

Cloud tomographic retrieval algorithms. I: Surrogate minimization method.

Adrian Doicu¹, Alexandru Doicu², Dmitry Efremenko¹,
and Thomas Trautmann¹

November 29, 2022

¹Remote Sensing Technology Institute, German Aerospace Centre, Oberpfaffenhofen, Wessling, Germany

²Institute of Mathematics, University of Augsburg, Germany

Abstract

A cloud tomographic retrieval algorithm relying on (i) the spherical harmonics discrete ordinate method for computing the radiative transfer and (ii) the surrogate minimization method for solving the inverse problem has been designed. The retrieval algorithm uses regularization, accelerated projected gradient methods, and two types of surrogate functions. The performances of the retrieval algorithm are analyzed on a few synthetic two- and three-dimensional problems.

1 Introduction

Remote sensing of clouds in the Earth’s atmosphere, using space-borne optical sensors, usually assumes horizontally homogeneous clouds and is focused on the retrieval of simple parameters, such as the cloud optical thickness and cloud top height. For this purpose, one-dimensional radiative transfer codes are used. This very crude approximation of the real cloud structure causes biases of the retrieved parameters in most cases.

A promising approach for treating clouds as multi-dimensional objects is cloud tomography. In cloud tomography, the focus is on the retrieval of cloud extinction field, and under some assumptions, of the microphysical parameters of the cloud. Some important contributions in this field are summarized below.

1. Levis et al. [1] demonstrated a full three-dimensional tomographic cloud retrieval by using as radiative transfer model, the spherical harmonics discrete ordinate method (SHDOM) developed by Evans [2]. The inverse problem is solved by the surrogate minimization method in conjunction with the gradient descent method. Specifically, the authors used reflectance measurements for nine different viewing angles and reconstructed the three-dimensional cloud extinction field on a Cartesian grid. In a subsequent paper, Levis et al. [3] retrieved the extinction field while treating the effective radius and variance of the cloud particles as unknowns, but constrained to vary only along the vertical direction. This work was extended in Ref. [4] by adding polarization and deriving a framework for a full three-dimensional tomography of cloud droplets for both their mass concentration in space and their distribution across a range of sizes.
2. Martin and Hasekamp [5] used adjoint methods to retrieve two-dimensional cloud extinction fields. The inverse problem is solved with a gradient-based quasi-Newton approach, i.e., the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [6]. At each step, the value of the objective function to be minimized (the misfit function) and its gradient are computed with two calls of the radiative transfer model. In the first call, the value of 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. The radiative transfer model is a simplified two-dimensional model, in which, the set of directions is a circle defined by one angular variable, rather than a sphere defined

by two angular variables. Although the problem is simplified, the main concepts of cloud tomography are fully represented.

In a series of two papers we intend to describe cloud tomographic retrieval algorithms that use SHDOM for computing the radiative transfer and either the surrogate minimization method or the adjoint method for solving the inverse problem. In the present paper, the focus is on the surrogate minimization method. Actually, surrogate minimization algorithms are a family of algorithms that can be regarded as a generalization of expectation minimization algorithms [7]. Such an algorithm aims at turning an otherwise intractable minimization problem into a tractable one by iterating two steps. The surrogate step computes a tractable surrogate function to substitute the original objective function and the minimization step seeks to minimize this surrogate function. In the literature, surrogate minimization algorithms are also referred to as optimization transfer algorithms [8], iterative majorization [9], or auxiliary function method [10]. These algorithms are very efficient because they can (i) turn a non-differentiable problem into a smooth problem, (ii) separate the parameters of a problem, (iii) linearize an optimization problem, (iv) deal gracefully with equality and inequality constraints, and (v) generate an algorithm that avoids large matrix inversion. Iteration is the price we pay for simplifying the original problem.

Levis et al. [1] showed how to construct a surrogate function in cloud tomography. The key point is to consider a surrogate function, in which the source function and the surface radiance, computed by a three-dimensional radiative model, are kept constant during the minimization step. Our aim is to implement this idea in an efficient and versatile retrieval algorithm. The paper is organized as follows. After formulating the radiative transfer problem in Section 2, we provide a succinct presentation of SHDOM in Section 3. In Section 4, we formulate the inverse problem in cloud tomography, while in Section 5, we describe in detail the surrogate minimization method. The construction of the regularization term is discussed in Section 6, the performances of the retrieval algorithm are analyzed in Section 5, while the final section of our paper contains a few concluding remarks.

2 Problem formulation

We consider the solar radiative transfer in a rectangular prism of lengths L_x , L_y and L_z . The top and the bottom faces of the prism are denoted by S_t and S_b , respectively, while the lateral faces are denoted by S_{1x} ($x = 0$), S_{2x} ($x = L_x$), S_{1y} ($y = 0$), and S_{2y} ($y = L_y$). The boundary-value problem for the diffuse radiance at point \mathbf{r} in direction $\boldsymbol{\Omega}$ consists in (i) the inhomogeneous differential equation

$$\frac{dI}{ds}(\mathbf{r}, \boldsymbol{\Omega}) = -\sigma_{\text{ext}}(\mathbf{r})I(\mathbf{r}, \boldsymbol{\Omega}) + \sigma_{\text{ext}}(\mathbf{r})J(\mathbf{r}, \boldsymbol{\Omega}), \quad (1)$$

(ii) the boundary conditions at the top and bottom surfaces

$$I(\mathbf{r}_t, \boldsymbol{\Omega}^-) = 0, \quad \mathbf{r}_t \in S_t, \quad (2)$$

and (in the case of a Lambertian surface)

$$\begin{aligned} I(\mathbf{r}_b, \boldsymbol{\Omega}^+) &= \frac{A_s}{\pi} F_0 T(\mathbf{r}_{0b}, \mathbf{r}_b) \\ &+ \frac{A_s}{\pi} \int_{\Omega^-} |\mu^-| I(\mathbf{r}_b, \boldsymbol{\Omega}^-) d\Omega^-, \quad \mathbf{r}_b \in S_b, \end{aligned} \quad (3)$$

respectively, and (iii) the (periodic) boundary conditions at the lateral faces,

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

for $\mathbf{r}_{ix} \in S_{ix}$ and $\mathbf{r}_{iy} \in S_{iy}$ with $i = 1, 2$. Here,

$$\begin{aligned} J(\mathbf{r}, \boldsymbol{\Omega}) &= \frac{\omega(\mathbf{r})}{4\pi} \frac{F_0}{|\mu_0|} P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0) T(\mathbf{r}_0, \mathbf{r}) \\ &+ \frac{\omega(\mathbf{r})}{4\pi} \int_{\Omega} P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}') I(\mathbf{r}, \boldsymbol{\Omega}') d\Omega', \end{aligned} \quad (5)$$

is the source function,

$$T(\mathbf{r}_0, \mathbf{r}) = \exp\left(-\int_{\mathbf{r}_0}^{\mathbf{r}} \sigma_{\text{ext}}(\mathbf{r}') d\mathbf{s}'\right) \quad (6)$$

is the transmission along the characteristic starting at \mathbf{r}_0 and ending at \mathbf{r} , σ_{ext} and $\sigma_{\text{sct}} = \omega\sigma_{\text{ext}}$ are the extinction and scattering coefficients, respectively, ω is the single-scattering albedo, P is the phase function, A_s is the surface albedo, $\mathbf{\Omega}_0 = (\mu_0, \varphi_0)$ with $\mu_0 < 0$, is the solar direction, F_0 is the solar flux, $\mathbf{\Omega}^+$ 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, $\mathbf{r}_{2x} = \mathbf{r}_{1x} + L_x \mathbf{i}$, $\mathbf{r}_{2y} = \mathbf{r}_{1y} + L_y \mathbf{j}$, and $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ are the Cartesian unit vectors.

3 Spherical harmonics discrete ordinate method

In cloud tomography, we use the radiances measured by a detector from multiple locations and viewing angles. For each viewing direction $\mathbf{\Omega}_{mq}$, $q = 1, \dots, N_a$, we assume that the signal of the p th detector pixel, that collects the radiances around the location $\mathbf{r}_{tp} \in S_t$ with $p = 1, \dots, N_p$, can be modeled as

$$\mathcal{I}(\mathbf{r}_{tp}, \mathbf{\Omega}_{mq}) = \frac{1}{A} \int_{S_t} h(\mathbf{r}_t - \mathbf{r}_{tp}) I(\mathbf{r}_t, \mathbf{\Omega}_{mq}) dS_t, \quad (7)$$

where $h(\mathbf{r}_t - \mathbf{r}_{tp})$ is the characteristic function of the p th detector pixel projected on the top surface, A the area of the top face of the prism, and $I(\mathbf{r}_t, \mathbf{\Omega}_{mq})$ the radiance at point $\mathbf{r}_t \in S_t$ in direction $\mathbf{\Omega}_{mq}$.

In the present study, the top-of-atmosphere radiance $I(\mathbf{r}_t, \mathbf{\Omega}_{mq})$, which enters in Eq. (7), is computed by using our own implementation of SHDOM. Essentially, SHDOM uses discrete ordinates $\mathbf{\Omega}_{jk} = (\mu_j, \varphi_k)$, $j = 1, \dots, N_\mu$, $k = 1, \dots, N_\varphi$, and spherical harmonics $Y_{mn}(\mathbf{\Omega})$, $m = -M, \dots, M$, $n = |m|, \dots, N$, to represent the radiance field at a set of grid points \mathbf{r}_i , $i = 1, \dots, N_{\text{pts}}$, where N_μ is the number of Gaussian quadrature cosine zenith angles μ_j , N_φ the number of discrete azimuth angles φ_k , $N = N_\mu - 1$ the maximum expansion order, $M = N_\varphi/2 - 1$ the maximum number of azimuthal modes, and N_{pts} the number of grid points. Note that in SHDOM, an adaptive grid is implemented to add grid points in regions where more resolution is judged to be needed, while for solar problems with the delta-M method, the TMS method of Nakajima and Tanaka [11] is used to compute the source function. The output quantities of SHDOM, that are relevant in our analysis, are the grid point values of (i) the expansion coefficients of the source function in terms of spherical harmonics $J_{mn}(\mathbf{r}_i)$, and (ii) the surface radiance at the bottom surface in the upward discrete ordinate direction $\mathbf{\Omega}_{jk}^+$, $I(\mathbf{r}_{bi}, \mathbf{\Omega}_{jk}^+)$. The radiance at a point $\mathbf{r}_t \in S_t$ in direction $\mathbf{\Omega}_{mq}$ is computed by integrating the source function through the medium, that is,

$$I(\mathbf{r}_t, \mathbf{\Omega}_{mq}) = I(\mathbf{r}_b, \mathbf{\Omega}_{mq}) T(\mathbf{r}_b, \mathbf{r}_t) + \int_{\mathbf{r}_b}^{\mathbf{r}_t} \sigma_{\text{ext}}(\mathbf{r}) J(\mathbf{r}, \mathbf{\Omega}_{mq}) T(\mathbf{r}, \mathbf{r}_t) d\mathbf{s}. \quad (8)$$

In Eq. (8), the grid point values

$$J(\mathbf{r}_i, \mathbf{\Omega}_{mq}) = \sum_{m=-M}^M \sum_{n=|m|}^N J_{mn}(\mathbf{r}_i) Y_{mn}(\mathbf{\Omega}_{mq}) \quad (9)$$

are used to compute the integral of the source function $J(\mathbf{r}, \mathbf{\Omega}_{mq})$, while for a Lambertian surface, the grid point surface radiances

$$I(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}) = I(\mathbf{r}_{bi}, \mathbf{\Omega}_{jk}^+) \quad (10)$$

are used to compute the surface radiance $I(\mathbf{r}_b, \mathbf{\Omega}_{mq})$ by bilinear interpolation. A brief summary of SHDOM is given in Appendix A.

4 Inverse problem

For an atmosphere consisting of a cloud and air molecules, the extinction field is given by

$$\sigma_{\text{ext}}(\mathbf{r}) = \sigma_{\text{ext}}^{\text{cld}}(\mathbf{r}) + \sigma_{\text{ext}}^{\text{mol}}(\mathbf{r}), \quad (11)$$

where $\sigma_{\text{ext}}^{\text{cld}}(\mathbf{r})$ and $\sigma_{\text{ext}}^{\text{mol}}(\mathbf{r})$ are the extinction coefficients of the cloud and air molecules, respectively. In cloud tomography, we seek to retrieve $\sigma_{\text{ext}}(\mathbf{r})$ within the volume of the cloud. This is equivalent to seeking $\sigma_{\text{ext}}^{\text{cld}}(\mathbf{r})$ since $\sigma_{\text{ext}}^{\text{mol}}(\mathbf{r})$ is known. For a numerical retrieval, the extinction field is represented on a discrete grid $\{\mathbf{r}_i\}_{i=1}^{N_{\text{pts}}}$, and the grid point values of the extinction coefficient $\sigma_{\text{ext}}(\mathbf{r}_i)$ are encapsulated in the extinction vector

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

In this regard, cloud tomographic retrieval can be formulated as an estimation of $\boldsymbol{\sigma}_{\text{ext}}$ that minimizes the objective function

$$E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}}) = R(\boldsymbol{\sigma}_{\text{ext}}) + \alpha L(\boldsymbol{\sigma}_{\text{ext}}), \quad (12)$$

where

$$R(\boldsymbol{\sigma}_{\text{ext}}) = \frac{1}{2} \sum_{q=1}^{N_a} \sum_{p=1}^{N_p} [\mathcal{I}(\mathbf{r}_{tp}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}) - \mathcal{I}_{\text{mes}}(\mathbf{r}_{tp}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}^{\dagger})]^2 \quad (13)$$

with (cf. Eqs. (7) and (8))

$$\mathcal{I}(\mathbf{r}_{tp}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}) = \frac{1}{A} \int_{S_t} h(\mathbf{r}_t - \mathbf{r}_{tp}) I(\mathbf{r}_t, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}) dS_t, \quad (14)$$

$$\begin{aligned} I(\mathbf{r}_t, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}) &= I(\mathbf{r}_b, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}) T(\mathbf{r}_b, \mathbf{r}_t; \boldsymbol{\sigma}_{\text{ext}}) \\ &+ \int_{\mathbf{r}_b}^{\mathbf{r}_t} \sigma_{\text{ext}}(\mathbf{r}) J(\mathbf{r}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}) T(\mathbf{r}, \mathbf{r}_t; \boldsymbol{\sigma}_{\text{ext}}) d\mathbf{s}, \end{aligned} \quad (15)$$

is the residual, $L(\boldsymbol{\sigma}_{\text{ext}})$ a regularization term, α the regularization parameter, \mathcal{I} and \mathcal{I}_{mes} the simulated and measured signals, respectively, and $\boldsymbol{\sigma}_{\text{ext}}^{\dagger}$ the true extinction vector to be retrieved. Note that in Eqs. (14) and (15), the dependency on the extinction vector $\boldsymbol{\sigma}_{\text{ext}}$ is indicated explicitly.

