JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D16, PAGES 20,673–20,696, AUGUST 27, 2000

A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements

Oleg Dubovik[1]

Laboratory for Terrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland

Michael D. King

Earth Sciences Directorate, NASA Goddard Space Flight Center, Greenbelt, Maryland

Abstract. The problem of deriving a complete set of aerosol optical properties from Sun and sky radiance measurements is discussed. Algorithm development is focused on improving aerosol retrievals by means of including a detailed statistical optimization of the influence of noise in the inversion procedure. The methodological aspects of such an optimization are discussed in detail and revised according to both modern findings in inversion theory and practical experience in remote sensing. Accordingly, the proposed inversion algorithm is built on the principles of statistical estimation: the spectral radiances and various a priori constraints on aerosol characteristics are considered as multisource data that are known with predetermined accuracy. The inversion is designed as a search for the best fit of all input data by a theoretical model that takes into account the different levels of accuracy of the fitted data. The algorithm allows a choice of normal or lognormal noise assumptions. The multivariable fitting is implemented by a stable numerical procedure combining matrix inversion and univariant relaxation. The theoretical inversion scheme has been realized in the advanced algorithm retrieving aerosol size distribution together with complex refractive index from the spectral measurements of direct and diffuse radiation. The aerosol particles are modeled as homogeneous spheres. The atmospheric radiative transfer modeling is implemented with well-established publicly available radiative transfer codes. The retrieved refractive indices can be wavelength dependent; however, the extended smoothness constraints are applied to its spectral dependence (and indirectly through smoothness constraints on retrieved size distributions). The positive effects of statistical optimization on the retrieval results as well as the importance of applying a priori constraints are discussed in detail for the retrieval of both aerosol size distribution and complex refractive index. The developed algorithm is adapted for the retrieval of aerosol properties from measurements made by ground-based Sun-sky scanning radiometers used in the Aerosol Robotic Network (AERONET). The results of numerical tests together with examples of experimental data inversions are presented.

1. Introduction

Recently, there have been numerous studies focused on measuring and interpreting aerosol optical properties, for example, SCAR-B [*Kaufman et al.*, 1998], TARFOX [*Russell et al.*, 1999], ACE 1 [*Bates et al.*, 1998], ACE 2 [*Russell and Heintzenberg*, 2000], and INDOEX [*Ramanathan et al.*, 1996; *Satheesh et al.*, 1999]. Especially high expectations are associated with satellite and ground-based remote sensing [e.g., *King et al.*, 1999; *Kaufman et al.*, 1997]; however, not every required radiative property can be measured remotely. For example, the angular and spectral ranges of remote measurements of atmospheric radiation are always limited. Correspondingly, a core aspect of remote sensing is the inversion procedure, whereby aerosol optical and radiative properties are derived from the remote sensing measurements. In the past three decades a number of inversion methods have been proposed for interpreting the measured radiative characteristics of the cloud-free atmosphere. For example, the algorithms of *King et al.* [1978], *Nakajima et al.* [1983, 1996], and *Wang and Gordon* [1993], developed for deriving aerosol optical properties from atmospheric radiances, are well established. These methods differ in the set of retrieved aerosol parameters and/or set of required input radiative characteristics. This paper describes an inversion strategy focused on retrieving an extended set of aerosol parameters from multiangular and multispectral measurements of atmospheric radiances. The purpose is to maximize the retrieved aerosol information by inverting simultaneously all available measurements of atmospheric radiances. Namely, in this paper we pursue the simultaneous retrieval of aerosol particle size distribution and complex refractive index from spectral optical thickness measurements combined with the angular distribution of sky radiance measured at different wavelengths.

Our retrieval developments are consistent with the developments by *King et al.* [1978] and *Nakajima et al.* [1983, 1996] for retrieving the particle size distribution of aerosol in the total atmospheric column. The method of King et al. is used to invert spectral measurements of optical thickness only,

[1]Also at Science Systems and Applications, Inc., Lanham, Maryland.

Copyright 2000 by the American Geophysical Union.

Paper number 2000JD900282.
0148-0227/00/2000JD900282$09.00

whereas the method of Nakajima et al. is used to invert the angular distribution of sky radiance (with or without spectral optical thickness). It should be noted that the method of Nakajima et al. adequately accounts for multiple-scattering effects in the whole range of the scattering angles. This was an important improvement over earlier sky radiance and optical thickness inversion algorithms [cf. *Twitty*, 1975; *Shaw*, 1979; *O'Neill and Miller*, 1984] which were limited to the aureole region where single-scattering or quasi-single-scattering models can be applied. All of these methods model aerosol particles as homogeneous spheres with refractive indices assumed a priori. Concepts for the determination of the aerosol particle refractive index from multiangular radiance measurements were developed by *Wendisch and von Hoyningen-Huene* [1994] and *Yamasoe et al.* [1998]. These methods are based on the principle of partial separation of the effects of refractive index and size distribution on the angular variability of sky radiance. Our approach is significantly different from earlier studies in that we implement simultaneous retrieval of the particle size distribution and complex refractive index via simultaneous fitting of radiances measured in the entire available angular and spectral range. Such an approach should provide a higher retrieval accuracy through adoption of sophisticated mathematical procedures.

The inversion methodology considered in this paper addresses the simultaneous retrieval of a large number of significantly different parameters from multisource data. For example, direct Sun and diffuse sky radiance are measured by sensors with different sensitivities, and the accuracy requirements on measurements of direct Sun radiation and diffuse sky radiance are rather different. Such accuracy differences should be taken into account when making multisource data compatible. Similarly, the aerosol particle size distribution and complex refractive index are fundamentally different parameters. Correspondingly, the design of an algorithm for retrieving these characteristics should congruously rationalize the differences in units, ranges of variability, etc. Developing any inversion algorithm demands two kinds of effort from the developer. First of all, accurate forward modeling of measured atmospheric characteristics is required. The second necessary component of an inversion algorithm is a formal numerical procedure that utilizes a mathematical inverse transformation, which is not limited in its application to inversion of atmospheric radiances and can be used in any retrieval algorithm. In the following sections we will discuss both of these aspects.

For modeling atmospheric radiances we adopted standardized, publicly available software. This approach allows for the possibility of easily replacing one code with another as radiative transfer theory advances. In keeping with the strategy in forward modeling strategy, we pursued a similar goal of making the entire algorithm flexible and adjustable. In designing the algorithm, we tried to anticipate the possibilities of upgrading forward modeling codes with new advanced versions and expanding the code applicability for new applications (e.g., accounting for light polarization, detailed characteristics of surface reflectance, incorporating particle non sphericity, etc.).

We pursued a similar objective in implementing the numerical inversion transformations in our retrieval algorithm. However, in addition to this objective, designing a numerical inversion algorithm requires clarification of inversion principles. Indeed, forward models differ mainly in the accuracy of describing a physical phenomenon and the speed of calculation. Accordingly, for practical applications, one always chooses the most accurate model provided it satisfies the time constraints. Choosing the best inversion method, on the other hand, is a more complicated task, in that the evaluation of inversion accuracy is an ambiguous question, especially for a case of the simultaneous retrieval of several variables. For example, replacing a scalar model of light scattering by a model accounting for polarization doubtlessly results in accuracy improvements in scattered light estimation. In contrast, retrieval errors are not so directly responsive for different retrieved parameters. Retrieval accuracy may improve for one parameter but degrade for another parameter as the result of a change of inversion methods. Correspondingly, the preference between inversion methods is always rather uncertain.

Detailed reviews of currently used methods can be found in various books, for example, *Twomey* [1977], *Tikhonov and Arsenin* [1977], *Houghton et al.* [1984], *Tarantola* [1987]. However, the existence of a variety of different well-established inversion procedures creates an uncertainty for researchers in understanding how to choose the optimal technique for inversion implementation. For example, the widely used book by *Press et al.* [1992] proposes a diversity of inversion methods; however, it does not direct the reader with explanations as to which method and why it should be chosen for a particular application. Such a situation is partly a result of the fact that most innovations were proposed under pressure of different specific practical needs and derived in rather different ways. In the present paper we follow the inversion strategy proposed and refined in the previous studies by *Dubovik et al.* [1995, 1998a]. This strategy is focused on clarifying the connection between different inversion methods established in atmospheric optics and unifying the key ideas of these methods in a single inversion procedure. Correspondingly, this strategy is rather helpful for building optimized and flexible inversion techniques. For example, in sections 3 and 4.2 we outline the important connections of designed retrieval algorithms with the inversion methods widely adopted in the application of atmospheric optics and remote sensing, such as the methods given by *Phillips* [1962], *Twomey* [1963, 1977], *Tikhonov* [1963, 1977], *Chahine* [1968], *Turchin et al.* [1970], and *Rodgers* [1976].

The effort of algorithm development was initiated under the AERONET (Aerosol Robotic Network) project [*Holben et al.*, 1998] with the purpose of meeting the high requirements of aerosol parameter retrieval accuracy needed for satellite data validation and improved understanding of the radiative effects of aerosols. Therefore the discussion of the algorithm design and retrieval accuracy will be focused on the interpretation of radiances measured by AERONET ground-based Sun-sky scanning radiometers.

2. Forward Modeling

The AERONET network provides globally distributed near-real-time observations of aerosol spectral optical thickness and sky radiance as well as derived parameters such as particle size distributions in a manner suitable for integration with satellite data. This network has been developed to provide aerosol information from two kinds of ground-based measurements: spectral data of direct Sun radiation extinction (i.e., aerosol optical thickness) and angular distribution of sky radiance. An inversion algorithm is required for the retrieval of aerosol size distribution, complex refractive index, single-scattering albedo, and phase function. In section 2.1 we discuss the concept of

atmospheric radiance modeling, which we employ in our retrieval algorithm.

2.1. Radiative Transfer Modeling

The atmospheric sky radiance can be modeled by solving the radiative transfer equation for a plane-parallel atmosphere. The angular distribution of diffuse downward radiation can be described by

$$I(\Theta;\lambda) = F_0 m_0 \frac{[\exp(-m_0\tau) - \exp(-m_1\tau)]}{m_0 - m_1} \cdot (\omega_0 P(\Theta;\lambda) + G(\ldots)), \quad \text{if } \theta \neq \theta_0 \tag{1a}$$

$$I(\Theta;\lambda) = F_0 m_0 \exp(-m_0\tau) \cdot (\omega_0\tau P(\Theta;\lambda) + G(\ldots)), \quad \text{if } \theta = \theta_0, \tag{1b}$$

where $I(\Theta;\lambda)$ is the spectral sky radiance measured at different wavelengths and at different scattering angles Θ; F_0 is exoatmospheric flux; θ_0 is solar zenith angle; θ is observation zenith angle; m is air mass ($m_0 \approx 1/\cos\theta_0$, $m_1 \approx 1/\cos\theta$); $\tau = \tau_{ext}(\lambda)$ is spectral extinction optical thickness; $\omega_0 = \omega_0(\lambda)$ is single-scattering albedo; and $P(\Theta;\lambda)$ is phase function at different wavelengths. The term $G(\ldots) = G(\omega_0(\lambda); \tau_{ext}(\lambda); P(\Theta;\lambda); A(\lambda); \theta_0; \theta; \phi)$ describes the multiple-scattering effects, where ϕ is zenith angle of observations and $A(\lambda)$ is spectral surface reflectance. The above equation is written for a homogeneous atmosphere, without accounting for polarization effects and for angular independent ground reflectance (Lambertian approximation). At present, there are a number of well-established and publicly available codes to account for multiple scattering in diffuse radiance. For example, in our studies we have used two independent discrete ordinates codes developed by *Nakajima and Tanaka* [1988] and *Stamnes et al.* [1988]. These codes allow for the inclusion of the vertical variability of atmospheric properties by dividing the atmosphere into a number of homogeneous layers. In these models, different optical thickness, phase function, and single-scattering albedo characterize each layer.

The modeling of $\tau(\lambda)$, $\omega_0(\lambda)$, and $P(\Theta;\lambda)$ requires consideration of three main components under cloud-free conditions: gaseous absorption, molecular scattering, and aerosol scattering and absorption. These three atmospheric components comprise the total optical characteristics of an atmospheric layer as follows:

$$\tau_{ext}^{total}(\lambda) = \tau_{scat}^{aer}(\lambda) + \tau_{abs}^{aer}(\lambda) + \tau_{scat}^{mol}(\lambda) + \tau_{abs}^{gas}(\lambda), \tag{2}$$

$$\omega_0^{total}(\lambda) = \frac{\tau_{scat}^{aer}(\lambda) + \tau_{scat}^{mol}(\lambda)}{\tau_{ext}^{total}(\lambda)} = \frac{\tau_{scat}^{total}(\lambda)}{\tau_{ext}^{total}(\lambda)}, \tag{3}$$

$$P^{total}(\Theta;\lambda) = \frac{\tau_{scat}^{aer}(\lambda)}{\tau_{scat}^{total}(\lambda)} P^{aer}(\Theta;\lambda) + \frac{\tau_{scat}^{mol}(\lambda)}{\tau_{scat}^{total}(\lambda)} P^{mol}(\Theta;\lambda), \tag{4}$$

where $\tau_{ext}^{aer}(\lambda)$ is the aerosol optical thickness of the layer; $\omega_0^{aer}(\lambda)$ is the aerosol single scattering albedo; and $P^{aer}(\Theta;\lambda)$ is the aerosol phase function. In the case of ground-based measurements of solar radiation, strong gaseous absorption can be avoided by instrumental design. Molecular scattering can easily be calculated from the surface pressure at the time of measurements. For instance, the specified wavelengths of the four AERONET sky radiometer spectral channels (440, 670, 870, and 1020 nm) were carefully selected to avoid strong gaseous absorption [*Holben et al.*, 1998]. Slight ozone absorption is accounted for from climatological data. The values of surface reflectance $A(\lambda)$ are also accounted for a priori, in spite of the fact that $A(\lambda)$ can vary significantly depending on climatological and meteorological conditions. Indeed, uncertainty in a priori knowledge of surface reflectance $A(\lambda)$ is usually not critical for modeling of downward solar radiation for two primary reasons. First, in most situations, unreflected solar light dominates over reflected light in the downward radiation field, and accuracy requirements on a priori estimates of $A(\lambda)$ are modest. Second, it is expected that values of $A(\lambda)$ can, in some cases, be available from accompanying measurements of upward radiation. Thus local variability of atmospheric radiance $I(\Theta;\lambda)$ depends primarily on the optical properties of the aerosol particles, and for convenience of further discussion, we can write

$$I(\Theta;\lambda) = I(\tau_{ext}^{aer}(\lambda); \omega_0^{aer}(\lambda); P^{aer}(\Theta;\lambda)). \tag{5}$$

All of these properties ($\tau_{ext}^{aer}(\lambda)$, $\omega_0^{aer}(\lambda)$, $P^{aer}(\Theta;\lambda)$) are highly variable and will be considered below as unknown characteristics that can be retrieved from multiangular and multispectral radiance data. In reality, aerosol properties vary in the vertical direction and a multilayer model of the atmosphere is required, in order to account for the vertical variations in $\tau(\lambda)$, $\omega_0(\lambda)$, and $P(\Theta;\lambda)$. However, radiances measured at the ground are influenced by the whole atmospheric column and are not expected to be strongly dependent on the vertical distribution of aerosol. Consequently, most ground-based retrievals characterize the optical properties of the aerosol in the total atmospheric column (columnar aerosol). In our present study we focus accordingly on designing an algorithm for the vertically homogeneous atmosphere. The strategy of accounting for a vertical variability in the atmosphere is outlined in section 4.

This inversion of atmospheric radiance can naturally be designed for the retrieval of the optical characteristics of columnar aerosol ($\tau(\lambda)$, $\omega_0(\lambda)$, and $P(\Theta;\lambda)$). For instance, *Wang and Gordon* [1993] and *Box and Sendra* [1999] employ such an inversion strategy in their retrievals. Alternatively, the inversion can be focused on retrieving parameters of aerosol microstructure, such as particle size, number, etc. We will utilize this approach by extending the ideas previously developed in the papers of *King et al.* [1978] and *Nakajima et al.* [1983, 1996].

2.2. Microphysics Modeling of Aerosol Optical Properties

The modeling of optical parameters via parameters of microstructure is a rather common way of light-scattering characterization in both laboratory and remote sensing methods [cf. *Mc Cartney*, 1977]. For example, the aerosol optical parameters (phase function ($P(\Theta)$), optical thickness of aerosol extinction, scattering and absorption ($\tau_{ext}(\lambda)$; $\tau_{scat}(\lambda)$; $\tau_{abs}(\lambda)$)) can be modeled from microstructure parameters using the following approximations:

$$\tau_{scat}(\lambda) P(\Theta;\lambda) = \int_{r_{min}}^{r_{max}} K_{scat}(\Theta;\lambda;\tilde{m};r)\mathrm{n}(r)\,dr, \tag{6}$$

$$\tau_{\ldots}(\lambda) = \int_{r_{min}}^{r_{max}} K_{\tau\ldots}(\lambda;\tilde{m};r)\mathrm{n}(r)\,dr, \tag{7}$$

where r is particle radius, $\mathrm{n}(r) = dN(r)/dr$ denotes particle number size distribution in atmospheric column. $K_{scat}(\ldots)$ is a scattering cross section, and $K_{\tau\ldots}(\ldots)$ is an extinction cross section ($\pi r^2 Q_{ext}(\ldots)$, where $Q_{ext}(\ldots)$ is the Mie extinction

efficiency factor). In our studies we will assume that aerosol particles are spherical. Correspondingly, the functions $K_{scat}(\ldots)$ and $K_{\tau}(\ldots)$ will be approximated by Mie functions derived for spherical and homogeneous particles characterized by a complex refractive index:

$$\tilde{m}(\lambda) = n(\lambda) - ik(\lambda).$$

Equations (6) and (7) allow one to consider size distribution and refractive index of aerosol particles instead of directly considering $\tau(\lambda)$, $\omega_0(\lambda)$, and $P(\Theta; \lambda)$ of the aerosol.

Finally, atmospheric radiance $I(\Theta; \lambda)$ given by (5) can be defined via (6) and (7) as a function of the parameters of aerosol microstructure:

$$I(\Theta; \lambda) = I(dN(r)/dr; \tilde{m}(\lambda)). \quad (8)$$

Thus (5) and (8) represent two different strategies of atmospheric radiance modeling. We focus the inversion on retrieving parameters of the aerosol microstructure. In this case some relationship among optical thickness, single-scattering albedo, and phase function is implied by assuming that the aerosol particles are homogeneous spheres, as in (8). Additional discussion on details of atmospheric radiance modeling is given in section 4.

3. Inversion Strategy

To formulate the criteria of inversion optimization, we employ principles of statistical estimation theory [cf. *Edie et al.*, 1971]. Accordingly, in designing the retrieval algorithm we account for the character and level of uncertainties in the initial data. This is especially important when we invert the data measured under different experimental conditions (i.e., data from different sources). Therefore inversion of multisource data is a subject of particular consideration here.

Using a priori constraints is another key aspect, which requires a detailed deliberation. *Phillips* [1962], *Twomey* [1963], and *Tikhonov* [1963] have shown that applying a priori constraints (e.g., the smoothness of retrieved functions) is a critical component of designing a successful inversion with many parameters. Choosing the strength of a priori constraints is, however, an especially challenging problem [e.g., *Rodgers*, 1976; *Twomey*, 1977; *King*, 1982], which becomes even more challenging when such different parameters as particle size distribution and complex refractive index are retrieved simultaneously. Our strategy is to consider measurements and a priori knowledge together as a single set of multisource data. Thus the current section discusses the principles of inversion optimization, which are the same for both measured and a priori data. The specific questions of applying a priori constraints are discussed in detail in section 4.2.

3.1. Statistically Optimized Inversion of Multisource Data

The inversion is designed as a search for the best fit of all data considered by a theoretical model taking into account the different magnitudes of the accuracy of the fitted data. The errors in all inverted data are determined statistically. Both measured and a priori data are separated into groups assuming that data obtained from the same source (i.e., in the same way) have a similar error structure, independent of errors in the data obtained from other sources. For example, direct Sun and diffuse sky radiances have different magnitudes and are measured by sensors with different sensitivity; that is, errors should be independent (due to different sensors) and have different magnitudes.

Formally, both measured and a priori data can be written as follows:

$$\mathbf{f}_k^*(\mathbf{a}) = \mathbf{f}_k(\mathbf{a}) + \Delta_k \qquad (k = 1, 2, \ldots, K), \quad (9)$$

where the vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ relate to sky and Sun radiance measurements at four standard AERONET wavelengths. The vector $\mathbf{a}$ denotes the aerosol parameters which should be retrieved. The vectors $\mathbf{f}_{k>2}$ include the values of a priori constraints on aerosol parameters or possible accessory data. The asterisk denotes the data known with some uncertainties Δ_k.

