise ratio. It should be pointed out that in view of the approximation

$$\ln \mathcal{I}_{sim}^\delta(\lambda_k, \mathbf{X}^\dagger) \approx \ln \mathcal{I}_{sim}(\lambda_k, \mathbf{X}^\dagger) + \frac{\delta_k}{\mathcal{I}_{sim}(\lambda_k, \mathbf{X}^\dagger)},$$

the error in $\ln \mathcal{I}_{\text{sim}}^{\delta}(\lambda_k, \mathbf{X}^{\dagger})$ is $\delta_{\ln k} = \delta_k / \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}^{\dagger})$ yielding $\sigma_{\ln k} = 1/\text{SNR}$ for all k. In other words, the measurement error vector is white noise with the covariance matrix $(1/\text{SNR}^2)\mathbf{I}_m$, where $\mathbf{I}_m$ is the identity matrix. Because in our simulations we are mainly interested in multi-dimensional effects, we take $\text{SNR} = 10^4$, that is, we assume an almost perfect instrument and neglect the forward model errors.

5. We include the Ring correction spectrum $S_{\text{R}}(\lambda_k, \mathbf{X}_{\text{a}})$ illustrated in Figure 4 in the retrieval, and choose the a priori and true Ring amplitudes as $b_{\text{aR}} = 5 \times 10^{-2}$ and $b_{\text{R}}^{\dagger} = 2b_{\text{aR}}$, respectively. Note that the inelastic scattering is described by a first-order Rayleigh scattering model, i.e., by applying a first-order iteration scheme to the one-dimensional radiative transfer equation for inelastic scattering [31].

6. For DRMI, we compute the measured differential spectral signal as

$$R_{\text{mes}}^{\delta}(\lambda_k) = [\ln \mathcal{I}_{\text{sim}}^{\delta}(\lambda_k, \mathbf{X}^{\dagger}) - P_{\text{sim}}(\lambda, \mathbf{c}_{\text{sim}}(\mathbf{X}^{\dagger}))] + b_{\text{R}}^{\dagger} S_{\text{R}}(\lambda_k, \mathbf{X}_{\text{a}}),$$

while for DRME, we choose $\mathbf{c}^{\dagger} = 0.5\,\mathbf{c}_{\text{sim}}(\mathbf{X}^{\dagger})$ and compute the measured differential spectral signal as

$$R_{\text{mes}}^{\delta}(\lambda_k) = \ln \mathcal{I}_{\text{sim}}^{\delta}(\lambda_k, \mathbf{X}^{\dagger}) + b_{\text{R}}^{\dagger} S_{\text{R}}(\lambda_k, \mathbf{X}_{\text{a}}) - P(\lambda_k, \mathbf{c}^{\dagger}).$$

SHDOM is run with a solution accuracy of 10^{-4} and by using

1. an adaptive grid with a splitting accuracy of 10^{-4},
2. the principal component analysis with $M = 4$, and the derivative correction factor $f_{X_g}(\lambda_w)$ as in Equation (39),
3. periodic boundary conditions,
4. the delta-M approximation [32], and the untruncated phase function single-scattering solution (TMS correction of Nakajima and Tanaka [33].

Some parameters characterizing the iteratively regularized Gauss–Newton algorithm are chosen as follows.

1. The regularization parameters α_k are the terms of a geometric sequence with ratio $q = 0.1$ and initial value $\alpha_0 = 10^{-3}$. Thus, at the first iteration step, the regularization parameter is $\alpha_1 = 10^{-4}$.
2. The weighting factors specifying the contribution of each component of the state vector into the regularization matrix are $w_{\text{NO}_2} = 1.0$, $w_{\text{O}_3} = w_{\text{O}_4} = w_{\text{H}_2\text{O}} = 10^2$, $w_{\text{R}} = 10^{-3}$, and $w_{\text{c}p} = 1.0$ for all $p = 0, \ldots, N_{\text{p}} - 1$. By this choice, the total columns of the auxiliary gases are stronger constrained to the a priori than the total column of NO_2 and the Ring correction spectrum.
3. The control parameter τ in the discrepancy principle Equation (24) is $\tau = 1.2$.