A gradient-based optimization method requires the computation of the derivatives of the objective function with respect to the extinction field, or more precisely and in view of Eq. (15), of the derivatives of the source function J , the surface radiance $I(\mathbf{r}_b, \cdot)$, and the transmission function T . This computational process is complex with respect to time and memory requirements. However, in the framework of the surrogate minimization method, the complex optimization problem can be decomposed into a series of much simpler sub-problems that are easier and faster to optimize.

5 Surrogate minimization method

The surrogate minimization algorithm computes the minimum of an objective function via two steps. The first step is to construct a proper surrogate that bounds the objective function tightly. The second step is to optimize the surrogate whose optimum is much easier to obtain. Given an estimate $\boldsymbol{\sigma}_{\text{ext}l}$ at the l th iteration, the two steps of a typical surrogate minimization algorithm can be formulated as follows [7, 8].

Surrogate Step. Substitute $E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}})$ by a surrogate function $Q_{\alpha}(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l})$, such that

$$E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}}) \leq Q_{\alpha}(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) \quad (16)$$

with equality holding at $\boldsymbol{\sigma}_{\text{ext}} = \boldsymbol{\sigma}_{\text{ext}l}$, i.e., $E_{\alpha}(\boldsymbol{\sigma}_{\text{ext}l}) = Q_{\alpha}(\boldsymbol{\sigma}_{\text{ext}l}|\boldsymbol{\sigma}_{\text{ext}l})$.

Minimization Step. Compute the next parameter estimate $\boldsymbol{\sigma}_{\text{ext},l+1}$ by minimizing the surrogate function $Q_{\alpha}(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l})$ with respect to $\boldsymbol{\sigma}_{\text{ext}}$, i.e.,

$$\boldsymbol{\sigma}_{\text{ext},l+1} = \arg \min_{\boldsymbol{\sigma}_{\text{ext}}} Q_{\alpha}(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}). \quad (17)$$

Note that in the same spirit as the generalized expectation minimization algorithm [15], in a generalized surrogate minimization algorithm, instead of minimizing $Q_\alpha(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l})$, one attempts to find a $\boldsymbol{\sigma}_{\text{ext},l+1}$ such that

$$Q_\alpha(\boldsymbol{\sigma}_{\text{ext},l+1}|\boldsymbol{\sigma}_{\text{ext}l}) \leq Q_\alpha(\boldsymbol{\sigma}_{\text{ext}l}|\boldsymbol{\sigma}_{\text{ext}l}). \quad (18)$$

Since the surrogate constructed at the current estimator is a majorant to the objective function, each minimization step over the surrogate will decrease the objective function monotonically. Indeed, from Eqs. (16) and (17) it is apparent that

$$E_\alpha(\boldsymbol{\sigma}_{\text{ext},l+1}) \leq Q_\alpha(\boldsymbol{\sigma}_{\text{ext},l+1}|\boldsymbol{\sigma}_{\text{ext}l}) \leq Q_\alpha(\boldsymbol{\sigma}_{\text{ext}l}|\boldsymbol{\sigma}_{\text{ext}l}) = E_\alpha(\boldsymbol{\sigma}_{\text{ext}l}). \quad (19)$$

Convergence and convergence rate results for the surrogate minimization algorithm were derived for convex problems. For example, Mairal [16] showed that for strongly convex objective functions, the algorithm converges at a rate of $O(1/\sqrt{l})$ in a finite-sum setting and $O(1/l)$ in a stochastic setting. However, when the objective function is non-convex, there is no guarantee that an exact global minimum is reachable, and theoretical convergence results are hard to be obtained (the limit points of the sequence $\{\boldsymbol{\sigma}_{\text{ext}l}\}$ are only critical points of the objective function). Moreover, the step size can be very small, leading to a slower convergence rate in practice.

In the following we address the two major issues in devising a surrogate minimization algorithm, namely, how a surrogate function is defined and how the resultant surrogate function is minimized.

5.1 Surrogate function

There are mainly three approaches to the construction of surrogate functions, i.e., by using the Jensen's inequality, the first-order Taylor approximation, and the low quadratic bound principle [8]. Unfortunately, these methods cannot be applied in the case of cloud tomographic retrieval. The reasons are that whether the objective function is not convex, or the computation of the gradient of the objective function is a complex process. However, in Ref. [1] it was shown that for cloud tomographic retrieval, a good choice for the surrogate function is

$$Q_\alpha(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) = \hat{R}(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) + \alpha L(\boldsymbol{\sigma}_{\text{ext}}), \quad (20)$$

where

$$\hat{R}(\boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) = \frac{1}{2} \sum_{q=1}^{N_a} \sum_{p=1}^{N_p} [\hat{\mathcal{I}}(\mathbf{r}_{tp}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) - \mathcal{I}_{\text{mes}}(\mathbf{r}_{tp}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}^\dagger)]^2, \quad (21)$$

$$\hat{\mathcal{I}}(\mathbf{r}_{tp}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) = \frac{1}{A} \int_{S_t} h(\mathbf{r}_t - \mathbf{r}_{tp}) \hat{I}(\mathbf{r}_t, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) dS_t, \quad (22)$$

$$\begin{aligned} \hat{I}(\mathbf{r}_t, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l}) &= I(\mathbf{r}_b, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}l}) T(\mathbf{r}_b, \mathbf{r}_t; \boldsymbol{\sigma}_{\text{ext}}) \\ &\quad + \int_{\mathbf{r}_b}^{\mathbf{r}_t} \sigma_{\text{ext}}(\mathbf{r}) J(\mathbf{r}, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}l}) T(\mathbf{r}, \mathbf{r}_t; \boldsymbol{\sigma}_{\text{ext}}) d\mathbf{s}. \end{aligned} \quad (23)$$

From Eq. (23) we deduce that the source function and the surface radiance are kept constant during the minimization step. Thus, in the minimization step, the computation of the derivatives of the surrogate function Q with respect to the extinction field requires only the knowledge of the derivatives of the transmission function T . Because these derivatives can be computed analytically in a very simple manner, the complexity of the problem is substantially reduced.

Unfortunately, for the surrogate function (20), condition (16) is not at all obvious. All what we can say is that the result

$$I(\mathbf{r}_t, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}l}) = \hat{I}(\mathbf{r}_t, \boldsymbol{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}l}|\boldsymbol{\sigma}_{\text{ext}l}), \quad (24)$$

implies that the equality holds at $\boldsymbol{\sigma}_{\text{ext}} = \boldsymbol{\sigma}_{\text{ext}l}$. With regard to the inequality, we can only intuitively expect that, because $\hat{I}(\cdot; \boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l})$ is an approximation of $I(\cdot; \boldsymbol{\sigma}_{\text{ext}})$, the residual corresponding $\hat{\mathcal{I}}(\cdot; \boldsymbol{\sigma}_{\text{ext}}|\boldsymbol{\sigma}_{\text{ext}l})$ is larger than that corresponding to $\mathcal{I}(\cdot; \boldsymbol{\sigma}_{\text{ext}})$. As we do not have a theoretical justification of this result, we adopt a pragmatic strategy; the convergence of the surrogate minimization approach will in fact be a numerical proof of the fulfillment of condition (16).

Algorithm 1 Surrogate minimization algorithm. The input data are the initial guess $\sigma_{\text{ext}0}$, the initial value of the regularization parameter α_0 , the ratio of a geometric sequence of regularization parameters q , and the number of steps (of the outer iteration) $N_{\text{iter}}^{\text{out}}$.

Input: $\sigma_{\text{ext}0}$, (α_0, q) , $N_{\text{iter}}^{\text{out}}$;

Output: σ_{ext} ;

for $l = 0, \dots, N_{\text{iter}}^{\text{out}} - 1$ **do**

- compute $I(\mathbf{r}_{\text{bi}}, \Omega_{\text{mq}}, \sigma_{\text{ext}l})$ and $J_{mn}(\mathbf{r}_i, \sigma_{\text{ext}l})$ by calling SHDOM;
- compute the surrogate function $Q_{\alpha_l}(\sigma_{\text{ext}} | \sigma_{\text{ext}l})$ by using Eqs. (20)–(23);
- compute the new iterate $\sigma_{\text{ext}, l+1} = \arg \min_{\sigma_{\text{ext}}} Q_{\alpha_l}(\sigma_{\text{ext}} | \sigma_{\text{ext}l})$;
- compute the new regularization parameter $\alpha_{l+1} = q\alpha_l$;

if convergence test is satisfied **then**

- $\sigma_{\text{ext}} = \sigma_{\text{ext}, l+1}$;
- exit**;

end if

end for

The standard procedure of the surrogate minimization approach is summarized in Algorithm 1. The following peculiarities of the algorithm can be emphasized.

1. The regularization parameter is not kept constant during the iteration. From the theory of iterative regularization methods, it is known that the amount of regularization should be gradually decreased during the iteration. For this reason, we use a decreasing geometric sequence of regularization parameters $\alpha_{l+1} = q\alpha_l$, where $q < 1$ is the ratio of the sequence.
2. The iteration can be stopped according to (i) the relative residual convergence test: $R(\sigma_{\text{ext}, l}) - R(\sigma_{\text{ext}, l+1}) \leq \varepsilon_{\text{RF}} R(\sigma_{\text{ext}, l})$, or (ii) the absolute residual convergence test: $R(\sigma_{\text{ext}, l+1}) \leq \varepsilon_{\text{AF}} R(\sigma_{\text{ext}0})$, for some prescribed tolerances ε_{RF} and ε_{AF} .

A general rule states that the closer is the surrogate function to the objective function, the more efficient is the surrogate minimization algorithm. In this regard, we may construct a surrogate function in which the single-scattering radiance is changed during the minimization step, or equivalently, the multiple-scattering components of (i) the spherical harmonic coefficients of the source function $J_{mn}(\mathbf{r}_i; \sigma_{\text{ext}})$ and (ii) the grid point radiances at the bottom surface $I(\mathbf{r}_{\text{bi}}, \Omega_{\text{mq}}; \sigma_{\text{ext}})$ are kept constant during the minimization step. This computational step, which is described in Appendix B, is implemented in the post-processing stage of SHDOM and is performed on the final adaptive grid by using the TMS correction in conjunction with the delta-M scaling method. Although both surrogate minimization models involve only the derivatives of the transmission function, the model with a variable single-scattering radiance is more time consuming and memory demanding.

5.2 Minimization step

In principle, in the minimization step we intend to compute a minimum of the surrogate function $Q_{\alpha_l}(\sigma_{\text{ext}} | \sigma_{\text{ext}l})$. However, in the framework of a generalized surrogate minimization algorithm, instead of minimizing $Q_{\alpha_l}(\sigma_{\text{ext}} | \sigma_{\text{ext}l})$, we aim to find a $\sigma_{\text{ext}, l+1}$ such that condition (18) is satisfied. This can be done by applying a finite number steps, say $N_{\text{iter}}^{\text{inn}}$ of a standard optimization algorithm. Thus, we do not require convergence of iterates; a decrease of the surrogate function after $N_{\text{iter}}^{\text{inn}}$ steps is sufficient. The resulting algorithm is similar to a REG-ularization based on INexact Newton (REGINN) iteration method for nonlinear problems [17, 18]. In this method (i) an outer Newton iteration updates the current iterate, and (ii) an inner iteration provides the update by approximately solving a linearized version of the nonlinear equation. Moreover, the finite number of steps of the inner iteration plays the role of a regularization parameter. As a result, and because the amount of regularization should be decreased during the iterative process, the number of steps of the inner iteration is increased during the outer iteration. Adopting this idea, we use the following linear selection rule:

$$N_{\text{iter}l}^{\text{inn}} = \left[N_{\text{iter}0}^{\text{inn}} + l \frac{N_{\text{iter}1}^{\text{inn}} - N_{\text{iter}0}^{\text{inn}}}{N_{\text{iter}}^{\text{out}}} \right], \quad (25)$$

Algorithm 2 Projected gradient method. The input data are the initial guess \mathbf{x}_0 and the number of steps of the inner iteration K .

Input: \mathbf{x}_0, K ;

Output: \mathbf{x}_K ;

for $k = 0, \dots, K - 1$ **do**

- compute $\nabla f(\mathbf{x}_k)$;
- compute the step length $\tau_k = \arg \min_{\tau} f(\mathcal{P}_B(\mathbf{x}_k - \tau \nabla f(\mathbf{x}_k)))$ by a line search algorithm;
- $\mathbf{x}_{k+1} = \mathcal{P}_B(\mathbf{x}_k - \tau_k \nabla f(\mathbf{x}_k))$;

end for

where $N_{\text{iter}0}^{\text{inn}}$ and $N_{\text{iter}1}^{\text{inn}}$ with $N_{\text{iter}0}^{\text{inn}} < N_{\text{iter}1}^{\text{inn}}$ determine the interval of variation of $N_{\text{iter}l}^{\text{inn}}$, and $[x]$ denotes the integer part of x .