Numerous studies have shown that the normal (or Gaussian) distribution is the most expected and appropriate function for describing random noise (detailed discussions can be found in the books by *Edie et al.* [1971] and *Tarantola* [1987]). The normal probability density function (PDF) for each vector $\mathbf{f}_k^*$ of initial data can be written in the form

$$P(\mathbf{f}_k(\mathbf{a})|\mathbf{f}_k^*) = ((2\pi)^m \det(\mathbf{C}_k))^{-1/2} \cdot \exp\left(-\tfrac{1}{2}(\mathbf{f}_k(\mathbf{a}) - \mathbf{f}_k^*)^T(\mathbf{C}_k)^{-1}(\mathbf{f}_k(\mathbf{a}) - \mathbf{f}_k^*)\right), \quad (10)$$

where T denotes matrix transposition, $\mathbf{C}_k$ is the covariance matrix of the vector $\mathbf{f}_k$; det $(\mathbf{C}_k)$ denotes determinate of $\mathbf{C}_k$, and m is the dimension of vectors $\mathbf{f}_k$ and $\mathbf{f}_k^*$. The vectors $\mathbf{f}_k^*$ are obtained from different sources and accordingly statistically independent. This is why the joint PDF of all inverted data can be obtained by simple multiplication of the PDF of all vectors $\mathbf{f}_k^*$ as follows:

$$P(\mathbf{f}_1(\mathbf{a}), \ldots, \mathbf{f}_K(\mathbf{a})|\mathbf{f}_1^*, \ldots, \mathbf{f}_K^*) = \prod_{k=1}^{K} P(\mathbf{f}_k(\mathbf{a})|\mathbf{f}_k^*)$$

$$\sim \exp\left(-\frac{1}{2}\sum_{k=1}^{K}(\mathbf{f}_k(\mathbf{a}) - \mathbf{f}_k^*)^T(\mathbf{C}_k)^{-1}(\mathbf{f}_k(\mathbf{a}) - \mathbf{f}_k^*)\right). \quad (11)$$

According to the method of maximum likelihood (MML) the best estimates $\hat{\mathbf{a}}$ of the unknowns $\mathbf{a}$ correspond to the maximum of likelihood function (PDF); that is,

$$P(\mathbf{f}_1(\hat{\mathbf{a}}), \ldots, \mathbf{f}_K(\hat{\mathbf{a}})|\mathbf{f}_1^*, \ldots, \mathbf{f}_K^*) = \max. \quad (12)$$

The MML is one of the strategic principles of statistical estimation, and the solution $\hat{\mathbf{a}}$ is statistically the best in many senses [see *Edie et al.*, 1971]. The solution is asymptotically normal (since PDF is defined asymptotically) and optimum (most accurate; the retrieval errors have the smallest standard deviations). In addition, the MML solution keeps many optimum characteristics even in the case of a limited number of observations. The optimum properties of MML are closely connected with the Fisher information determination [see *Edie et al.*, 1971].

The maximum of the PDF exponential term given by (11) corresponds to the minimum of the quadratic form in the exponent. Therefore the best solution $\hat{\mathbf{a}}$, which can be derived from all given data $\mathbf{f}_k^*$, is a vector $\hat{\mathbf{a}}$ corresponding to the minimum of the following form:

$$\Psi(\mathbf{a}) = \frac{1}{2}\sum_{k=1}^{K}\gamma_k\Psi_k(\mathbf{a}) = \frac{1}{2}\sum_{k=1}^{K}\gamma_k[(\mathbf{f}_k^* - \mathbf{f}_k(\mathbf{a}))^T(\mathbf{W}_k)^{-1}(\mathbf{f}_k^* - \mathbf{f}_k(\mathbf{a}))]. \quad (13)$$

This equation is written via Lagrange multipliers γ_k and weight matrices $\mathbf{W}_k$ defined as

$$\mathbf{W}_k = \frac{1}{\varepsilon_k^2}\mathbf{C}_k, \quad (14)$$

where ε_k^2 denotes the variance of errors Δ_k in the data vector $\mathbf{f}_k^*$. Accordingly, Lagrange multipliers have a clear statistical interpretation as the ratios of variances:

$$\gamma_k = \frac{\varepsilon_1^2}{\varepsilon_k^2}, \quad (15)$$

where ε_1^2 denotes the variance of the first ($k = 1$) and probably most important data set. It should be noted that there is no need to know the absolute value of the variance ε_1^2, because the retrieval process is aimed at finding the global minimum of $\Psi(\mathbf{a})$ and does not depend on the value of this minimum. At the same time, it is known that the value of $\Psi(\mathbf{a})$ has a χ^2 distribution and that the minimum of $\Psi(\mathbf{a})$ statistically relates to ε_1^2 as follows:

$$2\Psi_{\min}(\mathbf{a}) = (N_f - N_a)\varepsilon_1^2, \quad (16)$$

where N_f is the number of values in all fitted vectors $\mathbf{f}_k^*$, and N_a is the number of retrieved parameters. The above relation is often used for estimation of measurement error ε_1^2.

It is important to emphasize that the MML only formulates the condition of optimality and it does not suggest the approach for achieving the minimum of $\Psi(\mathbf{a})$. Finding the minimum of quadratic form $\Psi(\mathbf{a})$ is a technical question, and choosing one or another procedure does not improve the solution, provided the problem is not ill-posed and the solution is unique. According to our strategy of designing the inversion algorithm the correct posing of the problem should be done at the stage of forming the initial data set given by (9). For example, in our case of inverting sky ($\mathbf{f}_1^*$) and Sun ($\mathbf{f}_2^*$) radiances these two basic data sets will be supplemented by some a priori data of corresponding $\mathbf{f}_k^*$ with $k > 2$. Therefore the formulation of initial data sets denoted by (9) is a critical question in inversion algorithm development. In contrast, minimization of $\Psi(\mathbf{a})$ is a technical question, which practically does not affect the accuracy of the solution. Nevertheless, a good design of a minimizing technique is important for liberating computer power requirements and, consequently, reducing the time consumption of the retrieval.

3.2. Minimization Procedure

Modern scientific literature [e.g., *Press et al.*, 1992] contains a variety of standardized mathematical methods and software for minimizing quadratic forms. As noted in the previous section, the choice of method for finding the minimum of $\Psi(\mathbf{a})$ (equation (13)) is not a critical issue and mainly depends on the complexity of the dependencies $\mathbf{f}_k(\mathbf{a})$ and the preference of the inversion algorithm developer. Nevertheless, below, we propose a generalized flexible scheme of minimization that can be easily reduced to different standard methods. The scheme shows the clear relationship between different standard methods. Therefore our expectations are that this scheme should be helpful for designing inversion algorithms for different applications.

For the general case of nonlinear functions $\mathbf{f}_k(\mathbf{a})$ the minimization is usually implemented by iteration:

$$\hat{\mathbf{a}}^{p+1} = \hat{\mathbf{a}}^p - \Delta\mathbf{a}^p, \quad (17a)$$

where the correction $\Delta\mathbf{a}^p$ can be approximated by the linear estimator $\Delta\hat{\mathbf{a}}^p$ as follows:

$$\Delta\mathbf{a}^p \approx t_p\Delta\hat{\mathbf{a}}^p. \quad (17b)$$

The multiplier $t_p \le 1$ (arbitrarily chosen) is typically used for providing monotonic convergence of nonlinear numerical algorithms [cf. *Ortega and Reinboldt*, 1970]. Assuming that $\Delta\hat{\mathbf{a}}^p$ is in the close neighborhood of the solution $\hat{\mathbf{a}}$, a Taylor expansion can be used:

$$\mathbf{f}_k(\hat{\mathbf{a}}) = \mathbf{f}_k(\mathbf{a}^p) + \mathbf{U}_{k,\mathbf{a}^p}(\hat{\mathbf{a}} - \mathbf{a}^p) + o(\hat{\mathbf{a}} - \mathbf{a}^p)^2 + \cdots, \quad (18)$$

where $\mathbf{U}_{k,\mathbf{a}^p}$ is the Jacobi matrix of the first derivatives in the near vicinity of the vector $\mathbf{a}^p$; that is, $\{\mathbf{U}_{k,\mathbf{a}^p}\}_{ji} = \partial(\{\mathbf{f}_k(\mathbf{a})\}_j)/\partial a_i|_{\mathbf{a}_p}$, $o(\hat{\mathbf{a}} - \mathbf{a}^p)^2$ denotes the function that approaches zero as $(\hat{\mathbf{a}} - \mathbf{a}^p)^2$ when $(\hat{\mathbf{a}} - \mathbf{a}^p) \to 0$. Now, neglecting all terms of second or higher order in (18), we can consider $\mathbf{f}_k(\mathbf{a})$ as linear functions in (13). Accordingly, the correction $\Delta\hat{\mathbf{a}}^p$ corresponds to the minimum of $\Psi(\mathbf{a})$ with $\mathbf{f}_k(\mathbf{a})$ linearly approximated. Correspondingly, $\Delta\hat{\mathbf{a}}^p$ can be found (accounting for noise optimization) as a solution of the so-called normal equation system, which for our case is the following (details are given in Appendices A and B):

$$\left(\sum_{k=1}^{K}\gamma_k(\mathbf{U}_{k,\hat{\mathbf{a}}^p})^T(\mathbf{W}_k)^{-1}(\mathbf{U}_{k,\hat{\mathbf{a}}^p}) + \gamma_{\Delta a}(\mathbf{W}_{\Delta a})^{-1}\right)\Delta\hat{\mathbf{a}}^p = \sum_{k-1}^{K}\gamma_k(\mathbf{U}_{k,\hat{\mathbf{a}}^p})^T(\mathbf{W}_k)^{-1}(\mathbf{f}_k(\hat{\mathbf{a}}^p) - \mathbf{f}_k^*) + \gamma_{\Delta a}(\mathbf{W}_{\Delta a})^{-1}(\Delta\hat{\mathbf{a}})^*. \quad (19a)$$

This normal equation system is the solution of the linear least squares method (LSM) [e.g., *Tarantola*, 1987], which gives the minimum of the quadratic form (13) for linear functions $\mathbf{f}_k(\mathbf{a})$. The terms with multiplier $\gamma_{\Delta\mathbf{a}}$ are added on to both the left and the right parts of (19a) for improving the convergence of the whole minimization procedure given by (17)–(19a) (details are given in Appendix B). These terms are incorporated statistically in a similar manner for all data associated with (9); that is, the a priori expected correction $(\Delta\hat{\mathbf{a}})^*$ is assumed statistically to be estimated by $(\Delta\hat{\mathbf{a}})^* = (\Delta\hat{\mathbf{a}}) + \Delta(\Delta\hat{\mathbf{a}})$ with covariance matrix $\mathbf{C}_{\Delta\mathbf{a}}$. It should be noted that both the a priori estimate $(\Delta\hat{\mathbf{a}})^*$ in (19) and the multiplier $t_p \le 1$ in (17b) are mainly invoked to decrease the length of $\Delta\hat{\mathbf{a}}^p$, because linear approximation may strongly overestimate the $\Delta\hat{\mathbf{a}}^p$ correction. Underestimation of $\Delta\hat{\mathbf{a}}^p$ does not affect the convergence, since underestimation may only slow down the arrival to the minimum and not to mislead the minimization.

The key question of implementing minimization by (17)–(19) is the solving of the linear system (19a), which in the compact form can be rewritten as follows:

$$\Phi_p\Delta\hat{\mathbf{a}}^p = \nabla\Psi(\hat{\mathbf{a}}^p), \quad (19b)$$

where matrix $\mathbf{\Phi}_p$ denotes the matrix on the left side of (19a). This matrix (at $\gamma_{\Delta\mathbf{a}} = 0$) relates to the Fisher information matrix, considered in statistical estimation theory [*Edie et al.*, 1971]. The vector $\nabla\Psi(\mathbf{a}^p)$ (i.e., vector on the right side of (19a)) represents the gradient of the quadratic form $\Psi(\mathbf{a})$ (equation (13)). This vector is essential for building optimum minimization [*Ortega*, 1988].

Thus (17)–(19) give a general and flexible form to the minimization of the quadratic form $\Psi(\mathbf{a})$ (equation (13)). This procedure can be easily transformed, by choosing a method for solving (19a), to many other well-established numerical procedures based on matrix inversion, relaxation, combined iterations methods, etc. In our opinion such freedom in incorporating different linear inversion techniques to the generalized nonlinear scheme (equations (17)–(19)) is very useful for both understanding the relationships between existing inversion algorithms and in developing our new algorithm.

In our algorithm for inverting atmospheric radiance we implement two alternative techniques: matrix inversion (using singular value decomposition) and relaxation quasi-gradient techniques. A brief introduction to these methods is given below.

3.2.1. Matrix inversion. The linear system given by (19) can be solved using matrix inversion operations. First of all, the fundamental formula for the linear LSM solution implies matrix inversion [e.g., *Press et al.*, 1992]. Correspondingly, a great number of the LSM-related inversion methods use matrix inversion. For example, *Phillips* [1962], *Twomey* [1963], *Tikhonov* [1963], *Turchin et al.* [1970], and *Rodgers* [1976] employ matrix inversion in their methods. All of these methods are well known in optical applications and differ with the basic LSM formula by using differing a priori constraints (additional discussion can be found in section 4 and in the papers of *Dubovik et al.* [1995, 1998a]).

The basic scheme of solving a nonlinear system is the traditional Newton-Gauss procedure [e.g., *Ortega and Reinboldt*, 1975], which implements the LSM principle in the nonlinear case. Equations (17)–(19) can easily be reduced to the Newton-Gauss procedure. Namely, if we define $t_p = 1$, $\gamma_{\Delta\mathbf{a}} = 0$ and $\gamma_k = 0$ (for $k \geq 2$) in these formulas, we obtain the Newton-Gauss method with statistical optimization at each p step:

$$\mathbf{a}^{p+1} = \mathbf{a}^p - (\mathbf{U}_p^T\mathbf{W}^{-1}\mathbf{U}_p)^{-1}(\mathbf{U}_p^T\mathbf{W}^{-1}(\mathbf{f}^p - \mathbf{f}^*)), \quad (20)$$

where for simplicity we denote the vectors and matrices as follows: $\mathbf{U}_p$ denotes Jacobi matrix $\mathbf{U}_{1,\mathbf{a}^p}$; $\mathbf{W}$ denotes weight matrix $\mathbf{W}_1$; vector $\mathbf{f}^p$ denotes vector $\mathbf{f}(\mathbf{a}^p)$. In this section we always assume $\gamma_k = 0$ (for $k \geq 2$) only because the standard numerical formulas are written for inverting a single data set.

Obviously, (20) incorporates the basic linear LSM formula. Indeed, (20) is reduced to the linear LSM by assuming linear dependence $\mathbf{f}(\mathbf{a}) = \mathbf{U}_\mathbf{a}$:

$$\begin{aligned}\mathbf{a}^{p+1} &= \mathbf{a}^p - (\mathbf{U}^T\mathbf{W}^{-1}\mathbf{U})^{-1}(\mathbf{U}^T\mathbf{W}^{-1}(\mathbf{U}\mathbf{a}^p - \mathbf{f}^*)) \\ &= (\mathbf{U}^T\mathbf{W}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{W}^{-1}\mathbf{f}^*. \quad (20a)\end{aligned}$$

In practice, Newton-Gauss iterations may not converge and need to be modified. The most established modification of (20) is widely known as the Levenberg-Marquardt method [e.g., *Ortega and Reinboldt*, 1970; *Press et al.*, 1992]. This method is also included in the scheme of (17)–(19). Namely, if we assume $t_p \leq 1$, $\gamma_{\Delta\mathbf{a}} > 0$ and $(\Delta\hat{\mathbf{a}})^* = \mathbf{0}$, then (17)–(19) can be reduced to the Levenberg-Marquardt method:

$$\mathbf{a}^{p+1} = \mathbf{a}^p - t_p(\mathbf{U}_p^T\mathbf{W}^{-1}\mathbf{U}_p + \gamma_{\Delta\mathbf{a}}\mathbf{D})^{-1}(\mathbf{U}_p^T\mathbf{W}^{-1}(\mathbf{f}^p - \mathbf{f}^*)), \quad (21)$$

where $\mathbf{D} = (\mathbf{W}_\Delta)^{-1}$ and $\gamma_{\Delta\mathbf{a}} = \varepsilon_0^2/\varepsilon_\Delta^2$. It should be noted that using the generalized inversion procedure of (17) and (19) helps to provide an additional simple interpretation of the Levenberg-Marquardt method. Indeed, an a priori assumption of $(\Delta\hat{\mathbf{a}})^* = \mathbf{0}$ means that we constrain the solutions $\Delta\hat{\mathbf{a}}^p$ to the smallest value (the closest to $(\Delta\hat{\mathbf{a}})^* = \mathbf{0}$). In addition, by assuming $(\Delta\hat{\mathbf{a}})^* \neq \mathbf{0}$ and varying $\mathbf{W}_\Delta$ in (19a), the convergence character can be adjusted in the scheme of (17)–(19) more flexibly than is possible with the standard Levenberg-Marquardt formula (21).

The main difficulty in using the matrix method occurs when the matrix $\mathbf{\Phi}_p$ is of quasi-degenerate nature and the inverse operator $(\mathbf{\Phi}_p)^{-1}$ is very unstable. The practical way of applying matrix inversion is to use matrix singular value decomposition.

Singular value decomposition (SVD) is an operation of linear algebra, which allows one to decompose matrix $\mathbf{\Phi}$ as $\mathbf{\Phi} = \mathbf{V}\mathbf{I}_{w_i}\mathbf{A}$, where matrices $\mathbf{V}$ and $\mathbf{A}$ are orthogonal in the sense that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ and $\mathbf{A}^T\mathbf{A} = \mathbf{I}$. Matrix $\mathbf{I}_{w_i}$ is diagonal with the elements on the diagonal equal to w_i. Inversion of matrix $\mathbf{\Phi}$ trivially follows from this decomposition as $\mathbf{\Phi}^{-1} = \mathbf{A}^T\mathbf{I}_{1/w_i}\mathbf{V}^T$. In the case of a near-singular matrix $\mathbf{\Phi}$, the inverse matrix of $\mathbf{\Phi}$ is uncertain, because some values w_i are equal or close to zero. By replacing $w_i = 0$ with a moderately small nonzero w_i, the singular matrix $\mathbf{\Phi}$ can be replaced by a reasonably close nonsingular matrix $\mathbf{\Phi}'$ which can be easily inverted. The details of this method can be found in the paper of *Press et al.* [1992]. In many practical situations, singular value decomposition is very helpful. Therefore we employ this procedure in our algorithm to implement matrix inversion. The main concern of applying this method comes from the fact that replacement of matrix $\mathbf{\Phi}$ with matrix $\mathbf{\Phi}'$ is formal and has no relation to the physics of an application.

3.2.2. Alternatives to matrix inversion methods. Many methods are known in the mathematical literature that can be used to solve linear systems of equations without using matrix inversion. Some examples are the Jacobi and Gauss-Seidel univariant iterations, the steepest descent method, the method of conjugated gradients, etc. Some of these methods can yield superior results over matrix inversion operations. For example, in our algorithm we employ linear iterations, which always give a result even if the linear system is singular and a solution is not unique. In contrast with inversions performed by means of singular value decomposition, iterations do not require any change of matrix $\mathbf{\Phi}$.

In the papers by *Dubovik et al.* [1995, 1998a] the solution of the p-step system (equation (19)) is implemented by means of linear q-iterations and the whole minimization process is represented via combined iterations (two kinds of iteration). Namely, $\Delta\hat{\mathbf{a}}^p$ is obtained from (19b) by means of q-linear iterations:

$$(\Delta\mathbf{a}^p)^{q+1} = (\Delta\mathbf{a}^p)^q - (\mathbf{H}_p)^q[\mathbf{\Phi}_p(\Delta\mathbf{a}^p)^q - \nabla\Psi(\mathbf{a}^p)]. \quad (22a)$$

Equations (17) and (19) and (22) formulate a search for the minimum $\mathbf{a}^p$ of the quadratic form $\Psi(\mathbf{a})$ (equation (13)) via combined p and q iterations. For each p iteration, a larger number of q iterations can be made. The matrix $\mathbf{H}_p$ and vector $(\Delta\mathbf{a}^p)^{q=0}$ can be chosen in various ways to assure that the iterations converge.

Such a combined iteration technique is helpful for realizing statistical optimization (which usually is associated with matrix methods) by means of relaxation iterations ($\mathbf{H}_p$ is diagonal matrix) in situations where matrix inversion is not efficient. In addition, the consideration of combined iterations helps one to

understand relationships between two categories of inversion methods: matrix inversion methods [*Phillips*, 1962; *Twomey*, 1963; *Tikhonov*, 1963; *Turchin et al.*, 1970; *Rodgers*, 1976] and relaxation techniques [*Chahine*, 1968; *Twomey*, 1975]. These two kinds of methods are popular in atmospheric optics and remote sensing, and they usually are considered as alternative methodologies.

The steepest descent method deserves particular attention among all other relaxation techniques. This method has been well described in the mathematical literature [e.g., *Forsythe and Wasow*, 1960; *Ortega*, 1988]. The basic idea of the steepest descent method (or gradient search method) is to minimize the quadratic form $\Psi(\mathbf{a})$ using its gradient as a direction of the strongest local change of $\Psi(\mathbf{a})$. The minimization procedure given by (17) and (19) and (22) can be easily reduced to the steepest descent method by assuming $\mathbf{H}_p = t_q\mathbf{1}$, $(\Delta\mathbf{a}^p)^{q=0} = \mathbf{0}$ in (22):

$$\mathbf{a}^{p+1} = \mathbf{a}^p - t_p\nabla\Psi(\mathbf{a}^p) = \mathbf{a}^p - t_p(\mathbf{U}_p^T\mathbf{W}^{-1}(\mathbf{f}^p - \mathbf{f}^*)). \quad (22b)$$

Also, only one q iteration is to be implemented for each p iteration in (22b); that is, $t_{p,q} = t_p$ and the combined iterations are reduced to only one kind of p iteration.