The simulations were performed on a computer Intel(R) Core(TM) i5-3340M CPU @ 2.70 GHz with 7858 MB RAM.
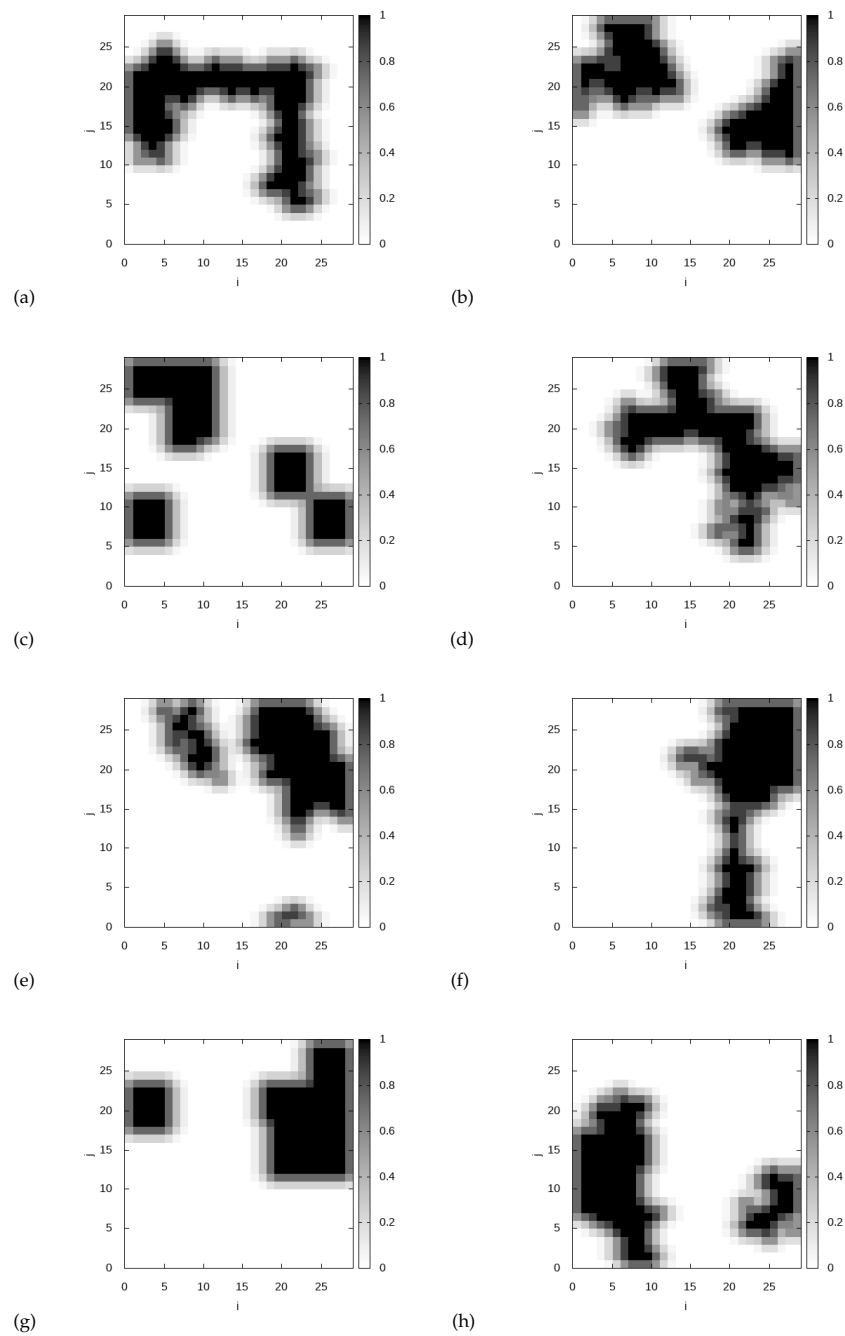
Figure 1. Indicator function $f(x_i, y_j)$ with $x_i = i\Delta x$ and $y_j = j\Delta y$ for the 8 cloudy scenes (cases (**a**–**h**)); i and j refer to the pixel index in the xy plane, while each pixel corresponds to a 500 m $\times$ 500 m area.
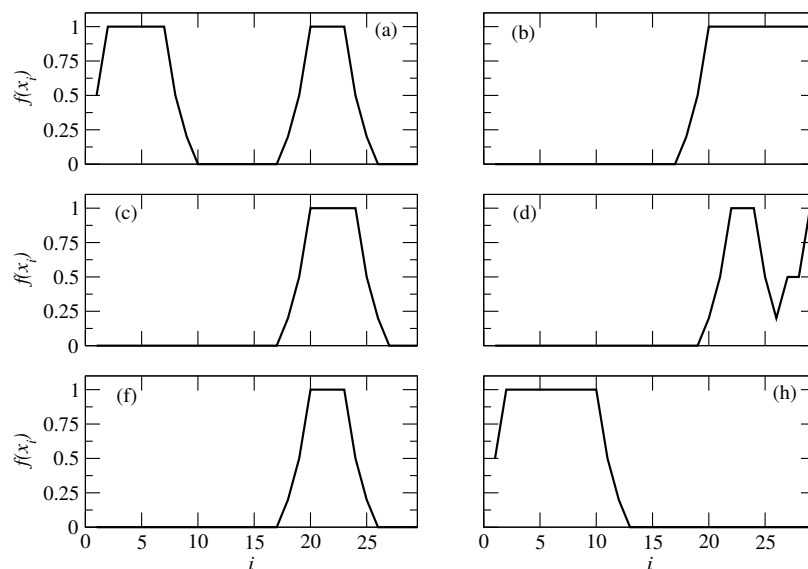
Figure 2. Slices of the indicator function $f(x_i)$ with $x_i = i\Delta x$ at $y = 7$ km for the cloudy scenes shown in Figure 1 (**a–d,f,h**).
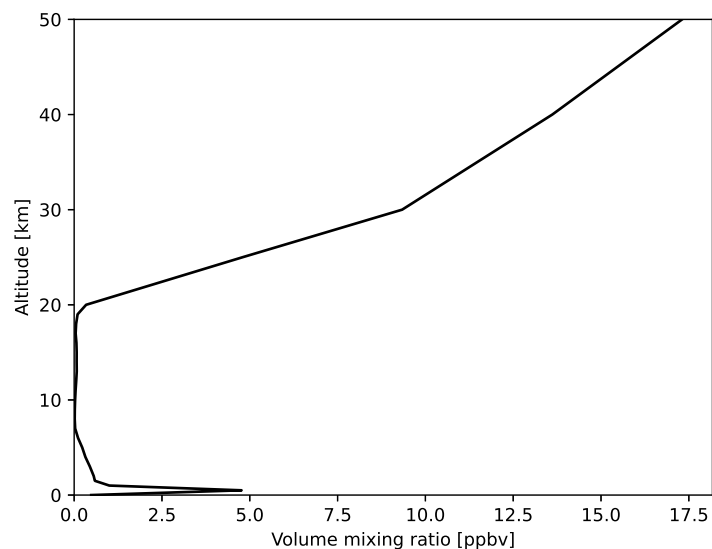
Figure 3. A priori vertical profile of NO_2 volume mixing ratio.
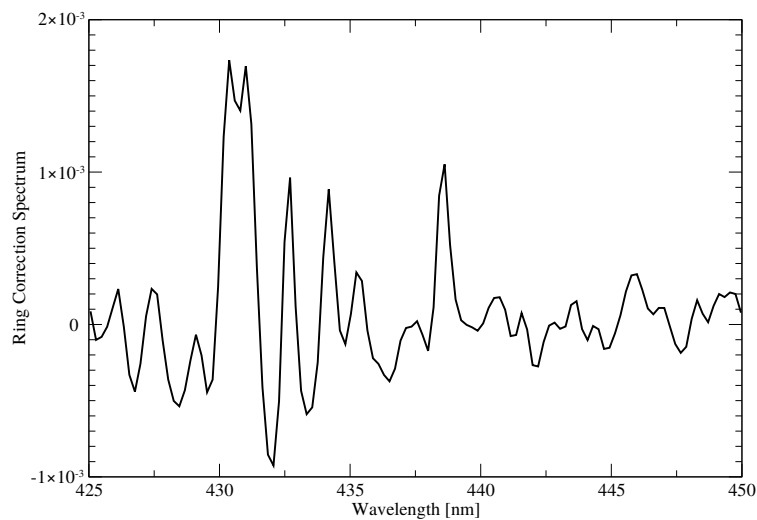
Figure 4. Ring correction spectrum included in the retrieval.

