This is a preprint of the paper doi: 10.1016/j.jqsrt.2022.108177

Cloud tomographic retrieval algorithms. II: Adjoint method

Adrian Doicu[1], Alexandru Doicu[2], Dmitry S. Efremenko[1], Thomas Trautmann[1]

[1] *Remote Sensing Technology Institute, German Aerospace Centre, Oberpfaffenhofen, Wessling, Germany*
[2] *Institute of Mathematics, University of Augsburg, Germany*

Abstract

A cloud tomographic retrieval algorithm relying on (i) the spherical harmonics discrete ordinate method for radiative transfer calculation and (ii) the adjoint radiative transfer theory for computing the gradient of the objective function has been designed. In order to escape local minima and to increase the efficiency of the retrieval algorithm, the computation of the gradient of the objective function by the adjoint method has been combined with that of the gradient of a surrogate function. The retrieval algorithm uses regularization and accelerated projected gradient methods endowed with a step length procedure. The performances of the retrieval algorithm as compared to those of a retrieval algorithm based on the surrogate minimization method are analyzed on a few synthetic problems.

Keywords: Cloud tomography, 3D radiative transfer, cloud retrieval

1. Introduction

Accurate information about the cloud optical thickness is crucial for trace gas retrieval algorithms, as well as for determining the radiative impact of clouds on the Earth radiative budget [1, 2]. The traditional one-dimensional radiative transfer models assume horizontally homogeneous clouds. In the Independent Pixel Approximation (IPA), the horizontally inhomogeneous scene is divided into columns and a one-dimensional radiative transfer model is applied to each of them. However, as the horizontal transport of photons is neglected [3], IPA leads to systematic errors, while some important three-dimensional effects cannot be captured (e.g., side leakage and illumination [3], and radiative smoothing effects [4]). For example, in Ref. [5], it was shown that due to horizontal radiative transport, the cloud optical thickness errors can reach 45%. In this regard, for accurate cloud retrievals, three-dimensional radiative transfer models should be used. A promising approach for treating clouds as multi-dimensional objects is cloud tomography, in which the spatial distribution of the cloud extinction field can be obtained.

In atmospheric remote sensing, we are faced with the retrieval of some atmospheric parameters by minimizing a scalar-valued objective function (the misfit function) that quantifies the agreement with data of the radiances measured by a detector. The inverse problem can be solved by using first- or second-order optimization methods

1. In a second-order optimization method, as for example, the Gauss-Newton method, the objective function is approximated by a quadratic model around the current iterate, and the Gauss-Newton approximation to the Hessian matrix is used. This method requires the knowledge of the partial derivatives of the radiances measured by a detector with respect to the atmospheric parameter of interest. For this purpose, (i) a linearized forward approach, relying on an analytical computation of the derivatives, or (ii) a linearized forward-adjoint approach, relying on the application of the adjoint radiative transfer theory, can be used. For a plane-parallel geometry, scalar and vector linearized forward approaches were described in Ref. [6, 7], while linearized forward-adjoint approaches were proposed in Ref. [8, 9, 10, 11, 12, 13, 14, 15]. For a multi-dimensional geometry, linearizations of the spherical harmonic discrete ordinate method (SHDOM) [16] by means of a forward and a forward-adjoint approach were presented in Ref. [17]. Essentially, in this work, SHDOM was specialized for derivative calculations and radiative transfer problems involving the delta-M approximation, the TMS correction [18], and the adaptive grid splitting, while practical formulas for computing the derivatives in the spherical harmonics space were derived. The analysis was focused on the case of a detector that measures the radiance from a single location and viewing direction. As compared to the linearized forward approach, the linearized forward-adjoint method is extremely efficient because only two radiative transfer calculations are required for derivative calculations.
2. In a first-order optimization method, as for example, a gradient-based method, the objective function is approximated by a linear model around the current iterate, and only the knowledge of the gradient of the objective function is required. The gradient of the objective function can be computed by adjoint methods. It should be pointed out that in medical imaging, adjoint methods were used to retrieve tissue properties in multi-dimensions [19, 20, 21]. As compared to the radiative transfer equation employed in medical imaging, the equations used in atmospheric remote sensing are challenging due to the angular singularity of the incident solar radiation and the (near) angular and spatial singularity of the detector response functions. Applications of adjoint methods for atmospheric remote sensing were reported for one-dimensional [22], spherical [23, 24], pseudo-spherical [25, 26], and three-dimensional domains [27]. The theoretical approach developed in Ref. [27] was latter used in cloud tomography for the retrieval of two-dimensional cloud extinction fields from radiances measured by a detector from multiple locations and viewing angles [28]. Although the analysis was restricted to two-dimensional problems

(in which the set of directions is a circle defined by one angular variable, rather than a sphere defined by two angular variables), the efficiency of adjoint methods in cloud tomography was demonstrated.

Based on a method developed by Levis et al. [29], we presented in Ref. [30], a cloud tomographic retrieval algorithm that uses SHDOM for radiative transfer calculation and the surrogate minimization method for solving the inverse problem. The algorithm includes a surrogate step, in which the original objective function is substituted by a surrogate function, and a minimization step, in which the surrogate function is minimized by means of gradient-based methods. The key point in this approach is the choice of a surrogate function that on the one hand, approximates the objective function, and on the other hand, has an easily computable gradient. If in the minimization stage, only one iteration step is considered, the resulting approach can be regarded as a gradient-based method that uses an approximate gradient of the objective function.

In this paper, we present a cloud tomographic retrieval algorithm based on the adjoint method. More precisely, the inverse problem is solved by means of a gradient-based method that fuses the "accurate" gradient computed by the adjoint radiative transfer theory and the more computationally efficient surrogate function gradient. Our development is based on the idea revealed in Ref. [28] and the extension of the formalism presented in Ref. [17] to the case of a detector that measures the radiance from multiple location and viewing directions.

The paper is organized as follows. In Section 2 we describe in detail the computation of the gradient of the objective function in the framework of the adjoint radiative transfer theory. The particulars of the retrieval algorithm are presented in Section 3, and its numerical performances are analyzed in Section 4, while the final section of our paper contains a few concluding remarks.

2. Adjoint radiative transfer theory

In this section we present the methodology for computing the gradient of the objective function in the framework of the adjoint radiative transfer theory.

2.1. Forward and adjoint radiative transfer equations

A mathematical derivation of the basic results in three-dimensional adjoint radiative transfer is given in Appendix A. These are summarized below.

We consider the solar radiative transfer in a three-dimensional domain in the shape of rectangular prism with lengths L_{x}, L_{y} and L_{z}, bottom and top faces $S_{\mathrm{b}}(z=0)$ and S_{t} $(z=L_{\mathrm{z}})$, respectively, and lateral faces $S_{1\mathrm{x}}$ $(x=0)$, $S_{2\mathrm{x}}$ $(x=L_{\mathrm{x}})$, $S_{1\mathrm{y}}$ $(y=0)$, and $S_{2\mathrm{y}}$ $(y=L_{\mathrm{y}})$. The boundary-value problem for the total radiance I at point $\mathbf{r}$ in direction $\mathbf{\Omega}$, consists in the integro-differential equation

$$\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r},\mathbf{\Omega}) = -\sigma_{\mathrm{ext}}(\mathbf{r})I(\mathbf{r},\mathbf{\Omega}) + \frac{\sigma_{\mathrm{sct}}(\mathbf{r})}{4\pi}\int_{\Omega} P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')I(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}', \qquad (1)$$

and the boundary conditions at the top, bottom, and lateral faces, that is,

$$I(\mathbf{r}_{\mathrm{t}}, \mathbf{\Omega}^-) = \frac{F_0}{|\mu_0|}\delta(\mathbf{\Omega}^- - \mathbf{\Omega}_0), \quad \mathbf{r}_{\mathrm{t}} \in S_{\mathrm{t}}, \tag{2}$$

$$I(\mathbf{r}_{\mathrm{b}}, \mathbf{\Omega}^+) = \frac{A_{\mathrm{s}}}{\pi}\int_{\Omega^-} |\mu^-| I(\mathbf{r}_{\mathrm{b}}, \mathbf{\Omega}^-)\,\mathrm{d}\mathbf{\Omega}^-, \quad \mathbf{r}_{\mathrm{b}} \in S_{\mathrm{b}}, \tag{3}$$

and

$$I(\mathbf{r}_{1\mathrm{x}}, \mathbf{\Omega}) = I(\mathbf{r}_{2\mathrm{x}}, \mathbf{\Omega}), \quad I(\mathbf{r}_{1\mathrm{y}}, \mathbf{\Omega}) = I(\mathbf{r}_{2\mathrm{y}}, \mathbf{\Omega}), \tag{4}$$

respectively. Here, $\mathrm{d}I/\mathrm{d}s = \mathbf{\Omega}\cdot\nabla I$, σ_{ext} is the extinction coefficient, $\sigma_{\mathrm{sct}} = \omega\sigma_{\mathrm{ext}}$ the scattering coefficient, ω the single scattering albedo, P the phase function, F_0 the solar flux, $\mathbf{\Omega}_0 = (\mu_0, \varphi_0)$ with $\mu_0 < 0$, the solar direction, δ the Dirac delta function, and A_{s} the surface albedo. Furthermore, $\mathbf{\Omega}^+$ and $\mathbf{\Omega}^-$ denote an upward and a downward direction, respectively, Ω is the unit sphere, Ω^+ and Ω^- stand for the upper and lower unit hemispheres, respectively, $\mathbf{r}_{2\mathrm{x}} = \mathbf{r}_{1\mathrm{x}} + L_{\mathrm{x}}\mathbf{i}$, $\mathbf{r}_{2\mathrm{y}} = \mathbf{r}_{1\mathrm{y}} + L_{\mathrm{y}}\mathbf{j}$, and $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ are the Cartesian unit vectors.

In order to express the radiative transfer equation in an operator form, we define the forward transport operator $\mathcal{L}$ and the forward source term Q by the relations

$$\begin{aligned}(\mathcal{L}I)(\mathbf{r}, \mathbf{\Omega}) = {} & \frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r}, \mathbf{\Omega}) + \sigma_{\mathrm{ext}}(\mathbf{r})I(\mathbf{r}, \mathbf{\Omega}) - \frac{\sigma_{\mathrm{sct}}(\mathbf{r})}{4\pi}\int_{\Omega} P(\mathbf{r}, \mathbf{\Omega}, \mathbf{\Omega}')I(\mathbf{r}, \mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}' \\ & - \frac{A_{\mathrm{s}}}{\pi}\delta(z)H(\mu)\mu\int_{\Omega} H(-\mu')|\mu'|I(\mathbf{r}, \mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}',\end{aligned} \tag{5}$$

and

$$Q(\mathbf{r}, \mathbf{\Omega}) = F_0\delta(z - L_{\mathrm{z}})\delta(\mathbf{\Omega} - \mathbf{\Omega}_0), \tag{6}$$

respectively, where H is the Heaviside step function. In this setting, it can be shown that if the (forward) radiance I solves the inhomogeneous operator equation

$$(\mathcal{L}I)(\mathbf{r}, \mathbf{\Omega}) = Q(\mathbf{r}, \mathbf{\Omega}), \tag{7}$$

with the periodic boundary conditions (4) and the homogeneous boundary conditions

$$I(\mathbf{r}_{\mathrm{t}}, \mathbf{\Omega}^-) = 0,\ \mathbf{r}_{\mathrm{t}} \in S_{\mathrm{t}},\ \text{and}\ I(\mathbf{r}_{\mathrm{b}}, \mathbf{\Omega}^+) = 0,\ \mathbf{r}_{\mathrm{b}} \in S_{\mathrm{b}}, \tag{8}$$

then I solves the radiative transfer equation (1) with the boundary conditions (2)–(4). The converse result is also true.

In adjoint radiative transfer theory, we consider the adjoint transport operator $\mathcal{L}^\dagger$, defined through the relation

$$\langle \mathcal{L}I, I^\dagger \rangle = \langle I, \mathcal{L}^\dagger I^\dagger \rangle, \tag{9}$$

where the scalar product of the fields I_1 and I_2 is given by

$$\langle I_1, I_2 \rangle = \int_{\Omega}\int_{D} I_1(\mathbf{r}, \mathbf{\Omega})I_2(\mathbf{r}, \mathbf{\Omega})\,\mathrm{d}V\mathrm{d}\mathbf{\Omega}, \tag{10}$$

and D is the domain of the rectangular prism. The expression of the adjoint operator $\mathcal{L}^\dagger$, under the assumptions that (i) the forward radiance I satisfies the periodic and homogeneous boundary conditions (4) and (8), respectively, and (ii) the adjoint radiance $I^\dagger$ satisfies the periodic boundary conditions (4) and the homogeneous boundary conditions

$$I^\dagger(\mathbf{r}_\mathrm{t}, \boldsymbol{\Omega}^+) = 0, \ \mathbf{r}_\mathrm{t} \in S_\mathrm{t}, \ \text{and} \ I^\dagger(\mathbf{r}_\mathrm{b}, \boldsymbol{\Omega}^-) = 0, \ \mathbf{r}_\mathrm{b} \in S_\mathrm{b}, \tag{11}$$

is given by

$$\begin{aligned}(\mathcal{L}^\dagger I^\dagger)(\mathbf{r}, \boldsymbol{\Omega}) = &-\frac{\mathrm{d}I^\dagger}{\mathrm{d}s}(\mathbf{r}, \boldsymbol{\Omega}) + \sigma_\mathrm{ext}(\mathbf{r}) I^\dagger(\mathbf{r}, \boldsymbol{\Omega}) - \frac{\sigma_\mathrm{sct}(\mathbf{r})}{4\pi} \int_\Omega P(\mathbf{r}, \boldsymbol{\Omega}', \boldsymbol{\Omega}) I^\dagger(\mathbf{r}, \boldsymbol{\Omega}') \, \mathrm{d}\boldsymbol{\Omega}' \\ &- \frac{A}{\pi} \delta(z) H(-\mu) |\mu| \int_\Omega H(\mu') \mu' I^\dagger(\mathbf{r}, \boldsymbol{\Omega}') \, \mathrm{d}\boldsymbol{\Omega}'. \end{aligned} \tag{12}$$

The main result of the adjoint radiative transfer theory states that if (i) the radiance I solves the forward problem consisting in the operator equation $\mathcal{L}I = Q$ and the boundary conditions (4) and (8), and (ii) for some adjoint source term $Q^\dagger$, the radiance $I^\dagger$ solves the adjoint problem consisting in the operator equation $\mathcal{L}^\dagger I^\dagger = Q^\dagger$ and the boundary conditions (4) and (11), then (cf. Eq. (9))

$$\left\langle Q^\dagger, I \right\rangle = \left\langle \mathcal{L}^\dagger I^\dagger, I \right\rangle = \left\langle I^\dagger, \mathcal{L} I \right\rangle = \left\langle I^\dagger, Q \right\rangle. \tag{13}$$

Note that the adjoint source term depends on the radiative transfer quantity to be modeled. For example, the signal of a pixel detector that collects the radiances around the location $\mathbf{r}_\mathrm{tm} \in S_\mathrm{t}$ in the viewing direction $\boldsymbol{\Omega}_\mathrm{m} = (\mu_\mathrm{m}, \varphi_\mathrm{m})$ with $\mu_\mathrm{m} > 0$, is given by $I_\mathrm{m} = \int_D \int_\Omega \delta(\mathbf{r} - \mathbf{r}_\mathrm{tm}) \delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_\mathrm{m}) I(\mathbf{r}, \boldsymbol{\Omega}) \mathrm{d}V \mathrm{d}\boldsymbol{\Omega}$. In this regard, defining the adjoint source term by the relation $Q^\dagger(\mathbf{r}, \boldsymbol{\Omega}) = \delta(\mathbf{r} - \mathbf{r}_\mathrm{tm}) \delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_\mathrm{m})$, we obtain $I_\mathrm{m} = \left\langle Q^\dagger, I \right\rangle = \left\langle I^\dagger, Q \right\rangle$; thus, the signal of a detector pixel can be computed by taking the scalar product between the adjoint radiance $I^\dagger$ and the forward source term Q.

The forward and adjoint radiative transfer equations $\mathcal{L}I = Q$ and $\mathcal{L}^\dagger I^\dagger = Q^\dagger$, respectively, are related to each other. Replacing $\boldsymbol{\Omega}$ by $-\boldsymbol{\Omega}$ in the expression of the adjoint operator $\mathcal{L}^\dagger$, gives $\mathcal{L}^\dagger(-\boldsymbol{\Omega}) I^\dagger(\mathbf{r}, -\boldsymbol{\Omega}) = Q^\dagger(\mathbf{r}, -\boldsymbol{\Omega})$. Defining the pseudo-forward radiance $\widehat{I}^\dagger$ by the relation $\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}) = I^\dagger(\mathbf{r}, -\boldsymbol{\Omega})$ and using the symmetry properties of the phase function $P(\mathbf{r}, -\boldsymbol{\Omega}, -\boldsymbol{\Omega}') = P(\mathbf{r}, \boldsymbol{\Omega}', \boldsymbol{\Omega})$, yields $\mathcal{L}^\dagger(-\boldsymbol{\Omega}) I^\dagger(\mathbf{r}, -\boldsymbol{\Omega}) = \mathcal{L}(\boldsymbol{\Omega}) \widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega})$. Thus, the pseudo-forward radiance $\widehat{I}^\dagger$ solves the same type of radiative transfer equation as the forward radiance I, i.e., $\mathcal{L}\widehat{I}^\dagger = \widehat{Q}^\dagger$, where the pseudo-forward source term is defined by $\widehat{Q}^\dagger(\mathbf{r}, \boldsymbol{\Omega}) = Q^\dagger(\mathbf{r}, -\boldsymbol{\Omega})$. In this work, the forward and adjoint radiative transfer equations are solved by using SHDOM [16]. A brief summary of this approach is given in Ref. [30].

2.2. Gradient of the objective function

In cloud tomography, we use the radiances measured by a detector from multiple locations and viewing angles to retrieve the extinction coefficient $\sigma_\mathrm{ext}(\mathbf{r}_i)$

at a set of grid points $\{\mathbf{r}_i\}_{i=1}^{N_{\text{pts}}}$, where N_{pts} is the number of grid points. The grid point values $\sigma_{\text{ext}}(\mathbf{r}_i)$ are encapsulated in the extinction vector

$$\boldsymbol{\sigma}_{\text{ext}} = (\sigma_{\text{ext}}(\mathbf{r}_1), \ldots, \sigma_{\text{ext}}(\mathbf{r}_{N_{\text{pts}}}))^T.$$

The inversion process consists in an estimation of $\boldsymbol{\sigma}_{\text{ext}}$ that minimizes the objective function

$$E_\alpha(\boldsymbol{\sigma}_{\text{ext}}) = R(\boldsymbol{\sigma}_{\text{ext}}) + \alpha L(\boldsymbol{\sigma}_{\text{ext}}), \tag{14}$$

where

$$R(\boldsymbol{\sigma}_{\text{ext}}) = \frac{1}{2}\sum_{q=1}^{N_{\text{a}}}\sum_{p=1}^{N_{\text{p}}}[\mathcal{I}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}_{\text{ext}}) - \mathcal{I}_{\text{mes}}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}^\star_{\text{ext}})]^2 \tag{15}$$

is the residual, $L(\boldsymbol{\sigma}_{\text{ext}})$ a regularization term, α the regularization parameter,

$$\mathcal{I}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}_{\text{ext}}) = \frac{1}{A}\int_{S_{\text{t}}} h(\mathbf{r}_{\text{t}} - \mathbf{r}_{\text{t}p}) I(\mathbf{r}_{\text{t}}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}_{\text{ext}})\,\mathrm{d}S_{\text{t}}, \tag{16}$$

the simulated signal of the pth detector pixel that collects the radiances around the location $\mathbf{r}_{\text{t}p} \in S_{\text{t}}$, $p = 1, \ldots, N_{\text{p}}$ in the viewing direction $\boldsymbol{\Omega}_{\text{m}q}$, $q = 1, \ldots, N_{\text{a}}$, $h(\mathbf{r}_{\text{t}} - \mathbf{r}_{\text{t}p})$ the characteristic function of the pth detector pixel projected on the top surface, $A = L_{\text{x}}L_{\text{y}}$ the area of the top face of the prism, $\mathcal{I}_{\text{mes}}$ the measured signal, and $\boldsymbol{\sigma}^\star_{\text{ext}}$ the true extinction vector to be retrieved.