As pointed out by *Press et al.* [1992], the steepest descent method is generalized by the Levenberg-Marquardt formula. Namely, (19a) can be reduced to (21) by defining matrix $\mathbf{D}$ in (19a) as the unit matrix $\mathbf{1}$ and prescribing a large value to the parameter $\gamma_{\Delta a}$. In Appendix D we show that the popular Twomey-Chahine relaxation technique proposed by *Twomey* [1975] can be considered to be the steepest descent method.

Equation (22b) can be used to solve both linear and nonlinear equations. Correspondingly, the nonlinear steepest descent iterations can be used directly for minimization of quadratic form in (13). However, such minimization can be very time consuming because, for the nonlinear case, each iteration requires a recalculation of the Jacobi matrix $\mathbf{U}_p$, and the steepest descent method converges to the solution only after a very large number of iterations. Therefore to reduce computation time, we use the steepest descent method only to solve linear p-step systems (equation (19b)), we assume $\mathbf{H}_p = t_{p,q}$, $(\Delta\mathbf{a}^p)^{q=0} = \mathbf{0}$ in (22a). Then we implement a large number (N_q) of the q iterations.

We choose the value of $t_{p,q}$, providing the fastest convergence of the process at each q iteration. *Forsythe and Wasow* [1960] and *Ortega* [1988] describe the principles of defining such a value.

4. Sun-Sky Radiance Inversion Algorithm

In sections 2 and 3 we described two complementary and necessary tools for realizing an inversion algorithm: a model of radiative transfer and a method of optimum inversion. Our intention was to structure and, in a certain sense, to standardize the process of designing an inversion algorithm. In section 3 we outlined the optimization strategy common to any numerical inversion and proposed the scheme (equations (17)–(19)) uniting a diversity of minimization methods. Our expectation is that the proposed inversion strategy enables one to create a flexible inversion algorithm that can be easily upgraded with new developments in forward modeling and/or numerical recipes.

At the same time, the ability to model radiance with available codes and to implement numerical inversions does not reduce the design of Sun-sky radiance inversion codes to a purely technical procedure. There are many small and specific questions that need to be resolved to create an inversion procedure that is efficient in practice. Definitively, the key question in inversion algorithm development is quantifying the a priori constraints (defining Lagrange multipliers, formulating smoothing matrices, etc.) In addition, the forward model may also require some adjustments. For instance, numerical inversion of (17)–(19) uses vectors of aerosol parameters, whereas the forward models (equations (1) and (6)–(8)) operate on continuous functions. Correspondingly, the vectors with a reasonable number of components should replace functions traditionally used in modeling. In this section we proceed with the detailed design of a Sun-sky radiance inversion algorithm, using the principles described in sections 2 and 3.

4.1. Adaptation of Forward Model to the Inversion

The scheme of numerical inversion given by (17)–(19) requires extensive forward calculations. Namely, each p step requires recalculation of fitted characteristics $\mathbf{f}(\mathbf{a})$ and Jacobi matrices $\mathbf{U}$ in the case of nonlinear dependence $\mathbf{f} = \mathbf{f}(\mathbf{a})$. Accordingly, the adoption of a fast technique forward calculation is very important for making the inversion algorithm practical and efficient. Possible ways of accelerating and adjusting the forward model for inversion purposes are discussed below.

4.1.1. Optical thickness and phase function simulations. Equation (8) summarizes the modeling concept that relates optical properties of the atmosphere with the size distribution $(dN(r)/dr)$ and complex refractive index $(\tilde{m}(\lambda))$ of the aerosol particles, which are assumed to be homogeneous spheres. Both size distribution $(dN(r)/dr)$ and refractive index $(\tilde{m}(\lambda))$ will be the focus of the retrieval in the designed algorithm. The retrieval of the particle size distribution from the measurements of light scattered by polydispersions of spheres is a well-developed optical application. The concept of size distribution retrieval from single-scattering measurements is particularly clear for a case of known refractive index. The integral equation (equations (6) or (7)) can be reduced to a linear system, then solved by standard algebraic methods. In our case, the situation is more complicated, because the refractive index is unknown and the contribution of multiple scattering to sky radiance is significant in some instances. Nevertheless, in our algorithm, replacing integral equations (6) and (7) with linear systems is essential for making radiance simulations more rapid. Also, (6) and (7) are written for the size distribution of columnar aerosol particle number concentration; however, practical algorithms are often designed to retrieve the size distribution of surface area or volume of aerosol particles since light scattering of a small single particle is a function of particle surface area or volume [cf. *Bohren and Huffman*, 1983] rather than number concentration. Thus for flexibility of our algorithm, we transform (6) and (7) using different kinds of size distributions: number, radius, area and volume particle size distributions. Then, to meet calculation speed requirements, we reduce the integral equations to a linear system as follows:

$$\tau_{\text{scat}}(\lambda)P(\Theta;\lambda) = \int_{r_{\min}}^{r_{\max}} \frac{K_{\text{scat}}(\Theta;\lambda;\tilde{m};r)}{g_n(r)} x_n(\ln r)\, d\ln r \approx \mathbf{K}_{\text{scat}}(\Theta;\lambda;n;k)\mathbf{x}_n.$$

$$\tau_{\dots}(\lambda) = \int_{r_{\min}}^{r_{\max}} \frac{K_{\tau\dots}(\lambda;\tilde{m};r)}{g_n(r)} x_n(\ln r)\, d\ln r \approx \mathbf{K}_{\tau\dots}(\lambda;n;k)\mathbf{x}_n. \quad (23)$$

Here the index n (n = 0, 1, 2, 3) denotes the type of distribution as follows:
for n = 0 (number)

$$x_0(\ln r) = \frac{dR^0(r)}{d\ln r} = r^0\frac{dN}{d\ln r} = \frac{dN}{d\ln r} \quad \text{(i.e., } g_0 = 1\text{)};$$

for n = 1 (radius)

$$x_1(\ln r) = \frac{dR^1(r)}{d\ln r} = r\frac{dN}{d\ln r} = \frac{dR}{d\ln r} \quad \text{(i.e., } g_1 = r\text{)};$$

for n = 2 (area) (24)

$$x_2(\ln r) = \frac{dR^2(r)}{d\ln r} = 2\pi r^2\frac{dN}{d\ln r} = \frac{dS}{d\ln r} \quad \text{(i.e., } g_2 = 2\pi r^2\text{)};$$

for n = 3 (volume)

$$x_3(\ln r) = \frac{dR^3(r)}{d\ln r} = \frac{4}{3}\pi r^3\frac{dN}{d\ln r} = \frac{dV}{d\ln r} \quad \text{(i.e., } g_3 = 4/3\pi r^3\text{)}.$$

The kernel functions of optical thickness $K_{\tau\dots}(\dots)$ and differential scattering coefficient $K_{\text{scat}}(\dots)$ are approximated in (23) and (24) by matrices $\mathbf{K}_{\tau\dots}(\dots)$ and $\mathbf{K}_{\text{scat}}(\dots)$. The vector $\mathbf{x}_n$ approximates size distribution $dR^n(r)/d\ln r$ by N_r elements corresponding to the points $\{\mathbf{x}_k\}_i = dR^n(r_i)/d\ln r$ chosen with equal step $\Delta\ln r = \ln r_{i+1} - \ln r_i$ = const. The calculations of the matrices $\mathbf{K}_{\tau\dots}(\dots)$ and $\mathbf{K}_{\text{scat}}(\dots)$ in our algorithm are implemented using two different ways of interpolating size distribution values between grid points r_i. First, the size distribution $dR^n(r)/d\ln r$ between points $\ln(r_i) - (\Delta\ln r)/2$ and $\ln(r_i) + (\Delta\ln r)/2$ can simply be assumed to be equal to $dR^n(r_i)/d\ln r$; that is, elements of the matrices are computed as

$$\{\mathbf{K}_{\dots}(\dots)\}_{ji} = \int_{\ln(r_i)-\Delta\ln r/2}^{\ln(r_i)+\Delta\ln r/2} \frac{K_{\dots}(\dots;r)}{g_n(r)}\, d\ln r. \quad (25a)$$

The trapezoidal approximation is another way of interpolating between points. In this case, the size distribution is approximated between grid points $\ln(r_{i+1})$ and $\ln(r_i)$ linearly by $dR^n(r)/d\ln r = a\ln r + b$, where a and b must be chosen to coincide with values $dR^n(r_{i+1})/d\ln r$ and $dR^n(r_i)/d\ln r$. The matrix elements for this case are computed according to *Twomey* [1977] as

$$\{\mathbf{K}_{\dots}(\dots)\}_{ji} = \int_{\ln(r_i)}^{\ln(r_{i+1})} \frac{\ln(r_{i+1})-\ln r}{\ln(r_{i+1})-\ln(r_i)} \frac{K_{\dots}(\dots;r)}{g_n(r)} d\ln r + \int_{\ln(r_{i-1})}^{\ln(r_i)} \frac{\ln r-\ln(r_{i-1})}{\ln(r_i)-\ln(r_{i-1})} \frac{K_{\dots}(\dots;r)}{g_n(r)} d\ln r. \quad (25b)$$

The index j in (25a) and (25b) relates to matrix elements with Sun radiance at different wavelengths and sky radiance at different wavelengths and angles.

The dependence of matrices $\mathbf{K}_{\tau\dots}(\dots)$ and $\mathbf{K}_{\text{scat}}(\dots)$ on the real n and imaginary k part of the refractive index are approximated from look-up tables over all possible n and k values. Namely, we compute matrices in N_n and N_k grid points, which cover the whole range of expected values. The matrices for the values of n and k between these grid points are computed using linear interpolation in a logarithmic scale.

It should be noted that in (23)–(25) the size distributions are written in the logarithmic scale ($dR^n(r)/d\ln r$) instead of the linear scale ($dN(r)/dr$) used in (6) and (7). This is because the kernel functions $K_{\dots}(\dots)$ show much smoother variability for equal relative steps $\Delta r/r$ (i.e., for equal logarithmic steps, since $dr/r = d\ln r$) than for equal absolute steps Δr. Correspondingly, the logarithmic scale is commonly preferred for viewing optically important details of the particle size distributions and for making faster integration over particle size.

According to (25a) and (25b) the elements of the kernel matrices $\mathbf{K}_{\dots}(\dots)$ are integrals of kernel functions over particle size. Such integration can be time consuming. Matrix approximations (equations (23) and (24)) are efficient in practice, because they allow prompt calculation of optical thickness $\tau_{\dots}$ (extinction and absorption optical thickness) and differential scattering coefficient $\tau_{\text{scat}}(\lambda)P(\Theta;\lambda)$, given a n vector of size distribution $\mathbf{x}_k$ and refractive index $\tilde{m}(\lambda)$.

All of the above mentioned approximations produce some error, even in so-called "error-free" conditions. According to our estimations (for N_r = 22 in the range 0.05 μm $\le r \le$ 15 μm; $N_n = N_k$ = 15 in the ranges $1.33 \le n \le 1.6$ and $0.0005 \le k \le 0.5$) these errors can be considered as relative random errors with variance less than 1% for the typical aerosol models given by *Tanré et al.* [1999]. For significantly narrower size distributions (which are rather unlikely for atmospheric aerosols) this error may increase to 2–3%.

4.1.2. Simulations of radiative transfer in the atmosphere. As it was mentioned in section 2.1, we have employed a scalar discrete ordinates radiative transfer code to simulate diffuse radiance $I(\Theta;\lambda)$ in the plane-parallel atmosphere approximation. To make possible internal checks of the algorithm, we adopted two independent radiative transfer codes, one by *Nakajima and Tanaka* [1988] and the other by *Stamnes et al.* [1988]. However, for practical reasons we mainly used the program by *Nakajima and Tanaka* [1988], since it employs a truncation approximation that allows fast and accurate calculation of downwelling radiance in the aureole angular range with a relatively small number of Gaussian quadrature points. At the same time, it should be noted that we use radiative transfer codes only for modeling fitted characteristics **f(a)**. Jacobi matrices $\mathbf{U}_{k,\mathbf{a}}$ of Sun/sky radiance derivatives are calculated in the single-scattering approximation, i.e., for k = 1, 2:

$$\mathbf{U}_{k,\mathbf{a}} \approx \mathbf{U}_{k,\mathbf{a}}\,(\text{single scattering}). \quad (26)$$

The elements of these matrices can be easily calculated from (1a) and (1b), assuming $G(\dots)$ is equal to zero. Our retrieval experience shows that neglecting multiple scattering in simulating first derivatives does not particularly affect the retrieval results.

Thus using (23)–(25), the aerosol optical thickness $\tau_{\dots}(\lambda)$, single-scattering albedo $\omega_0(\lambda) = \tau_{\text{scat}}(\lambda)/\tau_{\text{ext}}(\lambda)$, and phase function ($P(\Theta,\lambda)$) are generated from the refractive index $\tilde{m}(\lambda) = n(\lambda) - ik(\lambda)$ and the size distribution of aerosol

particles $dR^n(r)/d \ln r$ in the total atmospheric column. These aerosol characteristics weighted (as given by (2)–(4)) with molecular scattering and gas absorption compose a set of atmospheric layer optical characteristics, which are necessary for radiative transfer computations.

Regarding vertical variability of the atmosphere, we consider two approximations in our algorithm: (1) an atmosphere with vertically homogeneous optical properties and (2) an atmosphere with a known vertical profile of aerosol extinction coefficient. For the case of a vertically homogeneous atmosphere the optical thickness of molecular scattering and gas absorption are calculated as described by *Holben et al.* [1998]. If the vertical profile of the aerosol extinction coefficient is available, the radiative transfer calculations can be performed for a multilayered atmosphere. Therewith the profiles of water vapor and ozone absorption together with climatological profiles of temperature and pressure (for molecular scattering calculations) are required. However, we can hardly count on vertical distribution information of aerosol complex refractive index, single-scattering albedo, and shape of the particle size distribution. Therefore these optical characteristics are assumed to be constant for the aerosol in the whole atmospheric column.

We focus our attention on the simplest model of a homogeneous atmosphere. This is because information on aerosol vertical profiles is not currently available for the majority of AERONET Sun/sky radiometer locations. In addition, the effect of aerosol vertical variability on sky radiance ground measurements can often be neglected, because it is rather modest in comparison with effects caused by aerosol size distribution variability. In addition, to minimize possible retrieval uncertainty due to the assumption of a homogeneous atmosphere, we concentrate our analysis on inverting sky radiances measured in the solar almucantar (equation (1b)). In observations with such a scheme (zenith angle of observations is equal to the solar zenith angle) all atmospheric layers are always viewed with similar geometry. Accordingly, at least in single-scattering approximation, sky radiances in the solar almucantar are not sensitive to aerosol vertical variations.

4.2. Inversion Implementation

Implementing the inversion strategy (section 3) in a practical retrieval requires defining a number of values and parameters. First, the error statistics of Sun and sky radiance measurements should be quantified for incorporating covariance matrices in the inversion algorithm. Second, using a priori constraints should be clarified: what kind of a priori constraints should be used, and what values of the corresponding Lagrange multiplier are appropriate.

4.2.1. Measurement error statistics. The magnitudes of direct and diffuse radiance are very different and the sensors that measure them are different and use different calibration techniques. Therefore the values of errors in Sun and sky radiance measurements are also rather different. Correspondingly, in a retrieval algorithm we consider Sun and sky radiance measurements as two separate groups:

$$\begin{cases} I^*(\theta; \lambda) = I(\theta; dR^n(r)/d \ln r; n(\lambda); k(\lambda)) + \Delta_I(\theta; \lambda) \\ \tau^*(\lambda) = \tau(dR^n(r)/d \ln r; n(\lambda); k(\lambda)) + \Delta_\tau(\lambda) \end{cases} \quad (27)$$

In (27) and everywhere that follows, we consider spectral aerosol extinction optical thickness $\tau^*(\lambda)$ instead of Sun radiance as an initial data set for the retrieval. This is because aerosol extinction optical thickness is one of the standard products derived from AERONET Sun photometer measurements (since the instrument output is calibrated to retrieve $\tau^*(\lambda)$ rather than the absolute radiance) and operating with $\tau^*(\lambda)$ helps us to use both the extensive experience regarding the accuracy of AERONET-derived aerosol optical thickness and the existing knowledge of $\tau^*(\lambda)$ variability for atmospheric aerosol. Thus the two basic data sets in (9) correspond to sky radiance measurements ($k = 1$) and spectral aerosol optical thickness ($k = 2$). However, to define the elements of both the fitted vectors $\mathbf{f}_k$ and the vectors of the unknowns (including size distribution and complex refractive index), we need to outline the alternatives, namely, operating with logarithms or absolute values.

4.2.1.1. Logarithmic transformation (nonnegativity assumption): Retrieval of logarithms of physical characteristics, instead of their absolute values is an obvious way to avoid retrieval of negative values for fundamentally positive parameters (such as $dR^n(r_i)/d \ln r$). However, the literature devoted to inversion techniques tends to consider this apparently useful tactic as an artificial trick rather than a scientific technique to optimize solutions. Such misconception is probably caused by the fact that the pioneering efforts on inversion optimization by *Phillips* [1962], *Twomey* [1963], and *Tikhonov* [1963] were devoted to overcoming the difficulties in solving the Fredholm integral equation of the first kind, i.e., a linear system produced by quadrature. Problems of that nature involve the retrieval of size distribution by inverting spectrally dependent optical thickness (equation (23)) or by inverting angularily dependent sky radiance. Considering $\tau_{\ldots}(\lambda)$ and $\tau_{scat}(\lambda) P(\Theta; \lambda)$ as functions of the logarithm of the size distribution $\ln x_k(\ln r)$ (i.e., $d \ln R^n(r)/d \ln r$) instead of $x_n(\ln r)$ requires replacing initially linear equations (23) by nonlinear ones. On the face of it, such a transformation of linear problems to nonlinear ones is difficult to accept as an optimization. On the other hand, in cases when a forward model is a nonlinear function of parameters to be retrieved (e.g., atmospheric profiling), the retrieval of logarithms is more likely to be the logical approach.

In our studies we follow the concept proposed in earlier papers [e.g., *Dubovik et al.*, 1995]. According to that concept, using logarithms of measured and retrieved characteristics in the retrievals is often expedient due to both rigorous statistical considerations and practical experience. It is well known that the curve of the normal distribution is symmetrical. In other words, the assumption of a normal PDF necessarily implies the possibility of negative results arising even in the case of physically nonnegative values (e.g., intensities, fluxes, etc.). For nonnegative characteristics ($\tau^*(\lambda)$ and $I(\Theta; \lambda)$ in our studies) the choice of the lognormal distribution for the measurement noise (i.e., $\{\mathbf{f}_1\}_j = \ln I(\Theta_{j1}; \lambda_{j2})$ and $\{\mathbf{f}_2\}_j = \ln \tau^*(\lambda_{j2})$) seems more reasonable due to the following considerations: (1) lognormally distributed values $I(\Theta_{j1}; \lambda_{j2})$ and $\tau^*(\lambda_{j2})$ are positively defined, and (2) there are a number of theoretical and experimental reasons showing that for positively defined characteristics the lognormal curve (multiplicative errors, see *Edie et al.* [1971]) is closer to reality than normal noise (additive errors) (a statistical discussion can be found in the work of *Tarantola* [1987]). As well, the use of the lognormal PDF for noise optimization does not require any revision of normal concepts and can be implemented by simple transformation of the problem to the space of normally distributed logarithms.

A similar situation is found for retrieving logarithms of positively defined unknowns (e.g., $x_n(\ln r)$ in (23) and (24)) in-

stead of their absolute values. In fact, according to the statistical estimation theory, LSM estimates $\hat{\mathbf{a}}$ (obtained under the assumption of normal PDF) are also normally distributed. It is obvious, even without rigorous statistical consideration, that for nonnegative $x_i = x_n(\ln r_i)$, this statement can be applied only approximately, because the normal distribution cannot provide zero probability for $x_i < 0$. On the other hand, the retrieval of $\ln x_i$ instead of x_i illuminates the above contradiction, because the normal distribution of $\ln \hat{x}_i$ is a reasonable expectation for positively defined x_i.

Moreover, the analysis by *Dubovik et al.* [1995] has shown that the logarithmic transformation can be considered as one of the cornerstones of the practical efficiency of Chahine-like iterative procedures. These techniques are popular in atmospheric research in spite of the fact that they solve linear systems (equations (23) and (24)) by means of nonlinear iterations. In Appendices C and D we show that logarithm-based retrievals lead to the methods of *Chahine* [1968] and *Twomey* [1975]. The mathematical treatments given in Appendices C and D show the close relation of Chahine-like techniques to the steepest descent method (equation (22b)).