3.1. Test Example 1

In the first test example we consider a two-dimensional geometry and perform the retrieval by using the DRMI and DRME inverse models. It may in general be noted that

1. the inverse problem is severely ill-posed (for the DRMI model, the condition number of the Jacobian matrix at the initial guess is $\kappa_{\mathrm{DRMI}} = 2.92 \times 10^7$) and
2. there is a strong correlation between the NO_2 and Ring effect signatures (when the Ring correction spectrum is not included in the retrieval, the condition number is $\kappa_{\mathrm{DRMI}} = 2.16 \times 10^4$; thus, κ_{DRMI} decreases by three order of magnitude).

The histories of the residuals with respect to the iteration index are shown in Figure 5. The following conclusions can be drawn.

1. At the initial guess, the residual corresponding to DRMI is much smaller than that corresponding to DRME. This occurs because the discrepancies between the differential spectra are usually small.
2. In DRMI, the residual decreases very fast at the first iteration step and then more steadily, while in DRME, the residual gradually decreases.

The relative errors and the computation times for DRMI and DRME are given in the first two columns of Tables 1 and 2, respectively. Because the SNR is high, in both cases, the relative errors are smaller than 0.01%. On average, the computation times for DRMI and DRME are of about 40 and 35 CPU minutes, respectively. The higher computation time for DRMI is due to the fact that, according to Equation (13), the partial derivative of $\mathbf{c}_{\mathrm{sim}}(\mathbf{X})$ with respect to $\mathbf{X}$ needs also to be computed.

In the second two columns of Table 1, we list the relative errors and the computation times for DRMI when the partial derivative of the top of atmosphere radiance $\partial I(\cdot, \mathbf{r}_{\mathrm{t}}, \cdot)/\partial X_g$ is computed by taking into account only the contributions of the partial derivatives of the extinction field $\partial \sigma_{\mathrm{ext}}(\cdot, \mathbf{r})/\partial X_g$. The reduction of the computation time by about 25% does not justify the large relative errors. Thus, the simplified approach suggested in [26] for the retrieval of cloud extinction field, seems not to work in the case of total column trace gas retrieval.

In the second two columns of Table 2, we list the relative errors and the computation times for DRME when the iterative process is stopped after one iteration, that is, when the linear model (14) is considered. A reasonable accuracy can be obtained when the regularization parameter α is optimally chosen. Some hints regarding the selection of the optimal value of the regularization parameter α_{opt} can be found by analyzing the history of the residual curve in Figure 5. From this curve, we infer that for $\alpha = 10^{-4}$ the residual is of about 10^{-5}, for $\alpha = 10^{-5}$ of about 10^{-6}, and for $\alpha = 10^{-6}$ of about 10^{-8}. Therefore, by choosing α in the range $10^{-6}, \ldots, 10^{-5}$ we expect good reconstruction errors. However, as it can be seen in Table 2, α_{opt} depends on the cloud field. Moreover, for some cloudy scenes, α_{opt} should be precisely determined (for example, in the case of the cloudy scene 8, the relative errors are 8.30×10^{-2}, 1.87×10^{-2}, and 9.02×10^{-2} when α takes the values 1×10^{-5}, 2×10^{-5}, and 3×10^{-5}, respectively). In this regard, it should be mentioned that although less inefficient, the iteratively regularized Gauss–Newton method is the best option, i.e., when the initial value of regularization parameter α_0 is chosen sufficiently large, accurate results will always be found in comparatively small number of iteration steps (which depends on the initial value α_0 and the ratio q).
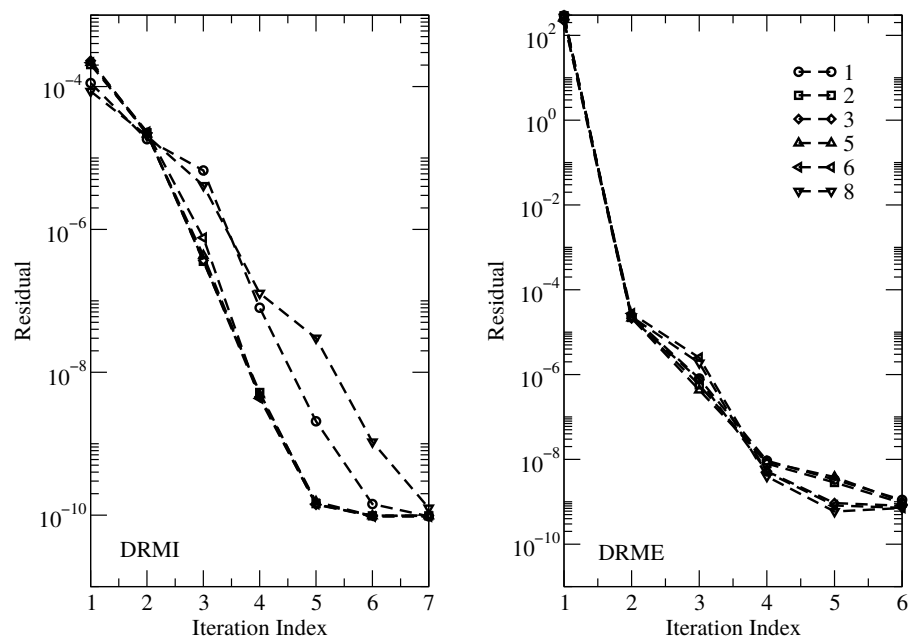
Figure 5. Residuals versus the iteration index for Differential Radiance Model with Internal closure (DRMI) (**left**) and Differential Radiance Model with External closure (DRME) (**right**).

Table 1. Relative errors and the computation times (CPU) in minutes: seconds for DRMI. The results correspond to a two-dimensional geometry.

