

# Improved identification of the solution space of aerosol microphysical properties derived from the inversion of profiles of lidar optical data, part 1: theory

ALEXEI KOLGOTIN,<sup>1</sup> DETLEF MÜLLER,<sup>2,3,\*</sup> EDUARD CHEMYAKIN,<sup>3</sup> AND ANTON ROMANOV<sup>4</sup>

<sup>1</sup>Physics Instrumentation Center, Troitsk, Moscow Region 142190, Russia

<sup>2</sup>University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK

<sup>3</sup>Science Systems and Applications, Inc. (SSAI), NASA LaRC, 1 Enterprise Parkway, Hampton, Virginia 23666, USA

<sup>4</sup>National University of Science and Technology, Leninskii av. 4, Moscow 119049, Russia

\*Corresponding author: d.mueller@herts.ac.uk

Received 29 August 2016; revised 12 September 2016; accepted 12 September 2016; posted 14 September 2016 (Doc. ID 266604); published 30 November 2016

Multiwavelength Raman/high spectral resolution lidars that measure backscatter coefficients at 355, 532, and 1064 nm and extinction coefficients at 355 and 532 nm can be used for the retrieval of particle microphysical parameters, such as effective and mean radius, number, surface-area and volume concentrations, and complex refractive index, from inversion algorithms. In this study, we carry out a correlation analysis in order to investigate the degree of dependence that may exist between the optical data taken with lidar and the underlying microphysical parameters. We also investigate if the correlation properties identified in our study can be used as *a priori* or *a posteriori* constraints for our inversion scheme so that the inversion results can be improved. We made the simplifying assumption of error-free optical data in order to find out what correlations exist in the best case situation. Clearly, for practical applications, erroneous data need to be considered too. On the basis of simulations with synthetic optical data, we find the following results, which hold true for arbitrary particle size distributions, i.e., regardless of the modality or the shape of the size distribution function: surface-area concentrations and extinction coefficients are linearly correlated with a correlation coefficient above 0.99. We also find a correlation coefficient above 0.99 for the extinction coefficient versus (1) the ratio of the volume concentration to effective radius and (2) the product of the number concentration times the sum of the squares of the mean radius and standard deviation of the investigated particle size distributions. Besides that, we find that for particles of any mode fraction of the particle size distribution, the complex refractive index is uniquely defined by extinction- and backscatter-related Ångström exponents, lidar ratios at two wavelengths, and an effective radius.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

**OCIS codes:** (010.1110) Aerosols; (010.1120) Air pollution monitoring; (010.3640) Lidar; (000.3860) Mathematical methods in physics; (010.0280) Remote sensing and sensors; (010.1100) Aerosol detection.

<http://dx.doi.org/10.1364/AO.55.009839>

## 1. INTRODUCTION

Several approaches for the retrieval of particle microphysical parameters from multiwavelength lidar measurements were developed in the past two decades [1–5]. Since we deal with an ill-posed, ill-conditioned problem, the derived solutions oscillate, and they are unstable to the point that we may obtain nonphysical results. One of the major sources of instability is the wide ranges of particle radii and complex refractive indices (CRI) that need to be considered. With regard to realistic aerosol conditions in the atmosphere, we think we need to consider the following domains

in our study: the radius domain needs to be [0.01; 20]  $\mu\text{m}$ , and the domains of the real and imaginary part of the CRI need to be [1.2; 2] and [0; 0.1], respectively. Even the use of advanced mathematical methods such as regularization, e.g., Tikhonov's regularization [6], does not guarantee that we find physically meaningful solutions unless we introduce additional constraints in the solution space that follows from data inversion. In a previous study, we showed that we can stabilize the solution space if we use whole sections of optical data (OD) profiles rather than OD sets at individual height bins, as input in our inversion scheme [7].