It should be noted that in many situations, retrieval of absolute values or their logarithms is practically similar. This is because narrow lognormal or normal noise distributions are almost equivalent. For example, for small variations of a nonnegative value of a the following relationship between Δa and $\Delta a/a$ is valid:

$$\Delta \ln a = \ln (a + \Delta a) - \ln (a) \approx \frac{\Delta a}{a}, \quad \text{if } \Delta \ln a \ll 1. \tag{28a}$$

Correspondingly, if only small relative variations of the value of a are allowed, the normal distribution of $\Delta \ln \mathbf{a}$ is almost equivalent to the normal distribution of absolute values $\Delta \mathbf{a}$. The covariance of these normal distributions are connected as follows:

$$\mathbf{C}_{\ln \mathbf{a}} \approx (\mathbf{1}_\mathbf{a})^{-1} \mathbf{C}_\mathbf{a} (\mathbf{1}_\mathbf{a})^{-1}, \tag{28b}$$

where $\mathbf{1}_\mathbf{a}$ is diagonal matrix with the elements $\{\mathbf{1}_\mathbf{a}\}_{ii} = a_i$.

To make our inversion algorithm flexible we allow two possibilities in its implementation, namely, using (1) absolute values or (2) logarithms for both measured characteristics (sky radiance and optical thickness) and retrieved parameters (size distribution, real and imaginary parts of complex refractive index). However, everywhere below, we focus our discussion on operating with logarithms. This is because all considered characteristics (both measured and retrieved) are positively defined. In addition, by using logarithms it is simple to operate simultaneously with characteristics that have different units and values varying over a wide range of magnitude. Thus the vectors of measurements are defined as follows:

$$\{\mathbf{f}_1^*\}_j = \ln I^*(\Theta_{j_1}; \lambda_{j_2}) \text{ and } \{\mathbf{f}_2^*\}_j = \ln \tau^*(\lambda_{j_2}). \tag{29a}$$

The vector $\mathbf{a}$ of unknowns unites the parameters of size distribution and complex refractive index as

$$\begin{aligned} \{\mathbf{a}\}_i &= \ln x_n(\ln r_{i_1}) && \text{for } i = 1, \ldots, N_r, \\ \{\mathbf{a}\}_i &= \ln n(\lambda_{i_2}) && \text{for } i = N_r + 1, \ldots, N_r + N_\lambda, \\ \{\mathbf{a}\}_i &= \ln k(\lambda_{i_2}) && \text{for } i = N_r + N_\lambda + 1, \ldots, N_r + 2N_\lambda, \end{aligned} \tag{29b}$$

where N_r is number of points used for the retrieval of size distribution, and N_λ is the number of wavelengths.

4.2.1.2. Weight matrices of measurement data sets: We consider a set of sky radiance measurements $I^*(\Theta; \lambda)$ as a critical piece of information that is absolutely necessary for the retrieval of size distribution and complex refractive index. Therefore we have assigned $k = 1$ (i.e., vector $\mathbf{f}_1^*$) in (29) and the Lagrange multipliers of all other data sets ($\mathbf{f}_k$, $k > 1$) according to (15) should be defined by rating the variance of corresponding errors to the variance of the errors in sky radiance. Hence the central question in the algorithm design is the comparison of errors in other data sets to sky radiance errors. Another question relates to the presence of error correlation for each set. In other words, should weight matrices $\mathbf{W}_k$ be assumed diagonal (no correlation) or nondiagonal (there is correlation). At present, we are not aware of any clear correlation between random errors in measurements of radiance at different wavelengths or angles. Therefore in our current study we consider the simplest case of diagonal weight matrices; that is, $\{\mathbf{W}_k\}_{j \neq i} = 0$. The diagonal elements of weight matrices reflect the spectral and angular changes of instrumental signal/noise ratio of atmospheric radiance.

The accuracy of sky channel radiance measurements is maintained by calibration of the sky radiometer with an integrating sphere radiance source at the level of 5% or better for all wavelengths [*Holben et al.*, 1998]. Therefore we assume the same 5% accuracy of sky radiance measurements for all wavelengths and angles of observation, independent of the magnitude of the sky radiance signal (i.e., relative accuracy is a constant). According to (28a), relative errors are approximately equal to logarithmic errors, i.e., for logarithms of measurements (equation (29)); that is, $\varepsilon_1 \sim \Delta_{\ln I(\Theta;\lambda)} \approx 0.05$ and the weight matrix is equal to unit matrix $\mathbf{W}_1 = \mathbf{1}$ (where $\mathbf{1}$ has diagonal elements equal to 1).

The calibration procedure of the Sun channels is expected to reduce the absolute uncertainty in $\tau(\lambda)$ to the level of about ± 0.01 for $\lambda \geq 440$ nm and ± 0.02 for $\lambda < 440$ nm wavelength dependence [*Holben et al.*, 1998; *Eck et al.*, 1999]. The studies by *Schmid et al.* [1999] have shown good agreement to this expected accuracy for an AERONET instrument in field experiment conditions. Thus we estimate $\tau(\lambda)$ with an absolute confidence interval of ± 0.01 for the wavelengths used in the retrieval which are all ≥ 440 nm. For the simplicity of further consideration we neglect any minor wavelength dependence. Correspondingly, relative error changes with $\tau(\lambda)$ and the value of the logarithmic error $\Delta \ln \tau(\lambda)$ depends on the magnitude of optical thickness. Indeed, applying (28a), we can use $0.01 = \Delta\tau(\lambda) \approx \tau(\lambda)\, \Delta \ln \tau(\lambda)$; that is, $\Delta \ln \tau(\lambda) \approx 0.01/\tau(\lambda)$. Thus to define the weight matrix $\mathbf{W}_2$, we normalize the covariance matrix of $\Delta \ln \tau(\lambda)$ by the variance of optical thickness logarithmic error at 440 nm ($\varepsilon_2^2 \sim \Delta^2_{\ln \tau(440)} \approx (0.01/\tau(440))^2$) and obtain the following diagonal elements (equation (14)):

$$\{\mathbf{W}_2\}_{jj} = (\tau(440)/\tau(\lambda_j))^2. \tag{30}$$

4.2.1.3. Values of the Lagrange multiplier: In the literature devoted to inversion techniques [e.g., *Twomey*, 1977; *Tikhonov and Arsenin*, 1977; *Tarantola*, 1987] the Lagrange multiplier is defined as a nonnegative multiplier that serves to weight the contribution of a priori (smoothness) constraints, relative to the contribution of the measurements. The value of this contribution is usually evaluated by correspondent sensitivity studies [*King*, 1982].

In our investigations we pursued a statistical optimization approach that defined the optimum inversion of multisource data as a minimization of the multiterm quadratic form given

by (13). This approach does not make any distinction between measured and a priori characteristics except for the different accuracy of each data set. The contribution of each data set is weighted by corresponding parameter γ_k related to the contribution of the basic set of measurements (i.e., for basic data set: $\gamma_1 = 1$). Hence we assign parameter γ_k for every set of a priori or measurement data and call this γ_k a Lagrange multiplier following the traditional terminology. The value of each Lagrange multiplier is clearly defined by (15) as a ratio of error variances. However, the practical choice of Lagrange multipliers is a challenging task, because accurate values of error variances are not typically available in practice. Nevertheless, (15) is very helpful in evaluating the expected range of γ_k values.

In our implementation the relative impacts of sky radiance and optical thickness measurements on the retrieval result are assumed to be comparable. Therefore in designing the current algorithm we focus especially on the control of fitting errors of both sky radiance and spectral optical thickness measurements. Namely, we anticipate that a successful retrieval should simultaneously satisfy the following criteria for all k:

$$\Psi_k \leq N_k \sigma_k^2, \tag{31}$$

where N_k is the number of values in the fitted vector $\mathbf{f}_k^*$, and σ_k is the measurement accuracy. Correspondingly, the values of Ψ_k, and the contribution of each term Ψ_k in the total value of Ψ, directly depend on N_k. However, this dependence on the number of measurements is not appropriate in practice, because a simple increase of N_k may lead to an increasing number of redundant measurements without an increase of information content. Therefore our strategy of combining data $\mathbf{f}_k^*$ is to consider sky radiance and optical thickness data sets as two critical pieces of information, and the importance of each piece of information is independent of the numbers of measurement N_k. Hence on the basis of our criteria given by (31) we define the Lagrange multipliers γ_k (for the measurements; that is, $k = 1, 2$) as the following function of the numbers of measurements:

$$\gamma_k = \frac{N_1 \sigma_1^2}{N_k \sigma_k^2} - \frac{N_1}{N_k} \gamma_k'. \tag{32}$$

Obviously, this definition of γ_k forces equal values of $\gamma_k \Psi_k$ and makes reasonable the consideration of parameters γ_k' (instead of γ_k), because of their independence of N_k in each data set. It should be emphasized that defining (32) is practically equivalent to the assumption that the expected accuracy σ_k of a single measurement is related to the uncertainty ε_k of the data set, which includes N_k measurements of radiance, as follows:

$$\varepsilon_k^2 = N_k \sigma_k^2. \tag{33}$$

This relationship assumes that ε_k increases with the number of measurements as $\sqrt{N_k}$. Such a result can be caused by the fact that the number of various random error sources may increase (as $\sqrt{N_k}$) with the increase of measurement number. For example, the increase of angular and spectral resolution of radiance measurements requires a longer measurement time resulting in an increase of errors due to natural temporal variability of sky radiance.

Thus the Lagrange multipliers γ_1' and γ_2' can be defined as follows: Obviously, γ_1' is always equal to unity and is included in (13) for identity in formulation of all terms. The multiplier γ_2' in our algorithm is the ratio of variances of sky radiance and optical thickness measurement errors and according to our assumptions about these errors ($\sigma_1 \approx \Delta_{\ln I(\Theta;\lambda)} \approx 0.05$ and $\sigma_2 \approx \Delta_{\ln \tau(440)} \approx 0.01/\tau(440)$), the value of γ_2' is the following from (32):

$$\gamma_2' \approx 25(\tau(440))^2. \tag{34}$$

It should be noted that the values used for σ_1 and σ_2 are rather approximate. Also, the correctness of the assumption in (33) needs validation (e.g., it may not work for rather small N_k). Therefore we consider (34) as an estimation of γ_2' that needs further verification.

4.2.2. A priori constraints. The retrieval of the aerosol size distribution from measurements of scattered light belongs to the class of so-called ill-posed inverse problems. Ill-posed problems tend to have an unstable nonunique solution, and using a priori constraints is essential for solving such problems successfully [e.g., *Tikhonov and Arsenin*, 1977]. Applying smoothness constraints on the variability of the size distribution (or other retrieved characteristics) is well established and a commonly accepted technique for eliminating unrealistic oscillations in the retrievals. *Twomey* [1977] gives the basic principles of solution smoothing for optical and remote sensing applications. In our algorithm we retrieve several functions (particle size distribution and complex refractive index) requiring different a priori constraints. The purpose of the current subsection is to introduce the specific limitations on retrieved $dR^n/d \ln r$, $n(\lambda)$, and $k(\lambda)$ by defining vectors $\mathbf{f}_{k>2}$ in (9).

We apply two basic methods of constraining the solution. The first method constrains the solution by a sample solution $\mathbf{a}^*$. This constraint has been proposed by *Twomey* [1963] and expanded in the scope of the statistical approach by *Strand and Westwater* [1968]. *Rodgers* [1976, 1990] accomplished further development and application of this method in atmospheric remote sensing applications. The second method constrains only the differences between elements of vector $\hat{\mathbf{a}}$ and does not restrict their values. In another words, this method applies pure smoothness constraints to eliminate only strong oscillations in the retrieved characteristics. *Twomey* [1977] and *Tikhonov and Arsenin* [1977] give the basic techniques of implementing such smoothing. This type of smoothing is commonly used in aerosol optical properties retrievals [e.g., *King et al.*, 1978; *Shaw*, 1979; *King*, 1982; *Nakajima et al.*, 1983, 1996; *Spinhirne and King*, 1995; *Dubovik et al.*, 1995].

4.2.2.1. Constraining the solution by a priori estimates: The most straightforward method of eliminating unrealistic values in the solution $\hat{\mathbf{a}}$ is to use an a priori estimate of the solution $\mathbf{a}^*$ (in another words, virtual measurements of retrieved characteristics). For example, the climatological data of $dR^n/d \ln r$, $n(\lambda)$, and $k(\lambda)$ (or of $d \ln R^n/d \ln r$, $\ln (n(\lambda))$, and $\ln (k(\lambda))$, if the lognormal statistic is applied), can be considered as a priori estimates. In this case, the kth equation system can simply be defined as

$$\mathbf{a}^* = \mathbf{a} + \Delta_{\mathbf{a}^*}, \tag{35}$$

where $\Delta_{\mathbf{a}^*}$ denotes the error in a priori estimates (climatological data) $\mathbf{a}^*$. Defining the covariance matrix $\mathbf{C}_{\mathbf{a}}$ of errors $\Delta_{\mathbf{a}^*}$ in an a priori estimate $\mathbf{a}^*$ should not be a problem (at least for climatological data). Since (35) is very simple, incorporating (35) into (17)–(19) is rather transparent, and we will not discuss it (for details, see *Dubovik et al.* [1995]).

In our algorithm we include the option of employing an a

priori estimate only for restricting the values of the real part of the refractive index. This is done for two reasons. First, the range of $n(\lambda)$ for aerosol is limited (for example, *Tanré et al.* [1999] give values of $n(\lambda)$ within a range of 1.40–1.55 for 470–1240 nm). Second, the information content of atmospheric radiance measured by AERONET is not sufficient for accurate retrieval of $n(\lambda)$ in some situations (for discussion, see *Dubovik et al.* [2000]). Nevertheless, we will not employ an a priori estimate of $n(\lambda)$ because we focus our efforts on the situation where $n(\lambda)$ can be retrieved without forcing retrievals by an a priori range of values of $n(\lambda)$. Examples of using a priori estimates of $n(\lambda)$ can be found in the paper by *Romanov et al.* [1999], where a similar approach has been applied. It should be noted that in contrast with a simple fixing of the refractive index the use of a priori estimates of refractive index through (35) gives some freedom (depending on corresponding γ_k) to obtain a refractive index different from its a priori value.

4.2.2.2. Smoothness constraints of the solution: For smoothing the solution the norm of the mth logarithmic derivatives of the retrieved characteristics $y(z)$ are restricted:

$$a_m = \int \left(\frac{d^m y(z)}{d^m z}\right)^2 dz \approx \sum_{i=1}^{n} \left(\frac{\Delta^m y(z_i)}{(\Delta z)^m}\right)^2 \Delta z$$

$$= (\Delta z)^{-2m+1} \mathbf{y}^T (\mathbf{S}_m)^T (\mathbf{S}_m)\mathbf{y}, \tag{36a}$$

where $\Delta^m y(z_i)$ denote mth differences ($m = 1, 2, 3, \ldots$), which are defined as

$$\Delta^1 y(z_i) = y(z_i) - y(z_{i+1}) = y_i - y_{i+1};$$

$$\Delta^2 y(z_i) = \Delta^1 y(z_i) - \Delta^1 y(z_{i+1}) = y_i - 2y_{i+1} + y_{i+2}; \tag{36b}$$

$$\Delta^3 y(z_i) = \Delta^2 y(z_i) - \Delta^2 y(z_{i+1}) = y_i - 3y_{i+1} + 3y_{i+2} - y_{i+3}.$$

Matrix $\mathbf{S}_m$ contains the coefficients for calculating vector $\mathbf{d}^m$ (with elements $\{\mathbf{d}^m\}_i = \Delta^m y(z_i)$) of mth differences of $y(x)$:

$$\mathbf{d}^m = \mathbf{S}_m \mathbf{y}, \tag{37}$$

where $y(z)$ is replaced by correspondent vectors $\mathbf{x}$ (with elements $z_i = z_1 + (i-1)\Delta_x$, $\Delta_x = \text{const}$) and $\mathbf{y}$ (with elements $y_i = y(z_i)$). For example, the matrix of second differences is given by

$$\mathbf{S}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \ldots & & & & \\ 0 & 1 & -2 & 1 & 0 & \ldots & & & \\ 0 & 0 & 1 & -2 & 1 & 0 & \ldots & & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 1 & -2 & 1 \end{pmatrix}. \tag{38}$$

This matrix is most commonly used for aerosol size distribution retrieval [cf. *Twomey*, 1977; *King et al.*, 1978; *King*, 1982; *Nakajima et al.*, 1996]. In our studies we are following the approach of *Dubovik et al.* [1995] and define smoothing constraints statistically in the form of the corresponding vector equation given by (9). In particular, we know that the mth derivatives of function $y(x)$ are limited and approach zero as the degree m increases: $\mathbf{d}^m = \mathbf{0} - \Delta_{dm}$ (Δ_{dm} defines deviations from zero). Correspondingly, we can write

$$\mathbf{0} = \mathbf{S}_m \mathbf{y} + \Delta_{dm}. \tag{39}$$

Using this equation in the case when $\mathbf{0} = \Delta_{dm}$ defines function $y(x)$ as follows (constant, straight line, parabola, etc.):

$$0 = \frac{dy(z)}{dz} \Rightarrow 0 = Y_1(z) = C,$$

$$0 = \frac{d^2 y(z)}{dz^2} \Rightarrow 0 = Y_2(z) = Bz + C, \tag{40}$$

$$0 = \frac{d^3 y(z)}{dz^3} \Rightarrow 0 = Y_3(z) = Az^2 + Bz + C,$$

where A, B, C are arbitrary constants.

Equations (39) and (40) are helpful for qualitative and, together with (36), quantitative evaluations of the effects of introducing smoothness constraints into retrieval algorithms. For example, we can always assume that we have a set of virtual measurements of mth degree derivatives of an unknown characteristic $y(z)$ (e.g., particle size distribution). The variance of the errors in such a data set can be easily estimated as follows:

$$\varepsilon_{dm}^2 = \langle (y_i - Y_m(z_i))^2 \rangle = \langle a_m \rangle (\Delta z)^{2m-1} \le a_m^{\max} (\Delta z)^{2m-1}, \tag{41}$$

where $Y_m(z)$ is given by (40). The value $a_m^{\max}$ is the maximum possible norm of the mth order derivatives of $y(x)$ and can be calculated according to (36) for most $y(z)$. For example, below we will estimate $a_m^{\max}$ for the aerosol size distribution on the basis of climatological data.

The difference between our algorithm and other known aerosol retrieval algorithms [*King et al.*, 1978; *Nakajima et al.*, 1983] is that we are restricting several functions simultaneously ($y(z) = \ln x(\ln r)$; $y(z) = \ln n(\lambda)$ and $y(z) = \ln k(\lambda)$) and for each function different values of a_m ($k = 1, 2, 3$) and ε_k are to be defined. Indeed, admissible variations of the size distribution $\ln x(\ln r) = d \ln R^n / d \ln r$ are expected to be much stronger than for spectral variations of the real $n(\lambda)$ and imaginary $k(\lambda)$ parts of the refractive index. Therefore we should define smoothness vectors $\mathbf{f}_k$ (and correspondent Lagrange multipliers) in (9) separately for particle size distribution and real and imaginary parts of the index of refraction; that is, each smoothness vector $\mathbf{f}_k$ should depend only on the part of the retrieved vector $\mathbf{a}$ corresponding to $x(\ln r)$, $n(\lambda)$, or $k(\lambda)$. Thus we have three vectors $\mathbf{f}_k$ ($k = 3, 4, 5$) defining a priori constrains:

$$\left.\begin{aligned} \mathbf{f}_3^* &= \mathbf{f}_3(\mathbf{a}) + \Delta_3 = \mathbf{U}_3\mathbf{a} + \Delta_3 \\ \mathbf{f}_4^* &= \mathbf{f}_4(\mathbf{a}) + \Delta_4 = \mathbf{U}_4\mathbf{a} + \Delta_4 \Rightarrow \\ \mathbf{f}_5^* &= \mathbf{f}_5(\mathbf{a}) + \Delta_5 = \mathbf{U}_5\mathbf{a} + \Delta_5 \end{aligned}\right.$$

$$\begin{pmatrix} \mathbf{0}^* \\ \mathbf{0}^* \\ \mathbf{0}^* \end{pmatrix} = \begin{pmatrix} \mathbf{S}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_k \end{pmatrix} \begin{pmatrix} \mathbf{a}_x \\ \mathbf{a}_n \\ \mathbf{a}_k \end{pmatrix} + \begin{pmatrix} \Delta_x \\ \Delta_n \\ \Delta_k \end{pmatrix}, \tag{42a}$$

where vectors $\mathbf{a}_x$, $\mathbf{a}_n$ and $\mathbf{a}_k$ denote parts of the complete vector $\mathbf{a}$ (i.e., $\mathbf{a}^T = (\mathbf{a}_x, \mathbf{a}_n, \mathbf{a}_k)^T$), matrices $\mathbf{S}_x$, $\mathbf{S}_n$, and $\mathbf{S}_k$ denote matrices of the corresponding differences, vectors $\mathbf{0}^*$ and matrices $\mathbf{0}$ denote, correspondingly, vectors and matrices with zero elements. The corresponding weight matrices are defined as unit matrices:

$$\mathbf{W}_k = \mathbf{1} \qquad (k = 3, 4, 5). \tag{42b}$$

Finally, to complete the description of smoothness constraints in our algorithm (i.e., to define $\mathbf{S}_x$, $\mathbf{S}_n$, $\mathbf{S}_k$, γ_3, γ_4, and γ_5), we need to evaluate the required orders of the derivatives

Table 1. Results of Evaluating Smoothness (Norm of First, Second, and Third Differences) of Modeled Particle Size Distributions and Corresponding Lagrange Multipliers (γ_3) for a Priori Constraints

Aerosol Size Distributions			Smoothness Characteristics			
Number of Volume Equivalent Modes	Standard Deviation	Volume Mode Radii r_i (μm)	Type of Differences	Norm of Differences	Standard Deviation (ε_3)	Lagrange Multipliers (γ_3)
Calculations for Retrieving Logarithms: $a_i = \ln (dR^n(r_i)/d \ln r)$						
2	0.6	0.33; 5	$\Delta^1 a_i$	25	2.6	0.0004*
			$\Delta^2 a_i$	68	1.2	0.002
			$\Delta^3 a_i$	524	0.9	0.003
3	0.4	0.1; 0.75; 5	$\Delta^1 a_i$	43	3.5	0.0001*
			$\Delta^2 a_i$	320	2.5	0.0004
			$\Delta^3 a_i$	4200	2.5	0.0004
3	0.15	0.1; 0.75; 5	$\Delta^1 a_i$	3400	30	3.0^{-6}*
			$\Delta^2 a_i$	30000	25	4.0^{-6}
			$\Delta^3 a_i$	490000	27	3.5^{-6}
Calculations for Retrieving Absolute Values of Volume Particle Size Distribution: $a_i = dV(r_i)/d \ln r$						
2	0.6	0.33; 5	$\Delta^1 a_i$	0.04	0.10	0.23
			$\Delta^2 a_i$	0.37	0.09	0.34
			$\Delta^3 a_i$	4.46	0.08	0.38
3	0.4	0.1; 0.75; 5	$\Delta^1 a_i$	0.53	0.38	0.02
			$\Delta^2 a_i$	4.85	0.31	0.025
			$\Delta^3 a_i$	56.4	0.29	0.03
3	0.15	0.1; 0.75; 5	$\Delta^1 a_i$	1.31	0.60	0.007
			$\Delta^2 a_i$	46.4	0.96	0.003
			$\Delta^3 a_i$	1945	1.70	0.001

*The results are given for volume particle size distribution $a_i = \ln (dV(r_i)/d \ln r)$.

and norms $a\ldots$ for the retrieved functions ($y(z) = \ln x(\ln r)$; $y(z) = \ln n(\lambda)$ and $y(z) = \ln k(\lambda)$).