Cloudy Scene	DRMI		DRMI (Approx. Derivatives)	
	Rel. Errors	CPU	Rel. Errors	CPU
1	1.66×10^{-5}	39:51	1.18×10^{-1}	32:24
2	1.10×10^{-5}	40:12	1.50×10^{-1}	32:48
3	9.22×10^{-6}	39:14	9.79×10^{-2}	31:06
5	1.08×10^{-5}	38:54	1.37×10^{-1}	30:44
6	1.03×10^{-5}	39:32	8.77×10^{-2}	31:16
8	1.94×10^{-5}	40:06	1.51×10^{-1}	32:33

Table 2. Relative errors and the computation times (CPU) in minutes:seconds for DRME. The numbers in parenthesis are the values of the regularization parameter α. The results correspond to a two-dimensional geometry.

Cloudy Scene	DRME		DRME (One Iteration)	
	Rel. Errors	CPU	Rel. Errors	CPU
1	6.94×10^{-5}	35:10	$2.84 \times 10^{-2} \ (3.0 \times 10^{-6})$	5:33
2	3.23×10^{-5}	36:47	$3.74 \times 10^{-3} \ (1.0 \times 10^{-6})$	5:54
3	1.23×10^{-5}	34:57	$7.74 \times 10^{-3} \ (1.0 \times 10^{-6})$	5:12
5	4.84×10^{-5}	34:23	$6.73 \times 10^{-3} \ (1.0 \times 10^{-6})$	5:01
6	1.08×10^{-5}	35:06	$2.60 \times 10^{-2} \ (5.0 \times 10^{-6})$	5:28
8	2.31×10^{-5}	36:23	$1.87 \times 10^{-2} \ (2.0 \times 10^{-5})$	5:48

3.2. Test Example 2

For the cloudy scenes considered in Figure 1, we illustrate, in Table 3, the relative errors and the computation times for DRME with several and one iteration steps. The solution accuracies are comparable to those corresponding to two-dimensional geometries. However, the computation time is extremely high; on average, it is of about 14 h and

15 min for DRME with several iteration steps, and 2 h and 30 min for DRME with one iteration step.

By taking a closer look at the cloudy scenes depicted in Figure 1, we see that the footprint of the detector is almost cloud-free. In this regard, we may try to retrieve the total column amount by using an one-dimensional radiative transfer model for a cloud-free atmosphere. The results given in the first two columns of Tables 4 and 5 show that in this case, the relative error is about 15%. In these simulations, the spectral signal measured by the instrument is computed by a three-dimensional radiative transfer model, while in the retrieval, the simulated spectral signal is computed by an one-dimensional radiative transfer model. The relative errors can be approximately halved when the differential correction spectrum due to three-dimensional effects (18) is included in the inverse model and its amplitude is retrieved. This is shown in the second two columns of Tables 4 and 5. Note that the computation time given in these tables does not take into account the time for computing the correction spectrum by a three-dimensional radiative transfer model. As shown in [27], this computation time is of about 1 h and 20 min, so that roughly speaking, this is the time for retrieving the total column amount by an one-dimensional inverse model that accounts on three-dimensional effects. The correction spectra included in the inverse model are shown in Figure 6, and it is apparent that they depend on the cloud field (some similitude can be observed for the cloudy scenes 2 and 7).

Table 3. The same as in Table 2 but for a three-dimensional geometry.

| Cloudy Scene	DRME		DRME (One Iteration)	
	Rel. Errors	CPU	Rel. Errors	CPU
1	4.74×10^{-4}	14:15:31	$4.05 \times 10^{-2} \, (4.0 \times 10^{-5})$	2:34:03
2	3.04×10^{-4}	14:21:43	$1.04 \times 10^{-2} \, (1.0 \times 10^{-6})$	2:36:42
3	1.35×10^{-4}	14:15:57	$3.01 \times 10^{-2} \, (2.0 \times 10^{-6})$	2:32:26
5	3.63×10^{-4}	14:15:23	$2.38 \times 10^{-2} \, (1.0 \times 10^{-6})$	2:32:11
6	1.46×10^{-4}	14:14:36	$3.56 \times 10^{-2} \, (5.0 \times 10^{-6})$	2:33:38
8	1.88×10^{-4}	14:20:13	$3.18 \times 10^{-2} \, (4.0 \times 10^{-5})$	2:34:16

Table 4. Relative errors and the computation times (CPU) in minutes:seconds for DRMI. The measured and simulated spectral signals are computed by a three- and an one-dimensional radiative transfer model, respectively.

Cloudy Scene	DRMI		DRMI (Correction. Spectrum)	
	Rel. Errors	CPU	Rel. Errors	CPU
1	1.21×10^{-1}	2:21	7.59×10^{-2}	2:42
2	5.18×10^{-2}	2:32	2.76×10^{-2}	2:58
3	7.03×10^{-2}	2:14	3.74×10^{-2}	2:36
4	1.51×10^{-1}	2:04	8.40×10^{-2}	2:24
5	6.08×10^{-2}	2:27	3.76×10^{-2}	2:47
6	9.82×10^{-2}	2:25	5.50×10^{-2}	2:45
7	5.41×10^{-2}	2:25	3.26×10^{-2}	2:46
8	7.15×10^{-2}	2:30	4.16×10^{-2}	2:50

Table 5. The same as in Table 4 but for DRME.