A gradient-based optimization method requires the computation of the gradient of the objective function with respect to the extinction field, that is, $\mathbf{g}(\boldsymbol{\sigma}_{\text{ext}}) = \nabla E_\alpha(\boldsymbol{\sigma}_{\text{ext}}) = \mathbf{g}_{\text{R}}(\boldsymbol{\sigma}_{\text{ext}}) + \alpha\mathbf{g}_{\text{L}}(\boldsymbol{\sigma}_{\text{ext}})$ with $\mathbf{g}_{\text{R}}(\boldsymbol{\sigma}_{\text{ext}}) = \nabla R(\boldsymbol{\sigma}_{\text{ext}})$ and $\mathbf{g}_{\text{L}}(\boldsymbol{\sigma}_{\text{ext}}) = \nabla L(\boldsymbol{\sigma}_{\text{ext}})$. The most challenging task is the computation of gradient of the residual $\mathbf{g}_{\text{R}}(\boldsymbol{\sigma}_{\text{ext}})$. Setting $\sigma_{\text{ext}u} = \sigma_{\text{ext}}(\mathbf{r}_u)$ for some $u = 1, \ldots, N_{\text{pts}}$, the uth component of the gradient $\mathbf{g}_{\text{R}}(\boldsymbol{\sigma}_{\text{ext}})$ is found to be (cf. Eq. (15))

$$\begin{aligned}
g_{\text{R}u}(\boldsymbol{\sigma}_{\text{ext}}) &= \frac{\partial R}{\partial \sigma_{\text{ext}u}}(\boldsymbol{\sigma}_{\text{ext}}) \\
&= \frac{1}{A}\int_{S_{\text{t}}}\sum_{q=1}^{N_{\text{a}}}\sum_{p=1}^{N_{\text{p}}}[\mathcal{I}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}_{\text{ext}}) - \mathcal{I}_{\text{mes}}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}^\star_{\text{ext}})] \\
&\quad \times h(\mathbf{r}_{\text{t}} - \mathbf{r}_{\text{t}p})\frac{\partial I}{\partial \sigma_{\text{ext}u}}(\mathbf{r}_{\text{t}}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}_{\text{ext}})\,\mathrm{d}S_{\text{t}} \\
&= \int_\Omega\int_V \Big\{\sum_{q=1}^{N_{\text{a}}}\Big[\sum_{p=1}^{N_{\text{p}}} f_{pq}(\boldsymbol{\sigma}_{\text{ext}})h(\mathbf{r}_{\text{t}} - \mathbf{r}_{\text{t}p})\Big]\delta(z - L_{\text{z}})\delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_{\text{m}q})\Big\} \\
&\quad \times \frac{\partial I}{\partial \sigma_{\text{ext}u}}(\mathbf{r}, \boldsymbol{\Omega}; \boldsymbol{\sigma}_{\text{ext}})\,\mathrm{d}V\mathrm{d}\boldsymbol{\Omega},
\end{aligned} \tag{17}$$

where

$$f_{pq}(\boldsymbol{\sigma}_{\text{ext}}) = \frac{1}{A}[\mathcal{I}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}_{\text{ext}}) - \mathcal{I}_{\text{mes}}(\mathbf{r}_{\text{t}p}, \boldsymbol{\Omega}_{\text{m}q}; \boldsymbol{\sigma}^\star_{\text{ext}})]. \tag{18}$$

From Eq. (17), it is apparent that $g_{\mathrm{R}u}$ can be written as

$$g_{\mathrm{R}u} = \Big\langle Q^\dagger, \frac{\partial I}{\partial \sigma_{\mathrm{ext}u}} \Big\rangle, \tag{19}$$

where the adjoint source term is defined by

$$Q^\dagger(\mathbf{r}, \mathbf{\Omega}; \boldsymbol{\sigma}_{\mathrm{ext}}) = \sum_{q=1}^{N_\mathrm{a}} F_{0q}^\dagger(\mathbf{r}_\mathrm{t}; \boldsymbol{\sigma}_{\mathrm{ext}}) \delta(z - L_\mathrm{z}) \delta(\mathbf{\Omega} - \mathbf{\Omega}_{\mathrm{m}q}), \tag{20}$$

$$F_{0q}^\dagger(\mathbf{r}_\mathrm{t}; \boldsymbol{\sigma}_{\mathrm{ext}}) = \sum_{p=1}^{N_\mathrm{p}} f_{pq}(\boldsymbol{\sigma}_{\mathrm{ext}}) h(\mathbf{r}_\mathrm{t} - \mathbf{r}_{\mathrm{t}p}) \tag{21}$$

and f_{pq} is given by Eq. (18). Consequently and by taking into account that from $\mathcal{L}I = Q$ and $\partial Q / \partial \sigma_{\mathrm{ext}u} = 0$, the relation $\mathcal{L}(\partial I / \partial \sigma_{\mathrm{ext}u}) = -(\partial \mathcal{L} / \partial \sigma_{\mathrm{ext}u}) I$ readily follows, we obtain

$$g_{\mathrm{R}u} = \Big\langle Q^\dagger, \frac{\partial I}{\partial \sigma_{\mathrm{ext}u}} \Big\rangle = \Big\langle \mathcal{L}^\dagger I^\dagger, \frac{\partial I}{\partial \sigma_{\mathrm{ext}u}} \Big\rangle = \Big\langle I^\dagger, \mathcal{L} \frac{\partial I}{\partial \sigma_{\mathrm{ext}u}} \Big\rangle = -\Big\langle I^\dagger, \frac{\partial \mathcal{L}}{\partial \sigma_{\mathrm{ext}u}} I \Big\rangle. \tag{22}$$

Thus, identifying the adjoint source term from Eq. (19) and solving the corresponding adjoint radiative transfer equation $\mathcal{L}^\dagger I^\dagger = Q^\dagger$, we can compute $g_{\mathrm{R}u}$ by means of Eq. (22). The computation of $g_{\mathrm{R}u}$ by using SHDOM is described in Appendix B. Essentially, we extended the formalism presented in Ref. [17] to cloud tomography, that is, to the case of a detector that measures the radiances from multiple locations and viewing angles.

3. Retrieval algorithm

The retrieval algorithm uses gradient-based methods. In the standard gradient descent method, we start with a guess $\boldsymbol{\sigma}_{\mathrm{ext}0}$ for a local minimum of E_α, and consider the sequence $\boldsymbol{\sigma}_{\mathrm{ext}0}$, $\boldsymbol{\sigma}_{\mathrm{ext}1}$, $\boldsymbol{\sigma}_{\mathrm{ext}2}$,..., such that

$$\boldsymbol{\sigma}_{\mathrm{ext}k+1} = \boldsymbol{\sigma}_{\mathrm{ext}k} - \tau_k \nabla E_\alpha(\boldsymbol{\sigma}_{\mathrm{ext}k}),$$

where τ_k is the step length. At each iteration step k, the value of the objective function and its gradient are computed with two calls of the radiative transfer model. In the first call, $E_\alpha(\boldsymbol{\sigma}_{\mathrm{ext}k})$ is computed by solving the forward radiative transfer equation, while in the second call, $\mathbf{g}(\boldsymbol{\sigma}_{\mathrm{ext}k}) = \nabla E_\alpha(\boldsymbol{\sigma}_{\mathrm{ext}k})$ is computed by solving the adjoint radiative transfer equation. The forward and adjoint problems are solved successively on the same grid; in this way, the interpolation between different grids is avoided. Actually, in the first step, the forward problem is solved by using the adaptive grid procedure with a prescribed splitting accuracy, and in the second step, the adjoint problem is solved on the resulting grid without splitting.

The design of the retrieval algorithm is based on the following considerations.

1. In cloud tomography, the objective function may have several local minima. The simplest techniques for escaping from local minima are (i) restarting and (ii) using non-improving steps. In the first case, the search is re-initialized whenever a local optimum is encountered. Specifically, when the line search algorithm fails to find a step length for which the descent condition is satisfied, the algorithm is restarted with a non-improving step. In the second case, solutions with approximate gradient values are used. In machine learning, this technique is equivalent to the use of stochastic gradient descent, mini-batch gradient descent, and/or momentum. More precisely, for regression problems in which the entire data set is grouped in batches and each batch in samples, the loss function to be minimized is the sum of the squared errors over the entire data set (over all batches and samples). In stochastic gradient method, at each iteration step, the gradient of the loss function over the entire data set is approximated by the gradient corresponding to one batch and one sample, while in mini-batch gradient methods, the gradient of the loss function is approximated by the gradient corresponding to one batch and all samples contained in the batch. These methods jump around the critical point much more than gradient methods, so that they can more easily escape from a small domain around a critical point. On the other hand, momentum based methods, in which the step is not taken in the current gradient direction but in a direction that combines the current and the previous recent gradient direction, are more likely to move through a small domain without stopping (when they are on a side of a critical point, but still in the domain of attraction, they only decrease their momentum instead of switching the direction).
2. The computation of the gradient by the adjoint theory can be a time consuming process, especially when the number of grid cells is large. The reason is that two radiative transfer simulations and one integration step (according to Eqs. (B.14)–(B.19) of Appendix B) have to be performed. To speed up the retrieval, we may compute $g_{\mathrm{R}u}(\boldsymbol{\sigma}_{\mathrm{ext}})$ as in Eq. (17), but in which the derivative of the diffuse radiance at point $\mathbf{r}_{\mathrm{t}}$ in direction $\boldsymbol{\Omega}_{\mathrm{m}q}$,

$$I_{\mathrm{d}}(\mathbf{r}_{\mathrm{t}},\boldsymbol{\Omega}_{\mathrm{m}q};\boldsymbol{\sigma}_{\mathrm{ext}}) = I_{\mathrm{d}}(\mathbf{r}_{\mathrm{b}},\boldsymbol{\Omega}_{\mathrm{m}q};\boldsymbol{\sigma}_{\mathrm{ext}})T(\mathbf{r}_{\mathrm{t}},\mathbf{r}_{\mathrm{b}},\boldsymbol{\Omega}_{\mathrm{m}q};\boldsymbol{\sigma}_{\mathrm{ext}}) + \int_{\mathbf{r}_{\mathrm{b}}}^{\mathbf{r}_{\mathrm{t}}} \sigma_{\mathrm{ext}}(\mathbf{r})J(\mathbf{r},\boldsymbol{\Omega}_{\mathrm{m}q};\boldsymbol{\sigma}_{\mathrm{ext}};I_{\mathrm{d}})T(\mathbf{r}_{\mathrm{t}},\mathbf{r},\boldsymbol{\Omega}_{\mathrm{m}q};\boldsymbol{\sigma}_{\mathrm{ext}})\,\mathrm{d}s.$$

is calculated by assuming that the single-scattering radiance depends on the extinction fields [30]. Consequently, the computation of $g_{\mathrm{R}u}(\boldsymbol{\sigma}_{\mathrm{ext}})$ requires only the knowledge of the derivatives of the single-scattering radiance, which in turn, requires only the knowledge of the derivatives of transmission function T. Because these derivatives can be computed analytically in a very simple manner, the complexity of the problem, as well as the computational time can be substantially reduced. The resulting method is similar to the surrogate minimization method, provided that in the minimization stage only one iteration step is considered. In fact, $\mathbf{g}(\boldsymbol{\sigma}_{\mathrm{ext}}) = \mathbf{g}_{\mathrm{R}}(\boldsymbol{\sigma}_{\mathrm{ext}}) + \alpha\mathbf{g}_{\mathrm{L}}(\boldsymbol{\sigma}_{\mathrm{ext}})$ computed in this way is nothing

else than the gradient of the surrogate function $S_\alpha(\boldsymbol{\sigma}_{\rm ext}|\boldsymbol{\sigma}_{{\rm ext}k})$, that is, $\mathbf{g}(\boldsymbol{\sigma}_{\rm ext}) = \nabla S_\alpha(\boldsymbol{\sigma}_{\rm ext}|\boldsymbol{\sigma}_{{\rm ext}k})$. An important feature of the surrogate minimization method is that, like any method using approximate gradients, it does not suffer from the problem of getting stuck in local minima [29, 30].

The idea of the retrieval algorithm is to start with an initial guess $\boldsymbol{\sigma}_{\rm ext0}$ for an atmosphere with no cloud, and to switch from the gradient of the objective function (computed by the adjoint method) to the gradient of the surrogate function, when the relative decrease in the objective function is small, that is,

$$0 < E_\alpha(\boldsymbol{\sigma}_{{\rm ext}k}) - E_\alpha(\boldsymbol{\sigma}_{{\rm ext}k+1}) \leq \varepsilon_1 E_\alpha(\boldsymbol{\sigma}_{{\rm ext}k}), \tag{23}$$

for some prescribed tolerance ε_1. In other words, when there is no significant reduction of the objective function and there is a risk that the algorithm gets stuck in a local minimum, the iteration is continued with an approximate gradient. The approximate gradient is the gradient of the surrogate function, which physically corresponds to the single-scattering approximation. In this way, we expect to accelerate the computation and to escape from possible local minima.

The retrieval algorithm borrows many feature from that described in Ref. [30]; more precisely,

1. the regularization term is constructed by using averaging, Gaussian, and median low-pass filters from image processing;
2. the regularization method is the iteratively regularized Gauss-Newton method, in which the regularization parameter is gradually decreased during the iteration; specifically, a decreasing geometric sequence of regularization parameters $\alpha_{k+1} = q\alpha_k$, where $q < 1$ is the ratio of the sequence, is used;
3. the first-order optimization algorithms are the Nesterov acceleration for the projected gradient method [32] and the BFGS algorithm [33];
4. the optimization problem is solved by imposing simple bounds on the variables (box constraints);
5. a line search procedure relying on the descent condition is used to compute the step length τ, that is,

$$E_\alpha(\boldsymbol{\sigma}_{{\rm ext}k\tau}) \leq E_\alpha(\boldsymbol{\sigma}_{{\rm ext}k}), \tag{24}$$

   where $\boldsymbol{\sigma}_{{\rm ext}k\tau} = \mathcal{P}_B(\boldsymbol{\sigma}_{{\rm ext}k} + \tau\Delta\boldsymbol{\sigma}_{{\rm ext}k})$, $\Delta\boldsymbol{\sigma}_{{\rm ext}k}$ is the step delivered by the optimization algorithm, and $\mathcal{P}_B$ is the projection operator onto the box constraints B (note that the descent condition corresponds to the Armijo–Goldstein condition with a very small fraction of the rate of decrease in the negative direction of the gradient);
6. the iteration is stopped according to the absolute residual convergence test, that is, $R(\boldsymbol{\sigma}_{{\rm ext}k}) \leq \varepsilon_{\rm AF} R(\boldsymbol{\sigma}_{\rm ext0})$, for some prescribed tolerance $\varepsilon_{\rm AF}$.

Some comments are in order.

1. Like a momentum method, the Nesterov acceleration method does not use the current gradient direction, i.e., the gradient is not computed with respect to the current step but with respect to a feature step. As in machine learning when the stochastic gradient descent or the mini-batch gradient descent are used in conjunction with momentum, in the present approach, the approximate gradient of the surrogate function is used in conjunction with the Nesterov acceleration method.
2. In the standard implementation of BFGS, the curvature condition, used by the line search procedure, is disregarded.
3. By interpreting $\varepsilon_{\mathrm{AF}} R(\boldsymbol{\sigma}_{\mathrm{ext0}})$ as the noise level, the absolute residual convergence test is equivalent with the discrepancy principle, which is the standard stopping rule for iterative regularization methods.
4. The gradient of the surrogate function corresponds to the single-scattering approximation and we expect this approximation to be inaccurate at points where the radiance has undergo multiple scattering events. These points are at a large optical distance from the sensors, and determine the so-called "veiled core" of the cloud [31]. But in the veiled core, the multiple scattering process cause a loss of sensitivity to the extinction field, and so, a reduction of the magnitude of the gradient. Thus, in the veiled core, the absolute errors in the gradient are not large. Unfortunately, outside of the veiled core, there is no guarantee that the errors in the gradient are small, even when the cloud itself is optically thick. However, in the retrieval algorithm, the accuracy of the approximate gradient is not decisive. The reason is that we are dealing with gradient approximations and not with function approximations, and in the line search procedure (24), the step length τ is computed such that the descent condition is satisfied for the objective function E_α and not for the surrogate function S_α. From this point of view, the algorithm can be regarded as an iterative method that uses the decrease of the objective function as a natural measure of progress. In the next section, the validity of the method will be checked through a numerical analysis.

In the following, the above algorithm (in which the gradient of the objective function is replaced by the gradient of the surrogate function when no significant reduction of the objective function occurs) will be referred to as the hybrid algorithm, while he algorithm described in Ref. [30] (in which the objective function is entirely substituted by a surrogate function, and in the minimization step, several iteration steps are used to reduce the surrogate function) will be referred to as the surrogate-minimization algorithm.