4.2.2.3. Smoothness of the particle size distribution: The particle size distribution of tropospheric aerosols may contain several distinct modes and each mode is most commonly modeled by a lognormal function [*Whitby*, 1978; *Remer et al.*, 1997, 1998; *Remer and Kaufman*, 1998]. The norms $a\ldots$ of (36) thus evaluated using

$$y(z) = \ln x_n(\ln r) = \frac{d \ln R^n(r)}{d \ln r}$$

$$= \ln\left[\sum_{j=1}^{J} \frac{C_{n,j}}{\sqrt{2\pi}\,\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{\ln r - \ln r_{n,j}}{\sigma_j}\right)^2\right)\right], \quad (43)$$

To evaluate $a_m^{\max}$, we should estimate a_k for the most variable function $\ln x_n(\ln r)$. For the particle size distribution given by (43) the norm of the derivatives would increase with increasing number J of the components and with decreasing standard deviation σ_j for each component. Accordingly, the size distribution with the largest number of narrow components has the greatest value of $a_m^{\max}$ (the smaller σ_j the narrower the function). Physical processes in the atmosphere most frequently result in a bimodal structure of the aerosol size distribution [*Remer and Kaufman*, 1998]. At the same time, the appearance of a third mode is also realistic. For example, a volcanic eruption may produce optically thick stratospheric aerosol, which adds a stable third additional mode to the commonly appearing accumulation mode (small particles; $r < 0.6$ μm) and coarse mode (large particles; $r > 0.6$ μm) composing tropospheric aerosol [*Kaufman and Holben*, 1996]. The standard deviation σ_j of the aerosol size distribution varies, depending on the type of aerosol and the atmospheric conditions. *Tanré et al.* [1999] give $\sigma = 0.4$ for the narrowest aerosol modes. In practice the size distributions can most likely be even narrower than size distributions corresponding to $\sigma = 0.4$. However, we cannot expect resolution smaller than the interval $\Delta \ln r = \ln r_{i+1} - \ln r_i$ chosen for defining the linear systems in (23)–(26); that is, particle size distribution should not be narrower than $\Delta \ln r$ ($\sigma = \Delta \ln r$).

Thus to estimate the maximum norm $a_m^{\max}$, we calculate the norm of the first, second and third derivatives for a trimodal lognormal size distribution for two cases $\sigma = 0.4$ and $\sigma = \Delta \ln r$. The results of these calculations are summarized in Table 1. Corresponding to these calculations, the values of the Lagrange multiplier (γ_3), obtained by means of (41) and (15) assuming $\varepsilon_1 = 0.05$, are found to be in the range 3.0e-6–3.0e-3. It should be noted that Table 1 contains the results of calculations for the size distribution of particle volume $dV/d \ln r$. However the values of a_m for the logarithmic differences of the second and greater order are the same for all distributions $dR^n/d \ln r$. This is because the differences $\Delta \ln (dR^n/d \ln r)$ of the second and greater order are independent of n. Thus the same smoothness restrictions can be used for the distributions of particle number, radius, area, and volume if these restrictions are applied to the logarithms $\ln x$ ($\ln r$).

4.2.2.4. Smoothness of spectral dependence of complex refractive index: To define the parameters $\gamma_4 = (\varepsilon_1/\varepsilon_n)^2$ and $\gamma_5 = (\varepsilon_1/\varepsilon_k)^2$, we need to evaluate derivative norms of spectral dependencies $y(z) = \ln n(\lambda)$ and $y(z) = \ln k(\lambda)$. Spectral variability is usually not expected for both real and imaginary parts of the aerosol particle refractive index. For example, the widely cited paper by *Shettle and Fenn* [1979] shows practically wavelength-independent complex refractive indices in the spectral interval of interest (440–1020 nm) for the materials typically composing atmospheric aerosols. Similarly, aerosol models by *Tanré et al.* [1999] assume single constant values of complex refractive index for the spectral interval considered. However, in the scientific literature there are multiple indications of possible spectral selectivity of the refractive index for aerosol particles [e.g., *Ackerman and Toon*, 1981; *Patterson and McMahon*, 1984; *Horvath*, 1993; *Dubovik et*

al., 1998b, *Yamasoe et al.*, 1998]. Therefore we constrain the spectral variability of the retrieved complex refractive index to some practically reasonable ranges rather than to any particular model of the atmospheric aerosol.

To analyze derivative values, we approximate spectral dependencies $n(\lambda)$ and $k(\lambda)$ by exponential functions in a manner similar to *Dubovik et al.* [1998b]:

$$n(\lambda) \sim (\lambda)^{-\alpha_n} \text{ and } k(\lambda) \sim (\lambda)^{-\alpha_k}. \quad (44)$$

Obviously, the logarithmic derivatives $[d^m \ln n(\lambda)]/(d^m \ln \lambda)$ and $[d^m \ln k(\lambda)]/(d^m \ln \lambda)$ are equal to zero for $m > 1$. Therefore we will be using first derivatives for constraining the spectral variability of the complex refractive index. The norms of the first derivatives $a_1^{\max}$ directly relate to exponents: $a_{1,n}^{\max} = \alpha_n^{\max}(\ln \lambda_{\max} - \ln \lambda_{\min})$ and $a_{1,k}^{\max} = \alpha_k^{\max}(\ln \lambda_{\max} - \ln \lambda_{\min})$. We estimate the maximum spectral dependence of the real part of the refractive index as $a_{1,n}^{\max} = 0.2$, which corresponds to change from $n(440 \text{ nm}) = 1.6$ to $n(1020 \text{ nm}) = 1.33$. The value of $a_{1,k}^{\max} = 1.5$, given by *Dubovik et al.* [1998b] for biomass burning aerosol, is accepted in our studies as an indicator of the strongest spectral variability of imaginary part of the complex refractive index ($k(440 \text{ nm}) = 0.04$ to $k(1020 \text{ nm}) = 0.011$). The values of corresponding Lagrange multipliers (γ_4 and γ_5) are given in Table 3.

It should be noted that the traditional smoothness matrices with elements given by integer numbers (e.g., the matrix given by (38)) cannot be applied for constraining the spectral dependence of the refractive index. This is because the spectral interval $\Delta\lambda_i$ is not constant in our application. For example, sky radiances are measured by AERONET Sun photometers at four wavelengths: 440, 670, 870, 1020 nm, i.e., $\Delta\lambda_i = \lambda_{i+1} - \lambda_i \neq \text{const}$. Correspondingly, we use smoothness matrices $\mathbf{S}_n$ and $\mathbf{S}_k$ in (42) which are constructed for numerical derivatives $\Delta y(z)/\Delta z$ rather than for differences $\Delta y(z)$; that is, the matrices $\mathbf{S}_n$ and $\mathbf{S}_k$ account for the $\Delta\lambda_i$ in differences with matrices given by (38).

The restriction of second derivatives also can be applied for the retrieval of the spectral dependence of refractive index. Such a restriction would constrain the refractive index spectral variability by exponential functions (44). However, it would not restrict the values α_n and α_k. This might be insufficient in practice, because limited information content of the Sun/sky radiance [*Dubovik et al.*, 2000] may result in retrieval of unrealistically strong spectral selectivity of the refractive index.

4.2.2.5. Convergence improvements: The procedure given by (17)–(19) should provide monotonic and fast convergence of the iterations to the minimum of the quadratic form $\Psi(\mathbf{a})$ (equation (13)). Equation (19a) contains terms (on both the right and the left sides of the equations) which limit the length of the correction vector $\Delta\mathbf{a}^p$ and help to provide monotonic convergence of minimization in a similar manner to the Levenberg-Marquardt method.

As was mentioned in section 3.1.1., we implement this correction by assuming a priori constraints on the step correction $\Delta\mathbf{a}^p$. We constrain the parts $\Delta\mathbf{a}_x^p$, $\Delta\mathbf{a}_n^p$ and $\Delta\mathbf{a}_k^p$ of the vector $\Delta\mathbf{a}^p$ differently. Namely, we assume $\mathbf{0}^* = \Delta\mathbf{a}^p + \Delta_{\Delta\mathbf{a}}$ with the weight matrix:

$$\mathbf{W}_{\Delta\mathbf{a}} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_n\mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & g_k\mathbf{1} \end{pmatrix}, \quad (45a)$$

where $\mathbf{1}$ is a unit matrix, $g_n = (\varepsilon_{\Delta n}/\varepsilon_{\Delta x})^2$ and $g_k = (\varepsilon_{\Delta k}/\varepsilon_{\Delta x})^2$. The variances $\varepsilon_{\Delta x}^2$, $\varepsilon_{\Delta n}^2$, and $\varepsilon_{\Delta k}^2$ can be estimated on the ranges of the variability of particle size distribution, real and imaginary parts of complex refractive index as follows:

$$\begin{aligned} \varepsilon_{\Delta,x} &= 0.5(\ln x_{\max} - \ln x_{\min}) \approx 2.5, \\ \varepsilon_{\Delta,n} &= 0.5(\ln n_{\max} - \ln n_{\min}) \approx 0.05, \\ \varepsilon_{\Delta,k} &= 0.5(\ln k_{\max} - \ln k_{\min}) \approx 1. \end{aligned} \quad (45b)$$

In this equation we considered the interval $[\ln a_{\max}, \ln a_{\min}]$ as 68% confidence interval $[\ln a + \varepsilon, \ln a - \varepsilon]$. We used the following considerations for choosing the maximum and minimum values. The realistic maximum values of $x = dV/d \ln r$ (the size distribution of the particle volume in the total atmospheric column) can be easily expected to be in the range from 0.005 $(\mu\text{m})^3/(\mu\text{m})^2$ to 0.5 $(\mu\text{m})^3/(\mu\text{m})^2$ [*Dubovik et al.*, 2000]. For the real and imaginary parts of the aerosol complex refractive index we assume variability ranges from 1.60 to 1.40 and 0.05 to 0.005, respectively (for the spectral range 440–1020 nm). Also, the values in (45) are rounded off to number multiples of 5. It should be noted that these ranges are only for restricting $\Delta\mathbf{a}^p$; that is, the correspondent values of $\Delta\mathbf{a}^p$ are not expected to be larger than the length of the above mentioned intervals. Obviously, after several iterations, even greater changes can be achieved.

The definition of the Lagrange multipliers $\gamma_{\Delta\mathbf{a}}$ is similar to the one given by (15) with the difference that instead of ε_1, we use its estimate $\hat{\varepsilon}(\mathbf{a}^p)$ obtained from the residual (equation (15)); that is,

$$\gamma_{\Delta a}(\mathbf{a}^p) = \frac{\hat{\varepsilon}^2(\mathbf{a}^p)}{\varepsilon_{\Delta,x}^2}, \quad (46)$$

where $\hat{\varepsilon}^2(\mathbf{a}^p) = 2\Psi(\mathbf{a}^p)/(N_f - N_a)$. According to this equation the value of the Lagrange multiplier $\gamma_{\Delta\mathbf{a}}$ decreases with decreasing quadratic form $\Psi(\mathbf{a}^p)$. We have chosen this definition because the linear approximation in the small vicinity of the solution $\mathbf{a}'$ produces rather accurate $\Delta\mathbf{a}^p$, and any restriction on the solution correction $\Delta\mathbf{a}^p$ is not needed. Moreover, it may also slow down the convergence of the iterative process.

Thus the a priori constraints on the correction $\Delta\mathbf{a}^p$ help to attain a monotonic and fast convergence. The restrictions are in effect when $\mathbf{a}^p$ is far from the solution and they weaken when approaching the solution $\mathbf{a}'$.

5. Summary and Illustrations

In sections 2 and 3 we described the concept of forward modeling and inversion strategy. Section 4 described the details of organizing the inversion algorithm for deriving aerosol optical properties from atmospheric radiance measurements by AERONET Sun-sky scanning radiometers. Two aspects were discussed: the forward model optimization from an inversion viewpoint and choosing the values of parameters required for setting up the inversion scheme. The purpose of section 5 is to summarize and illustrate the result of our algorithm development.

The strategy of our development was to make a flexible algorithm that can be easily adapted to different practical needs and that also can easily be upgraded by new developments in radiative transfer modeling and numerical recipes. The possibility of upgrading an algorithm is assumed in many modern codes and is generally more interesting for the developer than for the user. Therefore we will not discuss this aspect here. We will emphasize the flexibility in choosing a number of

Table 2. Assumed Models of Error in Sky Radiance and Optical Thickness Measurements and Corresponding Weight Matrices and Lagrange Parameters Adopted in the Retrieval Algorithm

		Error Model			
Type of Data	Error Expectations	Type of Noise Distribution	Weight Matrix ($\mathbf{W}_k$)	Standard Deviation (e_k)	Lagrange Multiplier (γ_k)
$I(\Theta; \lambda)$	$\frac{\Delta I(\Theta; \lambda)}{I(\Theta; \lambda)} \le 0.05$	lognormal-recommended $f_1(\Theta; \lambda) = \ln(I(\Theta; \lambda))$	$\mathbf{W}_1 = \mathbf{1}$	$\varepsilon_1 = 0.05$	$\gamma_1 = 1$
		normal-alternative $f_1 = I(\Theta; \lambda)$	$\{\mathbf{W}_2\}_{jj} = I^2(\Theta; \lambda_j)$	$\varepsilon_1 = 0.05$	$\gamma_1 = 1$
$\tau(\lambda)$	$\Delta\tau(\lambda) \le 0.01$	lognormal-recommended $f_2(\lambda) = \ln(\tau(\lambda))$	$\{\mathbf{W}_2\}_{jj} = (\tau(440)/\tau(\lambda_j))^2$	$\varepsilon_2 = 0.01/\tau(440)$	$\gamma_2 = \frac{N_1}{N_2} 25(\tau(440))^2$
		normal-alternative $f_2(\lambda) = \tau(\lambda)$	$\{\mathbf{W}_2\} = \mathbf{1}$	$\varepsilon_2 = 0.01$	$\gamma_2 = \frac{N_1}{N_2} 25$

alternatives for implementing the inversion so that the inversion scheme can be easily used with other radiative transfer schemes, or even in other applications. Correspondingly, we have tried to make the forward modeling and inversion parts of the algorithm as independent of each other as possible, and we have put a significant effort into making the inversion part of our algorithm rather transparent and changeable. Therefore below we will identify the possible alternatives in implementing the inversion and illustrate the resulting differences.

5.1. Proposed Algorithm and Alternative Implementations

Here we will discuss the following main questions: (1) ways of representing measured radiances in the retrieval algorithm, (2) ways of representing optical characteristics of the aerosol in the retrieval algorithm, and (3) choosing a matrix or iterative inversion for implementing the minimization.

5.1.1. Radiances in the retrieval algorithm. As was described in section 4, we optimize the algorithm by accounting for measurement error while fitting aerosol optical thickness and sky radiances. The chosen settings are summarized in Table 2. The key point in these settings is the noise assumption. We also recommend utilization of lognormal statistics (i.e., we fit the logarithms of optical thickness and sky radiance). As for alternative noise statistics, the normal distribution of sky radiance and optical thickness with weight matrices given in Table 2 is the most reasonable alternative to the assumption of lognormal statistics (i.e., we fit the absolute values of optical thickness and sky radiance). The values of the weight matrices, covariances, and Lagrange multipliers given in Table 2 for normal distributions were not discussed in the text; however, they can easily be derived for the expected errors based on the same concepts.

5.1.2. Optical characteristics of aerosol in the retrieval algorithm. The questions of defining the retrieved aerosol characteristics were described in section 4, and Table 3 summarizes the chosen settings. This table shows two main possibilities we considered: to retrieve logarithms (recommended) or absolute (alternative) values of aerosol characteristics ($x(\ln r_i)$, $n(\lambda_i)$, and $k(\lambda_i)$). For each case, Table 3 describes the a priori constraints for all of the retrieved aerosol characteristics. For the particle size distribution and the wavelength dependence of the imaginary part of refractive index we indicate possibilities of constraining the differences (derivatives) of the first, second, or third orders. According to our analysis, these constraints are approximately equivalent. It is our expectation that the differences of the third order (for particle size distribution) can be more efficient in practice, because it allows for the highest variability of $x(\ln r_i)$. However, this statement should be verified in practice and by numerical tests.

5.1.3. Matrix and iterative inversion. As discussed in section 3, statistical optimization requires the minimization of the quadratic form, and various mathematical techniques can be employed for implementing this minimization (see section 3.2). In our algorithm we include two main alternatives: using matrix inversion by means of the SVD technique (section 3.2.1) or by using combined iterations as described in section 3.2.2. Also, we include the possibility of algorithm convergence improvement in a manner similar to the Levenberg-Marquardt method. Namely, we include a priori constraints on the solution correction $\Delta\mathbf{a}^p$ at each p step as described in section 4.

5.2. Illustrations

5.2.1. Numerical tests. The algorithm is focused on the simultaneous retrieval of particle size distribution and wavelength-dependent refractive index (real and imaginary parts). The principal difference with known approaches [e.g., *Wendisch and von Hoyningen-Huene*, 1994; *Yamasoe et al.*, 1998] is that we retrieve all aerosol characteristics ($x(\ln r_i)$, $n(\lambda_i)$, and $k(\lambda_i)$) at once by simultaneous fitting measurements of optical thickness and the angular distribution of sky radiances in the entire available spectral range. To succeed in such a global fitting, we had to employ an elaborated inversion scheme, which has been described and which allows us a significant degree of flexibility in realizing the inversion. The purpose of this section is accordingly to illustrate how well the inversion scheme works and what kind of results can be expected by using the different inversion options.

We have conducted a large number of numerical tests with the purpose of verifying the efficiency of the algorithm and checking the results regarding the settings of the inversion algorithm. Each illustration displayed below illustrates the phenomenon that was distinctly observed in a large number of numerical tests.