In the following, we simplify the notation by making the changes $\boldsymbol{\sigma}_{\text{ext}} \rightarrow \mathbf{x}$, $Q_{\alpha_l}(\boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l}) \rightarrow f(\mathbf{x})$, and $N_{\text{iter}l}^{\text{inn}} \rightarrow K$. Because the extinction is a positive quantity, we consider the box solution domain

$$B(\mathbf{x}) = \{\mathbf{x} \mid \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}, \quad (26)$$

where \mathbf{l} and \mathbf{u} are lower and upper bound vectors, respectively, and the vector inequality $\mathbf{x} \leq \mathbf{y}$ should be understood componentwise.

The simplest first-order optimization algorithm (that requires only gradient calculations) is the projected gradient method illustrated in Algorithm 2. Some peculiarities of the algorithm can be summarized as follows.

1. $\mathcal{P}_B(\mathbf{x})$ is the projection operator onto the box B that thresholds the values of \mathbf{x} at the boundaries of the box. It is given by

$$(\mathcal{P}_B(\mathbf{x}))_i = \begin{cases} l_i & x_i < l_i \\ x_i & l_i \leq x_i \leq u_i \\ u_i & x_i > u_i \end{cases}, \quad (27)$$

where x_i is the i th component of the vector \mathbf{x} .

2. In the standard gradient method, a step length τ_k , which guarantees a sufficient decrease in $f(\mathbf{x}_k - \tau \nabla f(\mathbf{x}_k))$ over $\tau \in \mathbb{R}_+$, is computed by means of a line search algorithm. Specifically, in a backtracking line search algorithm based on Wolfe conditions [6], the step length τ is reduced starting from an initial value τ_0 , until (i) the average rate of decrease from $f(\mathbf{x}_k)$ to $f(\mathbf{x}_k - \tau \nabla f(\mathbf{x}_k))$ is at least some prescribed fraction ε_1 of the initial rate of decrease in that direction, and (ii) the rate of decreases of f in the direction $-\nabla f(\mathbf{x}_k)$ at $\mathbf{x}_k - \tau \nabla f(\mathbf{x}_k)$ is larger than some prescribed fraction ε_2 of the rate of decrease in the direction $-\nabla f(\mathbf{x}_k)$ at \mathbf{x}_k . These two conditions, known as the Armijo–Goldstein condition and the curvature condition, respectively, provide an upper and lower bound on the admissible step length values. In our algorithm, we use only the Armijo–Goldstein condition, which in the case of the projected gradient method is

$$f(\mathbf{z}_{k\tau}) \leq f(\mathbf{x}_k) - \varepsilon_1 \nabla f(\mathbf{x}_k)^T (\mathbf{x}_k - \mathbf{z}_{k\tau}), \quad (28)$$

$$\mathbf{z}_{k\tau} = \mathcal{P}_B(\mathbf{x}_k - \tau \nabla f(\mathbf{x}_k)). \quad (29)$$

In order to increase the convergence rate of the projected gradient method several acceleration methods have been designed. Algorithms 3 and 4 [19, 20, 21] represent variants of the Nesterov acceleration method [22, 23] and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [24] for non-convex problems, respectively. As compared to the original versions for convex problems, the modified methods accept the new iterate when the corresponding function value is sufficiently decreased, and this fact leads to a more stable convergence behavior. Some comments can be made here.

1. Algorithms 3 and 4 achieve the $O(1/k^2)$ convergence rate for convex problems and converge to a critical point of f at linear rates.
2. Algorithm 3 requires one proximal step, while Algorithm 4 requires two; thus, the Nesterov acceleration method computes faster than FISTA.

Algorithm 3 Nesterov acceleration method for non-convex problems. The input data are the initial guess \mathbf{x}_0 , the initial step length τ_0 , the integer parameter β of the Nesterov method, e.g., $\beta = 3$, and the number of steps of the inner iteration K .

Input: $\mathbf{x}_0, \tau_0, \beta, K$;

Output: \mathbf{x}_K ;

• initialize $\mathbf{y}_0 = \mathbf{x}_0$;

for $k = 0, \dots, K - 1$ **do**

• compute $\nabla f(\mathbf{y}_k)$;

• compute the step length $\tau_k = \arg \min_{\tau} f(\mathcal{P}_B(\mathbf{y}_k - \tau \nabla f(\mathbf{y}_k)))$
by a line search algorithm or set $\tau_k = \tau_0$;

• $\mathbf{x}_{k+1} = \mathcal{P}_B(\mathbf{y}_k - \tau_k \nabla f(\mathbf{y}_k))$;

• $\mathbf{v}_{k+1} = \mathbf{x}_{k+1} + \frac{k}{k+\beta}(\mathbf{x}_{k+1} - \mathbf{x}_k)$;

• compute $f(\mathbf{x}_{k+1})$ and $f(\mathbf{v}_{k+1})$;

if ($f(\mathbf{x}_{k+1}) \leq f(\mathbf{v}_{k+1})$) **then**

• $\mathbf{y}_{k+1} = \mathbf{x}_{k+1}$;

else

• $\mathbf{y}_{k+1} = \mathbf{v}_{k+1}$;

end if

end for

3. Because Algorithms 3 and 4 guarantee a monotone decrease of the objective function, the computation of the step length by a line search algorithm is not absolutely required. However, our numerical analysis indicates that such an algorithm improves the convergence rate.

The fourth algorithm that is implemented in our retrieval tool is the BFGS algorithm that uses the gradient projection method to determine a set of active constraints at each iteration, a limited memory BFGS matrix to approximate the Hessian of the objective function, and a line search procedure relying on Wolfe conditions to compute the step length.

6 Regularization

The last problem that we address is the choice of the regularization term $L(\sigma_{\text{ext}})$. In order to simplify our presentation we consider a two-dimensional geometry. In cloud tomography, the role of regularization is to enforce the smoothness of the spatial variability of the extinction field. There are two reasons for which the use of smooth extinction fields is advantageous. The first one is that in SHDOM, a strongly varying extinction field may cause a significant increase of the number of grid points. The second one is that for non-smooth extinction fields it is more probable that the retrieval algorithm get stuck in a local minimum. The amount of smoothness is controlled by the regularization parameter α ; a small value of α allows for more roughness in the cloud extinction field, while a large value of α ensures more smoothness. Although smoothing regularization terms can be constructed by using the L_2 norm of the gradient vector, we decided to use spatial filtering techniques from image processing [12, 13, 14]. Spatial filters fall into two category: linear and nonlinear.

1. A linear filter replaces the value of the extinction field at the grid point (i, j) , $\sigma_{\text{ext}}(i, j)$ by a weighted sum of its neighbors. The matrix defining the weight associated to each neighbor is called convolution kernel, and is of the form

$$\mathbf{W} = \begin{bmatrix} w(i-1, j+1) & w(i, j+1) & w(i+1, j+1) \\ w(i-1, j) & w(i, j) & w(i+1, j) \\ w(i-1, j-1) & w(i, j-1) & w(i+1, j-1) \end{bmatrix}. \quad (30)$$

The filtered pixel value $\sigma_{\text{ext}}^{\text{filter}}(i, j)$ is given by

$$\sigma_{\text{ext}}^{\text{filter}}(i, j) = \frac{1}{C} \sum_{i_1=-1}^1 \sum_{j_1=-1}^1 w(i+i_1, j+j_1) \sigma_{\text{ext}}(i+i_1, j+j_1), \quad (31)$$

Algorithm 4 FISTA for non-convex problems. The input data are the initial guess \mathbf{x}_0 , the initial step length τ_0 , and the number of steps of the inner iteration K .

Input: \mathbf{x}_0, τ_0, K ;

Output: \mathbf{x}_K ;

• initialize $\mathbf{y}_0 = \mathbf{x}_0, \theta_0 = 1$;

for $k = 0, \dots, K - 1$ **do**

• compute $\nabla f(\mathbf{y}_k)$ and $\nabla f(\mathbf{x}_k)$;

• compute the step lengths $\tau_k = \arg \min_{\tau} f(\mathcal{P}_B(\mathbf{y}_k - \tau \nabla f(\mathbf{y}_k)))$ and $\eta_k = \arg \min_{\tau} f(\mathcal{P}_B(\mathbf{x}_k - \tau \nabla f(\mathbf{x}_k)))$ by a line search algorithm or set

$\tau_k = \eta_k = \tau_0$;

• $\mathbf{z}_{k+1} = \mathcal{P}_B(\mathbf{y}_k - \tau_k \nabla f(\mathbf{y}_k))$;

• $\mathbf{v}_{k+1} = \mathcal{P}_B(\mathbf{x}_k - \eta_k \nabla f(\mathbf{x}_k))$;

• compute $f(\mathbf{z}_{k+1})$ and $f(\mathbf{v}_{k+1})$;

if ($f(\mathbf{z}_{k+1}) \leq f(\mathbf{v}_{k+1})$) **then**

• $\mathbf{x}_{k+1} = \mathbf{z}_{k+1}$;

else

• $\mathbf{x}_{k+1} = \mathbf{v}_{k+1}$;

end if

• $\theta_{k+1} = \frac{1}{2}(1 + \sqrt{1 + 4\theta_k^2})$;

• $\mathbf{y}_{k+1} = \mathbf{x}_{k+1} + \frac{\theta_k - 1}{\theta_{k+1}}(\mathbf{x}_{k+1} - \mathbf{x}_k) + \frac{\theta_k}{\theta_{k+1}}(\mathbf{z}_{k+1} - \mathbf{x}_{k+1})$;

end for

where C is a normalization factor equal to $\sum_{i_1=-1}^1 \sum_{j_1=1}^{-1} w(i + i_1, j + j_1)$. From the category of linear filters, the averaging and Gaussian filters with the convolution kernels

$$\mathbf{W}_{\text{avr}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \mathbf{W}_{\text{Gauss}} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad (32)$$

respectively, are low-pass filters that smooths the extinction field.

2. A nonlinear filter assigns to $\sigma_{\text{ext}}(i, j)$ a value that is not a linear combination of the surrounding values. The median filter is a nonlinear, low-pass filter with a smoothing effect. For a 3×3 region around $\sigma_{\text{ext}}(i, j)$, the median pixel value $\sigma_{\text{ext}}^{\text{filter}}(i, j)$ is obtained as follows: (i) set $\hat{\sigma}_{\text{ext}}(k) = \sigma_{\text{ext}}(i + i_1, j + j_1)$, where $k = j_1 + 2 + 3(i_1 + 1)$ for $i_1, j_1 = -1, \dots, 1$, (ii) sort the set $\{\hat{\sigma}_{\text{ext}}(k)\}_{k=1}^9$ in ascending order, and finally, (iii) set $\sigma_{\text{ext}}^{\text{filter}}(i, j) = \hat{\sigma}_{\text{ext}}(5)$ (note that $5 = [9/2] + 1$ is the median index of the sequence).

Parenthetically we note that low-pass filters attenuate variations of the extinction field and have the tendency to eliminate details and blur the cloud edges. Using spatial filters, we construct the regularization term as

$$L(\boldsymbol{\sigma}_{\text{ext}}) = \sum_i \sum_j [\sigma_{\text{ext}}(i, j) - \sigma_{\text{ext}}^{\text{filter}}(i, j)]^2 \quad (33)$$

where the summation over i and j is performed over all permissible values of $\sigma_{\text{ext}}(i + i_1, j + j_1)$.

7 Numerical Simulations

In this section we analyze the performances of the retrieval algorithm on a few synthetic two-dimensional problems. The true cloud extinction field is chosen as $\sigma_{\text{ext}}^{\text{cld}\dagger}(x, z) = \sigma_{\text{max}}\chi(x, z)$, where $\chi(x, z)$ with $0 \leq \chi(x, z) \leq 1$ is the normalized extinction field of the cloud. Because at each outer iteration, the adaptive grid evolves from a base grid to a final grid and the values of the source functions are specified on the final grid, the base grid point values of the extinction field are the unknowns of the inverse problem. At each step of the outer iteration, the values of the extinction field on the final grid are interpolated from the base grid values. Note that the same interpolation rule is used when splitting cells during the adaptive grid procedure. The following cloud models are considered in our numerical analysis.

1. Two circular clouds with the extinction field [5]

$$\sigma_{\text{ext}}^{\text{cld}\dagger}(x, z) = \sigma_{\text{max}}[\chi_1(x, z) + \chi_2(x, z)], \quad (34)$$

$$\chi_i(x, z; x_0, z_0, R_c) = \cos\left[\frac{\pi}{2R_c}\sqrt{(x - x_{0i})^2 + (z - z_0)^2}\right]^{1.25}, \quad i = 1, 2, \quad (35)$$

where $\sigma_{\text{max}} = 6 \text{ km}^{-1}$, $x_{01} = 2 \text{ km}$, $x_{02} = 4 \text{ km}$, $z_0 = 1 \text{ km}$, and $R_c = 0.6 \text{ km}$. Thus, the maximum optical thickness is $\tau_{\text{max}} = 7.2$. The lengths of the domain of analysis are $L_x = 6 \text{ km}$ and $L_z = 2 \text{ km}$, the numbers of base grid points are $N_x = 31$ and $N_z = 11$, and the base grid spacings are $\Delta x = \Delta z = 200 \text{ m}$. Thus, the number of unknowns is $N_{\text{pts}} = 341$.