4. Numerical Simulations

In our numerical analysis, we choose the scene parameters as in Ref. [30], namely, (i) the cloud single-scattering albedo and the phase function are computed by Mie theory at a wavelength of 672 nm and for a Gamma size distribu-

tion

$$p(a) \propto a^{\alpha} \exp\left[-\alpha\left(\frac{a}{a_{\mathrm{mod}}}\right)\right] \tag{25}$$

with an effective radius $a_{\mathrm{eff}} = 10\ \mu$m, a modal radius $a_{\mathrm{mod}} = 2a_{\mathrm{eff}}/3$, and a size distribution parameter $\alpha = 6$, (ii) in addition to the cloud, molecular Rayleigh scattering is considered as background, (iii) the solar zenith angle is $\theta_0 = 0°$, and a Lambertian reflecting surface with the surface albedo $A_{\mathrm{s}} = 0.05$ is chosen, (iv) there are $N_{\mathrm{a}} = 9$ viewing zenith angles corresponding to the Multiangle Imaging SpectroRadiometer (MISR); these are given by $\pm 70.5°$, $\pm 60°$, $\pm 45.6°$, $\pm 26.1°$, and $0°$, where $\pm$ indicates the forward- and aftward-facing cameras, (v) the number of discrete ordinates is $(N_{\mu} = 32) \times (N_{\varphi} = 2N_{\mu} = 64)$, and the spherical harmonics truncation indices are $N = N_{\mu} - 1$ and $M = N_{\varphi}/2 - 1$, (vi) the delta-M scaling method, the TMS correction, and an adaptive grid with a splitting accuracy of 10^{-3} are used, and (vii) the solution and spherical harmonic series accuracies are 10^{-5} and 10^{-6}, respectively. The discrete domain of analysis has (i) the lengths $L_{\mathrm{x}} = L_{\mathrm{y}} = 6\,$km and $L_{\mathrm{z}} = 2\,$km, (ii) the numbers of base grid points $N_{\mathrm{x}} = N_{\mathrm{y}} = 31$ and $N_{\mathrm{z}} = 11$, and (iii) the base grid spacings $\Delta x = \Delta y = \Delta z = 200\,$m. In Ref. [30], the characteristic function of the detector $h(\mathbf{r}_{\mathrm{t}} - \mathbf{r}_{\mathrm{t}p})$ was chosen as a box function. For such a characteristic function with abrupt changes, the pseudo-forward direct beam $\widehat{I}^{\dagger}_{\odot}(\mathbf{r}, \mathbf{\Omega})$ (given by Eqs. (B.8)–(B.9) of Appendix B in conjunction with Eq. (21), that is, $F^{\dagger}_{0q}(\mathbf{r}_{\mathrm{t}}, \cdot) = \sum_{p=1}^{N_{\mathrm{p}}} f_{pq} h(\mathbf{r}_{\mathrm{t}} - \mathbf{r}_{\mathrm{t}p})$) will have spatial discontinuities and the solution of the adjoint problem will be a challenging task. The reason is that the adaptive grid will supply extra spatial resolution along the boundaries of the pseudo-forward direct beam, and so, the number of adaptive grid cells will significantly increase as compared to those which are already required for a fine discretization of the cloud extinction field [17]. Therefore, we prefer to use the Gaussian characteristic function

$$h(\mathbf{r}_{\mathrm{t}} - \mathbf{r}_{\mathrm{t}p}) = \begin{cases} \exp\left(-\frac{||\mathbf{r}_{\mathrm{t}} - \mathbf{r}_{\mathrm{t}p}||_2^2}{2\sigma_{\mathrm{d}}^2}\right), & ||\mathbf{r}_{\mathrm{t}} - \mathbf{r}_{\mathrm{t}p}||_2 \leq 2\sigma_{\mathrm{d}} \\ 0, & \text{otherwise} \end{cases},$$

where $\sigma_{\mathrm{d}} = \Delta x/2$ is the half-width of the characteristic function of a detector pixel. By choosing a smooth decreasing characteristic function, the number of adaptive grid cells is reduced.

Before proceeding we make a parenthetic remark. According to Hadamard, an operator equation is called well-posed if (i) the operator is surjective (existence of solutions). (ii) the operator is injective (uniqueness of the solution), and (iii) the inverse operator is continuous (stability of the solution). A fundamental result in regularization theory states that non-uniqueness can be a consequence of the discretization, or more precisely, of the underlying interpolation scheme. To demonstrate this result, we consider a box cloud with the extinction field

$$\sigma^{\mathrm{cld}\star}_{\mathrm{ext}}(x, y, z) = \sigma_{\max} \chi(x, y, z), \tag{26}$$

$$\chi(x, y, z) = \begin{cases} 1, & x_1 \leq x \leq x_2,\ y_1 \leq y \leq y_2 \text{ and } z_1 \leq z \leq z_2 \\ 0, & \text{rest} \end{cases} \tag{27}$$

where $\sigma_{\max} = 3\,\mathrm{km}^{-1}$, $x_1 = y_1 = 1.2\,\mathrm{km}$, $x_2 = y_2 = 4.4\,\mathrm{km}$, $z_1 = 0.7\,\mathrm{km}$, $z_2 = 1.65\,\mathrm{km}$, and $\tau_{\max} = 3.6$. We keep fixed the cloud top height z_2, and change the cloud geometrical thickness $h = z_2 - z_1$ and the cloud extinction coefficient $\sigma_{\max}$. For h, we take 21 equidistant values in the range $[0.4,\ 1.2\,\mathrm{km}]$, while for $\sigma_{\max}$, we take 101 equidistant values in the range $[0.5,\ 5.0\ \mathrm{km}^{-1}]$. Then, for each pair $(h, \sigma_{\max})$ we compute the residual $R(\boldsymbol{\sigma}_{\mathrm{ext}})$. Note that the computations are performed without adaptive grid splitting. As interpolation schemes, we use (i) the piecewise constant interpolation (nearest-neighbor interpolation), that is, for $z_{k-1} \le z_1 \le z_k$, we take $\chi(\cdot, z_k) = 1$ and $\chi(\cdot, z_{k-1}) = 0$, and (ii) the piecewise linear interpolation, that is, for $z_{k-1} \le z_1 \le z_k$, we take $\chi(\cdot, z_k) = 1$ and $\chi(\cdot, z_{k-1}) = (z_k - z_1)/(z_k - z_{k-1})$. Thus, we interpolate the characteristic function, which is equivalent with the interpolation of the extinction field. The plots in Fig. 1a, corresponding to the piecewise constant interpolation, illustrate that in addition to the global minimum at $(h = 0.95\ \mathrm{km}, \sigma_{\max} = 3\,\mathrm{km}^{-1})$, the residual function has two local minima, while the plots in Fig. 1b, corresponding to the piecewise linear interpolation, illustrate that the residual function is convex with a single minimum. The explanation is that in the first case, the rough base grid resolution of 200 m cannot reproduce the small changes in the cloud geometrical thickness of about 80 m. Fortunately, SHDOM uses a trilinear interpolation scheme for the extinction field, so that in the first instant, we expect that local minima will not occur.

In our numerical analysis, we consider the retrieval of two three-dimensional clouds:

1. a Gaussian cloud extinction field

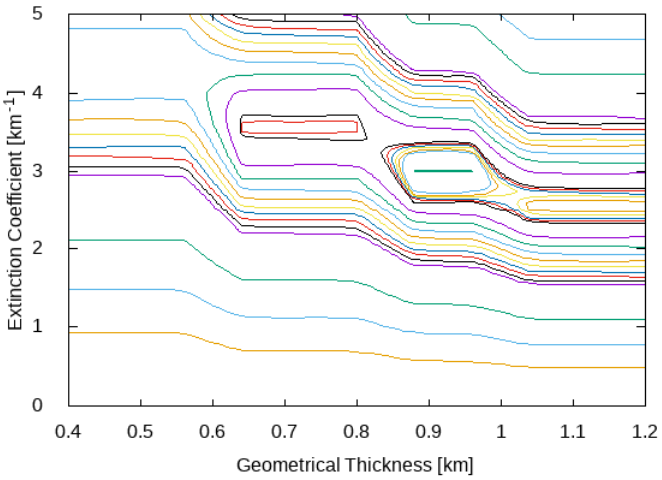
$$\sigma_{\mathrm{ext}}^{\mathrm{cld}\star}(x, y, z) = \sigma_{\max}\chi(x, y, z), \tag{28}$$

$$\chi(x, y, z) = \exp\Big[-\frac{(x - x_0)}{\sigma_{\mathrm{x}}^2} - \frac{(y - y_0)}{\sigma_y^2} - \frac{(z - z_0)}{\sigma_z^2}\Big], \tag{29}$$

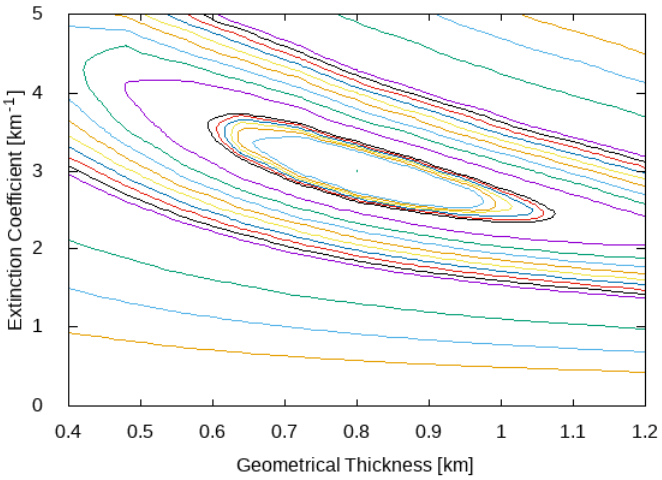
   with $\sigma_{\max} = 5\,\mathrm{km}^{-1}$, $x_0 = L_{\mathrm{x}}/2$, $y_0 = L_{\mathrm{y}}/2$, $z_0 = L_{\mathrm{z}}/2$, and $\sigma_{\mathrm{x,y,z}} = L_{\mathrm{x,y,z}}/5$, and
2. the three-dimensional cloud extinction field $\sigma_{\mathrm{ext}}^{\mathrm{cld}\star}(x, y, z) = \sigma_{\max}\chi(x, y, z)$ with $\sigma_{\max} = 8\,\mathrm{km}^{-1}$, obtained from a large eddy simulation (LES) of stratocumulus [16].

The number of unknowns is $N_{\mathrm{pts}} = 10571$. The following settings are used in the retrieval.

1. The Nesterov acceleration and the BFGS algorithm are used as optimization methods, and the iteration is stopped when the number of iteration steps exceeds 200 (the absolute residual convergence test is disregarded).
2. The regularization term corresponds to an averaging filter, the initial value of the regularization parameter is $\alpha_0 = 0.1$, and the ratio of the geometric sequence of regularization parameters is $q = 0.5$. In our simulations, we do not optimize the ratio of the geometric sequence for each test problem, because in practice, there is no error-free parameter choice method for



(a)



(b)

Figure 1: Residual function for a box cloud with different geometrical thicknesses h and extinction coefficient $\sigma_{\max}$. The results correspond to piecewise constant (a) and piecewise linear (b) interpolation schemes of the cloud geometrical thickness.

computing q (in the noise region, the residual cannot be used as a selection criterion).

3. In the hybrid method, the gradient of the surrogate function is computed by assuming that the single-scattering radiance changes during the minimization step, or equivalently, that the multiple-scattering components of the source function and the surface radiance are constant during the minimization step [30]. We switch to an approximate gradient when there is no significant reduction of the objective function, that is, when the condition (23) is satisfied with $\varepsilon_1 = 0.02$.
4. In the case of the Gaussian cloud, the initial extinction field corresponds to an atmosphere with no clouds. For the LES cloud, we use the ray data casting method to estimate the bounds of the cloud shape [34, 35]. Outside the shape bound, the extinction field is set to that of a Rayleigh atmosphere, while within the estimated shape bound, the extinction field is initialized as homogeneous with $\boldsymbol{\sigma}_{\text{ext0}} = 0.5\,\text{km}^{-1}$.
5. The initial and the minimum step lengths in the line search algorithm are $\tau_0 = 200$ and $\tau_{\min} = 2 \times 10^{-4}$, respectively.

Essentially, we perform an inverse crime retrieval, because the same theoretical ingredients (representation of the state, radiance field, and instrument sampling) are employed to synthesize as well as to invert the data of the inverse problem.

In the first numerical experiment, we demonstrate that for cloud extinction fields with uniform small-scale distributions, local minima may appear. For this purpose, we proceed as in Ref. [30], that is, we consider a perturbation of the Gaussian cloud extinction field $\sigma_{\text{ext}}^{\text{cld}\star}(x,y,z) = \sigma_{\max}\chi(x,y,z)$ given by Eqs. (28) and (29). Specifically, the perturbed cloud extinction field is $\sigma_{\text{ext}}^{\text{cld}}(x,y,z) = \sigma_{\max}\chi_\varepsilon(x,y,z)$, where $\chi_\varepsilon(x,y,z) = (1+\sigma_{\text{f}}\varepsilon)\chi(x,y,z)$, ε is a random number uniformly distributed in the interval $(-1,1)$, and $\chi(x,y,z)$ is given by Eq. (29). In Fig. 2, we plot the residual $R(\boldsymbol{\sigma}_{\text{ext}})$ versus the relative distance $d = \left\|\boldsymbol{\sigma}_{\text{ext}} - \boldsymbol{\sigma}_{\text{ext}}^\star\right\|_2 / \left\|\boldsymbol{\sigma}_{\text{ext}}^\star\right\|_2$ for different values of the amplitude σ_{f} and by considering 500 configurations for each σ_{f}. The plots reveal that (i) $R(\boldsymbol{\sigma}_{\text{ext}}) \to 0$ as $\sigma_{\text{f}} \to 0$, (ii) for a given value of the amplitude σ_{f}, the points are clustered in a domain, in which the residual may decrease by one order of magnitude, and (ii) the same residual may correspond to points situated in different clusters, that is, to points $\boldsymbol{\sigma}_{\text{ext}}$ situated at different distances with respect to the true solution $\boldsymbol{\sigma}_{\text{ext}}^\star$. The latter result implies that the residual $R(\boldsymbol{\sigma}_{\text{ext}})$ have critical points. Indeed, taking two distinct points $\boldsymbol{\sigma}_{\text{ext}}^{(a)}$ and $\boldsymbol{\sigma}_{\text{ext}}^{(b)}$ (from the same cluster or from different clusters) with $R(\boldsymbol{\sigma}_{\text{ext}}^{(a)}) \approx R(\boldsymbol{\sigma}_{\text{ext}}^{(b)})$, the mean value theorem $R(\boldsymbol{\sigma}_{\text{ext}}^{(a)}) - R(\boldsymbol{\sigma}_{\text{ext}}^{(b)}) = \nabla R(\boldsymbol{\sigma}_{\text{ext}}^{(c)}) \cdot (\boldsymbol{\sigma}_{\text{ext}}^{(a)} - \boldsymbol{\sigma}_{\text{ext}}^{(b)})$ shows that on the line segment between $\boldsymbol{\sigma}_{\text{ext}}^{(a)}$ and $\boldsymbol{\sigma}_{\text{ext}}^{(b)}$, there exist a point $\boldsymbol{\sigma}_{\text{extz}}^{(c)}$ such that $\nabla R(\boldsymbol{\sigma}_{\text{ext}}^{(c)}) \approx \mathbf{0}$. Thus, for non-smooth cloud extinction fields with uniform small-scale distributions, there is the risk that local minima occur. In principle, the regularization term, corresponding to an averaging filter, should avoid the appearance of non-smooth cloud extinction fields. However, because no infallible parameter choice method exists, or equivalently, the amount of regularization cannot be optimally con-

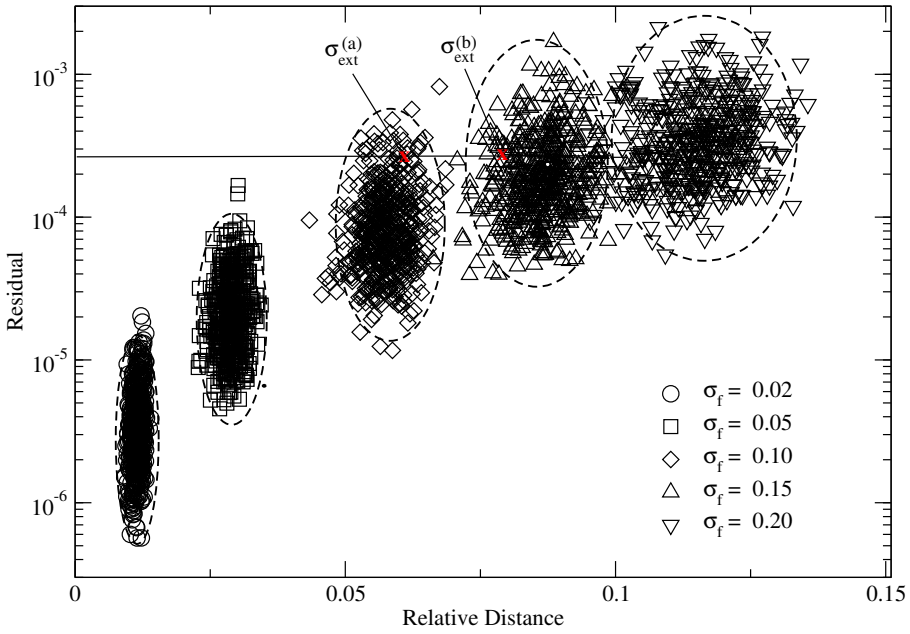


Figure 2: Residual $R(\boldsymbol{\sigma}_{\text{ext}})$ versus the relative distance $d = \|\boldsymbol{\sigma}_{\text{ext}} - \boldsymbol{\sigma}^{\star}_{\text{ext}}\|_2 / \|\boldsymbol{\sigma}^{\star}_{\text{ext}}\|_2$ for a non-smooth Gaussian cloud extinction field $\sigma^{\text{cld}}_{\text{ext}}(x, y, z) = \sigma_{\max}\chi_{\varepsilon}(x, y, z)$, where $\chi_{\varepsilon}(x, y, z) = (1+\sigma_{\text{f}}\varepsilon)\chi(x, y, z)$ and $\varepsilon \sim \text{U}(-1, 1)$. The results correspond to different values of the amplitude σ_{f} and 500 configurations for each σ_{f}. According to the mean value theorem, on the line segment between the points $\boldsymbol{\sigma}^{(a)}_{\text{ext}}$ and $\boldsymbol{\sigma}^{(b)}_{\text{ext}}$, which belong to different clusters and have approximately the same residual, there exist a point $\boldsymbol{\sigma}^{(c)}_{\text{ext}}$ such that $\nabla R(\boldsymbol{\sigma}^{(c)}_{\text{ext}}) \approx \mathbf{0}$.

trolled, it may happens that during the iterative process, such distributions occur and the algorithm gets stuck in a local minimum.

In Fig. 3 we illustrate the histories of the objective function for a restarted algorithm that is completely based on the adjoint method for gradient calculation and the hybrid algorithm. In the first case, the objective function drops quickly for some iteration steps and then decreases slowly toward a local minimum. When the line search algorithm fails to find a step length $\tau_k \geq \tau_{\min} = 2 \times 10^{-4}$ for which the descent condition $E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}k+1}) < E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}k})$ is satisfied, the algorithm is restarted with a step length chosen as a fraction of the initial step, e.g., $\tau_0 = 0.5 \times 200 = 100$. Compared to that, in the second case, when the relative decrease in the objective function is small, i.e., $E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}k}) - E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}k+1}) \leq \varepsilon_1 E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}k})$ with $\varepsilon_1 = 0.02$, the algorithm is continued with a worsening step corresponding to the gradient of the surrogate function. The results show that the restarted algorithm has a lower decreasing rate of the objective function than the hybrid algorithm. Note that by further numerical experiments, we found that the decreasing rate of the restarted algorithm cannot be significantly improved for any τ_0 in the range between 0 and 200. Taking into account that for these simulations, the computational time of the restarted algorithm is on

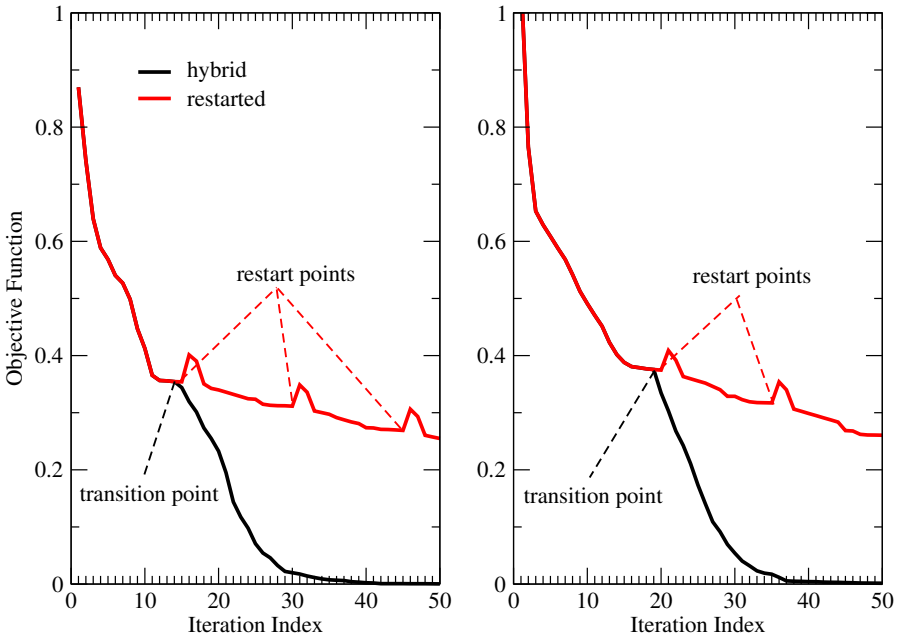


Figure 3: History of the objective function for the Gaussian cloud (left) and the LES cloud (right). The results are computed with the Nesterov acceleration method by using the hybrid and a restarted algorithm. The restarted algorithm uses at each restart the initial step length $\tau_0 = 100$. For the hybrid algorithm, the transition points are the points at which the relative decrease in the objective function is small, i.e., $E_\alpha(\boldsymbol{\sigma}_{\text{ext}k}) - E_\alpha(\boldsymbol{\sigma}_{\text{ext}k+1}) \leq \varepsilon_1 E_\alpha(\boldsymbol{\sigma}_{\text{ext}k})$ with $\varepsilon_1 = 0.02$, while for the restarted algorithm, the restart points are the points at which the descent condition $E_\alpha(\boldsymbol{\sigma}_{\text{ext}k+1}) < E_\alpha(\boldsymbol{\sigma}_{\text{ext}k})$ is not satisfied for all step lengths $\tau_k \geq \tau_{\min} = 2 \times 10^{-4}$. Observe that a transition point appears before a restart point.

average 3-4 times higher than that of the hybrid algorithm, we may conclude that the restarted algorithm is not only less accurate, but also less efficient than the hybrid algorithm.