Cloudy Scene	DRME		DRME (Correction. Spectrum)	
	Rel. Errors	CPU	Rel. Errors	CPU
1	1.21×10^{-1}	2:10	8.60×10^{-2}	2:16
2	5.19×10^{-2}	2:12	2.70×10^{-2}	2:18
3	6.99×10^{-2}	2:05	3.70×10^{-2}	2:11
4	1.53×10^{-1}	1:52	8.45×10^{-2}	2:01
5	6.12×10^{-2}	2:16	3.82×10^{-2}	2:22
6	8.82×10^{-2}	2:14	5.53×10^{-2}	2:18
7	5.37×10^{-2}	2:13	3.21×10^{-2}	2:19
8	7.11×10^{-2}	2:21	4.03×10^{-2}	2:17
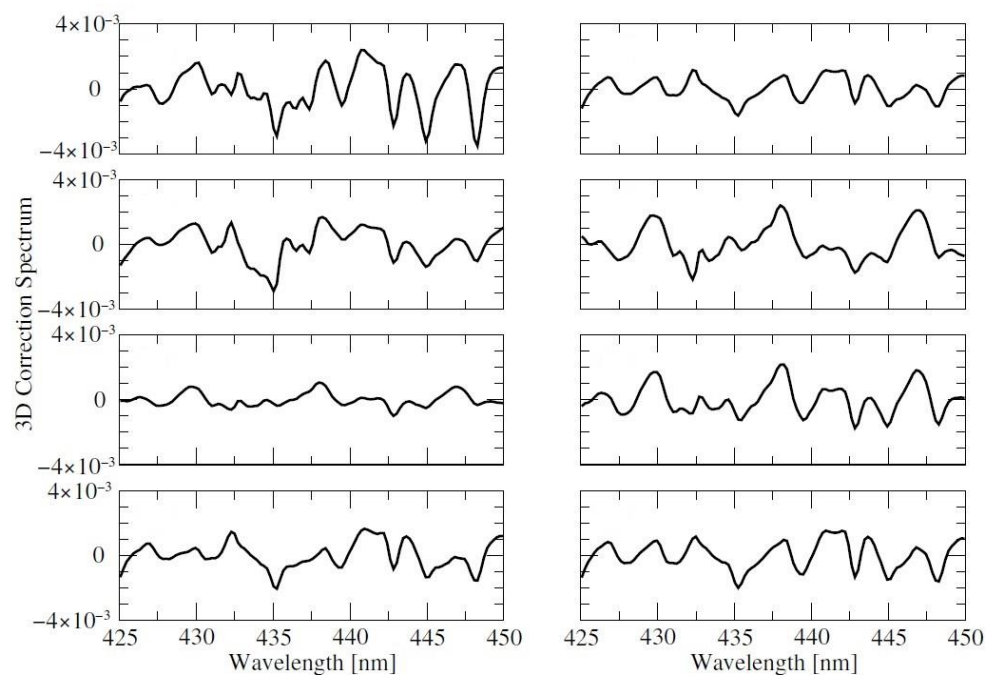
Figure 6. Correction spectra due to three-dimensional effects.

4. Discussion

The main concepts of an algorithm for the retrieval of the total column amount of trace gases in a multi-dimensional atmosphere have been presented. The algorithm uses (i) as inverse models, certain differential radiance models with internal and external closures; (ii) as regularization tool, the iteratively regularized Gauss-Newton method; and (iii) as linearized radiative transfer model, the spherical harmonics discrete ordinate method with a spectral acceleration approach that combines the correlated k-distribution method with the principal component analysis.

The algorithm has been applied to the retrieval of the total column amount of NO_2 under cloudy conditions. For two-dimensional geometries, we come to the following findings.

1. The differential radiance models with internal and external closures yield accurate results with reasonable computation times (of about 35–40 min).
2. An inverse model based on an approximate computation of the partial derivative leads to a reduction of the computation time by about 25%, but to large relative errors.
3. Provided that the regularization parameter is optimally chosen, reasonable accurate results with a computation time of about 6 min can be obtained when the iterative process, corresponding to the differential radiance model with external closure, is

stopped after one iteration step. The fact that the optimal value of the regularization parameter depends on the cloudy scene, makes it more difficult to apply this one-step retrieval algorithm, or equivalently, DOAS-type models.

For three-dimensional geometries, we come to the following results.

1. Although accurate, the retrievals based on the differential radiance model with external closures are inefficient; the computation time is of about 14 h and 15 min for a full-step retrieval algorithm, and 2 h and 30 min for an one-step retrieval algorithm. The one-step retrieval algorithm is less accurate than in the case of two-dimensional geometries, but the results are still acceptable.

2. The application of a fast one-dimensional radiative transfer model to retrieve the NO_2 column amount for a three-dimensional cloudy scene leads to relative errors up to 15%. These errors can be reduced when a differential correction spectrum due to three-dimensional effects is included in the retrieval.

The main conclusion of our analysis is that although for three-dimensional geometries the computational time is high, the main concepts of the algorithm are correct and the retrieval results are accurate.

For operational retrievals, we recommend

1. to construct a database for the spectral signal $\mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}_{\mathrm{a}})$ and its derivatives $\partial \ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}_{\mathrm{a}})/\partial X_g$, and to use these spectra in an one-step, three-dimensional retrieval algorithm, or

2. to construct a database for the differential correction spectra accounting on three-dimensional effects $S_{\mathrm{3D}}(\lambda_k, \mathbf{X}_{\mathrm{a}})$, and to use these spectra in a full-step, one-dimensional retrieval algorithm.

The databases should be built up for different cloud scenarios, surface albedo, and solar and satellite geometries. The main problem which has to be solved is the design of a classification algorithm for broken clouds. One option is to use a nearest neighbor classifier based on the two-dimensional principal component analysis. Another option is to use a classification algorithm in which only the distributions of the cloud fields in the two-dimensional domains around the azimuth planes of the solar and viewing directions are of interest. The width of these domains corresponds to the instrument footprint. Note that these domains are also relevant in the tilted independent pixel approximation [34]: the direct solar beam is computed at all grid points which belong to the vertical columns inside the instrument footprint and for all characteristics within the entrance (solar direction) domain, while the measured radiance is computed by integrating the source function along all characteristics within the exit (viewing direction) domain. This will be the topic of a forthcoming paper.

Author Contributions: Conceptualization, A.D.; Software, A.D. and D.S.E.; formal analysis, D.S.E. and T.T.; writing—original draft preparation, A.D.; writing—review and editing, D.S.E. and T.T. All authors have read and agreed to the published version of the manuscript.

Appendix A

In this appendix we describe two inversion models for tropospheric column retrieval.