2. A box cloud with the extinction field

$$\sigma_{\text{ext}}^{\text{cld}\dagger}(x, z) = \sigma_{\text{max}}\chi(x, z), \quad (36)$$

$$\chi(x, z) = \begin{cases} 1 & x_1 \leq x \leq x_2 \text{ and } z_1 \leq z \leq z_2 \\ 0 & \text{rest} \end{cases} \quad (37)$$

where $\sigma_{\text{max}} = 3 \text{ km}^{-1}$, $x_1 = 1.2 \text{ km}$, $x_2 = 4.4 \text{ km}$, $z_1 = 0.4 \text{ km}$, $z_2 = 1.6 \text{ km}$, and $\tau_{\text{max}} = 3.6$. The cloud extinction field is smoothed at the boundaries in order to avoid discontinuities between the cloudy and clear regions. The lengths of the domain of analysis, the numbers of base grid points, and the base grid spacings are as in the case of two circular clouds.

3. Two two-dimensional cloud fields obtained from the cloud liquid water content output from a large eddy simulation (LES) of stratocumulus [2]. The two-dimensional slices are selected from the three-dimensional cloud liquid water content. The native extinction field is scaled and interpolated on a discrete domain with (i) lengths $L_x = 6 \text{ km}$ and $L_z = 1 \text{ km}$, (ii) numbers of base grid points $N_x = 61$ and $N_z = 11$, and (iii) base grid spacings $\Delta x = \Delta z = 100 \text{ m}$. The interpolated values are normalized to 1 to obtain the normalized extinction field $\chi(x, z)$, and the cloud extinction fields are chosen as $\sigma_{\text{ext}}^{\text{cld}\dagger}(x, z) = \sigma_{\text{max}}\chi(x, z)$ with $\sigma_{\text{max}} = 10 \text{ km}^{-1}$ (LES1) and $\sigma_{\text{max}} = 12 \text{ km}^{-1}$ (LES2). The variation of the optical thickness τ with respect to the horizontal coordinate x is illustrated in Fig. 1 and the number of unknowns is $N_{\text{pts}} = 671$.

Other scene parameters are listed below.

1. 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 distribution

$$P(a) \propto a^\alpha \exp\left[-\alpha\left(\frac{a}{a_{\text{mod}}}\right)\right] \quad (38)$$

of parameters $a_{\text{eff}} = 10 \text{ }\mu\text{m}$, $a_{\text{mod}} = 2a_{\text{eff}}/3$, and $\alpha = 6$. Here, a is the particle radius, and the droplet size ranges between 0.02 and 50.0 μm . In addition to the cloud, molecular Rayleigh scattering is considered as background.

2. The solar zenith angle is $\theta_0 = 0^\circ$, and a Lambertian reflecting surface with the surface albedo $A_s = 0.05$ is chosen.
3. There are $N_a = 9$ viewing zenith angles corresponding to the Multiangle Imaging SpectroRadiometer (MISR). These are given by $\pm 70.5^\circ$, $\pm 60^\circ$, $\pm 45.6^\circ$, $\pm 26.1^\circ$, and 0° , where \pm indicates the forward- and aftward-facing cameras. Note that the domain size was chosen sufficiently large to avoid artifacts caused by periodic boundary conditions, even for the most oblique MISR viewing angles ($\pm 70.5^\circ$).
4. The characteristic function of the p th detector pixel is the box function

$$h(x - x_p) = \begin{cases} 1 & |x - x_p| \leq \Delta x/2 \\ 0 & \text{rest} \end{cases} \quad (39)$$

5. 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 $M = N_\varphi/2 - 1$.

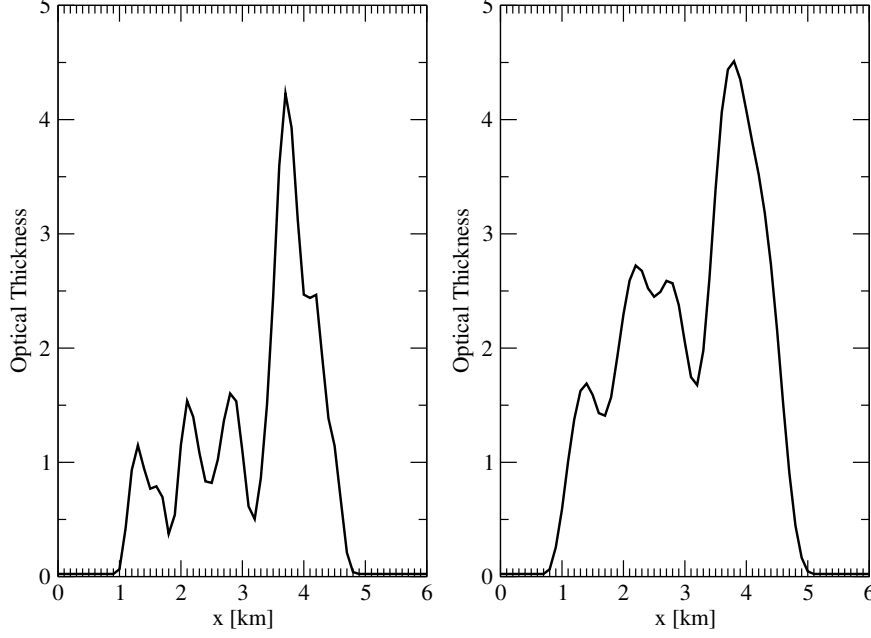


Figure 1: Variation of the optical thickness τ with respect to the horizontal coordinate x for the LES1 cloud (left) and the LES2 cloud (right).

6. The delta-M scaling method, the TMS correction, and an adaptive grid with a splitting accuracy of 10^{-3} are used.

Some input parameters of the retrieval algorithm are chosen as follows.

1. The initial extinction field $\sigma_{\text{ext}0}$ correspond to an atmosphere with no cloud.
2. The components of the lower and upper bound vectors that constrain the extinction field are $l_i = 10^{-6} \text{ km}^{-1}$ and $u_i = 10^3 \text{ km}^{-1}$ for all $i = 1, \dots, N_{\text{pts}}$.
3. Unless otherwise specified, the standard surrogate function is that in which the source function and the surface radiance are kept constant during the minimization step (cf. Eqs. (20)–(23)).
4. The number of steps of the outer iteration is $N_{\text{iter}}^{\text{out}} = 300$, and if not stated otherwise, the number of steps of the inner iteration varies between 5 and 7, that is, $N_{\text{iter}0}^{\text{inn}} = 5$ and $N_{\text{iter}1}^{\text{inn}} = 7$. The algorithm stops when the number of steps of the outer iteration exceeds $N_{\text{iter}}^{\text{out}} = 300$.
5. The regularization terms 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 $q = 0.8$.
6. The initial step in the line search algorithm is $\tau_0 = 250$.

The simulations were performed on a computer Intel(R) Core(TM) i5-3340M CPU @ 2.70GHz with 7858Mb RAM.

In Figs. 2–5, we show the true extinction field $\sigma_{\text{ext}}^{\dagger}(\mathbf{r}_i)$, the retrieved extinction field $\sigma_{\text{ext}}(\mathbf{r}_i)$, and the absolute error in extinction field at each base grid point

$$\varepsilon_{\text{ext}}(\mathbf{r}_i) = |\sigma_{\text{ext}}(\mathbf{r}_i) - \sigma_{\text{ext}}^{\dagger}(\mathbf{r}_i)| \quad (40)$$

for all test examples. From these plots, the same conclusions as in Refs. [1, 5] can be drawn: the errors are larger in opaque regions within the clouds, and the retrieval algorithm has the tendency to blur the boundary of the cloud (especially, in the case of the LES2 cloud).

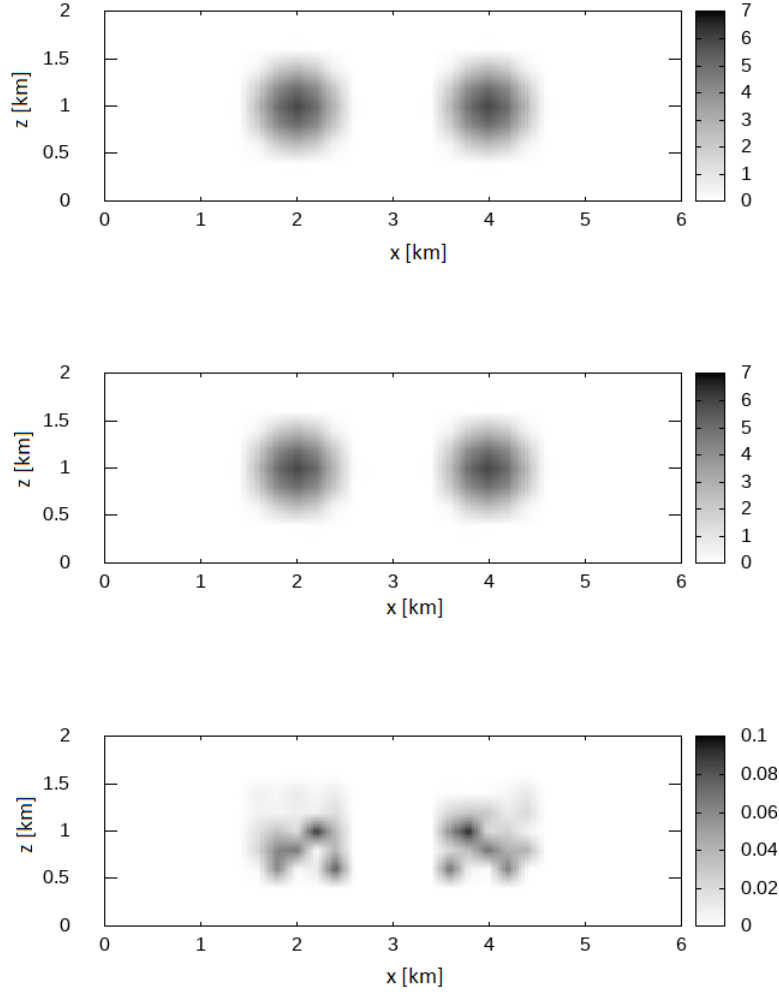


Figure 2: True extinction field $\sigma_{\text{ext}}^{\dagger}(\mathbf{r}_i)$ (top), retrieved extinction field $\sigma_{\text{ext}}(\mathbf{r}_i)$ (middle), and absolute error in extinction field at each base grid point $\epsilon_{\text{ext}}(\mathbf{r}_i) = |\sigma_{\text{ext}}(\mathbf{r}_i) - \sigma_{\text{ext}}^{\dagger}(\mathbf{r}_i)|$ (bottom) for two circular clouds. The results are computed with the Nesterov acceleration method.

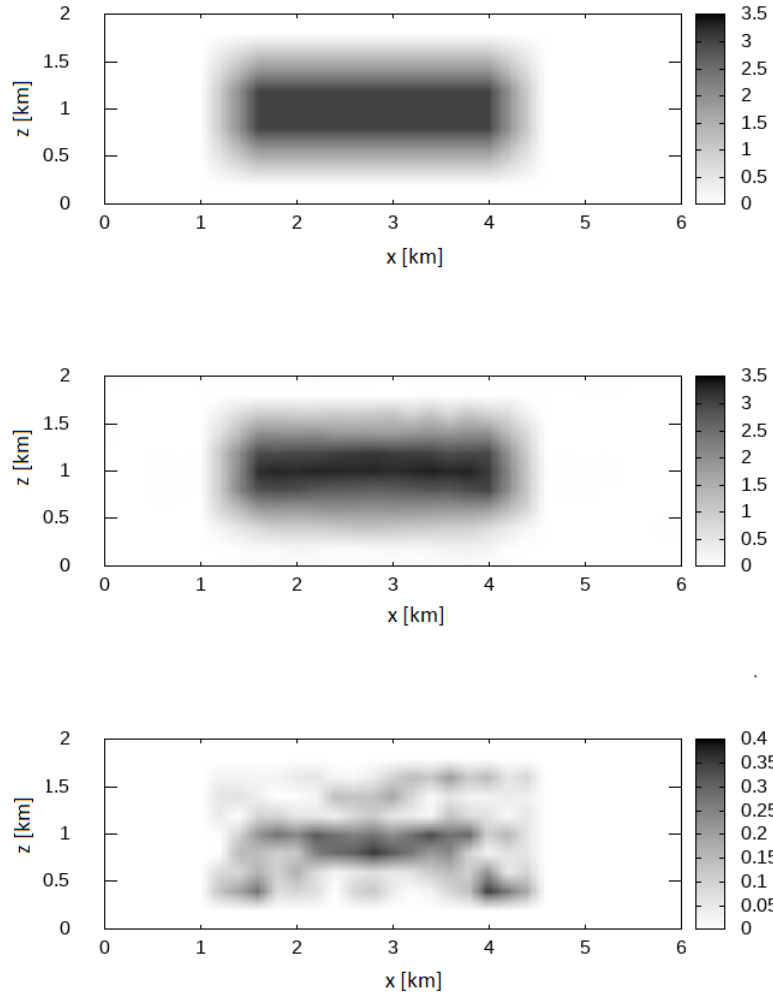


Figure 3: The same as in Fig. 2 but for a box cloud. The results are computed with the Nesterov acceleration method.