In Table 1 we illustrate the relative error in the extinction field

$$\epsilon_{\text{ext}} = \frac{\left\| \sigma_{\text{ext}} - \boldsymbol{\sigma}^\star_{\text{ext}} \right\|_2}{\left\| \boldsymbol{\sigma}^\star_{\text{ext}} \right\|_2}$$

and the computational time for the hybrid and surrogate-minimization algorithms. The variation of the relative error in the extinction field with respect to the iteration index are shown in Fig. 4 for the hybrid and surrogate-minimization algorithms, and in Fig. 5 for the Nesterov acceleration method and the BFGS algorithm. The following conclusions can be drawn.

1. The relative errors corresponding to the hybrid algorithm are smaller than those corresponding to the surrogate-minimization algorithm. In the hybrid algorithm, the relative error decreases faster at the beginning of the iterative process.

Cloud Model	Retrieval Algorithm	Relative Error in Extinction		Comput. Time (h:min)	
		Nesterov	BFGS	Nesterov	BFGS
Gaussian Cloud	Hybrid	18.43×10^{-2}	19.13×10^{-2}	3:07	3:23
	Surrogate-Minimization	20.96×10^{-2}	21.13×10^{-2}	2:16	2:34
LES Cloud	Hybrid	19.21×10^{-2}	19.51×10^{-2}	3:47	4:05
	Surrogate-Minimization	23.08×10^{-2}	23.44×10^{-2}	2:55	3:16

Table 1: Relative error in extinction field and computational time for the hybrid and surrogate-minimization algorithms. Each retrieval is performed by using the Nesterov acceleration method and the BFGS algorithm.

2. The Nesterov acceleration method and the BFGS algorithm have almost the same accuracies and convergence rates.
3. The surrogate-minimization algorithm is faster than the hybrid algorithm.

In Figs. 6, we illustrate the true extinction field $\boldsymbol{\sigma}^{\star}_{\rm ext}(\mathbf{r}_i)$ for the Gaussian cloud at each base grid point, and in Fig. 7, the corresponding scatter plots of the retrieved extinction field $\boldsymbol{\sigma}_{\rm ext}(\mathbf{r}_i)$ in the xz- and yz-planes. In the xz-plane, the mean absolute error MAE $= (1/N_{\rm pts}) \sum_{i=1}^{N_{\rm pts}} [\boldsymbol{\sigma}_{\rm ext}(\mathbf{r}_i) - \boldsymbol{\sigma}^{\star}_{\rm ext}(\mathbf{r}_i)]$ and the root mean square error RMSE $= \sqrt{(1/N_{\rm pts}) \sum_{i=1}^{N_{\rm pts}} [\boldsymbol{\sigma}_{\rm ext}(\mathbf{r}_i) - \boldsymbol{\sigma}^{\star}_{\rm ext}(\mathbf{r}_i)]^2}$ are 5.169×10^{-2} km^{-1} and 4.474×10^{-1} km^{-1}, respectively, while in the yz-plane, these are 1.988×10^{-2} km^{-1} and 2.793×10^{-1} km^{-1}, respectively.

In Figs. 8 and 9 we show the true extinction field $\boldsymbol{\sigma}^{\star}_{\rm ext}(\mathbf{r}_i)$ and the retrieved extinction field $\boldsymbol{\sigma}_{\rm ext}(\mathbf{r}_i)$ for the LES cloud at each base grid point, and in Figs. 10 and 11, the corresponding absolute errors $\boldsymbol{\sigma}_{\rm ext}(\mathbf{r}_i) - \boldsymbol{\sigma}^{\star}_{\rm ext}(\mathbf{r}_i)$ in the xz- and yz-planes. The results correspond to three values for $\sigma_{\rm max}$, namely 8, 10, and 12 km^{-1}, for which, the relative errors in the extinction field are 0.19, 0.24, and 0.28, respectively, and the maximum vertically-integrated extinctions are 5, 6.25, and 7.5, respectively. The plots show that by increasing $\sigma_{\rm max}$ the retrieval becomes less accurate, and the algorithm has the tendency to blur the boundary of the cloud [30]. Actually, in the xz-plane, the extinction curve located at $x \approx 2.8$ km becomes smoother along the z-axis when $\sigma_{\rm max}$ increases, while in the yz-plane, the extinction curves located at $x \approx 2$ km and $x \approx 3.8$ km become smoother along the z-axis.

5. Conclusions and outlook

A cloud tomographic retrieval algorithm using the adjoint radiative transfer theory for gradient calculations has been designed. The following special features of the algorithm can be mentioned.

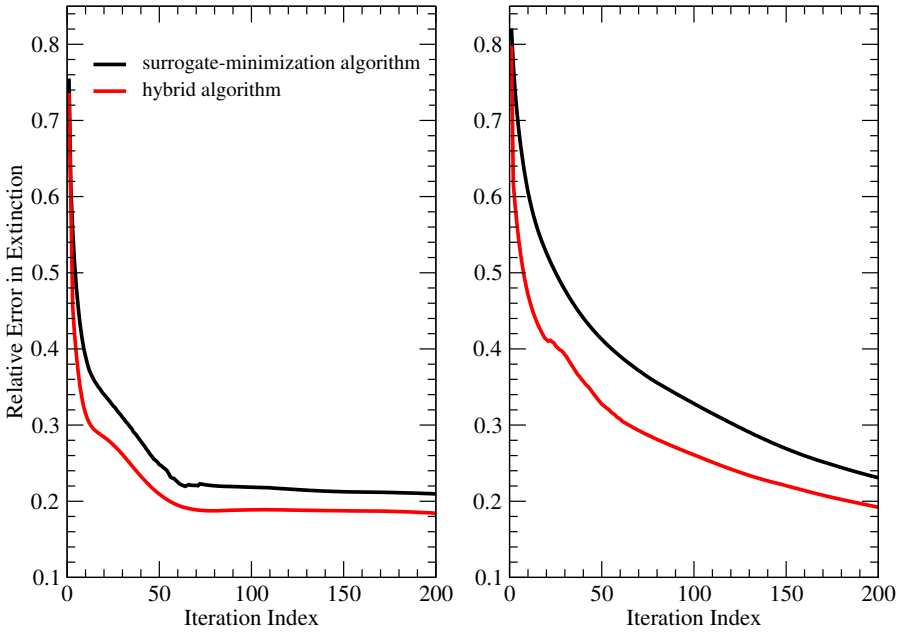


Figure 4: Relative error in extinction field versus the iteration index for the Gaussian cloud (left) and the LES cloud (right). The results correspond to the hybrid and surrogate-minimization algorithms and are computed with the Nesterov acceleration method.

1. The retrieval algorithm uses SHDOM with the delta-M approximation, the TMS correction, and the adaptive grid splitting for derivative calculations.
2. At each iteration step of a gradient-based optimization method, the value of the objective function and its gradient are computed with two calls of the radiative transfer model. In the first call, the objective function is computed by solving the forward radiative transfer equation, while in the second call, the gradient of the objective function is computed by solving the adjoint radiative transfer equation.
3. In order to escape local minima and to increase the efficiency, the retrieval algorithm is designed as a hybrid algorithm that combines the computation of the gradient of the objective function by the adjoint method with that of the gradient of a surrogate function. The decision of changing the gradients depends on the relative decrease in the objective function.

The retrieval algorithm shares many features of the algorithm based on the surrogate minimization method. These include (i) the construction of the regularization term by using spatial filtering techniques from image processing, (ii) the use of the iteratively regularized Gauss-Newton method, according to which, the amount of regularization is gradually decreased during the iteration, (iii) the application of the Nesterov acceleration method and the BFGS algorithm in combination with a step-length procedure relying on the descent condition, and (iv) the use of a stopping rule based on the absolute residual convergence

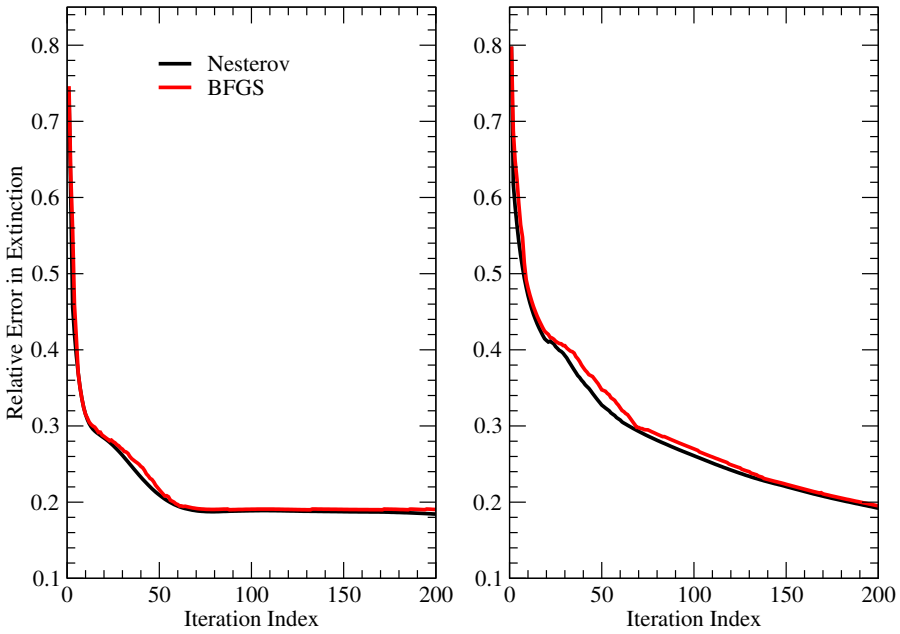


Figure 5: Relative error in extinction field versus the iteration index for the Gaussian cloud (left) and the LES cloud (right). The results correspond to the hybrid algorithm and are computed with the Nesterov acceleration method and the BFGS algorithm.

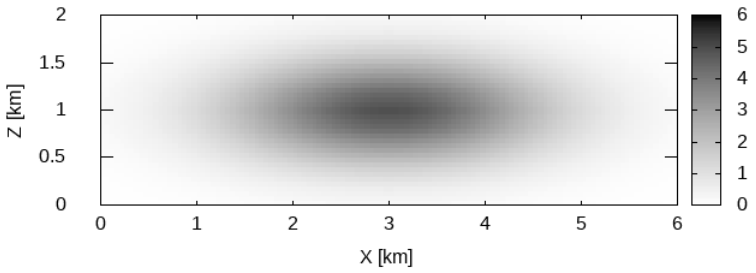


Figure 6: True extinction field $\boldsymbol{\sigma}^{\star}_{\mathrm{ext}}(\mathbf{r}_i)$ in km^{-1} for the Gaussian cloud.

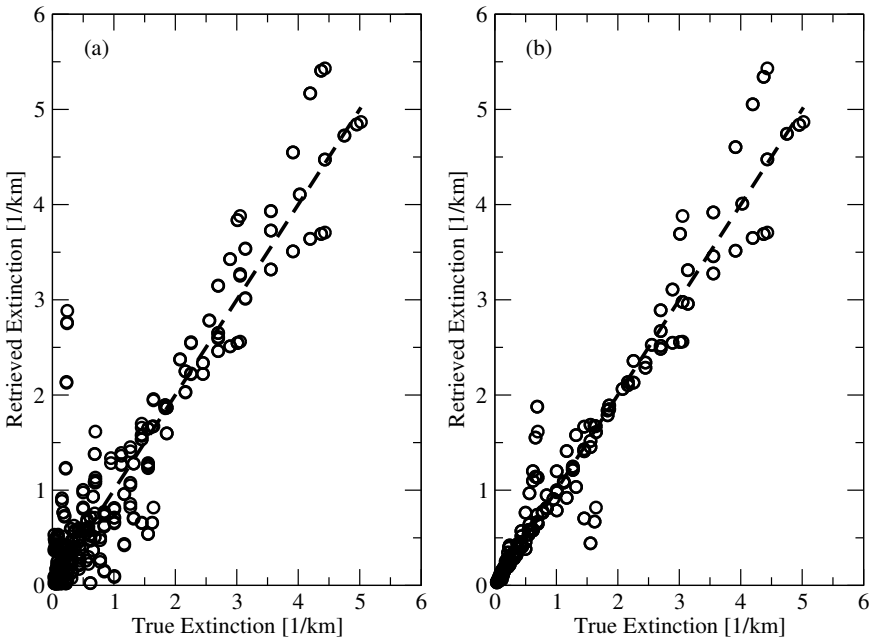


Figure 7: Scatter plot of retrieved extinction field $\boldsymbol{\sigma}_{\text{ext}}(\mathbf{r}_i)$ versus the true extinction field $\boldsymbol{\sigma}^{\star}_{\text{ext}}(\mathbf{r}_i)$ in the xz-plane (left) and the yz-plane (right) for the Gaussian cloud.

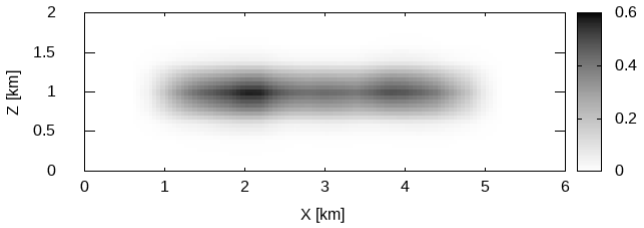
test.

Our numerical analysis has shown that the hybrid algorithm has

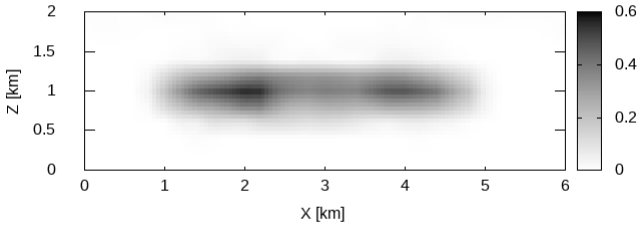
1. a significantly better decreasing rate of the objective function and a higher efficiency than a restarted algorithm that is completely based on the adjoint method for gradient calculation, and
2. a better accuracy, but a lower efficiency than a surrogate-minimization algorithm.

In a series of two papers, we presented our first attempt to design cloud tomographic retrieval algorithms based on the adjoint and surrogate minimization methods. Cloud tomography is a very complex research field that requires more detailed investigations. In order to draw pertinent conclusions about the retrieval performances, further experiments for scattering regimes that differ from those investigated in this paper (more realistic clouds with higher opacity and less smooth structure) should be performed. Moreover, the cloud tomographic retrieval algorithm should be extended by including additional features. These are listed below.

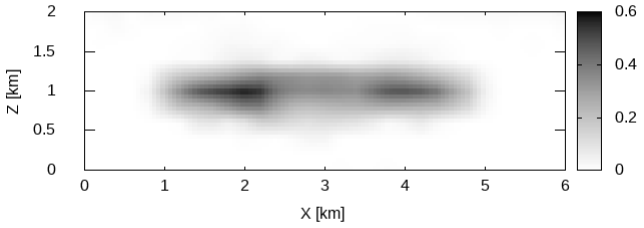
1. *Initial guess.* The chance of getting stuck in local minima can be weakened by using an appropriate initial guess. The ray data casting method overestimates the cloud volume especially at the cloud top and bottom. To reduce overestimation at cloud top, the cloud-top height derived by



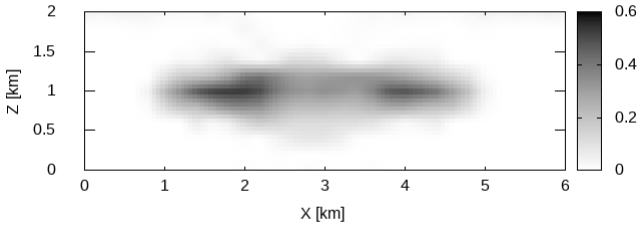
(a)



(b)

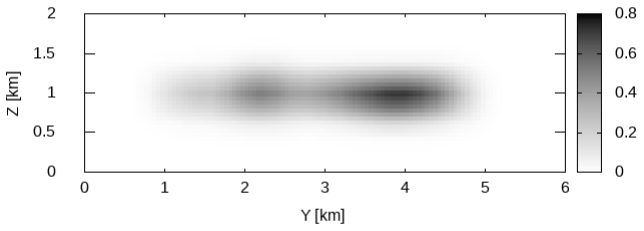


(c)

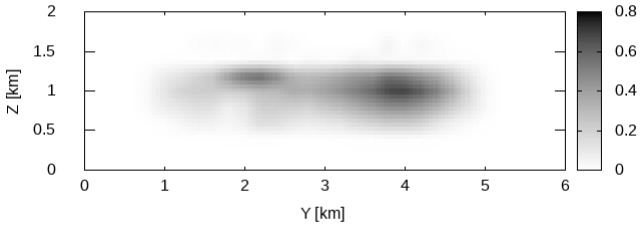


(d)

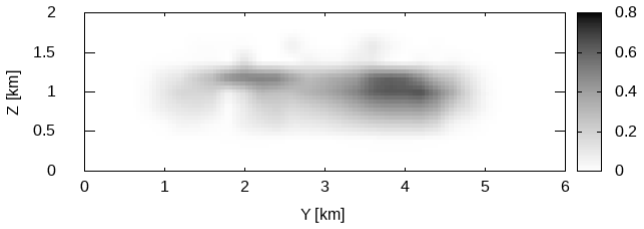
Figure 8: True extinction field $\boldsymbol{\sigma}^{\star}_{\mathrm{ext}}(\mathbf{r}_i)$ (a) and retrieved extinction field $\boldsymbol{\sigma}_{\mathrm{ext}}(\mathbf{r}_i)$ in the xz-plane for the LES cloud with $\sigma_{\max} = 8$ km^{-1} (b), $\sigma_{\max} = 10$ km^{-1} (c), and $\sigma_{\max} = 12$ km^{-1} (d). The extinction fields are normalized by $\sigma_{\max}$.