Assuming that the troposphere extends from level 1 to N_{t}, and the stratosphere from level N_{t} to N_{z}, the tropospheric and stratospheric total columns are given, respectively, by

$$X_{\mathrm{t}g} = \sum_{j=1}^{N_{\mathrm{t}}-1} x_{g,j} \tag{A1}$$

and

$$X_{sg} = \sum_{j=N_t}^{N_z-1} x_{g,j}. \tag{A2}$$

Obviously, we have $X_g = X_{tg} + X_{sg}$. The retrieval of the tropospheric column of a gas $\overline{g}$ (specifically, NO$_2$) is performed under the assumption that we have some a priori knowledge about the stratospheric column. In other words, we suppose that $X_{s\overline{g}}$ can be approximated by $X_{s\overline{g}} \approx X^{\star}_{s\overline{g}}$, where $X^{\star}_{s\overline{g}}$ is delivered for example, by the reference sector method. In this method, the stratospheric NO$_2$ columns is estimated from measurements over the remote regions, under the assumptions of low longitudinal variations of stratospheric NO$_2$, and negligible tropospheric NO$_2$ [35–37]. Denoting by $\mathbf{X}_{-\overline{g}}$ the set of all total columns excepting $X_{\overline{g}}$, i.e., $\mathbf{X} = \{X_{\overline{g}}\} \cup \mathbf{X}_{-\overline{g}}$, a tropospheric column retrieval algorithm can be summarized as follows [17].

1. Solve the nonlinear Equation (8) of the DRMI inversion model for $\mathbf{x} = [\mathbf{X}, \mathbf{b}]^T$, or the nonlinear Equation (11) of the DRME inversion model for $\mathbf{x} = [\mathbf{X}, \mathbf{b}, \mathbf{c}]^T$.

2. With $X^{\star}_{s\overline{g}}$ delivered by the reference sector method, and $\mathbf{X}_{-\overline{g}}$ and $\mathbf{b}$ determined at Step 1, solve the nonlinear equation

$$R^{\delta}_{\mathrm{mes}}(\lambda_k) = R_{\mathrm{sim}}(\lambda_k, X_{t\overline{g}}, X^{\star}_{s\overline{g}}, \mathbf{X}_{-\overline{g}}) + \sum_{j=1}^{N_s} b_j S_j(\lambda_k, \mathbf{X}_a), \quad k = 1, \ldots, N_\lambda, \tag{A3}$$

of the DRMI inversion model for $\mathbf{x} = [X_{t\overline{g}}]$, or the nonlinear equation

$$R^{\delta}_{\mathrm{mes}}(\lambda_k) = \ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, X_{t\overline{g}}, X^{\star}_{s\overline{g}}, \mathbf{X}_{-\overline{g}}) + \sum_{j=1}^{N_s} b_j S_j(\lambda_k, \mathbf{X}_a) - P(\lambda_k, \mathbf{c}), \quad k = 1, \ldots, N_\lambda, \tag{A4}$$

of the DRME inversion model for $\mathbf{x} = [X_{t\overline{g}}, \mathbf{c}]^T$.

Appendix B

In this appendix we present a simplified derivation of the standard DOAS equation. For a more rigorous treatment, we refer to the work in [19].

We start with the linearized model (14) written as

$$\ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}) = \sum_{g=1}^{N_g} X_g W_g(\lambda_k, \mathbf{X}_a) + \ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}_a) - \sum_{g=1}^{N_g} X_{ag} W_g(\lambda_k, \mathbf{X}_a) + P_p(\lambda_k, \mathbf{c}_p), \tag{A5}$$

where

$$W_g(\lambda_k, \mathbf{X}_a) = \frac{\partial \ln \mathcal{I}_{\mathrm{sim}}}{\partial X_g}(\lambda_k, \mathbf{X}_a). \tag{A6}$$

Further, we consider the first-order Taylor expansion

$$\ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}) = \ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}_a) + \sum_{g=1}^{N_g} (X_g - X_{ag}) \frac{\partial \ln \mathcal{I}_{\mathrm{sim}}}{\partial X_g}(\lambda_k, \mathbf{X}_a) + \varepsilon_{\mathrm{lin}}(\lambda_k, \mathbf{X} - \mathbf{X}_a) \tag{A7}$$

at $\mathbf{X} = 0$, that is,

$$\ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, 0) = \ln \mathcal{I}_{\mathrm{sim}}(\lambda_k, \mathbf{X}_a) - \sum_{g=1}^{N_g} X_{ag} W_g(\lambda_k, \mathbf{X}_a) + \varepsilon_{\mathrm{lin}}(\lambda_k, -\mathbf{X}_a), \tag{A8}$$

where $\mathcal{I}_{\mathrm{sim}}(\lambda_k, 0)$ corresponds to an atmosphere without gaseous absorption and $\varepsilon_{\mathrm{lin}}(\lambda_k, -\mathbf{X}_a)$ is the linearization error. Defining

$$\ln \overline{\mathcal{I}}_{\text{sim}}(\lambda_k, 0) = \ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}_{\text{a}}) - \sum_{g=1}^{N_{\text{g}}} X_{\text{a}g} W_g(\lambda_k, \mathbf{X}_{\text{a}}), \tag{A9}$$

which yields

$$\ln \mathcal{I}_{\text{sim}}(\lambda_k, 0) = \ln \overline{\mathcal{I}}_{\text{sim}}(\lambda_k, 0) + \varepsilon_{\text{lin}}(\lambda_k, -\mathbf{X}_{\text{a}}), \tag{A10}$$

we express Equation (A8) as

$$\ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}) = \sum_{g=1}^{N_{\text{g}}} X_g W_g(\lambda_k, \mathbf{X}_{\text{a}}) + \ln \overline{\mathcal{I}}_{\text{sim}}(\lambda_k, 0) + P_{\text{p}}(\lambda_k, \mathbf{c}_{\text{p}}). \tag{A11}$$