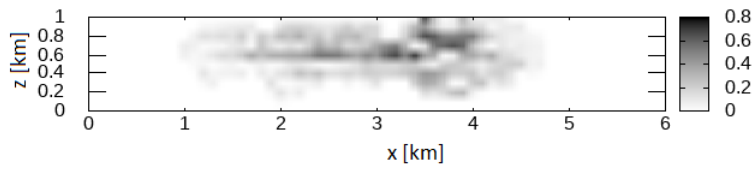
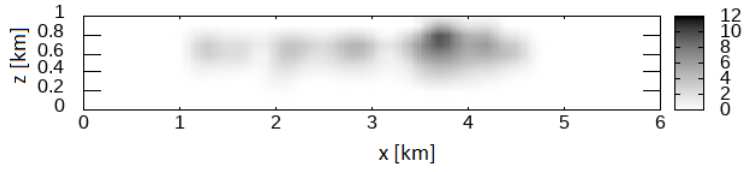
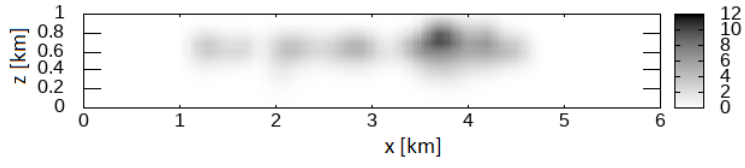


Figure 4: The same as in Fig. 2 but for the LES1 cloud. The results are computed with the BFGS method.

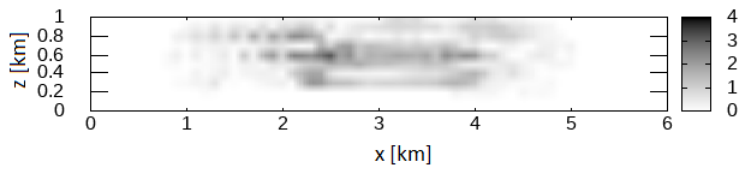
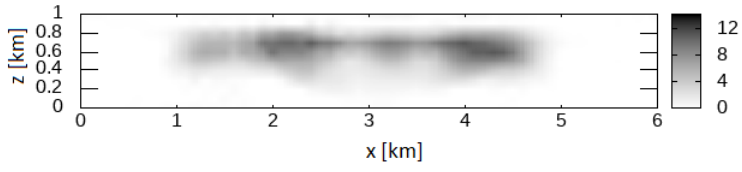
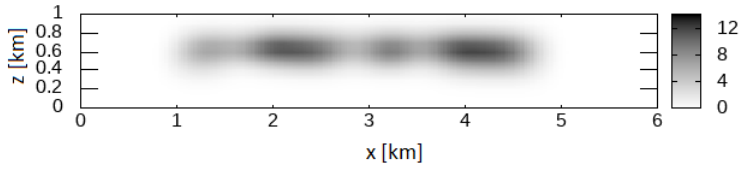


Figure 5: The same as in Fig. 2 but for the LES2 cloud. The results are computed with the BFGS method.

Cloud Model	Optimization Method	Residual Decrease Ratio	Relative Error in Extinction	Comput. Time (h:min)
Circular Clouds	Proj. Grad.	4.27×10^{-7}	1.88×10^{-2}	0:36
	Nesterov	1.62×10^{-7}	1.23×10^{-2}	0:39
	FISTA	2.18×10^{-7}	1.34×10^{-2}	0:46
	BFGS	6.85×10^{-7}	3.66×10^{-2}	0:33
Box Cloud	Proj. Grad.	1.51×10^{-6}	8.43×10^{-2}	0:39
	Nesterov	8.56×10^{-7}	7.78×10^{-2}	0:41
	FISTA	9.51×10^{-7}	7.90×10^{-2}	0:48
	BFGS	2.31×10^{-6}	9.68×10^{-2}	0:35
LES1 Cloud	Proj. Grad.	7.45×10^{-6}	10.03×10^{-2}	1:15
	Nesterov	3.08×10^{-6}	8.76×10^{-2}	1:23
	FISTA	3.78×10^{-6}	9.06×10^{-2}	1:40
	BFGS	1.72×10^{-6}	7.64×10^{-2}	1:20
LES2 Cloud	Proj. Grad.	1.20×10^{-4}	25.89×10^{-2}	1:23
	Nesterov	7.75×10^{-5}	23.79×10^{-2}	1:34
	FISTA	8.74×10^{-5}	24.34×10^{-2}	1:53
	BFGS	1.92×10^{-5}	20.41×10^{-2}	1:29

Table 1: Residual decrease ratio, relative error in extinction field, and computational time for the considered cloud models and optimization methods.

In Fig. 6 we plot the variation of the (total) relative error in the extinction field

$$\epsilon_{\text{ext}} = \frac{\|\sigma_{\text{ext}} - \sigma_{\text{ext}}^{\dagger}\|_2}{\|\sigma_{\text{ext}}^{\dagger}\|_2} \quad (41)$$

with respect to the outer iteration index, while in Table 1 we illustrate the residual decrease ratio $R(\sigma_{\text{ext}})/R(\sigma_{\text{ext}0})$, the relative error in the extinction field ϵ_{ext} , and the computational time for the considered cloud models and optimization methods. Excepting the BFGS results, which require special attention, the following conclusions are apparent.

1. The relative error decreases very fast at the beginning of the iterative process and very slowly after that (the step size becomes very small). The Nesterov and FISTA acceleration methods have a better convergence rate than the projected gradient method. In this regard it should be pointed that for a LES2 cloud, the projected gradient method attains a relative error of 24.85×10^{-2} after 400 iterations and of 24.03×10^{-2} after 500 iterations (in this case, the computational time is 2h : 38min).
2. The Nesterov acceleration method is the most accurate, and faster than FISTA.
3. The behavior of the residual decrease ratio reproduces that of the relative error; smaller values of the residual decrease ratio correspond to smaller values of the relative error. In this regard, we may conclude that a good stopping rule of the outer iteration is the absolute residual convergence test, i.e., $R(\sigma_{\text{ext},l+1}) \leq \varepsilon_{\text{AF}} R(\sigma_{\text{ext}0})$.

In the minimization step, the BFGS method yields the largest decrease of the objective function. However, this fact does not have a positive effect on the stability of the surrogate minimization algorithm. To obtain a stable algorithm, the amount of regularization should be increased, or equivalently, the number of steps of the inner loop should be decreased. From Fig. 7, we see that for the LES1 cloud, both the residual $R(\sigma_{\text{ext}})$ and the relative error in extinction field ϵ_{ext} decrease monotonically during the iterative process when the number of steps of the inner iteration varies between $N_{\text{iter}0}^{\text{inn}} = 2$ and $N_{\text{iter}1}^{\text{inn}} = 4$. Note that for the LES clouds, the results in Fig. 6 and Table 1 correspond to $N_{\text{iter}0}^{\text{inn}} = 3$ and $N_{\text{iter}1}^{\text{inn}} = 5$ (in this case, although $R(\sigma_{\text{ext}})$ and ϵ_{ext} do not decrease monotonically, the relative error is the smallest one).

The algorithm has several features that increase its performances.

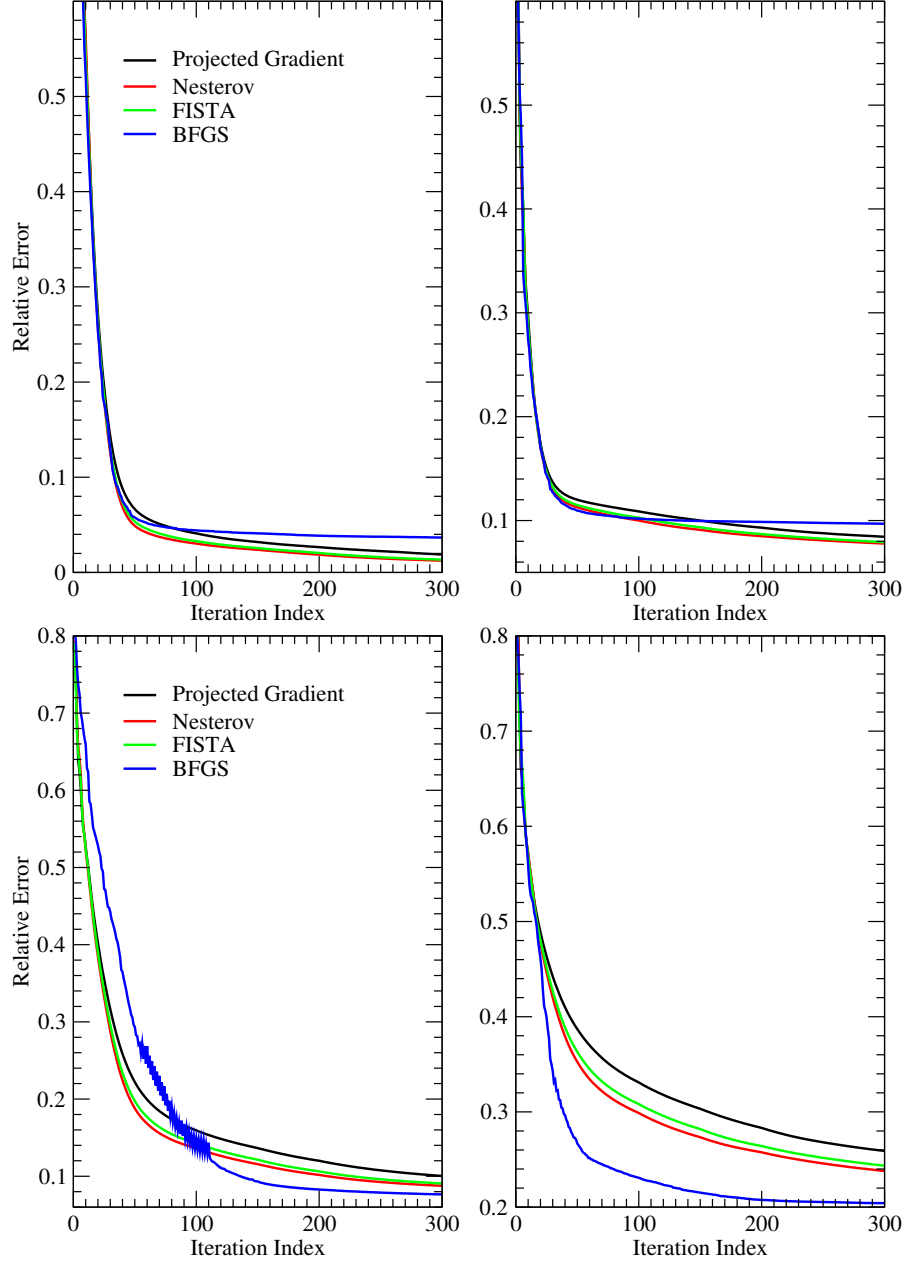


Figure 6: Relative error in extinction field versus the outer iteration index for two circular clouds (top left), box cloud (top right), LES1 cloud (bottom left), and LES2 cloud (bottom right).

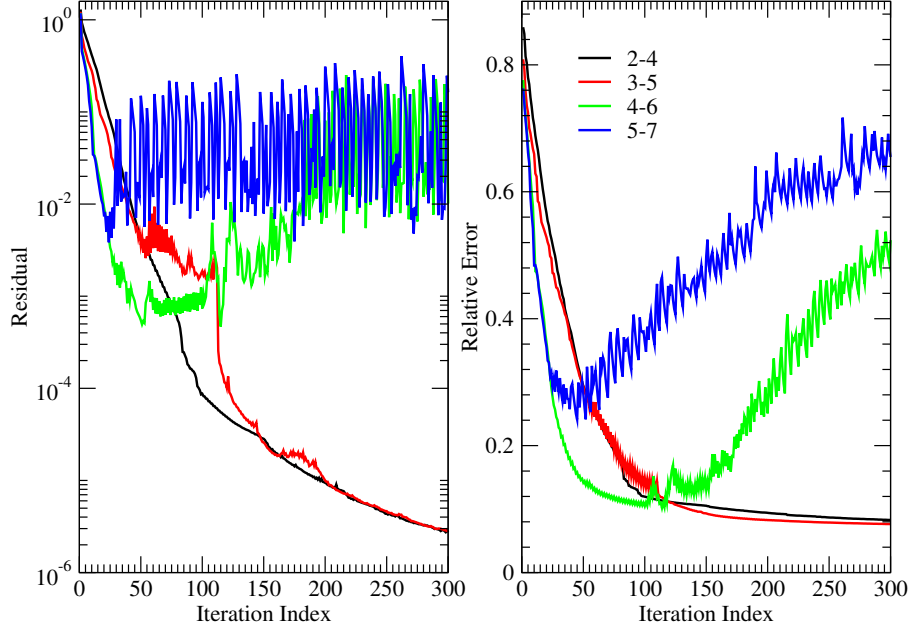


Figure 7: Residual (left) and relative error in extinction field (right) versus the outer iteration index for the LES1 cloud. The results are computed with the BFGS method for the following intervals of variation of the number of steps of the inner iteration: $(N_{\text{iter0}}^{\text{inn}} = 2, N_{\text{iter1}}^{\text{inn}} = 4)$, $(N_{\text{iter0}}^{\text{inn}} = 3, N_{\text{iter1}}^{\text{inn}} = 5)$, $(N_{\text{iter0}}^{\text{inn}} = 4, N_{\text{iter1}}^{\text{inn}} = 6)$, and $(N_{\text{iter0}}^{\text{inn}} = 5, N_{\text{iter1}}^{\text{inn}} = 7)$.