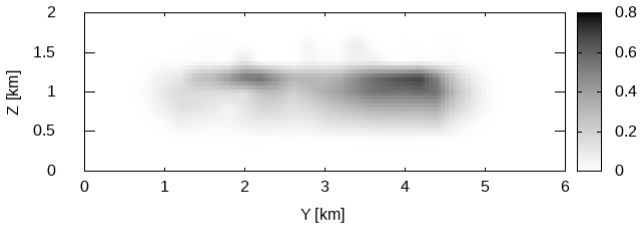
(a)



(b)

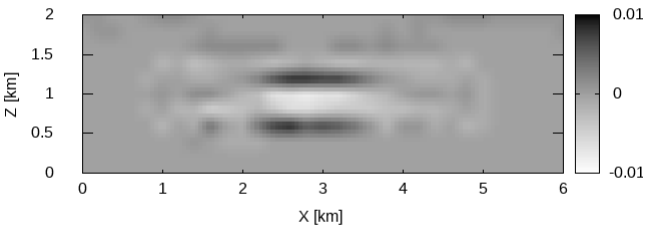


(c)

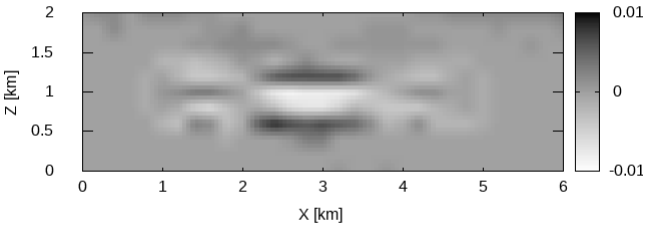


(d)

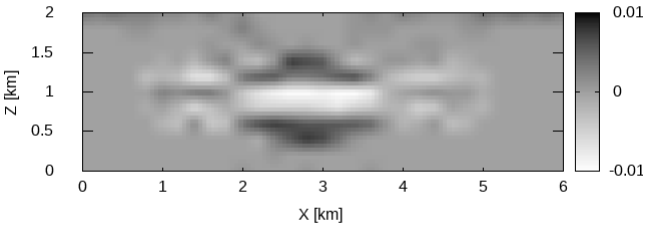
Figure 9: The same as in Fig. 8 but for the yz-plane.



(a)



(b)



(c)

Figure 10: Absolute error field $\boldsymbol{\sigma}_{\text{ext}}(\mathbf{r}_i) - \boldsymbol{\sigma}^{\star}_{\text{ext}}(\mathbf{r}_i)$ in the xz-plane for the LES cloud with $\sigma_{\max} = 8$ km^{-1} (a), $\sigma_{\max} = 10$ km^{-1} (b), and $\sigma_{\max} = 12$ km^{-1} (c). The error fields are normalized by $\sigma_{\max}$.

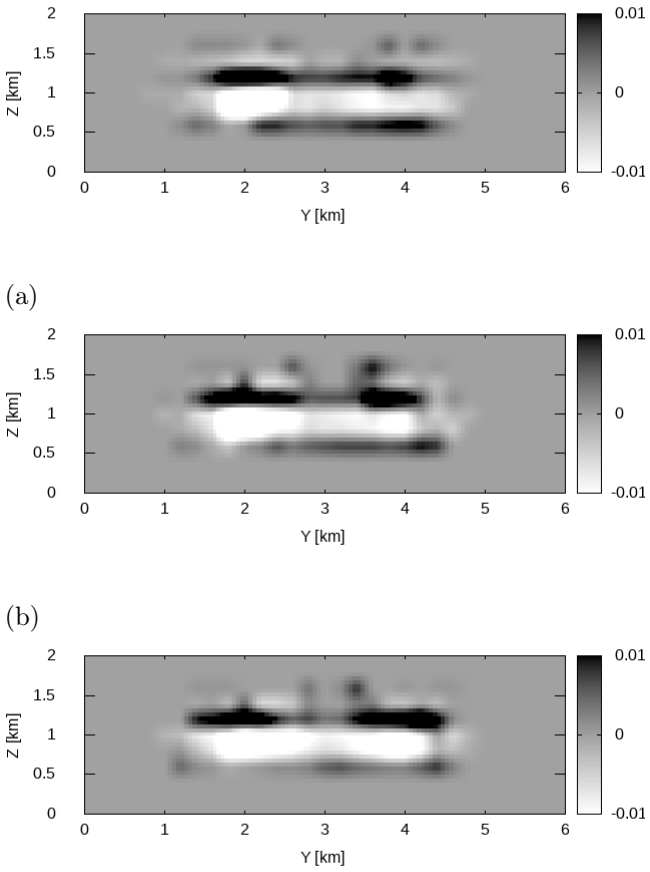


Figure 11: The same as in Fig. 10 but for the yz-plane.

a stereo-photogrammetric technique can be used, while overestimation at cloud bottom can be reduced by considering the lifting level from reanalysis data or ground-based measurements of cloud-base height.

2. *Optimization methods.* Instead of a line search procedure for computing the step length, adaptive learning rate (step length) methods that are frequently used in machine learning, as for example, Adagrad [36], Adadelta [37], Adam [38], AdaMax [38], and Nadam [39], should be tested. On the other hand, global optimization method relying on a combination of deterministic and stochastic methods (evolutionary strategy and simulated annealing) should be implemented. The approach used by Martin and Hasekamp [28], in which the regularization parameter is decreased when no improvement of the objective function is achieved, is also an interesting alternative.
3. *Regularization term.* The regularization term corresponds to averaging, Gaussian, and median low-pass filters from image processing. These low-pass filters, which are similar to the L_2-norm regularization of the gradient, attenuate variations of the extinction field and have the tendency to eliminate details and blur the cloud edges. In contrast, high-pass filters, as for example, gradient, Laplacian, Roberts, Sobel, and Prewitt, emphasize significant variations of the extinction field at the cloud boundaries, and help isolate varying patterns that correspond to sharp edges, details, and noise. This type of filters, which are similar to total variation regularization, should be also analyzed.
4. *Regularization method.* The regularization method used in this study is the iteratively regularized Gauss-Newton method. For trace gas retrievals, it was found that this method is superior to the method of Tikhonov regularization, because it is insensitive to overestimation of the regularization parameter [40]. For cloud tomography, its superiority is not yet proven. In this regard, the method of Tikhonov regularization using dynamic parameter choice methods, as for example, the L-curve method, generalized cross-validation, and maximum likelihood estimation, should be examined. Note that in this case, the amount of regularization is controlled at each iteration step.
5. *Diffusion approximation.* In cloud tomography, the spatial information originates primarily from the outer shell of the cloud, whereas the amount of information conveyed about the inner core of the cloud decreases significantly with depth due to multiple scattering. In Ref. [31] it was suggested that a hybrid three-dimensional radiative transfer solver, that treats the outer shell of the cloud with high accuracy and the veiled core under the diffusion approximation, will increase the efficiency of radiative transfer simulations. This idea should be transposed and implemented in practice. Note that in SHDOM, the use of a diffusion-type approximation is not a difficult task: in the domain of the veiled core, only the spherical harmonics expansion coefficients corresponding $n = 0$ and $n = 1$ will be used to represent the radiance (the rest of them will be set to zero), while the continuity of the radiance field will be automatically satisfied in the

discrete ordinate space.

6. *Retrieval of microphysical parameters.* The adjoint theory can be used for the retrieval of cloud microphysical parameters. If ξ is a parameter that determines the extinction field $\sigma_{\text{ext}}(\mathbf{r})$, the scattering field $\sigma_{\text{sct}}(\mathbf{r})$, and the phase function $P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}') = P(\mathbf{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}')$, e.g., a cloud size-distribution parameter or the refractive index of a cloud droplet, the derivative of the residual R with respect to the parameter ξ can be computed as

$$g_{\text{R}\xi} = \frac{\partial R}{\partial \xi} = -\Big\langle I^{\dagger}, \frac{\partial \mathcal{L}}{\partial \xi} I \Big\rangle,$$

where

$$\Big(\frac{\partial \mathcal{L}}{\partial \xi} I\Big)(\mathbf{r}, \boldsymbol{\Omega}) = I(\mathbf{r}, \boldsymbol{\Omega}) \frac{\partial \sigma_{\text{ext}}}{\partial \xi}(\mathbf{r}) - \frac{1}{4\pi} \frac{\partial \sigma_{\text{sct}}}{\partial \xi}(\mathbf{r}) \int_{\Omega} P(\mathbf{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') I(\mathbf{r}, \boldsymbol{\Omega}') \, \mathrm{d}\boldsymbol{\Omega}'$$
$$- \frac{\sigma_{\text{sct}}(\mathbf{r})}{4\pi} \int_{\Omega} \frac{\partial P}{\partial \xi}(\mathbf{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') I(\mathbf{r}, \boldsymbol{\Omega}') \, \mathrm{d}\boldsymbol{\Omega}',$$

$$\frac{\partial P}{\partial \xi}(\mathbf{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') = \sum_{n=1}^{N_{\text{rank}}} \frac{\partial \chi_n}{\partial \xi}(\mathbf{r}) P_n(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}'),$$

and for $P(\mathbf{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') = \sum_{n=1}^{N_{\text{rank}}} \chi_n(\mathbf{r}) P_n(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}')$, $P_n(\cos\Theta)$ are the unnormalized Legendre polynomials, $\chi_n(\mathbf{r})$ the Legendre phase function coefficients, and N_{rank} the maximum expansion order of the phase function. The derivative of the residual $g_{\text{R}\xi}$ can be calculated by means of SHDOM like the gradient of the residual $\mathbf{g}_{\text{R}}(\boldsymbol{\sigma}_{\text{ext}})$. Moreover, because the derivatives of the single-scattering radiance with respect to ξ can be computed analytically in a very simple manner, it seems to be possible to design a hybrid algorithm for microphysical parameters retrieval. Along this line it should be pointed out that in the framework of the surrogate minimization method, the retrieval of cloud size-distribution parameters from unpolarized and polarimetric measurements has been reported in Ref. [41] and [42], respectively.

Appendices

Appendix A Basic results in three-dimensional adjoint radiative transfer

In this self-contained appendix we intend to present in a pedagogical manner the fundamentals of the three-dimensional adjoint radiative transfer.

A.1 Radiative transfer equation in operator from

As in Section 2, we consider the boundary-value problem for the total radiance in a three-dimensional domain D in the shape of rectangular prism with

lengths L_x, L_y and L_z, bottom and top faces $S_\mathrm{b}(z=0)$ and S_t $(z=L_\mathrm{z})$, respectively, and lateral faces $S_{1\mathrm{x}}$ $(x=0)$, $S_{2\mathrm{x}}$ $(x=L_\mathrm{x})$, $S_{1\mathrm{y}}$ $(y=0)$, and $S_{2\mathrm{y}}$ $(y=L_\mathrm{y})$. This consists in the integro-differential equation

$$\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r},\boldsymbol{\Omega}) = -\sigma_\mathrm{ext}(\mathbf{r})I(\mathbf{r},\boldsymbol{\Omega}) + \frac{\sigma_\mathrm{sct}(\mathbf{r})}{4\pi}\int_\Omega P(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}')I(\mathbf{r},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}', \quad \mathbf{r}\in D, \tag{A.1}$$

the boundary conditions at the top and bottom faces

$$I(\mathbf{r}_\mathrm{t},\boldsymbol{\Omega}^-) = \frac{F_0}{|\mu_0|}\delta(\boldsymbol{\Omega}^- - \boldsymbol{\Omega}_0), \quad \mathbf{r}_\mathrm{t}\in S_\mathrm{t}, \tag{A.2}$$

and

$$I(\mathbf{r}_\mathrm{b},\boldsymbol{\Omega}^+) = \frac{A_\mathrm{s}}{\pi}\int_{\Omega^-}\rho(\mathbf{r}_\mathrm{b},\boldsymbol{\Omega}^+,\boldsymbol{\Omega}^-)|\mu^-|I(\mathbf{r}_\mathrm{b},\boldsymbol{\Omega}^-)\,\mathrm{d}\boldsymbol{\Omega}^-, \quad \mathbf{r}_\mathrm{b}\in S_\mathrm{b}, \tag{A.3}$$

respectively, and the periodic boundary conditions at the lateral faces

$$I(\mathbf{r}_{1\alpha},\boldsymbol{\Omega}) = I(\mathbf{r}_{2\alpha},\boldsymbol{\Omega}), \quad \mathbf{r}_{1\alpha}\in S_{1\alpha},\ \mathbf{r}_{2\alpha}\in S_{2\alpha},\ \alpha = x,y. \tag{A.4}$$

Here, A_s is the spherical albedo and $\rho(\mathbf{r}_\mathrm{b},\boldsymbol{\Omega}^+,\boldsymbol{\Omega}^-)$ the normalized bi-directional reflection function, while the meaning of the other quantities is as in Section 2.

Let Σ_b and Σ_t be the planes $z=0$ and $z=L_\mathrm{z}$, respectively, and Δ the domain of the plane-parallel layer bounded by the planes Σ_b and Σ_t. We extend the optical parameter functions $\sigma_\mathrm{ext}(\mathbf{r})$, $\sigma_\mathrm{sct}(\mathbf{r})$, and $P(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}')$ by periodicity along the x- and y-directions in the domain Δ, and similarly, the normalized reflection function $\rho(\mathbf{r}_\mathrm{b},\boldsymbol{\Omega}^+,\boldsymbol{\Omega}^-)$ in the plane Σ_b. By this construction it is apparent that *if the radiance I satisfies the boundary-value problem (A.1)–(A.4), then I satisfies the so-called extended boundary-value problem consisting in the radiative transfer equation (A.1) in Δ, the top boundary condition (A.2) on Σ_t, and the bottom boundary condition (A.3) on Σ_b. The converse results is also true.* Thus, by periodicity extension, the initial boundary-value problem for a finite domain D can be reformulated as a boundary-value problem for a plane-parallel layer Δ. In this case, *if the radiance I satisfies the extended boundary-value problem for a plane-parallel layer, then I solves the integral equation*

$$\begin{aligned} I(\mathbf{r},\boldsymbol{\Omega}^-) = {} & \frac{F_0}{|\mu_0|}\delta(\boldsymbol{\Omega}^- - \boldsymbol{\Omega}_0)T(\mathbf{r}_t,\mathbf{r},\boldsymbol{\Omega}_0) \\ & + \int_{\mathbf{r}_t}^{\mathbf{r}} \mathcal{J}(\mathbf{r}',\boldsymbol{\Omega}^-;I)T(\mathbf{r}',\mathbf{r},\boldsymbol{\Omega}^-)\,ds', \quad \mathbf{r}\in\Delta,\ \mathbf{r}_t\in\Sigma_t, \end{aligned} \tag{A.5}$$

for any downward directions $\boldsymbol{\Omega}^-$, and

$$\begin{aligned} I(\mathbf{r},\boldsymbol{\Omega}^+) = {} & I(\mathbf{r}_b,\boldsymbol{\Omega}^+)T(\mathbf{r}_b,\mathbf{r},\boldsymbol{\Omega}^+) \\ & + \int_{\mathbf{r}_b}^{\mathbf{r}} \mathcal{J}(\mathbf{r}',\boldsymbol{\Omega}^+;I)T(\mathbf{r}',\mathbf{r},\boldsymbol{\Omega}^+)\,ds', \quad \mathbf{r}\in\Delta,\ \mathbf{r}_b\in\Sigma_b, \end{aligned} \tag{A.6}$$

for any upward directions $\mathbf{\Omega}^+$*, where*

$$\mathcal{J}(\mathbf{r},\mathbf{\Omega};I)=\frac{\sigma_{\rm sct}(\mathbf{r})}{4\pi}\int_{\Omega}P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')I(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}'$$

is the source function,

$$T(\mathbf{r}_1,\mathbf{r}_2,\mathbf{\Omega})=\exp\Big(-\int_{\mathbf{r}_1}^{\mathbf{r}_2}\sigma_{\rm ext}(\mathbf{r})\,\mathrm{d}s\Big)$$

the transmission along the characteristic $\mathbf{\Omega}=(\mu,\varphi)$ *starting at* $\mathbf{r}_1$ *and ending at* $\mathbf{r}_2$*, and*

$$I(\mathbf{r}_b,\mathbf{\Omega}^+)=\frac{A_s}{\pi}\int_{\Omega^-}\rho(\mathbf{r}_b,\mathbf{\Omega}^+,\mathbf{\Omega}^-)|\mu^-|I(\mathbf{r}_b,\mathbf{\Omega}^-)\,d\mathbf{\Omega}^-$$

the upward bottom radiance. The converse results is also true. Equation (A.5) represents the integral form of the radiative transfer equation for downward radiances, while Eq. (A.6) represents the integral form of the radiative transfer equation for upward radiances.

Let us define the forward operator

$$\begin{aligned}(\mathcal{L}I)(\mathbf{r},\mathbf{\Omega})=&\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r},\mathbf{\Omega})+\sigma_{\rm ext}(\mathbf{r})I(\mathbf{r},\mathbf{\Omega})-\frac{\sigma_{\rm sct}(\mathbf{r})}{4\pi}\int_{\Omega}P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')I(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}'\\&-\frac{A_{\rm s}}{\pi}\delta(z)H(\mu)\mu\int_{\Omega}\rho(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')H(-\mu')|\mu'|I(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}',\end{aligned}\tag{A.7}$$

and the forward source term

$$Q(\mathbf{r},\mathbf{\Omega})=F_0\delta(z-L_{\rm z})\delta(\mathbf{\Omega}-\mathbf{\Omega}_0),\tag{A.8}$$

where H is the Heaviside step function. The last term in Eq. (A.7) represents the contribution at the bottom surface, while the boundary condition at the top surface is encapsulated in the expression of the forward source term (A.8). With respect to the Dirac delta functions $\delta(z)$ and $\delta(z-L_{\rm z})$ in Eqs. (A.7) and (A.8), respectively, we make the following remark. In general, the Dirac delta function $\delta(z-z_0)$ can be defined on $(-\infty,+\infty)$ as the limit of the sequence of functions

$$h_a(z-z_0)=\begin{cases}\frac{1}{2a}, & z_0-a\le z\le z_0+a\\ 0, & \text{otherwise}\end{cases},$$

i.e., $\delta(z-z_0)=\lim_{a\to 0}h_a(z-z_0)$, so that the filter property of $\delta(z-z_0)$ on the

interval $[z_0, z_1]$ is

$$\begin{aligned}\int_{z_0}^{z_1} f(z)\delta(z-z_0)\,\mathrm{d}z &= \lim_{a\to 0}\int_{z_0}^{z_1} f(z)h_a(z-z_0)\,\mathrm{d}z \\ &= \lim_{a\to 0}\frac{1}{2a}\int_{z_0}^{z_0+a} f(z)\,\mathrm{d}z \\ &= \lim_{a\to 0}\frac{1}{2a}[af(\zeta_a)] \\ &= \frac{1}{2}f(z_0),\end{aligned}$$

where, according to the mean-value theorem, $z_0 \le \zeta_a \le z_0 + a$, yielding $\zeta_a \to z_0$ as $a \to 0$. Analogously, the filter property of $\delta(z - z_1)$ on the interval $[z_0, z_1]$ is $\int_{z_0}^{z_1} f(z)\delta(z-z_1)\,\mathrm{d}z = (1/2)f(z_1)$. However, in Eq. (A.7) , $\delta(z - z_0)$ is defined on the interval $[z_0, +\infty)$ as the limit of the sequence of functions

$$h_a(z-z_0) = \begin{cases} \frac{1}{a}, & z_0 \le z \le z_0 + a \\ 0, & \text{otherwise} \end{cases},$$

in which case, $\int_{z_0}^{z_1} f(z)\delta(z-z_0)\,\mathrm{d}z = f(z_0)$, while in Eq. (A.8), $\delta(z - z_1)$ is defined on the interval $(-\infty, z_1]$ as the limit of the sequence of functions

$$h_a(z-z_1) = \begin{cases} \frac{1}{a}, & z_1 - a \le z \le z_1 \\ 0, & \text{otherwise} \end{cases},$$

in which case, $\int_{z_0}^{z_1} f(z)\delta(z-z_1)\,\mathrm{d}z = f(z_1)$.