In general, $\ln \mathcal{I}_{\text{sim}}(\lambda_k, 0)$ can be approximated by a polynomial, but $\ln \overline{\mathcal{I}}_{\text{sim}}(\lambda_k, 0)$ not. However, if $\varepsilon_{\text{lin}}(\lambda_k, -\mathbf{X}_{\text{a}})$ is negligible, then we have $\ln \mathcal{I}_{\text{sim}}(\lambda_k, 0) \approx \ln \overline{\mathcal{I}}_{\text{sim}}(\lambda_k, 0)$; thus, $\ln \overline{\mathcal{I}}_{\text{sim}}(\lambda_k, 0)$ can be approximated by a polynomial, which can be included in $P_{\text{p}}(\lambda_k, \mathbf{c}_{\text{p}})$. Consequently, we obtain

$$\ln \mathcal{I}_{\text{sim}}(\lambda_k, \mathbf{X}) = \sum_{g=1}^{N_{\text{g}}} X_g W_g(\lambda_k, \mathbf{X}_{\text{a}}) + P_{\text{p}}(\lambda_k, \mathbf{c}_{\text{p}}), \tag{A12}$$

so that by putting $P(\lambda_k, \mathbf{c}) - P_{\text{p}}(\lambda_k, \mathbf{c}_{\text{p}}) \to P(\lambda_k, \mathbf{c})$ we are led to the following representation for the DRME model (11),

$$R_{\text{mes}}^{\delta}(\lambda_k) = \sum_{g=1}^{N_{\text{g}}} X_g W_g(\lambda_k, \mathbf{X}_{\text{a}}) + \sum_{j=1}^{N_{\text{s}}} b_j S_j(\lambda_k, \mathbf{X}_{\text{a}}) - P(\lambda_k, \mathbf{c}). \tag{A13}$$

In view of the computational formula (cf. Equation (29))

$$W_g(\lambda_k, \mathbf{X}_{\text{a}}) = \frac{\partial \ln \mathcal{I}_{\text{sim}}}{\partial X_g}(\lambda_k, \mathbf{X}_{\text{a}}) = \frac{1}{X_{\text{a}g}} \sum_{j=1}^{N_{\text{z}}} C_{\text{abs}g}(\lambda_k, z_j) n_{\text{a}g,j} \frac{\partial \ln \mathcal{I}_{\text{sim}}}{\partial \sigma_{\text{abs}g,j}^{\text{gas}}}(\lambda_k, \mathbf{X}_{\text{a}}) \tag{A14}$$

we define

$$\overline{C}_{\text{abs}g}(\lambda_k) = \frac{\sum_{j=1}^{N_{\text{z}}} C_{\text{abs}g}(\lambda_k, z_j) n_{\text{a}g,j} \dfrac{\partial \ln \mathcal{I}_{\text{sim}}}{\partial \sigma_{\text{abs}g,j}^{\text{gas}}}(\lambda_k, \mathbf{X}_{\text{a}})}{\sum_{j=1}^{N_{\text{z}}} n_{\text{a}g,j} \dfrac{\partial \ln \mathcal{I}_{\text{sim}}}{\partial \sigma_{\text{abs}g,j}^{\text{gas}}}(\lambda_k, \mathbf{X}_{\text{a}})}, \tag{A15}$$

and rewrite Equation (A13) as

$$R_{\text{mes}}^{\delta}(\lambda_k) = -\sum_{g=1}^{N_{\text{g}}} A_g(\lambda_k, \mathbf{X}_{\text{a}}) \overline{C}_{\text{abs}g}(\lambda_k) X_g + \sum_{j=1}^{N_{\text{s}}} b_j S_j(\lambda_k, \mathbf{X}_{\text{a}}) - P(\lambda_k, \mathbf{c}), \tag{A16}$$

where

$$A_g(\lambda_k, \mathbf{X}_{\text{a}}) = -\frac{1}{\overline{C}_{\text{abs}g}(\lambda_k)} W_g(\lambda_k, \mathbf{X}_{\text{a}}) \tag{A17}$$

is the air mass factor. Equation (A16) represents the standard DOAS equation. Note that if the computation of the partial derivatives in Equation (A15) is an expensive process, we may define $\overline{C}_{\text{abs}g}(\lambda_k)$ by the relation

$$\overline{C}_{\text{abs}g}(\lambda_k) = \frac{\sum_{j=1}^{N_{\text{z}}} C_{\text{abs}g}(\lambda_k, z_j) n_{\text{a}g,j}}{\sum_{j=1}^{N_{\text{z}}} n_{\text{a}g,j}}. \tag{A18}$$