First, we have tested the efficiency of algorithm convergence and the sufficiency of information content for successful retrieval of all aerosol characteristics ($x(\ln r_i)$, $n(\lambda_i)$, and $k(\lambda_i)$). In this test we simulated aerosol optical thicknesses and the angular distribution of sky radiances at several wavelengths for an assumed particle size distribution and complex refractive index. Then we inverted the simulated optical thick-

Table 3. Summary of a Priori Constraints on the Smoothness of Aerosol Particle Size Distribution and on the Wavelength Dependence of Real and Imaginary Parts of Refractive Index

Aerosol Characteristics	Retrieved Parameters	Smoothness Constraints: Type of Constraint Differences	Standard Deviations (ε_k)	Lagrange Multiplier γ_k (k = 3, 4, 5)
$x(\ln r_i) = \frac{dR(r_i)}{d \ln r}$	$a_i = \ln x(\ln r_i)$ recommended	$\Delta^1 a_i$	$\varepsilon_3 = 2.6$*	$\gamma_3 = 4.0^{-4}$*
		$\Delta^2 a_i$	$\varepsilon_3 = 1.2$	$\gamma_3 = 2.0^{-3}$
		$\Delta^3 a_i$	$\varepsilon_3 = 0.9$	$\gamma_3 = 3.0^{-3}$
	$a_i = x(\ln r_i)$ alternative	$\Delta^1 a_i$	$\varepsilon_3 = 0.10$*	$\gamma_3 = 0.23$*
		$\Delta^2 a_i$	$\varepsilon_3 = 0.09$*	$\gamma_3 = 0.34$*
		$\Delta^3 a_i$	$\varepsilon_3 = 0.08$*	$\gamma_3 = 0.38$*
$n(\lambda_i)$	$a_i = \ln n(\lambda_i)$ recommended	$\frac{\Delta^1 a_i}{\Delta \ln \lambda_i}$	$\varepsilon_4 = 0.2$	$\gamma_4 = 0.0625$
	$a_i = n(\lambda_i)$ alternative	$\frac{\Delta^1 a_i}{\Delta \ln \lambda_i}$	$\varepsilon_4 = 0.125$	$\gamma_4 = 0.16$
$k(\lambda_i)$	$a_i = \ln k(\lambda_i)$ recommended	$\frac{\Delta^1 a_i}{\Delta \ln \lambda_i}$	$\varepsilon_5 = 1.25$	$\gamma_5 = 0.0016$
		$\frac{\Delta^2 a_i}{(\Delta \ln \lambda_i)^2}$	$\varepsilon_5 = 0$	$\gamma_5 = 0.1$†
	$a_i = k(\lambda_i)$ alternative	$\frac{\Delta^1 a_i}{\Delta \ln \lambda_i}$	$\varepsilon_5 = 1$	$\gamma_5 = 0.025$
		$\frac{\Delta^2 a_i}{(\Delta \ln \lambda_i)^2}$	$\varepsilon_5 = 0$	$\gamma_5 = 0.3$†

*Results are given for volume particle size distribution $x_i = dV(r_i)/d \ln r$.
†Lagrange multipliers used for the tests (they are not related to given ε_5).

ness and sky radiance and compared the retrieved particle size distribution and complex refractive index with the assumed values. Given the importance of aerosol absorption to issues of radiative forcing [see *Kaufman et al.*, 1997], we thought it is of interest to evaluate the agreement between values of single-scattering albedo ($\omega_0^{aer}(\lambda) = \tau_{scat}^{aer}(\lambda)/\tau_{ext}^{aer}(\lambda)$) obtained for assumed and retrieved aerosol characteristics $x(\ln r_i)$, $n(\lambda_i)$, and $k(\lambda_i)$. All tests were conducted for the measurement scheme (wavelengths, zenith and azimuth angles of observation, etc.) established for AERONET radiometers (for details, see *Holben et al.* [1998]). The tests have shown that both real and imaginary parts of the complex refractive index can be successfully retrieved together with particle size distribution, if no noise is introduced in the simulated radiance. In a majority of cases the errors did not exceed 20% for $k(\lambda_i)$, 0.02 for $n(\lambda_i)$, 0.015 for $\omega_0^{aer}(\lambda)$, and 10% for $dV/d \ln r$ for particles in the size range from 0.1 to 7 μm (the errors increase in the tails of the retrieved particle size distribution). The results remain good even in the presence of random noise. For example, Figures 1 and 2 illustrate the results of our test for retrieving biomass burning aerosol optical properties modeled with wavelength-dependent real and imaginary parts of the refractive index. A bimodal lognormal size distribution was assumed for this illustration according to the biomass burning aerosol model given by *Remer et al.* [1998]. The wavelength dependence of $n(\lambda_i)$ was assumed according to the values of the real part of the refractive index retrieved by *Yamasoe et al.* [1998] for smoke in Brazil. The wavelength dependence of the imaginary part of the refractive index was assumed accordingly to *Dubovik et al.* [1998b] for $k(\lambda_i)$ of Brazilian smoke with pronounced wavelength dependence of absorption ("artificial soot"). The algorithm computed the retrievals shown in Figures 1 and 2 with the settings recommended in Tables 2 and 3.

According to performance tests the use of the logarithmic transformation is a critical aspect of our algorithm (for both fitting the logarithm of radiance and retrieving logarithms of $x(\ln r_i)$, $n(\lambda_i)$, and $k(\lambda_i)$). By using absolute values (i.e., settings suggested in Tables 2 and 3 as alternatives) we could

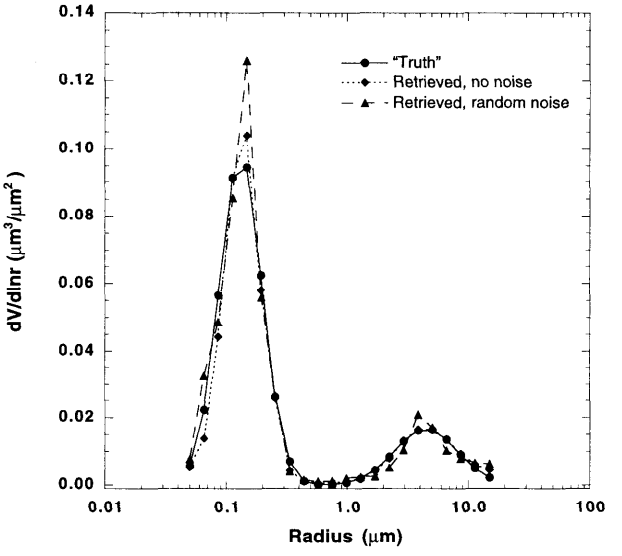

Figure 1. Results (particle size distribution) of the sensitivity test on aerosol optical properties retrieval from simulated sky radiance and optical thickness both without and with random noise added. Particle size distribution $dV/d \ln r$ for biomass burning aerosol [*Remer et al.*, 1998] is modeled by a bimodal lognormal function with parameters $r_{v1} = 0.132$ μm; $r_{v2} = 4.5$ μm; $\sigma_1 = 0.4$, $\sigma_2 = 0.6$; $C_{v1}/C_{v2} = 4$ ($\tau_{ext}(440) = 0.5$).

not obtain a stable convergence. In this case, the success of the retrieval required a special choice of the initial guess for each different combination of the retrieved parameters. In contrast, by using logarithms, we achieved good retrievals starting with the same initial guess ($dV(r)/d \ln r = 0.0001$, $n(\lambda_i) = 1.50$, and $k(\lambda_i) = 0.005$) in all cases.

Figures 3–5 illustrate the results of the retrievals, where the refractive index is known and fixed, for three different cases: large particles dominate (Figure 3), small particles dominate (Figure 4), and the presence of small and large particles is comparable with a third minor mode present in the middle range of particle size (Figure 5). In such situations, using both recommended and alternative settings gave good retrievals for the case with no noise added. However, if some random noise was added to the simulated radiances, the retrieval using logarithms was superior for both very small and large particle size ranges. In this respect, it is important to note that the widely used inversion code of *Nakajima et al.* [1983, 1996] does not use the logarithmic transformation (see analysis by *Dubovik et al.* [1998b]). This is probably one of the reasons for the inherent difficulties of the Nakajima et al. method in reproducing size distributions in the range of very small and large particles (see the discussion in the papers by *Remer et al.* [1997, 1998]).

It is important to notice that Figures 3–5 also illustrate the fact that in the presence of noise we were obtaining, in general, more stable retrievals when both size distribution and complex refractive index were retrieved than by retrieving only size distribution (with refractive index fixed to the correct value). This result can be explained by the fact that when refractive index is fixed only the size distribution can be changed during the retrieval. Thus the fitting of noisy data forces the size distribution to compensate for all of the errors in radiance. Alternatively, if both size distribution and refractive index are retrieved simultaneously, then errors in measured radiances will only be partially tied to errors in the size distribution, because some compensation or error redistribution will occur due to retrieval errors in refractive index. These errors in the refractive index retrieved under noisy conditions are acceptable. For example, the errors in refractive index for the tests

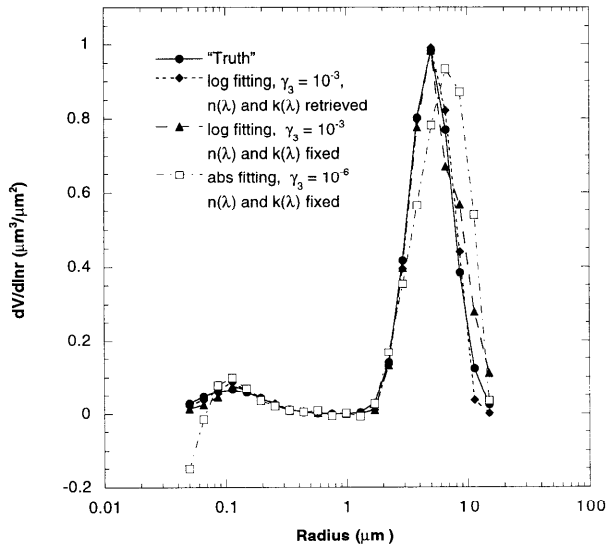

Figure 3. Numerical test results of the comparison of retrievals of size distribution by three different approaches (size distribution dominated by large particles): aerosol particle size distribution retrieval (refractive index is fixed) without logarithmic transformation ($f_1(\Theta; \lambda) = I(\Theta; \lambda)$); $f_2(\lambda) = \tau(\lambda)$ and $a_i = dV(r_i)/d \ln r$); aerosol particle size distribution retrieval (refractive index is fixed) under logarithmic transformation ($f_2(\Theta; \lambda) = \ln I(\Theta; \lambda)$); $f_2(\lambda) = \ln \tau(\lambda)$ and $a_i = \ln (dV(r_i)/d \ln r)$); aerosol particle size distribution retrieval (refractive index is retrieved) under logarithmic transformation ($f_1(\Theta; \lambda) = \ln I(\Theta; \lambda)$); $f_2(\lambda) = \ln \tau(\lambda)$ and $a_i = \ln (dV(r_i)/d \ln r)$). The radiance is perturbed by random noise (variances: 0.05% for $\Delta I(\Theta; \lambda)/I(\Theta; \lambda)$ and 0.01 for $\Delta\tau(\lambda)$).

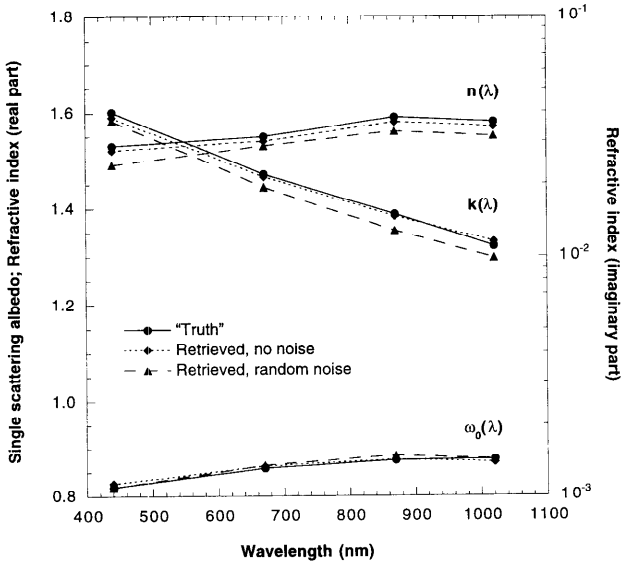

Figure 2. Results (single-scattering albedo, real and imaginary parts of refractive index) of the sensitivity test on aerosol optical properties retrieval from simulated sky radiance and optical thickness both without and with random noise added. Real part of the real part of refractive index for biomass burning aerosol is modeled according to the results by *Yamasoe et al.* [1998]: $n(440) = 1.53$, $n(670) = 1.55$, $n(870) = 1.59$, $n(1020) = 1.58$.

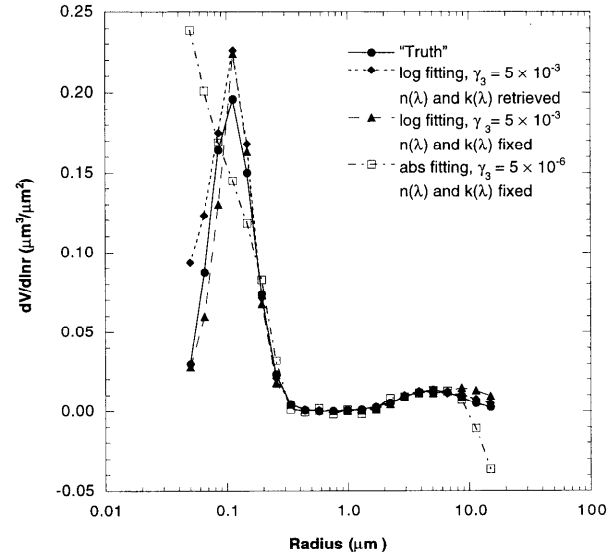

Figure 4. Same as Figure 3 but for an aerosol size distribution dominated by small particles.

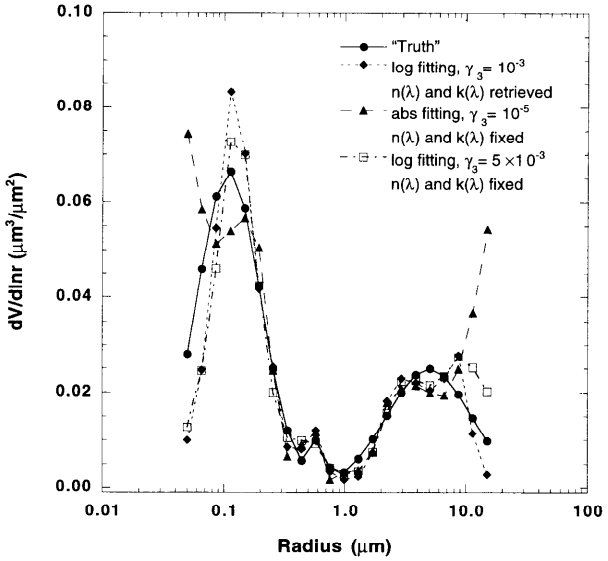

Figure 5. Same as Figure 3 but with an aerosol size distribution where large and small particles are comparably represented with a minor presence of particles in the middle size range.

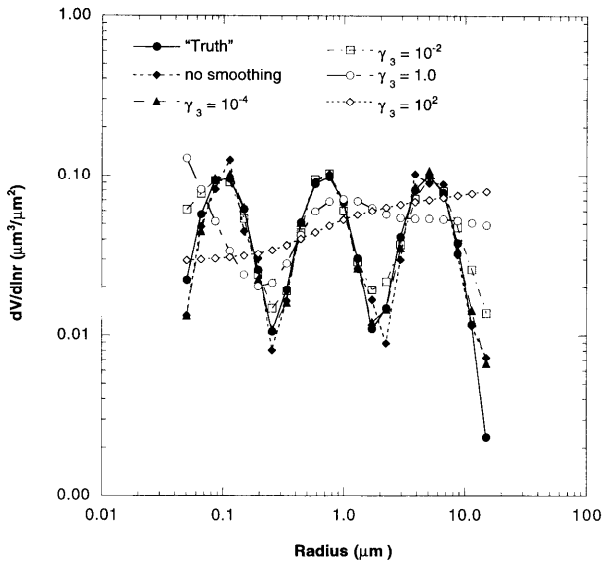

Figure 7. Illustration of size distribution retrieval results with constraining the second differences of $a_i = \ln (dV(r_i)/d \ln r)$.

shown in Figures 3–5 did not exceed 20% for $k(\lambda_i)$ and 0.02 for $n(\lambda_i)$.

We conducted a series of tests to verify our algorithm and settings regarding the smoothness constraints. Indeed, using overdetermined and/or inadequate constraints may result in smoothing out real (and possibly important) features of the retrieved aerosol characteristics (in particular, the particle size distribution). The tests have shown that the values of the Lagrange multipliers, recommended in Table 3, allow one to obtain satisfactory results for any monomodal, bimodal, or trimodal aerosol particle size distribution. Every mode of particle size distribution which was employed in our tests was assumed to be as narrow as the narrowest mode given by *Tanré et al.* [1999]. For example, Figure 5 shows a successful retrieval of a small feature in the size distribution (a third intermediate size aerosol mode), which was obtained using constraints similar to the ones applied in the tests without this feature (Figures 3 and 4).

Figures 6–8 illustrate the retrievals of particle size distribution with constraining first, second, or third derivatives. The results look good for the values of the Lagrange multipliers given in Table 3. Moreover, even if significantly higher values of γ_3 are used (up to $\gamma_3 = 0.01$) acceptable results for all cases were obtained. It is interesting to note that the intercomparison of retrieval results obtained with different constraints (in terms of variations of first, second, or third differences) did not show any dramatic difference for $\gamma_3 \leq 0.01$. For higher values of the Lagrange multiplier γ_3, a priori constraints forced the retrieved particle size distribution to assume an a priori prescribed shape (see (40)): horizontal line (for first differences), an arbitrary straight line (for second differences), and parabola

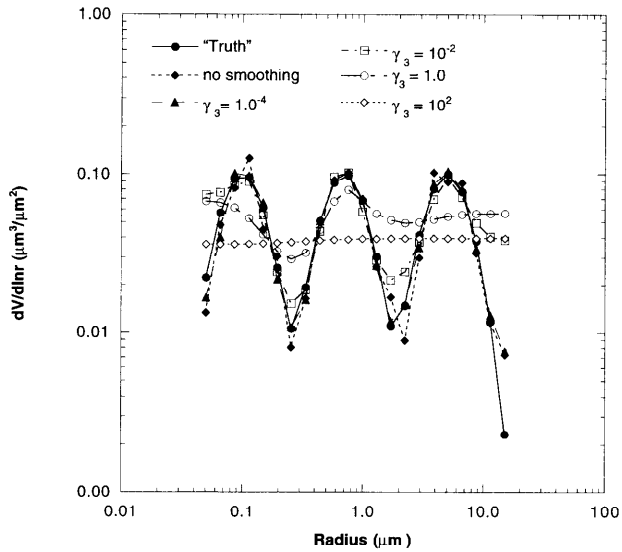

Figure 6. Illustration of size distribution retrieval results with constraining the first differences of $a_i = \ln (dV(r_i)/d \ln r)$.

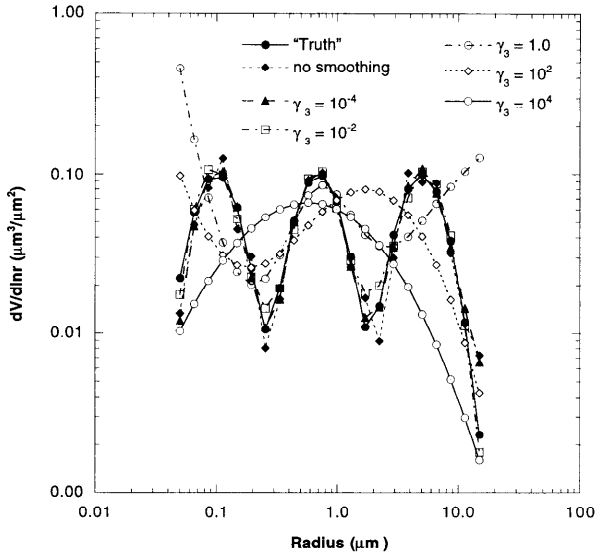

Figure 8. Illustration of size distribution retrieval results with constraining the third differences of $a_i = \ln (dV(r_i)/d \ln r)$.

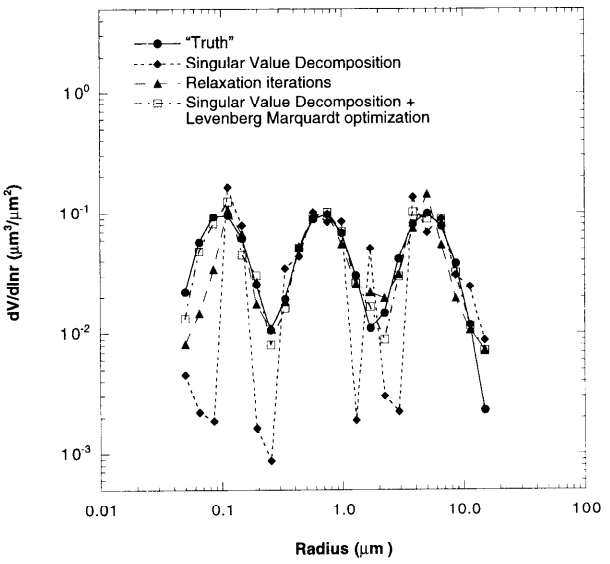

Figure 9. An illustration of using different mathematical techniques for minimization (no a priori constraints is used).