The following result states the operator form of the radiative transfer equation: *Let the forward radiance I solves the operator equation*

$$(\mathcal{L}I)(\mathbf{r}, \mathbf{\Omega}) = Q(\mathbf{r}, \mathbf{\Omega}), \quad \mathbf{r} \in D, \tag{A.9}$$

with the boundary conditions

$$\begin{aligned} &I(\mathbf{r}_t, \mathbf{\Omega}^-) = 0, \quad \mathbf{r}_t \in S_t, \\ &I(\mathbf{r}_b, \mathbf{\Omega}^+) = 0, \quad \mathbf{r}_b \in S_b, \\ &I(\mathbf{r}_{1\alpha}, \mathbf{\Omega}) = I(\mathbf{r}_{2\alpha}, \mathbf{\Omega}), \quad \mathbf{r}_{1\alpha} \in S_{1\alpha},\, \mathbf{r}_{2\alpha} \in S_{2\alpha},\, \alpha = x, y. \end{aligned} \tag{A.10}$$

Then I solves the boundary-value problem (A.1)-(A.4), and conversely. To prove this assertion we take into account that due to the periodicity extension, it is sufficient to show that if I solve the operator equation (A.9) in Δ with the homogeneous boundary condition $I(\mathbf{r}_\mathrm{t}, \mathbf{\Omega}^-) = 0$ for $\mathbf{r}_\mathrm{t} \in \Sigma_\mathrm{t}$ and $I(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}^+) = 0$ for $\mathbf{r}_\mathrm{b} \in \Sigma_\mathrm{b}$, then I solves the extended boundary value problem, and conversely. For a downward direction $\mathbf{\Omega} = \mathbf{\Omega}^-$, we have $H(\mu) = H(\mu^-) = 0$, and Eq. (A.9) can be written explicitly as

$$\begin{aligned}\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r}, \mathbf{\Omega}^-) &= -\sigma_\mathrm{ext}(\mathbf{r})I(\mathbf{r}, \mathbf{\Omega}^-) + \mathcal{J}(\mathbf{r}, \mathbf{\Omega}^-; I) \\ &\quad + F_0\delta(z - L_\mathrm{z})\delta(\mathbf{\Omega}^- - \mathbf{\Omega}_0).\end{aligned} \tag{A.11}$$

Integrating Eq. (A.11) along the direction $\mathbf{\Omega}^-$, using the boundary condition $I(\mathbf{r}_\mathrm{t}, \mathbf{\Omega}^-) = 0$ for $\mathbf{r}_\mathrm{t} \in \Sigma_\mathrm{t}$, and taking into account that $\int_z^{L_\mathrm{z}} f(z')\delta(z' - L_\mathrm{z})\,\mathrm{dz}' = f(L_\mathrm{z})$, which yields

$$
\begin{aligned}
&F_0 \int_{\mathbf{r}_\mathrm{t}}^{\mathbf{r}} \delta(z' - L_\mathrm{z})\delta(\mathbf{\Omega}^- - \mathbf{\Omega}_0) T(\mathbf{r}', \mathbf{r}, \mathbf{\Omega}^-)\,\mathrm{d}s' \\
&= F_0 \frac{1}{|\mu_0|} \int_z^{L_\mathrm{z}} \delta(z' - L_\mathrm{z})\delta(\mathbf{\Omega}^- - \mathbf{\Omega}_0) T(\mathbf{r}'(z'), \mathbf{r}, \mathbf{\Omega}^-)\,\mathrm{d}z' \\
&= \frac{F_0}{|\mu_0|} \delta(\mathbf{\Omega}^- - \mathbf{\Omega}_0) T(\mathbf{r}_\mathrm{t}, \mathbf{r}, \mathbf{\Omega}_0)
\end{aligned}
$$

with $\mathbf{r}'(L_\mathrm{z}) = \mathbf{r}_\mathrm{t}$, we are led to the integral form of the radiative transfer equation for the downward radiances (A.5). For an upward direction $\mathbf{\Omega} = \mathbf{\Omega}^+$, we have $H(\mu) = H(\mu^+) = 1$, and Eq. (A.9) read as

$$
\begin{aligned}
\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r}, \mathbf{\Omega}^+) &= -\sigma_\mathrm{ext}(\mathbf{r}) I(\mathbf{r}, \mathbf{\Omega}^+) + \mathcal{J}(\mathbf{r}, \mathbf{\Omega}^+; I) \\
&\quad + \frac{A_\mathrm{s}}{\pi} \delta(z) \mu^+ \int_\Omega \rho(\mathbf{r}, \mathbf{\Omega}^+, \mathbf{\Omega}') H(-\mu') |\mu'| I(\mathbf{r}, \mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}'. \qquad \text{(A.12)}
\end{aligned}
$$

Integrating Eq. (A.12) along the direction $\mathbf{\Omega}^+$, using the boundary condition $I(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}^+) = 0$ for $\mathbf{r}_\mathrm{b} \in \Sigma_\mathrm{b}$, and the result $\int_0^z f(z')\delta(z')\,\mathrm{dz}' = f(0)$, which yields

$$
\begin{aligned}
&\frac{A_\mathrm{s}}{\pi} \int_{\mathbf{r}_\mathrm{b}}^{\mathbf{r}} \left[\delta(z') \mu^+ \int_\Omega \rho(\mathbf{r}', \mathbf{\Omega}^+, \mathbf{\Omega}') H(-\mu') |\mu'| I(\mathbf{r}', \mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}' \right] \\
&\times T(\mathbf{r}', \mathbf{r}, \mathbf{\Omega}^+) \mathrm{d}s' \\
&= \frac{A_\mathrm{s}}{\pi} \frac{1}{\mu^+} \int_0^z \left[\delta(z') \mu^+ \int_\Omega \rho(\mathbf{r}'(z'), \mathbf{\Omega}^+, \mathbf{\Omega}') H(-\mu') |\mu'| I(\mathbf{r}'(z'), \mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}' \right] \\
&\times T(\mathbf{r}'(z'), \mathbf{r}, \mathbf{\Omega}^+) \mathrm{d}z' \\
&= \frac{A_\mathrm{s}}{\pi} \left[\int_\Omega \rho(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}^+, \mathbf{\Omega}') H(-\mu') |\mu'| I(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}' \right] T(\mathbf{r}_\mathrm{b}, \mathbf{r}, \mathbf{\Omega}^+) \\
&= \frac{A_\mathrm{s}}{\pi} \left[\int_{\Omega^-} \rho(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}^+, \mathbf{\Omega}^-) |\mu^-| I(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}^-)\,\mathrm{d}\mathbf{\Omega}^- \right] T(\mathbf{r}_\mathrm{b}, \mathbf{r}, \mathbf{\Omega}^+) \\
&= I(\mathbf{r}_\mathrm{b}, \mathbf{\Omega}^+) T(\mathbf{r}_\mathrm{b}, \mathbf{r}, \mathbf{\Omega}^+)
\end{aligned}
$$

with $\mathbf{r}'(0) = \mathbf{r}_\mathrm{b}$, we obtain the integral form of the radiative transfer equation for the upward radiances (A.6).

A.2 The adjoint transport operator

The linear operator $\mathcal{L}$ can be expressed as a linear combination of four operators $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4$, where

$$
\begin{aligned}
(\mathcal{L}_1 I)(\mathbf{r}, \mathbf{\Omega}) &= \frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r}, \mathbf{\Omega}), \\
(\mathcal{L}_2 I)(\mathbf{r}, \mathbf{\Omega}) &= \sigma_{\mathrm{ext}}(\mathbf{r})\, I(\mathbf{r}, \mathbf{\Omega}), \\
(\mathcal{L}_3 I)(\mathbf{r}, \mathbf{\Omega}) &= -\frac{\sigma_{\mathrm{sct}}(\mathbf{r})}{4\pi} \int_{\Omega} P(\mathbf{r}, \mathbf{\Omega}, \mathbf{\Omega}') I(\mathbf{r}, \mathbf{\Omega}')\, \mathrm{d}\mathbf{\Omega}',
\end{aligned}
$$

and

$$
(\mathcal{L}_4 I)(\mathbf{r}, \mathbf{\Omega}) = -\frac{A_{\mathrm{s}}}{\pi} \delta(z) H(\mu) \mu \int_{\Omega} \rho(\mathbf{r}, \mathbf{\Omega}, \mathbf{\Omega}') H(-\mu') |\mu'| I(\mathbf{r}, \mathbf{\Omega}')\, \mathrm{d}\mathbf{\Omega}'.
$$

The adjoint transport operator $\mathcal{L}^\dagger$ of $\mathcal{L}$ is defined through the equation

$$
\left\langle \mathcal{L} I, I^\dagger \right\rangle = \left\langle I, \mathcal{L}^\dagger I^\dagger \right\rangle, \tag{A.13}
$$

where the scalar product of the radiation fields I_1 and I_2 is given by

$$
\left\langle I_1, I_2 \right\rangle = \int_D \int_{\Omega} I_1(\mathbf{r}, \mathbf{\Omega}) I_2(\mathbf{r}, \mathbf{\Omega})\, \mathrm{d}\mathbf{\Omega} \mathrm{d}V.
$$

In the following, we intend to compute the adjoint transport operator $\mathcal{L}^\dagger$ under the assumptions that (i) the forward radiance I satisfies the boundary conditions (A.10), and (ii) the adjoint radiance $I^\dagger$ satisfies the boundary conditions

$$
\begin{aligned}
&I^\dagger(\mathbf{r}_{\mathrm{t}}, \mathbf{\Omega}^+) = 0, \quad \mathbf{r}_{\mathrm{t}} \in S_{\mathrm{t}}, \\
&I^\dagger(\mathbf{r}_{\mathrm{b}}, \mathbf{\Omega}^-) = 0, \quad \mathbf{r}_{\mathrm{b}} \in S_{\mathrm{b}}, \\
&I^\dagger(\mathbf{r}_{1\alpha}, \mathbf{\Omega}) = I^\dagger(\mathbf{r}_{2\alpha}, \mathbf{\Omega}), \quad \mathbf{r}_{1\alpha} \in S_{1\alpha},\, \mathbf{r}_{2\alpha} \in S_{2\alpha},\, \alpha = x, y.
\end{aligned} \tag{A.14}
$$

Since $\mathcal{L}$ is a linear combination of four operators we consider each term separately.

1. For the linear operator $\mathcal{L}_1$, we integrate by parts and obtain

$$\begin{aligned}
\left\langle \mathcal{L}_1 I, I^\dagger \right\rangle &= \int_\Omega \int_D \frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r},\mathbf{\Omega}) I^\dagger(\mathbf{r},\mathbf{\Omega})\,\mathrm{d}V\mathrm{d}\mathbf{\Omega} \\
&= \int_\Omega \int_D \left[\mathbf{\Omega}\cdot\nabla I(\mathbf{r},\mathbf{\Omega})\right] I^\dagger(\mathbf{r},\mathbf{\Omega})\,\mathrm{d}V\mathrm{d}\mathbf{\Omega} \\
&= \int_\Omega \int_D \left[\nabla\cdot\mathbf{\Omega} I(\mathbf{r},\mathbf{\Omega})\right] I^\dagger(\mathbf{r},\mathbf{\Omega})\,\mathrm{d}V\mathrm{d}\mathbf{\Omega} \\
&= \int_\Omega \int_D \nabla\cdot\left[\mathbf{\Omega} I(\mathbf{r},\mathbf{\Omega}) I^\dagger(\mathbf{r},\mathbf{\Omega})\right]\,\mathrm{d}V\mathrm{d}\mathbf{\Omega} \\
&- \int_\Omega \int_D I(\mathbf{r},\mathbf{\Omega})\left[\mathbf{\Omega}\cdot\nabla I^\dagger(\mathbf{r},\mathbf{\Omega})\right]\,\mathrm{d}V\mathrm{d}\mathbf{\Omega} \\
&= \int_\Omega \int_S \mathbf{n}\cdot\left[\mathbf{\Omega} I(\mathbf{r},\mathbf{\Omega}) I^\dagger(\mathbf{r},\mathbf{\Omega})\right]\,\mathrm{d}S\mathrm{d}\mathbf{\Omega} \\
&- \int_\Omega \int_D I(\mathbf{r},\mathbf{\Omega})\frac{\mathrm{d}I^\dagger}{\mathrm{d}s}(\mathbf{r},\mathbf{\Omega})\,\mathrm{d}V\mathrm{d}\mathbf{\Omega}.
\end{aligned}$$

   The homogeneous boundary conditions in Eqs. (A.10) and (A.14) imply that on the top boundary S_t we have $I(\mathbf{r}_\mathrm{t},\mathbf{\Omega}^-) = I^\dagger(\mathbf{r}_\mathrm{t},\mathbf{\Omega}^+) = 0$, while on the bottom boundary S_b we have $I(\mathbf{r}_\mathrm{b},\mathbf{\Omega}^+) = I^\dagger(\mathbf{r}_\mathrm{b},\mathbf{\Omega}^-) = 0$. Moreover, on the lateral boundaries, the dot product will have the same magnitude but opposite signs on opposite ends of the finite domain, and therefore, the integration will evaluate to zero, e.g., from $\mathbf{n}_{1\mathrm{x}}\cdot\mathbf{\Omega} = -\mathbf{n}_{2\mathrm{x}}\cdot\mathbf{\Omega}$ and the periodic boundary conditions in Eqs. (A.10) and (A.14), we infer that $\mathbf{n}_{1\mathrm{x}}\cdot\mathbf{\Omega} I(\mathbf{r}_{1x},\mathbf{\Omega}) I^\dagger(\mathbf{r}_{1\mathrm{x}},\mathbf{\Omega}) = -\mathbf{n}_{2\mathrm{x}}\cdot\mathbf{\Omega} I(\mathbf{r}_{2x},\mathbf{\Omega}) I^\dagger(\mathbf{r}_{2\mathrm{x}},\mathbf{\Omega})$. Hence, the surface integral is zero and we obtain $\left\langle \mathcal{L}_1 I, I^\dagger \right\rangle = \left\langle I, \mathcal{L}_1^\dagger I^\dagger \right\rangle$ with

$$\left(\mathcal{L}_1^\dagger I^\dagger\right)(\mathbf{r},\mathbf{\Omega}) = -\frac{\mathrm{d}I^\dagger}{\mathrm{d}s}(\mathbf{r},\mathbf{\Omega}).$$

2. The linear operator L_2 is the identity operator multiplied by the extinction cross-section and it is apparent that

$$(\mathcal{L}_2^\dagger I^\dagger)(\mathbf{r},\mathbf{\Omega}) = \sigma_\mathrm{ext}(\mathbf{r}) I^\dagger(\mathbf{r},\mathbf{\Omega}).$$

3. To compute the adjoint transport operator $\mathcal{L}_3^\dagger$ we consider the scalar product

$$\begin{aligned}
\left\langle \mathcal{L}_3 I, I^\dagger \right\rangle = &-\frac{1}{4\pi}\int_D \sigma_\mathrm{sct}(\mathbf{r})\,\mathrm{d}V \\
&\times \int_\Omega \int_\Omega P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}') I(\mathbf{r},\mathbf{\Omega}') I^\dagger(\mathbf{r},\mathbf{\Omega})\,\mathrm{d}\mathbf{\Omega}'\mathrm{d}\mathbf{\Omega},
\end{aligned}$$

   interchange the order of integration with respect to the variables $\mathbf{\Omega}$ and $\mathbf{\Omega}'$, and obtain $\left\langle \mathcal{L}_3 I, I^\dagger \right\rangle = \left\langle I, \mathcal{L}_3^\dagger I^\dagger \right\rangle$ with

$$(\mathcal{L}_3^\dagger I^\dagger)(\mathbf{r},\mathbf{\Omega}) = -\frac{\sigma_\mathrm{sct}(\mathbf{r})}{4\pi}\int_\Omega P(\mathbf{r},\mathbf{\Omega}',\mathbf{\Omega}) I^\dagger(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}'.$$

4. For the linear operator $\mathcal{L}_4$ we proceed analogously. In the integral

$$\begin{aligned}\left\langle \mathcal{L}_4 I, I^\dagger \right\rangle = -\frac{A_\mathrm{s}}{\pi} \int_D \delta(z)\,\mathrm{d}V \int_\Omega \int_\Omega H(\mu)\mu H(-\mu')|\mu'| \\ \times \rho(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}')I(\mathbf{r},\boldsymbol{\Omega}')I^\dagger(\mathbf{r},\boldsymbol{\Omega})\,\mathrm{d}\boldsymbol{\Omega}'\mathrm{d}\boldsymbol{\Omega}\end{aligned}$$

we interchange the order of integration with respect to the variables $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}'$, and obtain $\left\langle \mathcal{L}_4 I, I^\dagger \right\rangle = \left\langle I, \mathcal{L}_4^\dagger I^\dagger \right\rangle$ with

$$\begin{aligned}(\mathcal{L}_4 I^\dagger)(\mathbf{r},\boldsymbol{\Omega}) = -\frac{A_\mathrm{s}}{\pi}\delta(z)H(-\mu)|\mu| \\ \times \int_\Omega \rho(\mathbf{r},\boldsymbol{\Omega}',\boldsymbol{\Omega})H(\mu')\mu' I^\dagger(\mathbf{r},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}'.\end{aligned}$$

Collecting all results we find that the adjoint transport operator can be explicitly written as

$$\begin{aligned}(\mathcal{L}^\dagger I^\dagger)(\mathbf{r},\boldsymbol{\Omega}) = -\frac{\mathrm{d}I^\dagger}{\mathrm{d}s}(\mathbf{r},\boldsymbol{\Omega}) + \sigma_\mathrm{ext}(\mathbf{r})I^\dagger(\mathbf{r},\boldsymbol{\Omega}) - \frac{\sigma_\mathrm{sct}(\mathbf{r})}{4\pi}\int_\Omega P(\mathbf{r},\boldsymbol{\Omega}',\boldsymbol{\Omega})I^\dagger(\mathbf{r},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ - \frac{A_\mathrm{s}}{\pi}\delta(z)H(-\mu)|\mu|\int_\Omega \rho(\mathbf{r},\boldsymbol{\Omega}',\boldsymbol{\Omega})H(\mu')\mu' I^\dagger(\mathbf{r},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}'. \qquad \text{(A.15)}\end{aligned}$$

The main result of the adjoint radiative transfer theory states that *if (i) the radiance I solves the forward problem consisting in the operator equation $\mathcal{L}I = Q$ and the boundary conditions (A.10), and (ii) for some adjoint source term $Q^\dagger$, the radiance $I^\dagger$ solves the adjoint problem consisting in the operator equation $\mathcal{L}^\dagger I^\dagger = Q^\dagger$ and the boundary conditions* (A.14)*, then (cf. Eq. (A.13))*

$$\left\langle Q^\dagger, I \right\rangle = \left\langle \mathcal{L}^\dagger I^\dagger, I \right\rangle = \left\langle I^\dagger, \mathcal{L} I \right\rangle = \left\langle I^\dagger, Q \right\rangle. \qquad \text{(A.16)}$$

A.3 Solution of the adjoint radiative transfer equation

The forward and adjoint radiative transfer equations $\mathcal{L}I = Q$ and $\mathcal{L}^\dagger I^\dagger = Q^\dagger$, respectively, are related to each other.