References

1. Efremenko, D.S.; Doicu, A.; Loyola, D.; Trautmann, T. Fast Stochastic Radiative Transfer Models for Trace Gas and Cloud Property Retrievals Under Cloudy Conditions. In *Springer Series in Light Scattering*; Kokhanovsky, A., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 231–277. [CrossRef]
2. Belov, V.V.; Kirnos, I.V.; Tarasenkov, M.V. Estimation of the influence of cloudiness on the Earth observation from space through a gap in a cloudy field. In Proceedings of the 21st International Symposium Atmospheric and Ocean Optics: Atmospheric Physics, SPIE, Tomsk, Russia, 22–26 June 2015; Romanovskii, O.A., Ed. [CrossRef]
3. Tarasenkov, M.V.; Kirnos, I.V.; Belov, V.V. Observation of the Earth's surface from the space through a gap in a cloud field. *Atmos. Ocean. Opt.* **2017**, *30*, 39–43. [CrossRef]
4. Zhuravleva, T.B.; Nasrtdinov, I.M.; Russkova, T.V. Influence of 3D cloud effects on spatial-angular characteristics of the reflected solar radiation field. *Atmos. Ocean. Opt.* **2017**, *30*, 103–110. [CrossRef]
5. Schwaerzel, M.; Emde, C.; Brunner, D.; Morales, R.; Wagner, T.; Berne, A.; Buchmann, B.; Kuhlmann, G. Three-dimensional radiative transfer effects on airborne and ground-based trace gas remote sensing. *Atmos. Meas. Tech.* **2020**, *13*, 4277–4293. [CrossRef]
6. Evans, K. The spherical harmonic discrete ordinate method for three-dimensional atmospheric radiative transfer. *J. Atmos. Sci.* **1998**, *55*, 429–446. [CrossRef]
7. Cracknell, A.; Hayes, L. *Introduction to Remote Sensing*, 2nd ed.; CRC Press: New York, NY, USA, 2007.
8. Chandrasekhar, S. Radiative Transfer. In *Dover Books on Intermediate and Advanced Mathematics*; Dover Publications: New York, New York, USA, 1960.
9. Sobolev, V.V. *Light Scattering in Planetary Atmospheres*; Pergamon Press: New York, USA, 1972.
10. Doicu, A.; Efremenko, D.; Trautmann, T. An analysis of the short-characteristic method for the spherical harmonic discrete ordinate method (SHDOM). *J. Quant. Spectrosc. Radiat. Transf.* **2013**, *119*, 114–127. [CrossRef]
11. Bohren, C.; Huffman, D. *Absorption and Scattering of Light by Small Particles*; Wiley: New York, NY, USA, 1998. doi:10.1002/9783527618156. [CrossRef]
12. Kokhanovsky, A. *Cloud Optics*; Springer: Dordrecht, The Netherlands, 2006. [CrossRef]
13. Bodhaine, B.; Wood, N.; Dutton, E.; Slusser, J. On Rayleigh optical depth calculations. *J. Atmos. Ocean. Technol.* **1999**, *16*, 1854–1861. [CrossRef]
14. Alexandrov, M.; Marshak, A.; Ackerman, A. Cellular Statistical Models of Broken Cloud Fields. Part I: Theory. *J. Atmos. Sci.* **2010**, *67*, 2125–2151. [CrossRef]
15. Doicu, A.; Trautmann, T.; Schreier, F. *Numerical Regularization for Atmospheric Inverse Problems*; Springer: Berlin/Heidelberg, Germany, 2010. [CrossRef]
16. Efremenko, D.; Schüssler, O.; Doicu, A.; Loyola, D. A stochastic cloud model for cloud and ozone retrievals from UV measurements. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *184*, 167–179. [CrossRef]
17. Liu, S. Inversion Models for the Retrieval of Total and Tropospheric NO2 Columns. *Atmosphere* **2019**, *10*, 607. [CrossRef]
18. Platt, U.; Stutz, J. *Differential Optical Absorption Spectroscopy*; Springer: Berlin/Heidelberg, Germany, 2008. [CrossRef]
19. Rozanov, V.V.; Rozanov, A.V. Differential optical absorption spectroscopy (DOAS) and air mass factor concept for a multiply scattering vertically inhomogeneous medium: theoretical consideration. *Atmos. Meas. Tech.* **2010**, *3*, 751–780. [CrossRef]
20. Bakushinskii, A. The problem of the convergence of the iteratively regularized Gauss–Newton method. *Comput. Math. Math. Phys.* **1992**, *32*, 1353–1359.
21. Xu, J.; Rao, L.; Schreier, F.; Efremenko, D.; Doicu, A.; Trautmann, T. Insight into Construction of Tikhonov-Type Regularization for Atmospheric Retrievals. *Atmosphere* **2020**, *11*, 1052. [CrossRef]
22. Tikhonov, A. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR* **1963**, *151*, 501–504.
23. Morozov, V. On the solution of functional equations by the method of regularization. *Dokl. Akad. Nauk SSSR* **1966**, *167*, 510–512.
24. Spherical Harmonic Discrete Ordinate Method (SHDOM) for Atmospheric Radiative Transfer. Available online: http://coloradolinux.com/shdom/ (accessed on 3 January 2020).
25. Doicu, A.; Efremenko, D. Linearizations of the Spherical Harmonic Discrete Ordinate Method (SHDOM). *Atmosphere* **2019**, *10*, 292. [CrossRef]
26. Levis, A.; Schechner, Y.; Aides, A.; Davis, A. Airborne Three-Dimensional Cloud Tomography. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015. [CrossRef]
27. Doicu, A.; Efremenko, D.; Trautmann, T. A Spectral Acceleration Approach for the Spherical Harmonics Discrete Ordinate Method. *Remote Sens.* **2020**, *12*, 3703. [CrossRef]
28. Ambartzumyan, V. The effect of the absorption lines on the radiative equilibrium of the outer layers of the stars. *Publ. Obs. Astron. Univ. Leningrad* **1936**, *6*, 7–18.
29. Efremenko, D.; Doicu, A.; Loyola, D.; Trautmann, T. Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals. *J. Quant. Spectrosc. Radiat. Transfer.* **2014**, *133*, 128–135. [CrossRef]
30. Henyey, L.C.; Greenstein, J.L. Diffuse radiation in the Galaxy. *Astrophys. J.* **1941**, *93*, 70. [CrossRef]
31. Rozanov, V.; Vountas, M. Radiative transfer equation accounting for rotational Raman scattering and its solution by the discrete-ordinates method. *J. Quant. Spectrosc. Radiat. Transf.* **2014**, *133*, 603–618. [CrossRef]

32. Wiscombe, W. The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. *J. Atmos. Sci.* **1977**, *34*, 1408–1422. [CrossRef]

33. Nakajima, T.; Tanaka, M. Algorithms for radiative intensity calculations in moderately thick atmos using a truncation approximation. *J. Quant. Spectrosc. Radiat. Transfer.* **1988**, *40*, 51–69. [CrossRef]

34. Várnai, T.; Davies, R. Effects of Cloud Heterogeneities on Shortwave Radiation: Comparison of Cloud-Top Variability and Internal Heterogeneity. *J. Atmos. Sci.* **1999**, *56*, 4206–4224. [CrossRef]

35. Richter, A.; Burrows, J. Tropospheric NO_2 from GOME measurements. *Adv. Space Res.* **2002**, *29*, 1673–1683. [CrossRef]

36. Bucsela, E.J.; Krotkov, N.A.; Celarier, E.A.; Lamsal, L.N.; Swartz, W.H.; Bhartia, P.K.; Boersma, K.F.; Veefkind, J.P.; Gleason, J.F.; Pickering, K.E. A new stratospheric and tropospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI. *Atmos. Meas. Tech.* **2013**, *6*, 2607–2626. [CrossRef]

37. Beirle, S.; Hörmann, C.; Jöckel, P.; Liu, S.; de Vries, M.; Pozzer, A.; Sihler, H.; Valks, P.; Wagner, T. The STRatospheric Estimation Algorithm from Mainz (STREAM): Estimating stratospheric NO_2 from nadir-viewing satellites by weighted convolution. *Atmos. Meas. Tech.* **2016**, *9*, 2753–2779. [CrossRef]