(for third differences). In spite of the fact that all constraints yielded satisfactory retrievals (for $\gamma_3 \leq 0.01$), we have concluded that using second or third differences is more appropriate for the retrieval of the particle size distribution. First, the restricting of first differences is the most severe restriction on particle size distribution (since this a priori assumes that the solution is a horizontal straight line). Second, the values of the Lagrange multiplier γ_3 for constraining the second- or higher-order differences are the same for size distributions of volume, area, radius, or number of particles (for the case when we retrieve the logarithms of $dR^n/d\ln r$ in the grid points r_i chosen with any equal step $\Delta \ln r = \ln r_{i+1} - \ln r_i = \text{const}$). This can be easily illustrated using (24) and (36b) on an example of the size distributions of particle volume and number:

$$\ln (dV(r_{i+1})/d\ln r)$$

$$= \ln (4/3\pi) + 3\ln (r_i) + \Delta \ln r + \ln (dN(r_{i+1})/d\ln r)$$

$$\Rightarrow \Delta^1 \ln (dV(r_i)/d\ln r) = -\Delta \ln r + \Delta^1 \ln (dN(r_i)/d\ln r)$$

$$\Rightarrow \Delta^{m\geq 2} \ln (dV(r_i)/d\ln r) = \Delta^{m\geq 2} \ln (dN(r_i)/d\ln r)$$

It should be noted that all illustrations show the results for the retrieval of volume particle size distribution because $dV/d\ln r$ is a standard product of AERONET [*Holben et al.*, 1998]. However, the retrieval of any other kind of particle size distribution $dR^n/d\ln r$ is also assumed in the algorithm and can be employed depending on user needs. Equation (24) can also be applied rather successfully for transforming the $dR^i/d\ln r$ to any other distribution $dR^n/d\ln r$; however, in general, the direct retrieval of the required $dR^n/d\ln r$ gives slightly better accuracy.

The final illustration of the results of our numerical tests relates to the use of iterative versus matrix inversion (the methods outlined in section 3.2.1). Figure 9 shows the retrievals of particle size distribution obtained by applying an iterative inversion and a SVD technique for matrix inversion (with and without applying constraints on $\Delta\mathbf{a}^p$). The inversions were obtained without using any a priori smoothness constraints on the solution and without adding any noise to the simulated radiance. We obtained good convergence of $\Psi(\mathbf{a}^p)$ to a minimum in all three cases, and the results were equally good for retrieval of $k(\lambda_i)$ and $n(\lambda_i)$. However, the results of particle size distribution retrievals were significantly different. In spite of the fact that the SVD inversion always gives an inverse matrix, it forces the appearance of physically unrealistic (but optically indistinguishable) oscillations (Figure 9). Using an iterative inversion always gives an appropriate solution without any inversion modification while requiring a longer time for convergence. The SVD technique coupled with the Levenberg-Marquardt-type constraints on $\Delta\mathbf{a}^p$ (included according to (45) and (46)) appeared to be practically the most efficient way of implementing the inversion. Indeed, the retrieval result is rather smooth, and the retrieval is faster than for the iterative inversion. Thus we have adopted the SVD technique with constraints on $\Delta\mathbf{a}^p$ (equations (45) and (46)) as the recommended way of implementing the inversion in our algorithm.

5.2.2. Application to real measurements. The purpose of our development is to make the code perform a reliable inversion of the measurements. However, we have thus far illustrated the performance of the inversion by inverting simulated atmospheric radiances. The difference between simulated and real measurements may contain various uncertainties that can affect the retrieval results. The random noise used in our tests does not reflect the diversity of all uncertainties present in real data. To understand the accuracy of inverting real data, some special analysis is needed. However, such analysis requires extensive studies related to information content of particular measurements rather than to the design of the inversion. The quality assessments of aerosol optical properties retrieved using AERONET spectral optical thickness and atmospheric radiance measurements are given in the paper by *Dubovik et al.* [2000]. In the current paper we limit ourselves to a single example, showing the practical capability of simultaneous retrievals of aerosol particle size distribution and wavelength-dependent refractive index from Sun and sky radiance obtained using AERONET radiometers. For this illustration we have chosen observations of different kinds of aerosols (biomass burning and urban aerosol) with similar wavelength dependence of optical thickness, $\alpha = 1.5$ ($\tau(\lambda) \sim \lambda^{-\alpha}$). Figures 10 and 11 show the retrieval results for urban aerosol measured in hazy conditions at the Goddard Space Flight Center (GSFC) and for biomass burning smoke measured in Cuiabá (Brazil) in different years (1993 and 1995). The particle size distribution is dominated by fine particles in all cases. At the same time, some differences in $dV/d\ln r$ can also be clearly seen. It is important to note that retrievals show very strong differences between biomass burning and urban aerosols in the values of real and imaginary parts of the refractive index. Indeed, n for urban aerosol at GSFC ranges between 1.33 and 1.40 (i.e., close to the values of n for water), whereas smoke-retrieved values of n are significantly higher than 1.4. This may be the results of much greater hygroscopic growth of particles with increasing humidity for mid-Atlantic U.S. pollution versus Brazilian smoke [*Kotchenruther and Hobbs*, 1998]. As expected, the values of the imaginary part of the refractive index are more than 10 times higher for smoke than for urban aerosol. The values of single-scattering albedo are close to unity for urban aerosol and significantly smaller for smoke. Moreover, the wavelength dependencies of $\omega_0^{\text{aer}}(\lambda)$, obtained for smoke in 1993 and 1995, are different for some cases (slightly increases with wavelength for aged smoke in 1995). This result qualitatively agrees with the results of the $\omega_0^{\text{aer}}(\lambda)$ retrievals obtained by independent

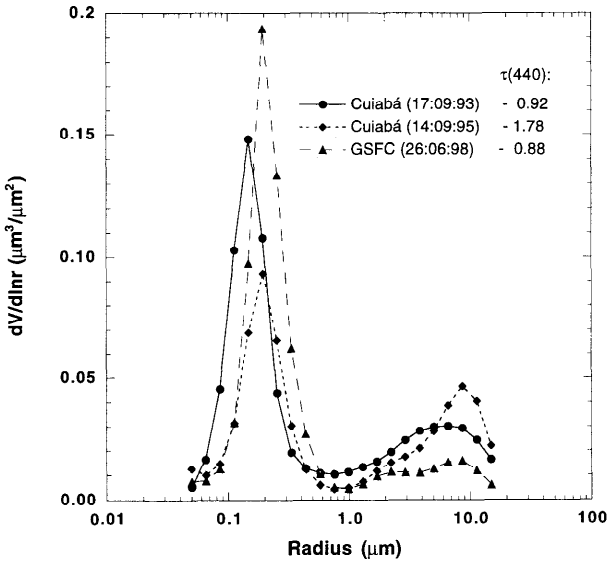

Figure 10. An application of the algorithm for particle size distribution retrieval from sky radiance and optical thickness measured by AERONET. The values of plotted particle size distribution are scaled to the values corresponding to $\tau(440) = 1$. The illustrated retrievals were obtained for the observations with similar wavelength dependence of optical thickness ($\alpha = 1.5$).

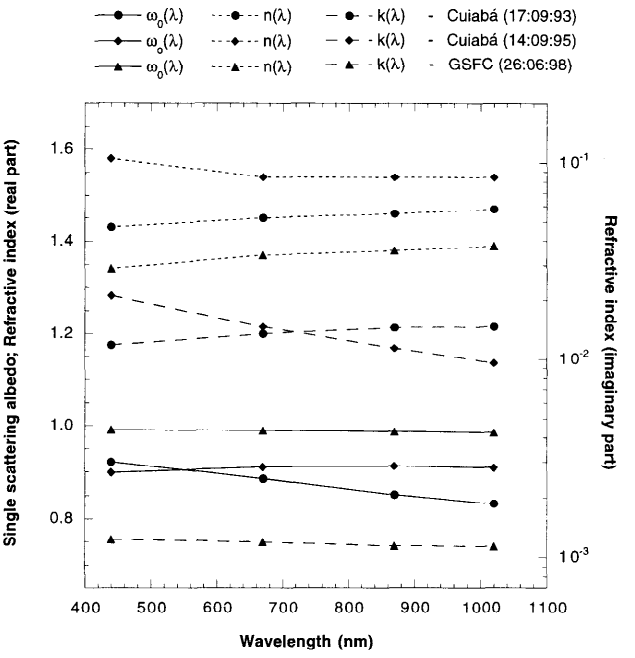

Figure 11. An application of the algorithm for single-scattering albedo, real and imaginary parts of refractive index retrieval from sky radiance, and optical thickness measured by AERONET.

techniques also for Cuiabá, Brazil, in 1995 [*Chu et al.*, 1999; J. V. Martins et al., personal communication, 1999; *Dubovik et al.*, 1998b]. The retrieved $k(\lambda_i)$ for observations of smoke at Cuiabá, Brazil, in 1995 show a strong decrease with wavelength. This is in good agreement with the results of the discussion given in the paper by *Dubovik et al.* [1998b]. It should be noted that retrievals using our algorithm for Cuiabá, Brazil, in 1995 for some days other than those illustrated in Figures 10–11 also show $k(\lambda_i)$ decreasing with wavelengths for aged smoke. At the same time, $\omega_0^{aer}(\lambda)$ is almost wavelength independent.

These examples thus show that by applying our inversion algorithm we were able to derive more detailed information from Sun and sky radiance measurements from AERONET radiometers than with procedures that were previously employed for the retrieval of aerosol optical properties from AERONET measurements [see *Holben et al.*, 1998].

6. Conclusion

A flexible algorithm for inverting complex sets of measured radiative and a priori known aerosol characteristics has been developed and implemented for the interpretation of ground-based measurements of Sun and sky radiance. The algorithm retrieves the particle size distribution over a wide range of sizes (0.05–15 μm) together with spectrally dependent complex refractive index and single-scattering albedo.

To achieve flexibility of the algorithm, we considered forward modeling and numerical inversion as two complementary but relatively independent components of the retrieval algorithm. The modeling of atmospheric radiance is performed by publicly available discrete ordinates radiative transfer codes for a multilayered plane-parallel atmosphere. Aerosol microstructure is incorporated in the inversion scheme by assuming homogeneous Mie-scattering spheres. The possibility of retrieving different kinds of particle size distribution (volume, area, radius, or number) was discussed and included in the algorithm.

The strategy of statistical optimization of multisource data, such as different types of measurements as well as a priori knowledge, was elaborated point by point. Accounting for different levels of input data accuracy and the nonnegativity of measured and retrieved parameters in the optimized inversion was discussed. We outlined the operational alternatives of assuming either normal or lognormal noise distributions in the radiance measurements, i.e., using either absolute values of Sun and sky radiance or their logarithms. The associated covariance matrices were presented. Similarly, we emphasized the differences of retrieving logarithms or absolute values of particle size distribution and real and imaginary parts of the refractive index.

The statistical concept of evaluating values of the Lagrange multiplier for including both accessory measurements and a priori constraints was described. This concept has been applied to determining measurement weights of spectral optical thickness and angular measurements of sky radiance in our procedure of simultaneous fitting of these characteristics. The results of this analysis are summarized in Table 2. On the basis of the same concept we defined values of the Lagrange multiplier for all a priori constraints employed in the algorithm. Namely, we have utilized constraints of variability on the particle size distribution and constraints on the spectral variability of real and imaginary parts of refractive index. For this purpose we applied limitations on the norm of the first, second, and third differences of the particle size distribution. In the same way we restricted the norm of the first and second derivatives of the variability of refractive index with wavelength. To evaluate the values of the corresponding Lagrange multiplier, we analyzed the maximum changes in atmospheric particle size distribution, as well as maximum spectral variability of the refractive index

(both real and imaginary parts). Table 3 summarizes applying a priori constraints.

Tables 2 and 3 show the recommended and the alternative setting for inverting measurements of spectral optical thickness and sky radiance together with a priori constraints. Table 3 also shows alternative a priori constraints for limiting differences (derivatives) of different orders. According to our results these constraints provide almost equivalent retrieval efficiencies. Nevertheless, for a number of reasons, we recommend using second or third differences for the smoothing of retrieved particle size distributions.

We have examined the practical efficiency of implementing numerical fitting by diverse mathematical techniques. Particular attention has been devoted to considering differences between methods using matrix and iterative inversion. Improving the convergence of nonlinear fitting by applying Levenberg-Marquardt or steepest descent types of iterations was studied. As a result we have outlined two alternatives: (1) combined linear iterations or (2) matrix inversion using singular value decomposition. Both of these methods give reliable convergence. The matrix inversion is more rapid but requires organizing the Levenberg-Marquardt-type iterations to obtain a stable result.

We have done a series of numerical tests for both checking the efficiency of the algorithm in general and for each particular algorithm setting. In the tests we inverted simulated ground-based measurements of Sun and sky radiance at the wavelengths and angles defined according to the measurement protocol established for AERONET radiometers. The results have shown that both the particle size distribution and the spectrally dependent parts of the complex refractive index can be derived, with reasonable accuracy, from the ground-based measurements of Sun and sky radiance. Moreover, these tests have shown that the method is sufficiently sensitive to observe important minor features in spectral dependencies of the real and imaginary parts of the aerosol refractive index and, accordingly, in the spectral dependence of single-scattering albedo.

The retrieval algorithm is currently being employed for operational use by the AERONET project. The results of these retrievals can be found on the AERONET project web page (http://aeronet.gsfc.nasa.gov:8080), and some illustrations are given in the text. The paper by *Dubovik et al.* [2000] discusses the stability of retrieval results to the diverse errors occurring in AERONET measurements.

Appendix A: Derivation of Linear Correction $\Delta\hat{\mathbf{a}}^p$ With Noise Optimization

To define a linear correction $\Delta\hat{\mathbf{a}}^p$, we can consider $\Delta\mathbf{f}_k(\Delta\mathbf{a}^p)$ as a linear functions of $\Delta\mathbf{a}^p$. Neglecting all terms of second or higher order in (18), we can write

$$\mathbf{f}_k^*(\hat{\mathbf{a}}) - \mathbf{f}_k(\mathbf{a}^p) \approx \mathbf{U}_{k,\mathbf{a}^p}(\hat{\mathbf{a}} - \mathbf{a}^p) \Rightarrow \Delta\mathbf{f}_k^* \approx \mathbf{U}_{k,\mathbf{a}^p}\Delta\mathbf{a}^p. \quad \text{(A1)}$$

The correction $\Delta\hat{\mathbf{a}}^p$ can be found with accounting for presenting noise as a value $\Delta\hat{\mathbf{a}}^p$ corresponding to the minimum of the quadratic form $\Psi(\Delta\hat{\mathbf{a}}^p)$ (defined in a similar manner to (13)):

$$\Psi(\Delta\mathbf{a}^p) = \frac{1}{2}\sum_{k=1}^{K}\gamma_k\Psi_k(\Delta\mathbf{a}^p) = \frac{1}{2}\sum_{k=1}^{K}\gamma_k[(\Delta\mathbf{f}_k^* - \mathbf{U}_{k,\mathbf{a}^p}\Delta\mathbf{a}^p)^T(\mathbf{W}_k)^{-1}(\Delta\mathbf{f}_k^* - \mathbf{U}_{k,\mathbf{a}^p}\Delta\mathbf{a}^p)]. \quad \text{(A2)}$$

The minimum of this quadratic form corresponds to the vector $\Delta\hat{\mathbf{a}}^p$ which yields a zero gradient vector $\nabla\Psi(\Delta\mathbf{a}^p)$:

$$\frac{\partial\Psi(\Delta\mathbf{a}^p)}{\partial(\Delta a_i^p)} = 0,\ (i = 1, \ldots, N_i) \Rightarrow \nabla\Psi(\Delta\mathbf{a}^p) = \mathbf{0}. \quad \text{(A3)}$$

The gradient of the quadratic form $\Psi(\Delta\hat{\mathbf{a}}^p)$ is a sum of the gradients of the following K terms:

$$\nabla\Psi(\Delta\mathbf{a}^p) = \frac{1}{2}\sum_{k=1}^{K}\gamma_k\nabla\Psi_k(\Delta\mathbf{a}^p). \quad \text{(A4)}$$

The gradient of each quadratic form $\nabla\Psi_k(\Delta\mathbf{a}^p)$ can be written as follows:

$$\nabla\Psi_k(\Delta\mathbf{a}^p) = 2(\mathbf{U}_{k,\mathbf{a}^p})^T(\mathbf{W}_k)^{-1}(\mathbf{U}_{k,\mathbf{a}^p})\Delta a^p - 2(\mathbf{U}_{k,\mathbf{a}^p})^T(\mathbf{W}_k)^{-1}(\Delta\mathbf{f}_k^*). \quad \text{(A5)}$$

Using (A4) and (A5), we can write (A3) as

$$\sum_{k=1}^{K}\gamma_k[(\mathbf{U}_{k,\mathbf{a}^p})^T(\mathbf{W}_k)^{-1}(\mathbf{U}_{k,\mathbf{a}^p})]\Delta\mathbf{a}^p - \sum_{k=1}^{K}\gamma_k[(\mathbf{U}_{k,\mathbf{a}^p})^T(\mathbf{W}_k)^{-1}(\Delta\mathbf{f}_k^*)] = \mathbf{0}. \quad \text{(A6)}$$

The detailed derivation of (A6) (for the case of $K = 1$) can be found elsewhere in numerous books on statistical estimation [cf. *Serber*, 1977; *Tarantola*, 1987].

Thus deriving $\Delta\hat{\mathbf{a}}^p$ from (A6) and using it to obtain $\hat{\mathbf{a}}^{p+1}$ by means of (17a) permits the definition of a nonlinear process for deriving a statistically optimum solution of (9).

Appendix B: Including a Priori Estimates $\Delta\hat{\mathbf{a}}^*$ in Retrieval of $\Delta\hat{\mathbf{a}}^p$

To improve the convergence of the retrieval process (given by (17) and (A6)) we can limit the length of $\Delta\hat{\mathbf{a}}^p$ by assuming a vector of a priori estimates for $\Delta\hat{\mathbf{a}}^*$; that is, we add one more constraining equation:

$$\Delta\hat{\mathbf{a}}^* = \Delta\hat{\mathbf{a}}^p + \Delta_{\Delta\mathbf{a}}, \quad \text{(B1)}$$

where $\Delta_{\Delta\mathbf{a}}$ are normally distributed errors with zero means and covariance matrix $\mathbf{C}_{\Delta a}$. Therefore the PDF of the estimates $\Delta\hat{\mathbf{a}}^*$ is defined as

$$P(\Delta\hat{\mathbf{a}}^p|\Delta\hat{\mathbf{a}}^*) \sim \exp\left(-\tfrac{1}{2}(\Delta\hat{\mathbf{a}}^p - \Delta\hat{\mathbf{a}}^*)^T(\mathbf{C}_{\Delta a})^{-1}(\Delta\hat{\mathbf{a}}^p - \Delta\hat{\mathbf{a}}^*)\right). \quad \text{(B2)}$$

Since (B1) restricts only the value of the correction $\Delta\hat{\mathbf{a}}^p$ but not the value of the unknown parameter $\hat{\mathbf{a}}^p$ itself, this constraint is only important for obtaining corrections $\Delta\hat{\mathbf{a}}^p$. To be consistent with this added constraint, we add an additional $K + 1$th term to the quadratic form $\Psi(\Delta\hat{\mathbf{a}}^p)$ and instead of (A4) we can write

$$\nabla\Psi(\Delta\hat{\mathbf{a}}^p) = \frac{1}{2}\sum_{k=1}^{K}\gamma_k\nabla\Psi_k(\Delta\hat{\mathbf{a}}^p) + \tfrac{1}{2}\gamma_{\Delta a}\nabla\Psi_{\Delta a}(\Delta\hat{\mathbf{a}}^p), \quad \text{(B3)}$$

where

$$\Psi_{\Delta a}(\Delta\hat{\mathbf{a}}) = (\Delta\hat{\mathbf{a}}^p - \Delta\hat{\mathbf{a}}^*)^T(\mathbf{W}_{\Delta a})^{-1}(\Delta\hat{\mathbf{a}}^p - \Delta\hat{\mathbf{a}}^*). \quad \text{(B4)}$$

The gradient of this quadratic form can be obtained using an expression similar to (A5):

$$\nabla\Psi_{\Delta a}(\Delta\hat{\mathbf{a}}^p) = 2(\mathbf{W}_{\Delta a})^{-1}\Delta\hat{\mathbf{a}}^p - 2(\mathbf{W}_{\Delta a})^{-1}(\Delta\hat{\mathbf{a}}^*). \quad \text{(B5)}$$

Thus the vector $\Delta\hat{\mathbf{a}}^p$ which minimizes the quadratic form $\Psi(\Delta\hat{\mathbf{a}}^p)$ corresponds to the solution of the following equation:

$$\left(\sum_{k=1}^{K} \gamma_k[(\mathbf{U}_{k,\mathbf{a}^p})^T(\mathbf{W}_k)^{-1}(\mathbf{U}_{k,\mathbf{a}^p})] + \gamma_{\Delta a}(\mathbf{W}_{\Delta a})^{-1} \right) \Delta\hat{\mathbf{a}}^p$$

$$- \sum_{k=1}^{K} \gamma_k[(\mathbf{U}_{k,\mathbf{a}^p})^T(\mathbf{W}_k)^{-1}(\Delta\mathbf{f}_k^*)] - \gamma_{\Delta a}(\mathbf{W}_{\Delta a})^{-1}\Delta\hat{\mathbf{a}}^* = \mathbf{0}. \quad \text{(B6)}$$

Appendix C: Derivation of Chahine's Formula

The method of *Chahine* [1968] involves the solution of the linear system $\mathbf{I}(\mathbf{x}) = \mathbf{K}\mathbf{x}$ by nonlinear iterations ($x_i^{p+1} = x_i^p(I_i^*/I_i^p)$). The utility of this method is limited by the fact that the matrix $\mathbf{K}$ is square (i.e., the numbers of initial characteristics I_j and unknowns x_i are equal), when initial characteristics I_j and unknowns x_i are positively defined. Also, the matrix $\mathbf{K}$ must be diagonally dominant in order that convergence be achieved. In Chahine's iterative approach, the solution vector is restricted to positive and smooth values, thereby eliminating the negative and highly oscillatory solutions typical of linear matrix inversion.