Replacing $\boldsymbol{\Omega}$ by $-\boldsymbol{\Omega}$ in the expression of the adjoint transport operator $\mathcal{L}^\dagger$, we rewrite the adjoint radiative transfer equation as $(\mathcal{L}^\dagger I^\dagger)(\mathbf{r},-\boldsymbol{\Omega}) = Q^\dagger(\mathbf{r},-\boldsymbol{\Omega})$. Defining the pseudo-forward radiance $\widehat{I}^\dagger$ by the relation $\widehat{I}^\dagger(\mathbf{r},\boldsymbol{\Omega}) = I^\dagger(\mathbf{r},-\boldsymbol{\Omega})$, and using the identity $H(\mu)|-\mu| = H(\mu)\mu$, we express $(\mathcal{L}^\dagger I^\dagger)(\mathbf{r},-\boldsymbol{\Omega})$ as

$$\begin{aligned}(\mathcal{L}^\dagger I^\dagger)(\mathbf{r},-\boldsymbol{\Omega}) = \frac{\mathrm{d}I^\dagger}{\mathrm{d}s}(\mathbf{r},-\boldsymbol{\Omega}) + \sigma_\mathrm{ext}(\mathbf{r})I^\dagger(\mathbf{r},-\boldsymbol{\Omega}) \\ - \frac{\sigma_\mathrm{sct}(\mathbf{r})}{4\pi}\int_\Omega P(\mathbf{r},\boldsymbol{\Omega}',-\boldsymbol{\Omega})I^\dagger(\mathbf{r},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ - \frac{A_\mathrm{s}}{\pi}\delta(z)H(\mu)\mu\int_\Omega \rho(\mathbf{r},\boldsymbol{\Omega}',-\boldsymbol{\Omega})H(\mu')\mu' I^\dagger(\mathbf{r},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}'. \qquad \text{(A.17)}\end{aligned}$$

Further, using the symmetry properties of the phase function $P(\mathbf{r}, -\boldsymbol{\Omega}', -\boldsymbol{\Omega}) = P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}')$ and of the bi-directional reflection function $\rho(\mathbf{r}, -\boldsymbol{\Omega}', -\boldsymbol{\Omega}) = \rho(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}')$, yielding

$$\begin{aligned}\int_\Omega P(\mathbf{r}, \boldsymbol{\Omega}', -\boldsymbol{\Omega})\widehat{I}^\dagger(\mathbf{r}, -\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' &= \int_\Omega P(\mathbf{r}, -\boldsymbol{\Omega}', -\boldsymbol{\Omega})\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ &= \int_\Omega P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}')\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}',\end{aligned}$$

and $(H(-\mu')|-\mu'| = H(-\mu')|\mu'|)$

$$\begin{aligned}&\int_\Omega \rho(\mathbf{r}, \boldsymbol{\Omega}', -\boldsymbol{\Omega})H(\mu')\mu'\widehat{I}^\dagger(\mathbf{r}, -\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ &= \int_\Omega \rho(\mathbf{r}, \boldsymbol{\Omega}', -\boldsymbol{\Omega})H(\mu')|\mu'|\widehat{I}^\dagger(\mathbf{r}, -\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ &= \int_\Omega \rho(\mathbf{r}, -\boldsymbol{\Omega}', -\boldsymbol{\Omega})H(-\mu')|-\mu'|\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ &= \int_\Omega \rho(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}')H(-\mu')|\mu'|\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}',\end{aligned}$$

respectively, we obtain

$$\begin{aligned}(\mathcal{L}^\dagger I^\dagger)(\mathbf{r}, -\boldsymbol{\Omega}) = &\frac{\mathrm{d}\widehat{I}^\dagger}{\mathrm{d}s}(\mathbf{r}, \boldsymbol{\Omega}) + \sigma_{\mathrm{ext}}(\mathbf{r})\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}) \\ &- \frac{\sigma_{\mathrm{sct}}(\mathbf{r})}{4\pi}\int_\Omega P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}')\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}' \\ &- \frac{A_{\mathrm{s}}}{\pi}\delta(z)H(\mu)\mu\int_\Omega \rho(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}')H(-\mu')|\mu'|\widehat{I}^\dagger(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}',\end{aligned}$$

showing that $(\mathcal{L}^\dagger I^\dagger)(\mathbf{r}, -\boldsymbol{\Omega}) = (\mathcal{L}\widehat{I}^\dagger)(\mathbf{r}, \boldsymbol{\Omega})$. Thus, defining the pseudo-forward source term by the relation $\widehat{Q}^\dagger(\mathbf{r}, \boldsymbol{\Omega}) = Q^\dagger(\mathbf{r}, -\boldsymbol{\Omega})$, we see that the pseudo-forward radiance $\widehat{I}^\dagger$ (i) solves the same type of radiative transfer equation as the forward radiance I, i.e., $(\mathcal{L}\widehat{I}^\dagger)(\mathbf{r}, \boldsymbol{\Omega}) = \widehat{Q}^\dagger(\mathbf{r}, \boldsymbol{\Omega})$), and (ii) satisfies the boundary conditions (A.10).

Appendix B Gradient computation by means of SHDOM

In this appendix, we present a method for computing the gradient in the framework of SHDOM. To simplify the writing we will omit to indicate explicitly the dependency on $\boldsymbol{\sigma}_{\mathrm{ext}}$.

Rewriting the forward transport operator (5) as

$$\begin{aligned}(\mathcal{L}I)(\mathbf{r}, \boldsymbol{\Omega}) = &\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r}, \boldsymbol{\Omega}) + \sigma_{\mathrm{ext}}(\mathbf{r})I(\mathbf{r}, \boldsymbol{\Omega}) - \sigma_{\mathrm{ext}}(\mathbf{r})J(\mathbf{r}, \boldsymbol{\Omega}; I) \\ &- \frac{A_{\mathrm{s}}}{\pi}\delta(z)H(\mu)\mu\int_\Omega H(-\mu')|\mu'|I(\mathbf{r}, \boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}',\end{aligned}$$

where

$$J(\mathbf{r},\mathbf{\Omega};I)=\frac{\omega(\mathbf{r})}{4\pi}\int_{\Omega}P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')I(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}'$$

is now the source function, and taking its derivative with respect to $\sigma_{\mathrm{ext}u}$, yields

$$\Big(\frac{\partial\mathcal{L}}{\partial\sigma_{\mathrm{ext}u}}I\Big)(\mathbf{r},\mathbf{\Omega})=I(\mathbf{r},\mathbf{\Omega})\frac{\partial\sigma_{\mathrm{ext}}}{\partial\sigma_{\mathrm{ext}u}}(\mathbf{r})-J(\mathbf{r},\mathbf{\Omega};I)\frac{\partial\sigma_{\mathrm{ext}}}{\partial\sigma_{\mathrm{ext}u}}(\mathbf{r}).$$

On the adaptive grid, we consider a cell with the global index c and a grid point with the global index j, which belongs to cell c. If $\bar{j}$ is the local index of this grid point within cell c, we define the map g as $(\bar{j},c)\xrightarrow{g}j$, or equivalently, $j=g(\bar{j},c)$. For $\mathbf{r}\in D_c$, where D_c is the domain of cell c, we assume that the extinction and extinction/source function product vary linearly within the cell, i.e.,

$$\sigma_{\mathrm{ext}}(\mathbf{r})=\sum_{\bar{i}}L_{\bar{i}}(\mathbf{R})\sigma_{\mathrm{ext}}(\mathbf{r}_{g(\bar{i},c)}) \tag{B.1}$$

and

$$\sigma_{\mathrm{ext}}(\mathbf{r})J(\mathbf{r},\mathbf{\Omega};I)=\sum_{\bar{i}}L_{\bar{i}}(\mathbf{R})\sigma_{\mathrm{ext}}(\mathbf{r}_{g(\bar{i},c)})J(\mathbf{r}_{g(\bar{i},c)},\mathbf{\Omega};I), \tag{B.2}$$

respectively. Here, $L_{\bar{i}}$ are the first-order interpolation basis functions for a rectangular prism element, $\mathbf{R}=\mathbf{r}-\boldsymbol{\rho}_c$ is the position vector of a point in a local coordinate system attached to cell c, $\boldsymbol{\rho}_c=(1/4)\sum_{\bar{i}}\mathbf{r}_{g(\bar{i},c)}$ the position vector of the center point of cell c, $\mathbf{r}_{g(\bar{i},c)}=\mathbf{r}_i$ the position vector of the grid point with global index $i=g(\bar{i},c)$, and the sum $\sum_{\bar{i}}$ is taken over all grid points $\bar{i}$ of a cell (for a two-dimensional grid, $\bar{i}$ ranges from 1 to 4, while for a three-dimensional grid, $\bar{i}$ ranges from 1 to 8). Setting $\sigma_{\mathrm{ext}g(\bar{i},c)}=\sigma_{\mathrm{ext}}(\mathbf{r}_{g(\bar{i},c)})$, gives

$$\Big(\frac{\partial\mathcal{L}}{\partial\sigma_{\mathrm{ext}u}}I\Big)(\mathbf{r},\mathbf{\Omega})=\sum_{\bar{i}}L_{\bar{i}}(\mathbf{R})[I(\mathbf{r},\mathbf{\Omega})-J(\mathbf{r}_{g(\bar{i},c)},\mathbf{\Omega};I)]\frac{\partial\sigma_{\mathrm{ext}g(\bar{i},c)}}{\partial\sigma_{\mathrm{ext}u}},$$

and further (cf. Eq. (22))

$$\begin{aligned}g_{\mathrm{R}u}=-\sum_{c=1}^{N_{\mathrm{cells}}}\sum_{\bar{i}}\int_{\Omega}\int_{D_c}&L_{\bar{i}}(\mathbf{R})\widehat{I}^{\dagger}(\mathbf{r},-\mathbf{\Omega})\\&\times[I(\mathbf{r},\mathbf{\Omega})-J(\mathbf{r}_{g(\bar{i},c)},\mathbf{\Omega};I)]\frac{\partial\sigma_{\mathrm{ext}g(\bar{i},c)}}{\partial\sigma_{\mathrm{ext}u}}\,\mathrm{d}V\mathrm{d}\mathbf{\Omega},\end{aligned} \tag{B.3}$$

where N_{cells} is the number of cells.

In the next step, we split the forward and pseudo-forward radiances $I(\mathbf{r},\mathbf{\Omega})$ and $\widehat{I}^{\dagger}(\mathbf{r},\mathbf{\Omega})$, respectively, into their diffuse and direct components, i.e.,

$$I(\mathbf{r},\mathbf{\Omega})=I_{\mathrm{d}}(\mathbf{r},\mathbf{\Omega})+I_{\odot}(\mathbf{r},\mathbf{\Omega}), \tag{B.4}$$

$$I_{\odot}(\mathbf{r},\mathbf{\Omega})=\delta(\mathbf{\Omega}-\mathbf{\Omega}_0)\mathcal{T}_0(\mathbf{r}), \tag{B.5}$$

$$\mathcal{T}_0(\mathbf{r})=\frac{F_0}{|\mu_0|}T(\mathbf{r}_0,\mathbf{r},\mathbf{\Omega}_0), \tag{B.6}$$

and

$$\widehat{I}^{\dagger}(\mathbf{r},\mathbf{\Omega}) = \widehat{I}^{\dagger}_{\mathrm{d}}(\mathbf{r},\mathbf{\Omega}) + \widehat{I}^{\dagger}_{\odot}(\mathbf{r},\mathbf{\Omega}), \tag{B.7}$$

$$\widehat{I}^{\dagger}_{\odot}(\mathbf{r},\mathbf{\Omega}) = \sum_{q=1}^{N_{\mathrm{a}}} \delta(\mathbf{\Omega} - \widehat{\mathbf{\Omega}}_{\mathrm{m}q})\widehat{\mathcal{T}}^{\dagger}_{\mathrm{m}q}(\mathbf{r}), \tag{B.8}$$

$$\widehat{\mathcal{T}}^{\dagger}_{\mathrm{m}q}(\mathbf{r}) = \frac{F^{\dagger}_{0q}(\mathbf{r}_{\mathrm{t}q})}{|\widehat{\mu}_{\mathrm{m}q}|} T(\mathbf{r}_{\mathrm{t}q},\mathbf{r},\widehat{\mathbf{\Omega}}_{\mathrm{m}q}) \tag{B.9}$$

respectively, where $\widehat{\mathbf{\Omega}}_{\mathrm{m}q} = -\mathbf{\Omega}_{\mathrm{m}q}$ and

$$T(\mathbf{r}_0,\mathbf{r},\mathbf{\Omega}_0) = \exp\Big(-\int_{\mathbf{r}_0}^{\mathbf{r}} \sigma_{\mathrm{ext}}(\mathbf{r}')\mathrm{d}s'\Big)$$

is the transmission along the characteristic $\mathbf{\Omega}_0$ starting at $\mathbf{r}_0$ and ending at $\mathbf{r}$. Inserting Eqs. (B.4)–(B.8) into Eq. (B.3), assuming the expansions

$$I_{\mathrm{d}}(\mathbf{r},\mathbf{\Omega}) = \sum_{m=-M}^{M}\sum_{n=|m|}^{N} I_{mn}(\mathbf{r})Y_{mn}(\mathbf{\Omega}), \tag{B.10}$$

$$\widehat{I}^{\dagger}_{\mathrm{d}}(\mathbf{r},\mathbf{\Omega}) = \sum_{m=-M}^{M}\sum_{n=|m|}^{N} \widehat{I}^{\dagger}_{mn}(\mathbf{r})Y_{mn}(\mathbf{\Omega}), \tag{B.11}$$

$$J(\mathbf{r},\mathbf{\Omega};I) = \sum_{m=-M}^{M}\sum_{n=|m|}^{N} J_{mn}(\mathbf{r})Y_{mn}(\mathbf{\Omega}), \tag{B.12}$$

where $Y_{mn}(\mathbf{\Omega})$ are the orthonormal real-valued spherical harmonic functions, and N and M are the maximum expansion order and number of azimuthal modes, respectively, using the linear approximation

$$f(\mathbf{r}) = \sum_{\bar{j}} L_{\bar{j}}(\mathbf{R}) f(\mathbf{r}_{g(\bar{j},c)}) = \sum_{\bar{j}} L_{\bar{j}}(\mathbf{R}) f_{g(\bar{j},c)}, \tag{B.13}$$

where $f(\mathbf{r})$ stands for $I_{mn}(\mathbf{r})$, $\widehat{I}^{\dagger}_{mn}(\mathbf{r})$, $\mathcal{T}_0(\mathbf{r})$, and $\widehat{\mathcal{T}}^{\dagger}_{\mathrm{m}q}(\mathbf{r})$, and employing the identity $Y_{mn}(-\mathbf{\Omega}) = (-1)^n Y_{mn}(\mathbf{\Omega})$, we end up with the computational formulas

$$g_{\mathrm{R}u} = T_A + T_B + T_C + T_D + T_E, \tag{B.14}$$

where

$$T_A = -\sum_{c=1}^{N_{\mathrm{cells}}}\sum_{\bar{i},\bar{j},\bar{k}}\sum_{m=-M}^{M}\sum_{n=|m|}^{N} (-1)^n \widehat{I}^{\dagger}_{mn,g(\bar{j},c)} I_{mn,g(\bar{k},c)} \frac{\partial \sigma_{\mathrm{ext}g(\bar{i},c)}}{\partial \sigma_{\mathrm{ext}u}} L_{\bar{i},\bar{j},\bar{k}}, \tag{B.15}$$

$$T_B = \sum_{c=1}^{N_{\mathrm{cells}}}\sum_{\bar{i},\bar{j}}\sum_{m=-M}^{M}\sum_{n=|m|}^{N} (-1)^n \widehat{I}^{\dagger}_{mn,g(\bar{j},c)} J_{mn,g(\bar{i},c)} \frac{\partial \sigma_{\mathrm{ext}g(\bar{i},c)}}{\partial \sigma_{\mathrm{ext}u}} L_{\bar{i},\bar{j}}, \tag{B.16}$$

and

$$T_C = -\sum_{c=1}^{N_{\text{cells}}} \sum_{\bar{i},\bar{j},\bar{k}} \widehat{I}_{\text{d}}^{\dagger}(\mathbf{r}_{g(\bar{j},c)}, -\mathbf{\Omega}_0) \mathcal{T}_{0g(\bar{k},c)} \frac{\partial \sigma_{\text{ext}g(\bar{i},c)}}{\partial \sigma_{\text{ext}u}} L_{\bar{i},\bar{j},\bar{k}}, \tag{B.17}$$

$$T_D = -\sum_{c=1}^{N_{\text{cells}}} \sum_{q=1}^{N_{\text{a}}} \sum_{\bar{i},\bar{j},\bar{k}} I_{\text{d}}(\mathbf{r}_{g(\bar{j},c)}, \mathbf{\Omega}_{\text{m}q}) \widehat{\mathcal{T}}^{\dagger}_{\text{m}q,g(\bar{k},c)} \frac{\partial \sigma_{\text{ext}g(\bar{i},c)}}{\partial \sigma_{\text{ext}u}} L_{\bar{i},\bar{j},\bar{k}}, \tag{B.18}$$

$$T_E = \sum_{c=1}^{N_{\text{cells}}} \sum_{q=1}^{N_{\text{a}}} \sum_{\bar{i},\bar{j}} \sum_{m=-M}^{M} \sum_{n=|m|}^{N} Y_{mn}(\mathbf{\Omega}_{\text{m}q}) \widehat{\mathcal{T}}^{\dagger}_{\text{m}q,g(\bar{j},c)} J_{mn,g(\bar{i},c)} \frac{\partial \sigma_{\text{ext}g(\bar{i},c)}}{\partial \sigma_{\text{ext}u}} L_{\bar{i},\bar{j}}. \tag{B.19}$$

Comments.

1. The input parameters of SHDOM are the extinction coefficient σ_{ext}, the single scattering albedo ω, and the expansion coefficients of the phase function χ_n, specified at all grid points on the base grid. The base grid point values of the extinction field are the unknowns of the inverse problem. An adaptive grid is implemented to add grid points in regions where the source function is changing more rapidly. The adaptive grid evolves from the base grid by splitting cells where more resolution is judged to be needed. Actually, a base grid cell, also called a "parent cell", is split into "child cells" to achieve higher spatial resolution. In this regard, the sums in Eqs. (B.15)–(B.19) are taken over all child cells. Moreover, the derivative $(\partial\sigma_{\text{ext}}/\partial\sigma_{\text{ext}u})(\mathbf{r}_i)$ is computed as (i) $(\partial\sigma_{\text{ext}}/\partial\sigma_{\text{ext}u})(\mathbf{r}_i) = \delta_{iu}$, if $\mathbf{r}_i$ is a point on the base grid, and (ii) by interpolating the base grid point values, if $\mathbf{r}_i$ is a point on the adaptive grid but not on the base grid. If the delta-M scaling method is applied, σ_{ext}, ω, and χ_n are scaled before their use. If $\overline{\sigma}_{\text{ext}}$, $\overline{\omega}$, and $\overline{\chi}_n$ are the scaled quantities, given respectively, by

$$\overline{\sigma}_{\text{ext}} = (1 - f\omega)\sigma_{\text{ext}}, \tag{B.20}$$

$$\overline{\omega} = \frac{1 - f}{1 - f\omega}\omega, \tag{B.21}$$

$$\frac{2}{2n+1}\overline{\chi}_n = \frac{1}{1-f}\Big(\frac{2}{2n+1}\chi_n - 2f\Big), \quad n = 0, \ldots, N, \tag{B.22}$$

where $f = \chi_{N+1}/(2N+3)$ is the truncation factor, we have

$$\frac{\partial \overline{\sigma}_{\text{ext}}}{\partial \sigma_{\text{ext}u}}(\mathbf{r}_i) = (1 - f\omega)\frac{\partial \sigma_{\text{ext}}}{\partial \sigma_{\text{ext}u}}(\mathbf{r}_i),$$

and the computational formulas (B.14)–(B.19) remain valid, but with $\partial\sigma_{\text{ext}}/\partial\sigma_{\text{ext}u}$ and $T(\mathbf{r}_0, \mathbf{r}, \mathbf{\Omega}_0)$ replaced by $\partial\overline{\sigma}_{\text{ext}}/\partial\sigma_{\text{ext}u}$ and

$$\overline{T}(\mathbf{r}_0, \mathbf{r}, \mathbf{\Omega}_0) = \exp\Big(-\int_{\mathbf{r}_0}^{\mathbf{r}} \overline{\sigma}_{\text{ext}}(\mathbf{r}')\mathrm{d}s'\Big),$$

respectively.