1. In the projected gradient method, the step length algorithm avoids guessing an optimal value for the step length τ . This is illustrated in Fig. 8 for two circular clouds. When τ is fixed, its optimal value, for which a relative error of 3.12×10^{-2} is obtained, is 200; for $\tau = 50$ and $\tau = 250$, the relative errors are much higher. In contrast, the step length algorithm directly yields a relative error of 1.88×10^{-2} .
2. The regularization improves significantly the convergence rate of the algorithm. This can be seen in Fig. 9 for a box cloud. The relative errors computed without regularization by using the projected gradient method, Nesterov acceleration, and FISTA are 14.61×10^{-2} , 15.43×10^{-2} , and 12.33×10^{-2} , respectively. When regularization is used, the relative errors decrease by a factor between 1.5 and 2; they are 8.43×10^{-2} , 7.78×10^{-2} , and 7.90×10^{-2} , respectively (see Table 1). In fact, the benefit of using regularization can be explained by a numerical experiment, in which a box cloud with the non-smooth extinction field $\sigma_{\text{ext}}^{\text{cld}}(x, z) = \sigma_{\text{max}} \chi_{\varepsilon}(x, z)$, where $\chi_{\varepsilon}(x, z) = (1 + \sigma_f \varepsilon) \chi(x, z)$, ε is a random number uniformly distributed in the interval $(-1, 1)$, and $\chi(x, z)$ is given by Eq. (37), is considered. In Fig. 10, we plot the residual $R(\sigma_{\text{ext}})$ versus the relative distance $d = \|\sigma_{\text{ext}} - \sigma_{\text{ext}}^{\dagger}\|_2 / \|\sigma_{\text{ext}}^{\dagger}\|_2 (= \epsilon_{\text{ext}})$ for different values of the standard deviation σ_f and by considering 1000 configurations for each σ_f . As expected, we see that $R(\sigma_{\text{ext}}) \rightarrow 0$ as $\sigma_f \rightarrow 0$. Moreover, we see that (i) for a given value of the standard deviation 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 situated at different distances with respect to the true solution (in other words, in the neighborhood of the solution, the residual function is not injective). Thus, the use of non-smooth extinction fields with uniform small-scale distributions may increase the risk that the retrieval algorithm get stuck in a local minimum, or that it recede an optimal point and starts to oscillate.
3. The use of a surrogate function, in which the single-scattering radiance is changed during the minimization step, slightly improves the convergence rate. In the case of the LES2 cloud (see Fig. 11), the relative errors corresponding to the surrogate function in which only the transmission in Eq. (23) is changed during the minimization step, are 25.89×10^{-2} for the projected gradient method, 23.79×10^{-2} for Nesterov acceleration, and 24.34×10^{-2} for FISTA. For the surrogate function, in which the source function and the surface radiance are kept constant during the minimization step, the relative errors

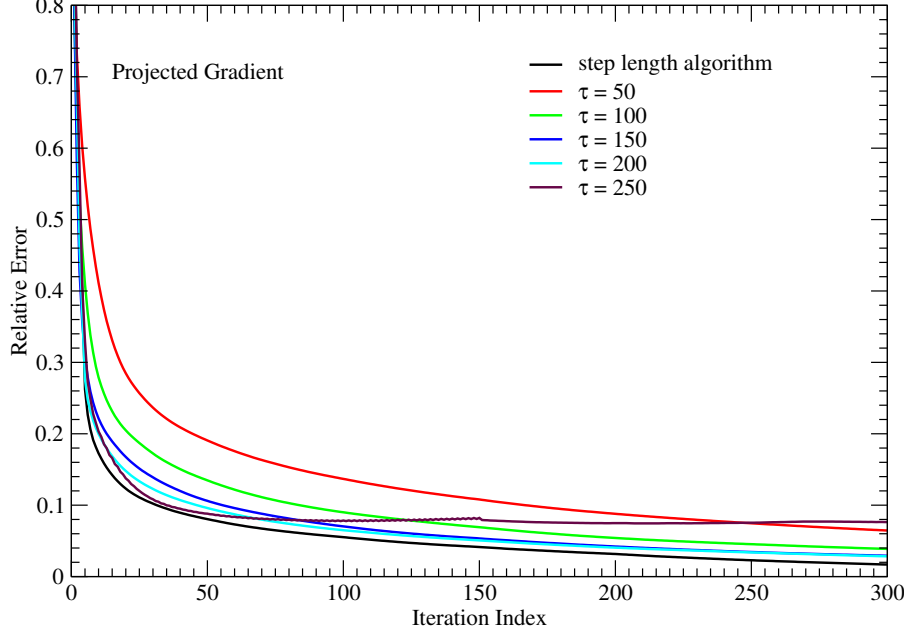


Figure 8: Relative error in extinction field versus the outer iteration index for two circular clouds. The results are computed with the projected gradient method and correspond to a step length algorithm for computing τ and a fixed values of τ .

are 23.63×10^{-2} , 21.58×10^{-2} , and 22.46×10^{-2} , respectively; thus, a reduction of about 2% is attained. However, in this case, the computational times increase by about 20%, and it is a decision question if the modified surrogate function should be used or not.

We conclude our numerical analysis by presenting the retrieval results for the three-dimensional cloud extinction field $\sigma_{\text{ext}}^{\text{cld}\dagger}(x, y, z) = \sigma_{\text{max}}\chi(x, y, z)$ with $\sigma_{\text{max}} = 8 \text{ km}^{-1}$, obtained from a large eddy simulation of stratocumulus [2] and from which the two-dimensional slices LES1 and LES2 were considered. The discrete domain of analysis has (i) the lengths $L_x = L_y = 6 \text{ km}$ and $L_z = 2 \text{ km}$, (ii) the numbers of base grid points $N_x = N_y = 31$ and $N_z = 11$, and (iii) the base grid spacings $\Delta x = \Delta y = \Delta z = 200 \text{ m}$. The number of unknowns is $N_{\text{pts}} = 10571$. The true and retrieved extinction fields computed with the Nesterov acceleration method are illustrated in Figs. 12 and 13. For these results, the residual decrease ratio is 2.10×10^{-4} , the relative error in the extinction field is 23.08×10^{-2} , and the computational time is 7h16min. In conclusion, the retrieval accuracy is acceptable and comparable to that obtained for the two-dimensional slice LES2, but the computational time is on average 5 times higher.

8 Conclusions

A retrieval algorithm for cloud tomography has been designed. For this purpose, we used our own implementation of SHDOM to compute the radiative transfer and the surrogate minimization method to solve the inverse problem. The following features of the algorithm deserve to be mentioned.

1. The objective function to be minimized includes, in addition to the residual, a regularization term. This is constructed by using spatial filtering techniques from image processing, and in particular, averaging, Gaussian, and median low-pass filters. The regularization parameter which controls the amount of smoothness is gradually decreased during the iteration.
2. The iteration is stopped according to the absolute residual convergence test, i.e., when the residual declines a specified fraction of its initial value.
3. The algorithm includes a surrogate step, in which the original objective function is substituted by a tractable surrogate function, and a minimization step, in which the surrogate function is minimized.

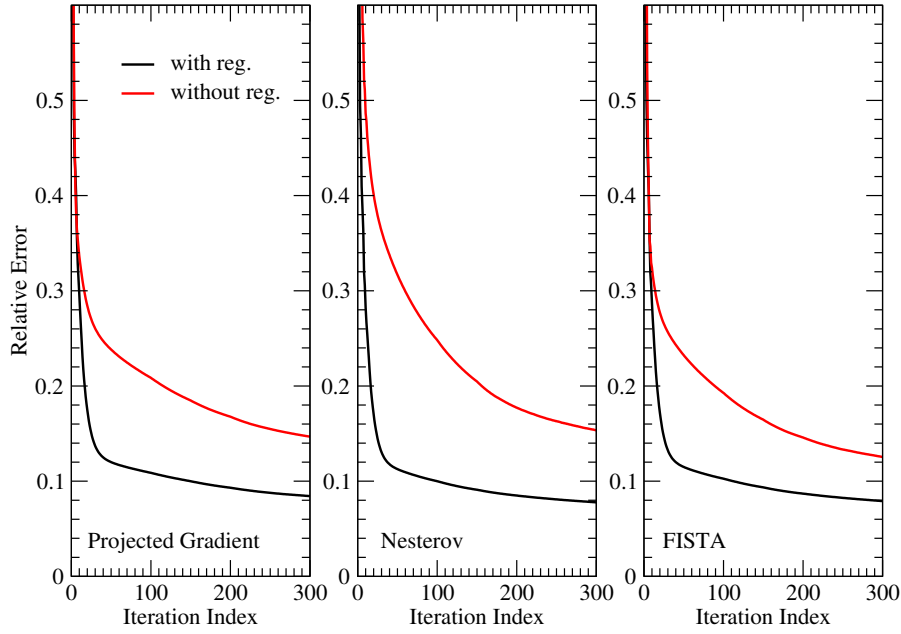


Figure 9: Relative error in extinction field versus the outer iteration index for a box cloud. The results are computed with and without regularization. The regularization terms corresponds to an averaging filter.

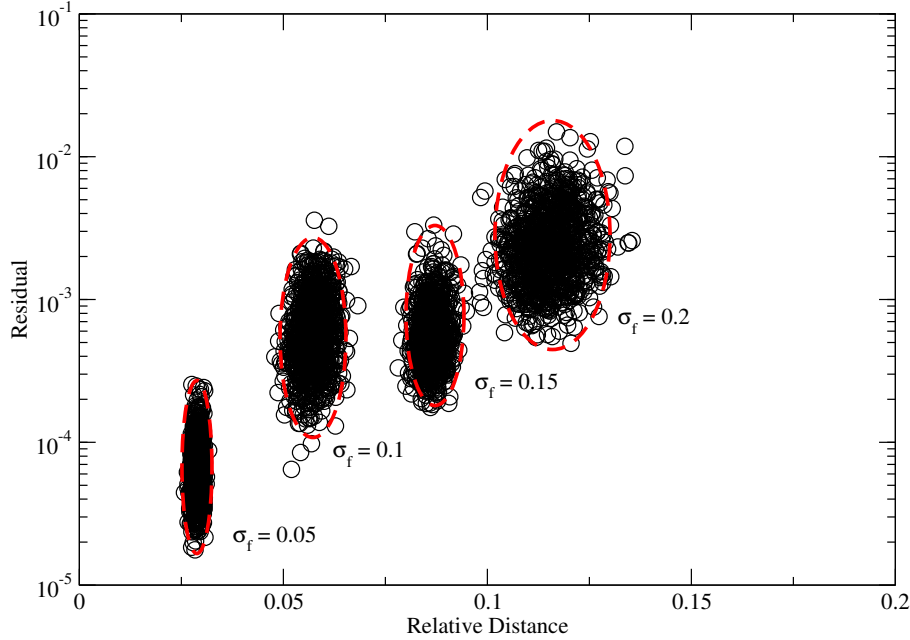


Figure 10: Residual $R(\sigma_{\text{ext}})$ versus the relative distance $d = \|\sigma_{\text{ext}} - \sigma_{\text{ext}}^{\dagger}\|_2 / \|\sigma_{\text{ext}}^{\dagger}\|_2$ for a box cloud with a non-smooth cloud extinction field $\sigma_{\text{ext}}^{\text{cld}}(x, z) = \sigma_{\text{max}} \chi_{\varepsilon}(x, z)$, where $\chi_{\varepsilon}(x, z) = (1 + \sigma_f \varepsilon) \chi(x, z)$ and $\varepsilon \sim \text{U}(-1, 1)$. The results correspond to different values of the standard deviation σ_f and 1000 configurations for each σ_f .

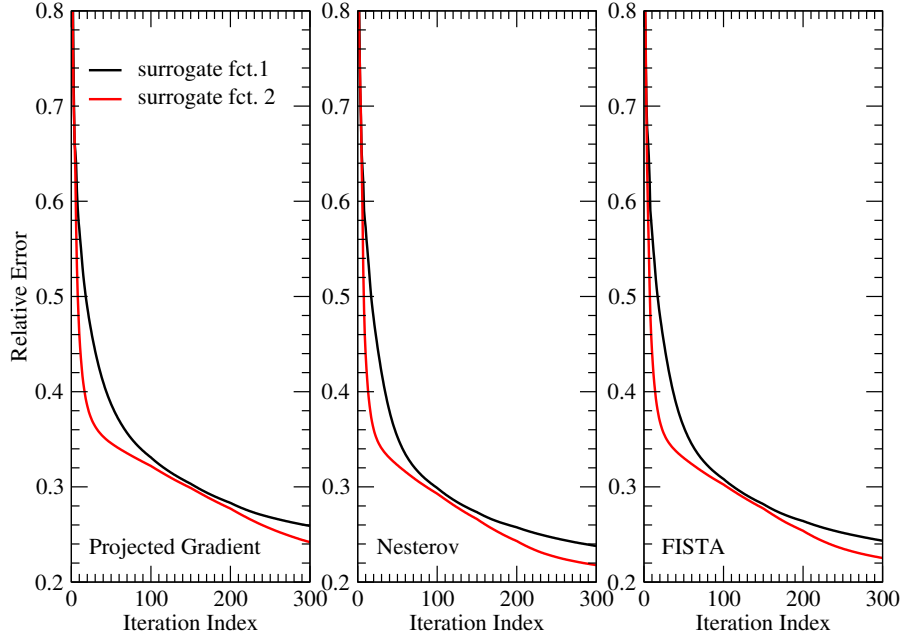


Figure 11: Relative error in extinction field versus the outer iteration index for the LES2 cloud. The results correspond to a surrogate function in which the source function and the surface radiance are kept constant during the minimization step (1) and a surrogate function in which the single-scattering radiance is changed during the minimization step (2).