Analyzing Chahine's formula, one can see that this formula is very different with both matrix inversion by (20) and (21) and linear iterations by (22). Namely, Chahine's formula is nonlinear and includes multiplication and division instead of addition and substruction in the linear methods. The concept of statistical optimization of the inversion and retrieval of nonnegative values (section 4.2.1) prescribes that the initially linear system should be solved in logarithmic space:

$$\mathbf{I}^* = \mathbf{K}\mathbf{x} \Rightarrow \ln I_j^* = \ln I_j(\ln x_1, \ln x_1, \ldots, \ln x_n). \quad \text{(C1)}$$

This nonlinear system can be solved by Newtonian iterations similar to (20a) $\mathbf{W}$ is canceled for the case of square $\mathbf{K}$:

$$\ln \hat{\mathbf{x}}^{p+1} = \ln \hat{\mathbf{x}}^p - \Delta \ln \hat{\mathbf{x}}^p,$$

$$\Delta \ln \hat{\mathbf{x}}^p = (\mathbf{U}_p)^{-1}(\ln \mathbf{I}^p - \ln \mathbf{I}^*). \quad \text{(C2)}$$

Matrix $\mathbf{U}_p$ contains the first derivatives, which for $\mathbf{I}(\mathbf{x}) = \mathbf{K}\mathbf{x}$ can be expressed as follows:

$$\{\mathbf{U}_p\}_{ji} = \frac{\partial \ln I_j}{\partial \ln x_i}\bigg|_{\hat{\mathbf{x}}^p} = \frac{K_{ji}}{I_j^p}x_i^p = \frac{K_{ji}x_i^p}{\sum_{k=1}^{n} K_{jk}x_k^p}. \quad \text{(C3)}$$

Using Chahine's condition of a diagonally dominant matrix $\mathbf{K}$, we can now approximate $\mathbf{U}_p$ by the unit matrix; that is,

$$\text{for } K_{jj} \gg K_{jj'\neq j},\ \mathbf{U}_p \approx \mathbf{1}. \quad \text{(C4)}$$

Substituting matrix (C4) in (C2), we arrive at the formula proposed by *Chahine* [1968]:

$$\ln \hat{\mathbf{x}}^{p+1} = \ln \hat{\mathbf{x}}^p - (\ln \mathbf{I}^p - \ln \mathbf{I}^*) \Rightarrow x_i^{p+1} = x_i^p\left(\frac{I_i^*}{I_i^p}\right). \quad \text{(C5)}$$

Chahine's method converges for any diagonally dominant matrix $\mathbf{K}$ (i.e., $K_{jj} > K_{jj'\neq j}$), although the approximation for (C4) is correct only for a diagonally dominant matrix $\mathbf{K}$ where the diagonal dominance is strong (i.e., $K_{jj} \gg K_{jj'\neq j}$). In this regard, the nonlinear univariate relaxation of Chahine is formally similar to the standard linear Gauss-Seidel algorithm used for solving systems of equations and which always converges if the matrix $\mathbf{K}$ is diagonally dominant [e.g., *Ortega*, 1988].

Appendix D: Statistical Derivation of Twomey-Chahine Formula

The generalization of Chahine's formula was the objective of a number of inversion studies, because the convergence conditions associated with that method (square and diagonally dominant matrix $\mathbf{K}$) seriously restrict its application. The absence of a clear strategy which exploits the added information content of a priori and accessory data is an additional reason for seeking out alternatives to the Chahine technique.

The first nonlinear Chahine-like formula (which is widely known in atmospheric studies) was proposed by *Twomey* [1975] for solving linear overdetermined system $\mathbf{I}(\mathbf{x}) = \mathbf{K}\mathbf{x}$ ($m > n$):

$$x_i^{p+1} = x_i^p \prod_{j=1}^{m} \left(1 + \left(\frac{I_j^*}{I_j^p} - 1\right)\tilde{K}_{ji}\right), \quad \text{(D1)}$$

where $\tilde{K}_{ji}$ denotes the elements of matrix $\mathbf{K}$ which are scaled to be less than unity. Below, we do not repeat the original methodology for deriving these iterations (which can be found in the work of *Twomey* [1975, 1979]). Rather, we try to understand the Chahine approach in a fashion consistent with the idea of the present paper (section 3) inasmuch as we consider the solution as a noise optimization procedure. For positively defined I_j and x_i we accordingly assume a lognormal noise distribution. The solution of the system $\mathbf{I}(\mathbf{x}) = \mathbf{K}_{(m\times n)}\mathbf{x} + \Delta$ in logarithmic space then corresponds to the minimum of the quadratic form

$$\Psi(\ln \mathbf{x}^p) = \tfrac{1}{2}(\ln \mathbf{I}^p - \ln \mathbf{I}^*)^T(\mathbf{W}_{\ln \mathbf{I}})^{-1}(\ln \mathbf{I}^p - \ln \mathbf{I}^*). \quad \text{(D2)}$$

According to the discussion in section 3.2 the minimum of the above residual can be obtained by the Levenberg-Marquardt procedure and can be easily reduced to the steepest descent method (22b):

$$\ln \hat{\mathbf{x}}^{p+1} = \ln \hat{\mathbf{x}}^p - t^p\nabla\Psi)x^p)$$

$$= \ln \hat{\mathbf{x}}^p - t^p\mathbf{U}_p^T(\mathbf{W}_{\ln \mathbf{I}}))^{-1}(\ln \mathbf{I}^p - \ln \mathbf{I}^*). \quad \text{(D3)}$$

Equation (D3) is already quite similar to Chahine-like iterations, since it restricts the solution to be positively defined and since no complicated matrix inversion is involved (the weight matrix is diagonal in most of the cases). To emphasize the similarity between (D3) and (D1), we rewrite (D3) in terms of x_i and I_j:

$$\hat{x}_i^{p+1} = \hat{x}_i^p \exp\left(t^p \sum_{j=1}^{m} \bar{K}_{ji}(\ln I_j^* - \ln I_j^p)\right)$$

$$= \hat{x}_i^p \prod_{j=1}^{m} \exp\left(t^p\bar{K}_{ji}(\ln I_j^* - \ln I_j^p)\right), \quad \text{(D4)}$$

where

$$\bar{\mathbf{K}} = \mathbf{U}_p^T(\mathbf{W}_{\ln \mathbf{I}})^{-1}. \quad \text{(D5)}$$

For an appropriate initial guess of $(\ln I_j^* - \ln I_j^p)$ (which must be <1), (28a) can be applied. We can, as well, approximate the exponents in (D5) by the two first terms of a Taylor expansion ($\exp(\Delta a) = 1 + \Delta a + o(\Delta a)^2$). Consequently, (D5) can be transformed into the form of (D1):

$$\hat{x}_i^{p+1} = \hat{x}_i^p \prod_{j=1}^{m} \left(1 + t^p \tilde{K}_{ji}\left(\frac{I_j^* - I_j^p}{I_j^p}\right)\right)$$

$$= \hat{x}_i^p \prod_{j=1}^{m} \left(1 + t^p \left(\frac{I_j^*}{I_j^p} - 1\right)\tilde{K}_{ji}\right). \quad \text{(D6)}$$

It should be noted that according to (D5) the matrix $\tilde{\mathbf{K}} = \mathbf{U}_p^T$, given the common assumption of a unity matrix being used as the weight matrix ($\mathbf{W}_{\ln \mathbf{I}} = \mathbf{1}$; see Table 2). The elements of this matrix are naturally restricted to be less that unity (see (C3)). The multiplier t^p can be considered as a Levenberg-Marquardt multiplier and can, accordingly, be chosen in a manner similar to how it is performed in the Levenberg-Marquardt method ($t^p \leq 1$) in order to provide monotonic convergence. Similar coefficients or operations restricting changes of parameters at each step were used in applying (D1) to concrete inversions [*Trakhovsky and Shettle*, 1986; *Dubovik et al.*, 1995].

Acknowledgments. The authors are grateful to Tom Eck, Brent N. Holben, Yoram J. Kaufman, Norm O'Neill, Alexander Smirnov, Robert S. Fraser, and Ilya Slutsker for valuable discussions and help in improving the paper. We thank Robert Curran of NASA headquarters for his support. The authors wish to thank Stefan Kinne and Boris Fomin for providing software simulating molecular and Mie scattering. Additional thanks to the AERONET staff for data collection, calibration, and processing.

References

Ackerman, T. P., and O. B. Toon, Absorption of visible radiation in atmosphere containing mixtures of absorbing and nonabsorbing particles, *Appl. Opt.*, *20*, 3661–3668, 1981.

Bates, T. S., B. J. Huebert, J. L. Gras, F. B. Griffiths, and P. A. Durkee, International Global Atmospheric Chemistry (IGAC) Project's First Aerosol Characterization Experiment (ACE 1): Overview, *J. Geophys. Res.*, *103*, 16,297–16,318, 1998.

Bohren, C. F., and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, 550 pp., John Wiley, New York, 1983.

Box, M. A., and C. Sendra, Retrieval of the albedo and phase function from exiting radiances with radiative perturbation theory, *Appl. Opt.*, *38*, 1636–1643, 1999.

Chahine, M. T., Determination of temperature profile in an atmosphere from its outgoing radiance, *J. Opt. Soc. Am.*, *12*, 1634–1637, 1968.

Chu, D. A., Y. J. Kaufman, D. Tanré, and B. N. Holben, Smoke optical properties derived from airborne measurements in Amazonia, paper presented at ALPS'99, Cent. Natl. d'Etudes Spatiales, Meribel, France, January 18–22, 1999.

Dubovik, O. V., T. V. Lapyonok, and S. L. Oshchepkov, Improved technique for data inversion: Optical sizing of multicomponent aerosols, *Appl. Opt.*, *34*, 8422–8436, 1995.

Dubovik, O., T. Yokota, and Y. Sasano, Improved technique for data inversion and its application to the retrieval algorithm for ADEOS/ILAS, *Adv. Space. Res.*, *21*, 397–403, 1998a.

Dubovik, O., B. N. Holben, Y. J. Kaufman, M. Yamasoe, A. Smirnov, D. Tanré, and I. Slutsker, Single-scattering albedo of smoke retrieved from the sky radiance and solar transmittance measured from ground, *J. Geophys. Res.*, *103*, 31,903–31,924, 1998b.

Dubovik, O., A. Smirnov, B. N. Holben, M. D. King, Y. J. Kaufman, T. F. Eck, and I. Slutsker, Accuracy assessments of aerosol optical properties retrieved from AERONET Sun- and sky-radiance measurements, *J. Geophys. Res.*, *105*, 9791–9806, 2000.

Eck, T. F., B. N. Holben, J. S. Reid, O. Dubovik, S. Kinne, A. Smirnov, N. T. O'Neill, and I. Slutsker, The wavelength dependence of the optical depth of biomass burning, urban and desert dust aerosols, *J. Geophys. Res.*, *104*, 31,333–31,350, 1999.

Edie, W. T., D. Dryard, F. E. James, M. Roos, B. Sadoulet, *Statistical Methods in Experimental Physics*, 155 pp., North-Holland, New York, 1971.

Forsythe, G., and W. Wasow, *Finite Difference Methods for Partial Differential Equations*, John Wiley, New York, 1960.

Holben, B. N., et al., AERONET—A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, *66*, 1–16, 1998.

Horvath, H., Atmospheric light absorption—A review, *Atmos. Environ.*, *27*, 293–317, 1993.

Houghton, J. T., F. W. Taylor, and C. D. Rodgers, *Remote Sounding of Atmospheres*, 343 pp., Cambridge Univ. Press, New York, 1984.

Kaufman, Y. J., and B. N. Holben, Hemispherical backscattering by biomass burning and sulfate particles derived from sky measurements, *J. Geophys. Res.*, *101*, 19,433–19,445, 1996.

Kaufman, Y. J., D. Tanré, H. R. Gordon, T. Nakajima, J. Lenoble, R. Frouin, H. Grassl, B. M. Herman, M. D. King, and P. M. Teillet, Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect, *J. Geophys. Res.*, *102*, 16,815–16,830, 1997.

Kaufman, Y. J., et al., Smoke, Clouds, and Radiation-Brazil (SCAR-B) experiment, *J. Geophys. Res.*, *103*, 31,783–31,808, 1998.

King, M. D., Sensitivity of constrained linear inversions to the selection of the Lagrange multiplier, *J. Atmos. Sci.*, *39*, 1356–1369, 1982.

King, M. D., D. M. Byrne, B. M. Herman, and J. A. Reagan, Aerosol size distributions obtained by inversion of spectral optical depth measurements, *J. Atmos. Sci.*, *21*, 2153–2167, 1978.

King, M. D., Y. J. Kaufman, D. Tanré, and T. Nakajima, Remote sensing of tropospheric aerosols from space: Past, present, and future, *Bull. Am. Meteorol. Soc.*, *80*, 2229–2259, 1999.

Kotchenruther, R. A., and P. V. Hobbs, Humidification factors of aerosols from biomass in Brazil, *J. Geophys. Res.*, *103*, 32,081–32,089, 1998.

McCartney, E. J., *Optics of the Atmosphere*, 200 pp., John Wiley, New York, 1977.

Nakajima, T., and M. Tanaka, Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation, *J. Quant. Spectrosc. Radiat. Transfer*, *40*, 51–69, 1988.

Nakajima, T., M. Tanaka, and T. Yamauchi, Retrieval of the optical properties of aerosols from aureole and extinction data, *Appl. Opt.*, *22*, 2951–2959, 1983.

Nakajima, T., G. Tonna, R. Rao, P. Boi, Y. Kaufman, and B. Holben, Use of sky brightness measurements from ground for remote sensing of particulate polydispersions, *Appl. Opt.*, *35*, 2672–2686, 1996.

O'Neill, N. T., and J. R. Miller, Combined solar aureole and solar beam extinction measurements, 2, Studies of inferred aerosol size distribution, *Appl. Opt.*, *23*, 3697–3704, 1984.

Ortega, J. M., *Introduction to Parallel and Vector Solution of Linear System*, 200 pp., Plenum, New York, 1988.

Ortega, J. M., and W. C. Reinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, 504 pp., Academic, San Diego, Calif., 1970.

Patterson, E. M., and C. K. McMahon, Absorption characteristic of forest fire particulate matter, *Atmos. Environ.*, *18*, 2541–2551, 1984.

Phillips, B. L., A technique for numerical solution of certain integral equation of first kind, *J. Assoc. Comput. Mach.*, *9*, 84–97, 1962.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN, The Art of Scientific Computing*, 965 pp., Cambridge Univ. Press, New York, 1992.

Ramanathan, V., et al., Indian Ocean Experiment (INDOEX), A multi-agency proposal for field experiment in the Indian Ocean, *C4 Publ. 162*, 83 pp., Scripps Inst. of Oceanogr., La Jolla, Calif., 1996.

Remer, L. A., and Y. J. Kaufman, Dynamic aerosol model: Urban/industrial aerosol, *J. Geophys. Res.*, *103*, 13,859–13,871, 1998.

Remer, L. A., S. Gasso, D. A. Hegg, Y. J. Kaufman, and B. N. Holben, Urban/industrial aerosol: Ground-based Sun/sky radiometer and in situ measurements, *J. Geophys. Res.*, *102*, 16,849–16,859, 1997.

Remer, L. A., Y. J. Kaufman, B. N. Holben, A. M. Thompson, and D. P. McNamara, Biomass burning aerosol size distribution and modeled optical properties, *J. Geophys. Res.*, *103*, 31,879–31,891, 1998.

Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys.*, *14*, 609–624, 1976.

Rodgers, C. D., Characterization and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, *95*, 5587–5595, 1990.

Romanov, P., N. T. O'Neill, and A. Royer, Simultaneous retrieval of aerosol refractive index and particle size distribution from ground

based measurements of direct and scattered solar radiation, *Appl. Opt.*, *38*, 7305–7320, 1999.

Russell, P., and J. Heintzenberg, An overview of the ACE 2 Clear Sky Column Closure Experiment (Clearcolumn), *Tellus*, *54*(2), 462–482, 2000.

Russell, P., P. V. Hobbs, and L. L. Stowe, Aerosol properties and radiative effects in the United States east coast haze plume: An overview of the Tropospheric Aerosol Radiative Forcing Observational Experiment (TARFOX), *J. Geophys. Res.*, *104*, 2213–2222, 1999.

Satheesh, S. K., V. Ramanathan, J.-J. Xu, J. M. Lobert, I. A. Podgorny, J. M. Prospero, B. N. Holben, and N. G. Loeb, A model for natural and anthropogenic aerosols over the tropical Indian Ocean derived from Indian Ocean Experiment data, *J. Geophys. Res.*, *104*, 27,421–27,440, 1999.

Serber, G. A. F., *Linear Regression Analysis*, 430 pp., John Wiley, New York, 1977.

Schmid, B., J. Michalsky, R. Halthore, M. Beauharnois, L. Harrison, J. Livingston, P. Russell, B. Holben, T. Eck, and A. Smirnov, Comparison of aerosol optical depth from four solar radiometers during fall 1997 ARM intensive observation period, *Geophys. Res. Lett.*, *26*, 2725–2728, 1999.

Shaw, G. E., Inversion of optical scattering and spectral extinction measurements to recover aerosol size spectra, *Appl. Opt.*, *18*, 988–993, 1979.

Shettle, E. P., and R. W. Fenn, Models of aerosols of lower troposphere and the effect of humidity variations on their optical properties, *AFCRL Tech. Rep. 79 0214*, 100 pp., Air Force Cambridge Res. Lab., Hanscom, Air Force Base, Mass., 1979.

Spinhirne, J. D., and M. D. King, Latitudinal variation of spectral optical thickness and columnar size distribution of the El Chichón stratospheric aerosol layer, *J. Geophys. Res.*, *90*, 10,607–10,619, 1985.

Stamnes, K., S. C. Tsay, W. Wiscombe, and K. Jayaweera, Numerically stable algorithm for discrete ordinate-method radiative transfer in multiple scattering and emitting layered media, *Appl. Opt.*, *27*, 2502–2509, 1988.

Strand, O. N., and E. R. Westwater, Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind, *J. Assoc. Comput. Mach.*, *15*, 104–114, 1968.

Tanré, D., L. R. Remer, Y. J. Kaufman, S. Mattoo, P. V. Hobbs, J. M. Livingston, P. B. Russell, and A. Smirnov, Retrieval of aerosol optical thickness and size distribution over ocean from the MODIS airborne simulator during TARFOX, *J. Geophys. Res.*, *104*, 2261–2278, 1999.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, 500 pp., Elsevier Sci., New York, 1987.

Tikhonov, A. N., On the solution of incorrectly stated problems and a method of regularization, *Dokl. Akad. Nauk*, *151*, 501–504, 1963.

Tikhonov, A. N., and V. Y. Arsenin, *Solution of Ill-Posed Problems*, 300 pp., John Wiley, New York, 1977.

Trakhovsky, E., and E. P. Shettle, Improved inversion procedure for the retrieval of aerosol size distributions using aureole measurements, *J. Opt. Soc. Am.*, *2*, 2054–2061, 1985.

Turchin, V. F., V. P. Kozlov, and M. S. Malkevich, The use of the methods of mathematical statistics for the solution of the incorrect problems, *Usp. Fiz. Nauk*, *102*, 345–386, 1970.

Twitty, J. T., The inversion of aureole measurements to derive aerosol size distributions, *J. Atmos. Sci.*, *32*, 584–591, 1975.

Twomey, S., On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, *J. Assoc. Comput. Mach.*, *10*, 97–101, 1963.

Twomey, S., Comparison of constrained linear inverse and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions, *J. Comput. Phys.*, *18*, 188–200, 1975.

Twomey, S., *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*, 243 pp., Elsevier Sci., New York, 1977.

Wang, M., and H. Gordon, Retrieval of the columnar aerosol phase function and single scatting albedo from sky radiance over the ocean: Simulations, *Appl. Opt.*, *32*, 4598–4609, 1993.

Wendisch, M., and W. von Hoyningen-Huene, Possibility of refractive index determination of atmospheric aerosol particles by ground-based solar extinction and scattering measurements, *Atmos. Environ.*, *28*, 785–792, 1994.

Whitby, K. T., The physical characteristics of sulfur aerosols, *Atmos. Environ.*, *12*, 135–159, 1978.

Yamasoe, M. A., Y. J. Kaufman, O. Dubovik, L. A. Remer, B. N. Holben, and P. Artaxo, Retrieval of the real part of the refractive index of smoke particles from Sun/sky measurements during SCAR-B, *J. Geophys. Res.*, *103*, 31,893–31,902, 1998.

O. Dubovik, NASA GSFC, SSAI, Code 923, Greenbelt, MD 20771. (dubovik@aeronet.gsfc.nasa.gov)

M. D. King, Earth Sciences Directorate, NASA GSFC, Greenbelt, MD 20771.

(Received October 22, 1999; revised April 27, 2000;
accepted April 30, 2000.)