2. The pseudo-forward source term, defined as $\widehat{Q}^{\,\dagger}(\mathbf{r},\mathbf{\Omega}) = Q^{\dagger}(\mathbf{r},-\mathbf{\Omega})$ with $Q^{\dagger}(\mathbf{r},\mathbf{\Omega})$ as in Eq. (20), is given by

$$\widehat{Q}^{\,\dagger}(\mathbf{r},\mathbf{\Omega}) = \sum_{q=1}^{N_{\mathrm{a}}} F_{0q}^{\dagger}(\mathbf{r}_{\mathrm{t}})\delta(z-L_{\mathrm{z}})\delta(\mathbf{\Omega}-\widehat{\mathbf{\Omega}}_{\mathrm{m}q}).$$

As a result, the pseudo-forward diffuse radiance $\widehat{I}_{\mathrm{d}}^{\dagger}(\mathbf{r},\mathbf{\Omega})$ satisfies (i) the radiative transfer equation

$$\frac{\mathrm{d}\widehat{I}_{\mathrm{d}}^{\dagger}}{\mathrm{d}s}(\mathbf{r},\mathbf{\Omega}) = -\sigma_{\mathrm{ext}}(\mathbf{r})\widehat{I}_{\mathrm{d}}^{\dagger}(\mathbf{r},\mathbf{\Omega}) + \sigma_{\mathrm{ext}}(\mathbf{r})J(\mathbf{r},\mathbf{\Omega};\widehat{I}_{\mathrm{d}}^{\dagger})$$

with the source function

$$\begin{aligned} J(\mathbf{r},\mathbf{\Omega};\widehat{I}_{\mathrm{d}}^{\dagger}) = {} & \frac{\omega(\mathbf{r})}{4\pi}\sum_{q=1}^{N_{\mathrm{a}}}\frac{F_{0q}^{\dagger}(\mathbf{r}_{\mathrm{t}q})}{|\widehat{\mu}_{\mathrm{m}q}|}T(\mathbf{r}_{\mathrm{t}q},\mathbf{r},\widehat{\mathbf{\Omega}}_{\mathrm{m}q}) \\ & + \frac{\omega(\mathbf{r})}{4\pi}\int_{\Omega} P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')\widehat{I}_{\mathrm{d}}^{\dagger}(\mathbf{r},\mathbf{\Omega}')\,\mathrm{d}\mathbf{\Omega}', \end{aligned}$$

and (ii) the boundary condition

$$\begin{aligned} \widehat{I}_{\mathrm{d}}^{\dagger}(\mathbf{r}_{\mathrm{b}},\mathbf{\Omega}^{+}) = {} & \frac{A}{\pi}\sum_{q=1}^{N_{\mathrm{a}}} F_{0q}^{\dagger}(\mathbf{r}_{\mathrm{t}q})T(\mathbf{r}_{\mathrm{t}q},\mathbf{r}_{\mathrm{b}},\widehat{\mathbf{\Omega}}_{\mathrm{m}q}) \\ & + \frac{A}{\pi}\int_{\Omega^{-}} |\mu^{-}|\widehat{I}_{\mathrm{d}}^{\dagger}(\mathbf{r}_{\mathrm{b}},\mathbf{\Omega}^{-})\,\mathrm{d}\mathbf{\Omega}^{-} \end{aligned}$$

at the bottom surface.

3. In Eqs. (B.15)–(B.19) the interpolations coefficients $L_{\bar{i},\bar{j},\bar{k}}$ and $L_{\bar{i},\bar{j}}$, given respectively, by

$$L_{\bar{i},\bar{j},\bar{k}} = \int_{D_c} L_{\bar{i}}(\mathbf{R})L_{\bar{j}}(\mathbf{R})L_{\bar{k}}(\mathbf{R})\,\mathrm{d}V, \tag{B.23}$$

$$L_{\bar{i},\bar{j}} = \int_{D_c} L_{\bar{i}}(\mathbf{R})L_{\bar{j}}(\mathbf{R})\,\mathrm{d}V, \tag{B.24}$$

are computed analytically in a local coordinate system attached to the grid cell and stored in a lookup table.

4. For $i = g(\bar{i},c)$, $\omega_i = \omega(\mathbf{r}_i)$, $\chi_{n,i} = \chi_n(\mathbf{r}_i)$, and $\mathcal{T}_{0i} = \mathcal{T}_0(\mathbf{r}_i)$, the expansion coefficients of the source function in the expressions of T_B and T_E given by Eqs. (B.16) and (B.19), respectively, are calculated as

$$J_{mn,i} = \omega_i\frac{\chi_{n,i}}{2n+1}I_{mn,i} + \omega_i\frac{\chi_{n,i}}{2n+1}Y_{mn}(\mathbf{\Omega}_0)\mathcal{T}_{0i}. \tag{B.25}$$

5. The diffuse radiance at a point $\mathbf{r}_{\mathrm{t}} \in S_{\mathrm{t}}$ in direction $\mathbf{\Omega}_{\mathrm{m}q}$ is computed by integrating the source function through the medium, that is,

$$\begin{aligned} I_{\mathrm{d}}(\mathbf{r}_{\mathrm{t}},\mathbf{\Omega}_{\mathrm{m}q}) = {} & I_{\mathrm{d}}(\mathbf{r}_{\mathrm{b}},\mathbf{\Omega}_{\mathrm{m}q})T(\mathbf{r}_{\mathrm{b}},\mathbf{r}_{\mathrm{t}},\mathbf{\Omega}_{\mathrm{m}q}) \\ & + \int_{\mathbf{r}_{\mathrm{b}}}^{\mathbf{r}_{\mathrm{t}}} \overline{\sigma}_{\mathrm{ext}}(\mathbf{r})J(\mathbf{r},\mathbf{\Omega}_{\mathrm{m}q};I_{\mathrm{d}})T(\mathbf{r},\mathbf{r}_{\mathrm{t}},\mathbf{\Omega}_{\mathrm{m}q})\,\mathrm{d}s. \end{aligned} \tag{B.26}$$

Similarly, the radiances $\widehat{I}_{\mathrm{d}}^{\dagger}(\mathbf{r}_{g(\overline{j},c)}, -\mathbf{\Omega}_0)$ and $I_{\mathrm{d}}(\mathbf{r}_{g(\overline{j},c)}, \mathbf{\Omega}_{\mathrm{m}q})$ in the expressions of T_C and T_D given by Eqs. (B.17) and (B.18), respectively, are also computed by the source integration method.

6. For solar problems with the delta-M method and when the TMS method is applied, we compute in a first step, the TMS correction of the source function as

$$\begin{aligned}\Delta J(\mathbf{r}_i, \mathbf{\Omega}_{\mathrm{m}q}, \boldsymbol{\sigma}_{\mathrm{ext}}) = \frac{F_0}{|\mu_0|}\overline{T}(\mathbf{r}_{0i}, \mathbf{r}_i, \mathbf{\Omega}_0; \boldsymbol{\sigma}_{\mathrm{ext}}) \\ \times \Big[\frac{\omega(\mathbf{r}_i)}{1 - f\omega(\mathbf{r}_i)}\frac{1}{4\pi}\sum_{n=1}^{N_{\mathrm{rank}}}\chi_n(\mathbf{r}_i)\widetilde{P}_n(\cos\Theta_q) \\ - \overline{\omega}(\mathbf{r}_i)\sum_{m=-M}^{M}\sum_{n=|m|}^{N}\frac{\overline{\chi}_n(\mathbf{r}_i)}{2n+1}Y_{mn}(\mathbf{\Omega}_0)Y_{mn}(\mathbf{\Omega}_{\mathrm{m}q})\Big],\end{aligned} \tag{B.27}$$

where $\widetilde{P}_n(\cos\Theta_q)$ are the unnormalized Legendre polynomials and $\cos\Theta_q = \mathbf{\Omega}_0 \cdot \mathbf{\Omega}_{\mathrm{m}q}$, and in a second step, the TMS correction of the signal as

$$\cdot\mathcal{I}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}) = \frac{1}{A}\int_{S_{\mathrm{t}}} h(\mathbf{r}_{\mathrm{t}} - \mathbf{r}_{\mathrm{t}p})\Delta I(\mathbf{r}_{\mathrm{t}}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}})\,\mathrm{d}S_{\mathrm{t}}, \tag{B.28}$$

$$\Delta I(\mathbf{r}_{\mathrm{t}}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}) = \int_{\mathbf{r}_{\mathrm{b}}}^{\mathbf{r}_{\mathrm{t}}} \overline{\sigma}_{\mathrm{ext}}(\mathbf{r})\triangle J(\mathbf{r}, \mathbf{\Omega}_{\mathrm{m}q}, \boldsymbol{\sigma}_{\mathrm{ext}})\overline{T}(\mathbf{r}, \mathbf{r}_{\mathrm{t}}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}})\,\mathrm{d}s, \tag{B.29}$$

where the grid point values $\Delta J(\mathbf{r}_i, \mathbf{\Omega}_{\mathrm{m}q}, \boldsymbol{\sigma}_{\mathrm{ext}})$ are used to compute the integral of the source function $\Delta J(\mathbf{r}, \mathbf{\Omega}_{\mathrm{m}q}, \boldsymbol{\sigma}_{\mathrm{ext}})$ in Eq. (B.29). Finally, we determine the corrected residual and the TMS correction of the gradient by using the relations

$$\begin{aligned}R(\boldsymbol{\sigma}_{\mathrm{ext}}) = \frac{1}{2}\sum_{q=1}^{N_{\mathrm{a}}}\sum_{p=1}^{N_{\mathrm{p}}}[\mathcal{I}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}) + \Delta\mathcal{I}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}) \\ - \mathcal{I}_{\mathrm{mes}}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}^{\star}_{\mathrm{ext}})]^2,\end{aligned} \tag{B.30}$$

and

$$\begin{aligned}\Delta g_{\mathrm{R}u}(\boldsymbol{\sigma}_{\mathrm{ext}}) = \sum_{q=1}^{N_{\mathrm{a}}}\sum_{p=1}^{N_{\mathrm{p}}}[\mathcal{I}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}) + \Delta\mathcal{I}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}) \\ - \mathcal{I}_{\mathrm{mes}}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}^{\star}_{\mathrm{ext}})]\frac{\partial\Delta\mathcal{I}}{\partial\sigma_{\mathrm{ext}u}}(\mathbf{r}_{\mathrm{t}p}, \mathbf{\Omega}_{\mathrm{m}q}; \boldsymbol{\sigma}_{\mathrm{ext}}),\end{aligned} \tag{B.31}$$

respectively. In Eq. (B.31), the partial derivatives of the signal correction $\partial\Delta\mathcal{I}/\partial\sigma_{\mathrm{ext}u}$ are computed analytically by means of Eqs. (B.27)–(B.29).

[1] Peris-Ferrus C, Gomez-Amo JL, Catalan-Valdelomar P, Scarlatti F, Emde C, Utrillas MP. Retrieval of cloud optical depth through radiative transfer and remote sensing: from 1D to 3D approach. Proc. SPIE 11859, Remote Sensing of Clouds and the Atmosphere XXVI, 2021. https://doi.org/10.1117/12.2599591.

[2] Cahalan R, Ridgway W, Wiscombe W, Bell TL, Snider JB. The Albedo of Fractal Stratocumulus Clouds. J Atmos Sci 1994;51:2434 –2455.

[3] Benner TC, Evans F. Three-dimensional solar radiative transfer in small tropical cumulus fields derived from high-resolution imagery. Journal of Geophysical Research: Atmospheres 2001;106(D14).

[4] Marshak A, Davis A, Wiscombe A, Cahalan R. Radiative smoothing in fractal clouds. Journal of Geophysical Research: Atmospheres 1995;100(D12):26247–26261.

[5] Chambers LH, Wielicki BA, Evans K F (1997). Accuracy of the independent pixel approximation for satellite estimates of oceanic boundary layer cloud optical depth. J Geophys Res 1997;102(D2):1779– 1794. doi:10.1029/96JD02995.

[6] Spurr RJD. A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. JQSRT 2001;68:689-735.

[7] Spurr RJD. LIDORT and VLIDORT. Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In Kokhanovsky AA editor. Light Scattering Reviews Vol. 3, Springer Verlag, Berlin, 2008. 229-71.

[8] Marchuk GI. Equation for the value of information from weather satellites and formulation of inverse problems. Cosmic Res 1964;2:394-409.

[9] Marchuk GI. Adjoint equations and analysis of complex systems. Amsterdam: Kluwer; 1995.

[10] Box MA. Radiative perturbation theory: a review. Environ Modelling Software 2002;17:95-106.

[11] Ustinov EA. Adjoint sensitivity analysis of radiative transfer equation: temperature and gas mixing ratio weighting functions for remote sensing of scattering atmospheres in thermal IR. JQSRT 2001;68:195-211.

[12] Ustinov EA. Atmospheric weighting functions and surface partial derivatives for remote sensing of scattering planetary atmospheres in thermal spectral region: general adjoint approach. JQSRT 2005;92:351-71

[13] Walter H, Landgraf J, Hasekamp OP. Linearization of a pseudo-spherical vector radiative transfer model. JQSRT 2004;85:251-83.

[14] Rozanov VV, Rozanov AV. Relationship between different approaches to derive weighting functions related to atmospheric remote sensing problems. JQSRT 2007;105:217-42.

[15] Landgraf J, Hasekamp OP, Box MA, Trautmann T. A linearized radiative transfer model for ozone profile retrieval using the analytical forward-adjoint perturbation theory approach. J Geophys Res 2001;106(D21):27291-305.

[16] Evans KF. The spherical harmonic discrete ordinate method for three-dimensional atmospheric radiative transfer. J Atmos Sci 1998;55:429-446.

[17] Doicu A, Efremenko D. Linearizations of the Spherical Harmonic Discrete Ordinate Method (SHDOM). Atmosphere 2019;10:292. doi:10.3390/atmos10060292.

[18] Nakajima T, Tanaka M. Algorithms for radiative intensity calculations in moderately thick atmos using a truncation approximation. JQSRT 1988;40:51–69.

[19] Ren K, Bal G, Hielscher A. Frequency Domain Optical Tomography Based on the Equation of Radiative Transfer. SIAM Journal on Scientific Computing 2006;28:1463–1489. doi:10.1137/040619193.

[20] Kim HK, A. Charette A. A sensitivity function-based conjugate gradient method for optical tomography with the frequency-domain equation of radiative transfer. JQSRT 2007;104:24–39. doi:10.1016/j.jqsrt.2006.08.007.[

[21] Kim HK, Hielscher AH. A PDE-constrained SQP algorithm for optical tomography based on the frequency-domain equation of radiative transfer. Inverse Problems 2009;25:015010. doi:10.1088/0266-5611/25/1/015010

[22] Hasekamp O, Landgraf J. Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing. Journal of Geophysical Research: Atmospheres 2005;110:D04203.

[23] Ukhinov SA, Yurkov DI. Computation of the parametric derivatives of polarized radiation and the solution of inverse atmospheric optics problems. Russian Journal of Numerical Analysis and Mathematical Modelling 2002;17:283–303.

[24] Postylyakov OV. Linearized vector radiative transfer model MCC++ for a spherical atmosphere. JQSRT 2004;88:297–317.

[25] Doicu A, Trautmann T. Adjoint problem of radiative transfer for a pseudo-spherical atmosphere and general viewing geometries. JQSRT 2009;110:464–476.

[26] Walter HH, Landgraf J. Linearization of radiative transfer in spherical geometry: an application of the forward-adjoint perturbation theory. in: Kokhanovsky AA (Ed.), Light Scattering Reviews 5, Springer Praxis Books, Springer Berlin Heidelberg, 2010, pp. 105–145.

[27] Martin W, Cairns B, Bal G. Adjoint methods for adjusting three-dimensional atmosphere and surface properties to fit multi-angle/multi-pixel polarimetric measurements. JQSRT 2014;144:68-85.

[28] Martin WGK, Hasekamp OP. A demonstration of adjoint methods for multi-dimensional remote sensing of the atmosphere and surface. JQSRT 2018;204:215–231.

[29] Levis A, Schechner YY, Aides A, Davis AB. Airborne three-dimensional cloud tomography. 2015 IEEE Int Conf on Computer Vision (ICCV), Santiago, Chile, IEEE, 2015;3379–3387. https://doi.org/10.1109/ICCV.2015.386.

[30] Doicu A, Doicu A, Efremenko D, Trautmann T. Cloud tomographic retrieval algorithms. I: Surrogate minimization method. JQSRT 2022;277. https://doi.org/10.1016/j.jqsrt.2021.107954.

[31] Forster L, Davis AB, Diner DJ, Mayer B. Toward Cloud Tomography from Space using MISR and MODIS: Locating the "Veiled Core" in Opaque Convective Clouds. Journal of the Atmospheric Sciences 2020;78:155–166. doi.org/10.1175/JAS-D-19-0262.1

[32] Nesterov Y. A method of solving a convex programming problem with convergence rate o(1/k2). Soviet Mathematics Doklady 1983;27:372–376.

[33] Byrd RH, Lu P, Nocedal J, Zhu C. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 1995;16:1190-1208.

[34] Lee B, Di Girolamo L, Zhao G, Zhan Y. Three-Dimensional Cloud Volume Reconstruction from the Multi-angle Imaging SpectroRadiometer. Remote Sensing 2018;10:1858. doi:10.3390/rs10111858

[35] Yu H, Ma J, Ahmad S, Sun E , Li C, Li Z, Hong J. Three-Dimensional Cloud Structure Reconstruction from the Directional Polarimetric Camera. Remote Sensing 2019;11:2894. doi:10.3390/rs11242894

[36] Duchi J, Hazan E, Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research 2011;12:2121–2159.

[37] Zeiler MD. ADADELTA: An Adaptive Learning Rate Method. 2012: http://arxiv.org/abs/1212.570

[38] Kingma DP, Ba JL. Adam: a Method for Stochastic Optimization. International Conference on Learning Representations 2015:1–13.

[39] Dozat T. Incorporating Nesterov Momentum into Adam. ICLR Workshop 2016;1:2013–2016.

[40] Doicu A, Trautmann T, Schreier F. Numerical Regularization for Atmospheric Inverse Problems. Berlin: Springer; 2010.

[41] Levis A, Schechner YY, Davis AB. Multiple-scattering microphysics tomography. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017; 6740–6749.

[42] Levis A, Schechner YY, Davis AB, Loveridge J. Multi-view polarimetric scattering cloud tomography and retrieval of droplet size. Remote Sensing 2020;12:2831.