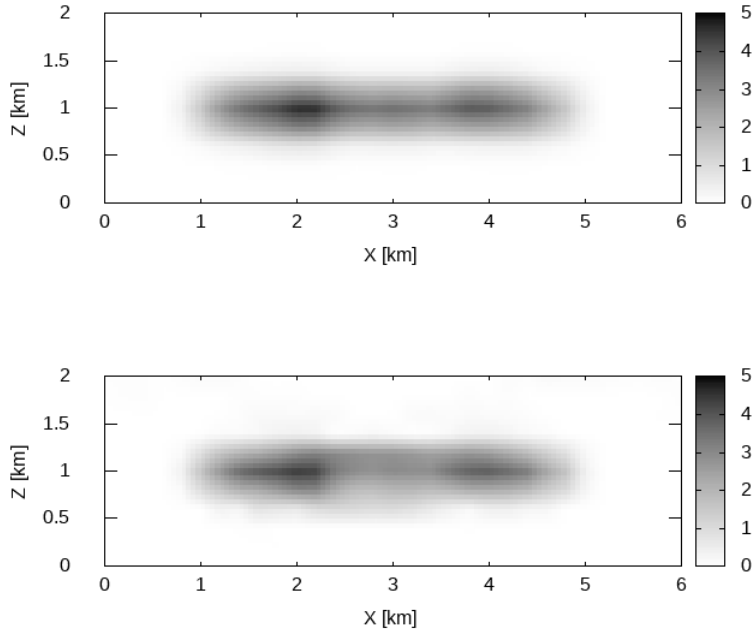


Figure 12: True extinction field (top) and retrieved extinction field (bottom) in the xz -plane for a three-dimensional cloud extinction field obtained from a large eddy simulation of stratocumulus. The results are computed with the Nesterov acceleration method.

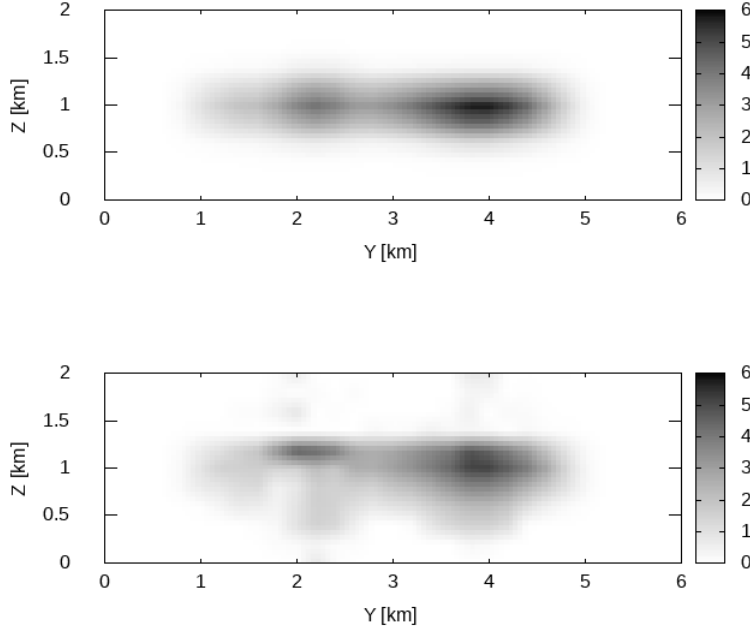


Figure 13: The same as in Fig. 12 but for the yz -plane.

Accordingly, at each step of an outer iteration, the objective function is approximated by a surrogate function, while at each step of an inner iteration, the surrogate function is reduced.

4. To construct a surrogate function, the single-scattering radiance can either be changed or kept constant during the minimization step.
5. To minimize the surrogate function, several first-order optimization methods are employed. These include the (i) projected gradient method, the non-convex versions of (ii) the Nesterov acceleration method and (iii) FISTA, and (iv) the BFGS algorithm. The optimization methods are used in combination with a step length algorithm based on Wolfe conditions. Analogous to REGINN iteration method, the number of steps of the inner iteration is increased during the outer iteration.

In fact, the algorithm performances are mainly increased by the use of (i) regularization and (ii) accelerated projected gradient methods. Depending on the application, the use of a surrogate function, in which the single-scattering radiance is changed during the minimization step, may also improve the algorithm efficiency.

In a companion paper we will describe the implementation of the adjoint method in the cloud tomographic retrieval algorithm. This method can compute a more accurate gradient, but the accuracy comes at the cost of requiring one extra call to the radiative transfer solver.

Appendix A

In SHDOM, discrete ordinates and spherical harmonics are used to represent the radiance field during different parts of the solution algorithm.

1. The discrete ordinates are a reduced Gaussian grid and are used to integrate the radiative transfer equation spatially. There are N_μ Gaussian quadrature cosine zenith angles, μ_j , and N_φ evenly spaced azimuth angles, φ_k . The discrete ordinate set is reduced by having fewer azimuth angles at larger $|\mu|_j$; thus, N_φ depends on j . The corresponding Gauss-Legendre quadrature weights and azimuthal integration weights (normalized appropriately) are w_{μ_j} and $w_{\varphi_{jk}}$, respectively.

2. The spherical harmonics are orthonormal real valued functions and are used to compute the source function including the scattering integral. They are defined by

$$Y_{mn}(\mu, \varphi) = P_n^{|m|}(\mu)u_m(\varphi), \quad (42)$$

where $P_n^{|m|}(\mu)$ are the normalized associated Legendre functions, and

$$u_m(\varphi) = \begin{cases} (1/\sqrt{\pi}) \cos(|m|\varphi) & m > 0 \\ (1/\sqrt{2\pi}) & m = 0 \\ (1/\sqrt{\pi}) \sin(|m|\varphi) & m < 0 \end{cases} \quad (43)$$

are the Fourier harmonics. Specifically, the source function (5) is expressed in the spherical harmonic space as

$$J(\mathbf{r}, \boldsymbol{\Omega}) = \sum_{m=-M}^M \sum_{n=|m|}^N J_{mn}(\mathbf{r}) Y_{mn}(\boldsymbol{\Omega}), \quad (44)$$

where $N = N_\mu - 1$ is the maximum expansion order and $M = N_\varphi/2 - 1$ is the maximum number of azimuthal modes. Moreover, by means of the addition theorem for the Legendre functions, the phase function, expanded in terms of unnormalized Legendre polynomials

$$\tilde{P}_n(\cos \Theta) = \sqrt{\frac{2}{2n+1}} P_n(\cos \Theta), \quad (45)$$

as

$$P(\mathbf{r}, \cos \Theta) = \sum_{n=1}^{N_{\text{rank}}} \chi_n(\mathbf{r}) \tilde{P}_n(\cos \Theta), \quad (46)$$

where $\chi_n(\mathbf{r})$ are the Legendre phase function coefficients and N_{rank} is the maximum expansion order of the phase function, is truncated in the spherical harmonic space as

$$P(\mathbf{r}, \cos \Theta) = 4\pi \sum_{m=-M}^M \sum_{n=|m|}^N \frac{\chi_n(\mathbf{r})}{2n+1} Y_{mn}(\boldsymbol{\Omega}) Y_{mn}(\boldsymbol{\Omega}'), \quad (47)$$

where $\cos \Theta = \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}'$.

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 . The quantities are specified at all grid points \mathbf{r}_i . If the delta-M scaling method is applied, σ_{ext} , ω , and χ_n are scaled before their use. The scaled quantities $\bar{\sigma}_{\text{ext}}$, $\bar{\omega}$, and $\bar{\chi}_n$ are given, respectively, by

$$\bar{\sigma}_{\text{ext}} = (1 - f\omega)\sigma_{\text{ext}}, \quad (48)$$

$$\bar{\omega} = \frac{1-f}{1-f\omega}\omega, \quad (49)$$

$$\frac{2}{2n+1}\bar{\chi}_n = \frac{1}{1-f} \left(\frac{2}{2n+1}\chi_n - 2f \right), \quad n = 0, \dots, N, \quad (50)$$

where

$$f = \frac{1}{2N+3}\chi_{N+1} \quad (51)$$

is the truncation fraction.

The solution method is based on Picard iterations. At the beginning of each iteration step, the expansion coefficients $J_{mn}(\mathbf{r}_i)$, $m = -M, \dots, M$, $n = |m|, \dots, N$, are assumed to be known at all grid points \mathbf{r}_i . They are updated according to the following computational steps.

1. The source function is transformed to discrete ordinates,

$$\begin{aligned}
J(\mathbf{r}_i, \boldsymbol{\Omega}_{jk}) &= \sum_{m=-M}^M \sum_{n=|m|}^N J_{mn}(\mathbf{r}_i) Y_{mn}(\boldsymbol{\Omega}_{jk}) \\
&= \sum_{m=-M}^M u_m(\varphi_k) \sum_{n=|m|}^N J_{mn}(\mathbf{r}_i) P_n^{[m]}(\mu_j),
\end{aligned} \tag{52}$$

where $\boldsymbol{\Omega}_{jk} = (\mu_j, \varphi_k)$.

2. The discrete ordinate radiance $I(\mathbf{r}_i, \boldsymbol{\Omega}_{jk})$ is computed from the source function $J(\mathbf{r}_i, \boldsymbol{\Omega}_{jk})$ by integrating the radiative transfer equation. Essentially, (i) the radiances are computed at all grid points \mathbf{r}_i and in all downward directions $\boldsymbol{\Omega}_{jk}^-$ by integrating the radiative transfer equation with the top boundary condition $I(\mathbf{r}_{ti}, \boldsymbol{\Omega}_{jk}^-) = 0$, i.e.,

$$I(\mathbf{r}_i, \boldsymbol{\Omega}_{jk}^-) = \int_{\mathbf{r}_{ti}}^{\mathbf{r}_i} \bar{\sigma}_{\text{ext}}(\mathbf{r}) J(\mathbf{r}, \boldsymbol{\Omega}_{jk}^-) \bar{T}(\mathbf{r}, \mathbf{r}_i) d\mathbf{s}, \tag{53}$$

- (ii) the radiances at all points on the bottom surface \mathbf{r}_{bi} and in all upward directions $\boldsymbol{\Omega}_{jk}^+$ are computed from the boundary condition, e.g., in the case of a Lambertian surface,

$$\begin{aligned}
I(\mathbf{r}_{bi}, \boldsymbol{\Omega}_{jk}^+) &= \frac{A_s}{\pi} F_0 \bar{T}(\mathbf{r}_{0bi}, \mathbf{r}_{bi}) \\
&+ \frac{A_s}{\pi} \sum_{j'=1}^{N_\mu/2} \sum_{k'=1}^{N_{\varphi j'}} w_{\mu j'} w_{\varphi j' k'} |\mu_{j'}^-| I(\mathbf{r}_{bi}, \boldsymbol{\Omega}_{j' k'}^-),
\end{aligned} \tag{54}$$

- (iii) the radiances are computed at all grid points \mathbf{r}_i and in all upward directions $\boldsymbol{\Omega}_{jk}^+$ by integrating the radiative transfer equation with the boundary condition (54), i.e.,

$$\begin{aligned}
I(\mathbf{r}_i, \boldsymbol{\Omega}_{jk}^+) &= I(\mathbf{r}_{bi}, \boldsymbol{\Omega}_{jk}^+) \bar{T}(\mathbf{r}_{bi}, \mathbf{r}_i) \\
&+ \int_{\mathbf{r}_{bi}}^{\mathbf{r}_i} \bar{\sigma}_{\text{ext}}(\mathbf{r}) J(\mathbf{r}, \boldsymbol{\Omega}_{jk}^+) \bar{T}(\mathbf{r}, \mathbf{r}_i) d\mathbf{s}.
\end{aligned} \tag{55}$$

Here and in the following, a bar notation stands for a delta-M scaled quantity.

3. The discrete ordinate radiance is transformed to the spherical harmonic space according to

$$I_{mn}(\mathbf{r}_i) = \sum_{j=1}^{N_\mu} w_{\mu j} P_n^{[m]}(\mu_j) \sum_{k=1}^{N_{\varphi j}} w_{\varphi j k} I(\mathbf{r}_i, \boldsymbol{\Omega}_{jk}) u_m(\varphi_k). \tag{56}$$

4. The source function is updated from the radiance in the spherical harmonic space as

$$\begin{aligned}
J_{mn}(\mathbf{r}_i) &= \bar{\omega}(\mathbf{r}_i) \frac{\bar{\chi}_n(\mathbf{r}_i)}{2n+1} I_{mn}(\mathbf{r}_i) \\
&+ \bar{\omega}(\mathbf{r}_i) \frac{\bar{\chi}_n(\mathbf{r}_i)}{2n+1} Y_{mn}(\boldsymbol{\Omega}_0) \frac{F_0}{|\mu_0|} \bar{T}(\mathbf{r}_{0i}, \mathbf{r}_i).
\end{aligned} \tag{57}$$

The following additional features of SHDOM should be mentioned.

1. 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. The criterion for splitting cells is based on how much the source function changes across a cell. A cell may be split in half in either of the three Cartesian directions, depending on whether any of them exceed the splitting criterion.

2. Because the solution method is similar to an order of scattering approach, convergence is slower for optically thicker or more scattering media. For these situations, an acceleration method, based on the observed geometric convergence of the solution, is used.

For solar problems with the delta-M method and when the TMS method [11] is used to compute the source function, the spherical harmonic representation of the source function $J_{mn}(\mathbf{r}_i)$ is transformed to a desired viewing direction $\mathbf{\Omega}_{mq}$ by using the relations

$$J(\mathbf{r}_i, \mathbf{\Omega}_{mq}) = \sum_{m=-M}^M \sum_{n=|m|}^N J_{mn}(\mathbf{r}_i) Y_{mn}(\mathbf{\Omega}_{mq}) + \Delta J(\mathbf{r}_i, \mathbf{\Omega}_{mq}), \quad (58)$$

$$\begin{aligned} \Delta J(\mathbf{r}_i, \mathbf{\Omega}_{mq}) = & \frac{F_0}{|\mu_0|} \bar{T}(\mathbf{r}_{0i}, \mathbf{r}_i) \left[\frac{\omega(\mathbf{r}_i)}{1 - f\omega(\mathbf{r}_i)} \frac{1}{4\pi} \sum_{n=1}^{N_{\text{rank}}} \chi_n(\mathbf{r}_i) \tilde{P}_n(\cos \Theta_q) \right. \\ & \left. - \bar{\omega}(\mathbf{r}_i) \sum_{m=-M}^M \sum_{n=|m|}^N \frac{\bar{\chi}_n(\mathbf{r}_i)}{2n+1} Y_{mn}(\mathbf{\Omega}_0) Y_{mn}(\mathbf{\Omega}_{mq}) \right], \end{aligned} \quad (59)$$

where $\cos \Theta_q = \mathbf{\Omega}_0 \cdot \mathbf{\Omega}_{mq}$. Note that according to Eqs. (58) and (59), the TMS method replaces the scaled, truncated Legendre phase function expansion for the singly scattered solar radiation by the full, unscaled phase function expansion, while the multiply scattered contribution still comes from the truncated phase function.

Appendix B

In this appendix we describe a procedure to construct a surrogate function in which only the single-scattering radiance is changed during the minimization step. Specifically, after the iterative process of SHDOM has converged, we perform the following computational steps.

Step 1. We use the spherical harmonic coefficients

$$J_{mn}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l}) = J_{mn}^{\text{ms}}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}l}) + J_{mn}^{\text{ss}}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}}), \quad (60)$$

$$J_{mn}^{\text{ms}}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}l}) = J_{mn}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}l}) - J_{mn}^{\text{ss}}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}l}), \quad (61)$$

$$J_{mn}^{\text{ss}}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}}) = \bar{\omega}(\mathbf{r}_i) \frac{\bar{\chi}_n(\mathbf{r}_i)}{2n+1} Y_{mn}(\mathbf{\Omega}_0) \frac{F_0}{|\mu_0|} \bar{T}(\mathbf{r}_{0i}, \mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}}), \quad (62)$$

to construct the surrogate function. This result implies that in the framework of the TMS method, the spherical harmonic representation of the source function $J_{mn}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l})$ transforms to direction $\mathbf{\Omega}_{mq}$ according to the relation

$$\begin{aligned} J(\mathbf{r}_i, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l}) = & \sum_{m=-M}^M \sum_{n=|m|}^N J_{mn}^{\text{ms}}(\mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}l}) Y_{mn}(\mathbf{\Omega}_{mq}) \\ & + \frac{F_0}{4\pi|\mu_0|} \bar{T}(\mathbf{r}_{0i}, \mathbf{r}_i; \boldsymbol{\sigma}_{\text{ext}}) \frac{\omega(\mathbf{r}_i)}{1 - f\omega(\mathbf{r}_i)} \\ & \times \sum_{n=1}^{N_{\text{rank}}} \chi_n(\mathbf{r}_i) \tilde{P}_n(\cos \Theta_q), \end{aligned} \quad (63)$$

where $\cos \Theta_q = \mathbf{\Omega}_0 \cdot \mathbf{\Omega}_{mq}$. The grid point values $J(\mathbf{r}_i, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l})$ are then used in Eq. (23), i.e.,

$$\begin{aligned} \hat{I}(\mathbf{r}_t, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l}) = & I(\mathbf{r}_b, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}l}) \bar{T}(\mathbf{r}_b, \mathbf{r}_t; \boldsymbol{\sigma}_{\text{ext}}) \\ & + \int_{\mathbf{r}_b}^{\mathbf{r}_t} \bar{\sigma}_{\text{ext}}(\mathbf{r}) J(\mathbf{r}, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l}) \bar{T}(\mathbf{r}, \mathbf{r}_t; \boldsymbol{\sigma}_{\text{ext}}) d\mathbf{s}. \end{aligned} \quad (64)$$

to compute the integral of the source function $J(\mathbf{r}, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{\text{ext}} | \boldsymbol{\sigma}_{\text{ext}l})$.

Step 2. We use the grid point surface radiance

$$I(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}}) = I_{\text{ms}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{extl}}) + I_{\text{ss}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}}), \quad (65)$$

$$I_{\text{ms}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{extl}}) = I(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{extl}}) - I_{\text{ss}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{extl}}) \quad (66)$$

$$I_{\text{ss}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}}) = \frac{A_s}{\pi} \mu_0 \bar{T}(\mathbf{r}_{0b}, \mathbf{r}_{bi}; \sigma_{\text{ext}}) \quad (67)$$

to compute the surface radiance $I(\mathbf{r}_b, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}})$ by bilinear interpolation.

Thus, the radiance $I(\mathbf{r}_t, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}})$, which determines the surrogate function $Q_\alpha(\sigma_{\text{ext}} | \sigma_{\text{extl}})$ according to Eqs. (20)–(22), is given by Eq. (64) in which $I(\mathbf{r}_b, \mathbf{\Omega}_{mq}; \sigma_{\text{extl}})$ and $J(\mathbf{r}, \mathbf{\Omega}_{mq}; \sigma_{\text{extl}})$ are replaced by $I(\mathbf{r}_b, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}})$ and $J(\mathbf{r}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}})$, respectively. From Eqs. (63) and (67), we infer that the derivatives of $J(\mathbf{r}_i, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}})$ and $I(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}} | \sigma_{\text{extl}})$ are also expressed in terms of the derivatives of the transmission function \bar{T} . It should be pointed out that in the retrieval, either Step 1, or both Steps 1 and 2 can be performed.

Some comments are in order.

1. The single-scattering radiance is usually defined as the solution of a boundary-value problem consisting in the inhomogeneous differential equation

$$\frac{dI_{\text{ss}}}{ds}(\mathbf{r}, \mathbf{\Omega}) = -\sigma_{\text{ext}}(\mathbf{r}) I_{\text{ss}}(\mathbf{r}, \mathbf{\Omega}) + \sigma_{\text{ext}}(\mathbf{r}) J_{\text{ss}}(\mathbf{r}, \mathbf{\Omega}), \quad (68)$$

$$J_{\text{ss}}(\mathbf{r}, \mathbf{\Omega}) = \frac{\omega(\mathbf{r})}{4\pi} \frac{F_0}{|\mu_0|} P(\mathbf{r}, \mathbf{\Omega}, \mathbf{\Omega}_0) T(\mathbf{r}_0, \mathbf{r}), \quad (69)$$

and the boundary conditions

$$I_{\text{ss}}(\mathbf{r}_t, \mathbf{\Omega}^-) = 0, \quad \mathbf{r}_t \in S_t, \quad (70)$$

$$I_{\text{ss}}(\mathbf{r}_b, \mathbf{\Omega}^+) = 0, \quad \mathbf{r}_b \in S_b. \quad (71)$$

In this case, by Step 1 above we ensure that this single-scattering radiance is changed during the minimization step.

2. We define the single-scattering radiance as the solution of a boundary-problem consisting in the inhomogeneous differential equation (68)–(69) and the boundary conditions (70) and

$$I_{\text{ss}}(\mathbf{r}_b, \mathbf{\Omega}^+) = \frac{A_s}{\pi} F_0 T(\mathbf{r}_{0b}, \mathbf{r}_b). \quad (72)$$

In this case, by Steps 1 and 2 above we ensure that this single-scattering radiance is changed during the minimization step. Actually, by this procedure, the multiple-scattering components of (i) the spherical harmonic coefficients of the source function $J_{mn}(\mathbf{r}_i; \sigma_{\text{ext}})$ and (ii) the grid point radiances at the bottom surface $I(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}})$ are kept constant during the minimization step.

3. To exploit a full analytical computation of the derivatives of the transmission function, we define the single-scattering radiance as the solution of a boundary-problem consisting in the inhomogeneous differential equation (68)–(69) and the boundary conditions (70) and

$$\begin{aligned} I_{\text{ss}}(\mathbf{r}_b, \mathbf{\Omega}^+) &= \frac{A_s}{\pi} F_0 T(\mathbf{r}_{0b}, \mathbf{r}_b) \\ &+ \frac{A_s}{\pi} \int_{\Omega^-} |\mu^-| I_{\text{ss}}(\mathbf{r}_b, \mathbf{\Omega}^-) d\Omega^-, \quad \mathbf{r}_b \in S_b, \end{aligned} \quad (73)$$

In this case, the surface radiance $I_{\text{ss}}(\mathbf{r}_b, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}})$ in Eq. (65), is computed from the grid point values

$$\begin{aligned} I_{\text{ss}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \sigma_{\text{ext}}) &= \frac{A_s}{\pi} F_0 \bar{T}(\mathbf{r}_{0b}, \mathbf{r}_{bi}; \sigma_{\text{ext}}) \\ &+ \frac{A_s}{\pi} \sum_{j'=1}^{N_\mu/2} \sum_{k'=1}^{N_{\varphi j'}} w_{\mu j'} w_{\varphi j' k'} |\mu_{j'}^-| I_{\text{ss}}(\mathbf{r}_{bi}, \mathbf{\Omega}_{j' k'}^-; \sigma_{\text{ext}}), \end{aligned} \quad (74)$$

with

$$I_{ss}(\mathbf{r}_{bi}, \mathbf{\Omega}_{j'k'}^-; \boldsymbol{\sigma}_{ext}) = \int_{\mathbf{r}_{ti}}^{\mathbf{r}_{bi}} \bar{\sigma}_{ext}(\mathbf{r}) J_{ss}(\mathbf{r}, \mathbf{\Omega}_{j'k'}^-; \boldsymbol{\sigma}_{ext}) \bar{T}(\mathbf{r}, \mathbf{r}_{bi}; \boldsymbol{\sigma}_{ext}) d\mathbf{s}, \quad (75)$$

$$J_{ss}(\mathbf{r}, \mathbf{\Omega}_{j'k'}^-; \boldsymbol{\sigma}_{ext}) = \sum_{m=-M}^M \sum_{n=|m|}^N \bar{\omega}(\mathbf{r}) \frac{\bar{\chi}_n(\mathbf{r})}{2n+1} Y_{mn}(\mathbf{\Omega}_0) Y_{mn}(\mathbf{\Omega}_{j'k'}^-) \times \frac{F_0}{|\mu_0|} \bar{T}(\mathbf{r}_0, \mathbf{r}; \boldsymbol{\sigma}_{ext}). \quad (76)$$

In principle, the computation of the derivatives of $I_{ss}(\mathbf{r}_{bi}, \mathbf{\Omega}_{j'k'}^-; \boldsymbol{\sigma}_{ext})$, and so, of $I_{ss}(\mathbf{r}_{bi}, \mathbf{\Omega}_{mq}; \boldsymbol{\sigma}_{ext})$, also requires only the computation of the derivatives of the transmission function \bar{T} . However, our numerical analysis showed that this computational step is extremely time consuming, and for this reason, less recommendable.

References

- [1] 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>.
- [2] Evans KF. The spherical harmonic discrete ordinate method for three-dimensional atmospheric radiative transfer. J Atmos Sci 1998;55:429–446.
- [3] Levis A, Schechner YY, Davis AB. Multiple-scattering microphysics tomography. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, IEEE, 2017;5797– 5806. <https://doi.org/10.1109/CVPR.2017.614>.
- [4] Levis A, Schechner YY, Davis AB, Loveridge J. Multi-View Polarimetric Scattering Cloud Tomography and Retrieval of Droplet Size. Remote Sensing. 2020; 12(17):2831. <https://doi.org/10.3390/rs12172831>
- [5] Martin WGK, Hasekamp OP. A demonstration of adjoint methods for multi-dimensional remote sensing of the atmosphere and surface. JQSRT 2018;204:215–231.
- [6] 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.
- [7] Meng X-L. Discussion on “optimization transfer using surrogate objective functions”. Journal of Computational and Graphical Statistics 2000;9:35–43.
- [8] Lange K, Hunter DR, Yang I. Optimization transfer using surrogate objective functions with discussion. Journal of Computational and Graphical Statistics 2000;9:1–59.
- [9] Borg I, Groenen P. Modern multidimensional scaling. New York: Springer; 1997.
- [10] Boyd S, Vandenberghe L. Convex optimization. Cambridge: Cambridge University Press; 2004.
- [11] Nakajima T, Tanaka M. Algorithms for radiative intensity calculations in moderately thick atmosphere using a truncation approximation. JQSRT 1988;40:51–69.
- [12] Klinger T. Image Processing with LabVIEW and IMAQ Vision. ISBN: 0130474150. 2005.
- [13] Sinha PK. Image Acquisition and Preprocessing for Machine Vision Systems. SPIE eBooks, <https://doi.org/10.1117/3.858360.ch13>. 2012.
- [14] Burger W, Burge MJ. Principles of Digital Image Processing: Fundamental Techniques. London: Springer; 2009.

- [15] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 1977;39:1–38.
- [16] Mairal J. Stochastic majorization-minimization algorithms for large-scale optimization. In *Advances in Neural Information Processing Systems* 2013, pp.2283–2291.
- [17] Rieder A. On the regularization of nonlinear ill-posed problems via inexact Newton iterations. *Inverse Problems* 1999;15:309–327.
- [18] Doicu A, Trautmann T, Schreier F. *Numerical Regularization for Atmospheric Inverse Problems*. Berlin: Springer; 2010.
- [19] Li H, Lin Z. Accelerated proximal gradient methods for nonconvex programming. In *Advances in Neural Information Processing Systems* 2015, pp. 379–387.
- [20] Li Q, Zhou Y, Liang Y, Varshney PK. Convergence Analysis of Proximal Gradient with Momentum for Nonconvex Optimization. *Proceedings of the 34th International Conference on Machine Learning*, PMLR 2017;70:2111–2119,.
- [21] Yao Q, Kwok JT. More efficient accelerated proximal algorithm for nonconvex problems. ArXiv preprint, December 2016. URL <https://arxiv.org/abs/1612.09069>.
- [22] Nesterov Y. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Soviet Mathematics Doklady* 1983;27:372–376.
- [23] Nesterov Y. *Lectures on convex optimization*, volume 137. Springer; 2018.
- [24] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal of Image Science* 2009;2:183–202.
- [25] 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