If the profiles of the OD do not change, or in other words if the profile gradient is close to 0, we expect that the corresponding profiles of the particle microphysical parameters (PMP) do not vary significantly either. On the contrary, if the OD profiles vary with measurement height, i.e., the “gradient” profile differs from zero, we expect that the PMPs vary with height too.

Certainly, compensation effects are possible. For example, the backscatter coefficient decreases if the imaginary part of the CRI increases, and it increases if the real part increases. Theoretically, if the real and imaginary parts increase (decrease) simultaneously, the lidar ratios (LRs) may remain constant, and we may not observe any vertical variation of the optical profiles, i.e., there may be no gradients in the lidar profiles. We do not consider those cases in our study for the moment, but we will investigate those cases in future.

In particular, simulations with synthetic OD (SOD) and data collected during field measurements show that the particle effective radius, which is a mean property of a given particle size distribution (PSD), and the extinction-related Ångström exponent are related to each other in an inversely proportional fashion [8]. Profiles of the extinction-related Ångström exponent can be measured with the type of multiwavelength lidar we are using, e.g., [9]. These profiles serve as the basis of our sensitivity study presented in this contribution. Currently such profiles are mainly used for a qualitative assessment of whether effective-radius profiles retrieved with inversion algorithms are reasonable. For instance, in a first approximation, we expect that the effective radius drops if the extinction-related Ångström exponent increases, e.g., [10,11].

Aside from such qualitative ways of using relationships between the particle size and extinction-related Ångström exponents, there are attempts on finding interdependencies between particle concentration and backscatter/extinction coefficients [12,13]. In Ref. [12], the surface-area and volume concentrations of mineral dust were retrieved from lidar measurements taken during the Saharan Mineral Dust Experiment (SAMUM). These parameters were subsequently analyzed in the context of a correlation study that involved extinction and backscatter coefficients. In the first case (surface-area concentration versus extinction coefficients), the correlation reached  $R^2 = 0.98$ , whereas in the second case (volume concentration versus backscatter coefficient), there was barely any correlation. In addition, it was shown [13] that the number concentration has only a low correlation with the measured extinction and backscatter coefficients.

In this contribution, we investigate in a systematic manner and on the basis of a special look-up table, which correlations between the lidar OD and PMPs exist, and if these correlations can be used in our inversion methodology for improving the retrieval quality of the microphysical data products.

In Section 2, we present the correlation relationships we found between the PMP and the OD. In Section 3, we show the statistical results for a set of SOD we generated for this study. The statistical results for the experimental data are shown in Section 4. Section 5 summarizes our results.

## 2. ASYMPTOTIC RELATIONS AND REGRESSION EQUATIONS

Lidar measurements deliver OD that are used for the retrieval of bulk PMPs ( $p$ ), such as mean radius ( $r_{\text{mean}}$ ) and effective

radius ( $r_{\text{eff}}$ ), mean width, i.e., geometrical standard deviation ( $\sigma$ ) of a PSD, and number ( $n$ ), surface-area ( $s$ ), and volume ( $v$ ) concentrations. Solving this problem is related to solving the Fredholm integral equation of the first kind;

$$\int_{r_{\min}}^{r_{\max}} K_g(\lambda, m, r) f(r) dr = g(\lambda), \quad g = \alpha, \beta. \quad (1)$$

The OD  $g(\lambda)$  are taken with lidar. In the case considered in our study, we focus on particle backscatter coefficients measured at 355, 532, and 1064 nm, and extinction coefficients measured at 355 and 532 nm. The unknown function  $f(r)$  describes the PSD, and  $r$  describes the particle radius on the domain  $[r_{\min}, r_{\max}]$ . The parameter  $r_{\min}$  describes the minimum particle radius, and  $r_{\max}$  describes the maximum particle radius of the investigated PSD. The kernel functions  $K_g(\lambda, m, r)$  are calculated from the respective extinction and backscatter efficiencies  $Q_g(\lambda, m, r)$  for individual particles weighted with their geometrical cross-section  $\pi r^2$  [14]. The parameter  $m = m_R - im_I$  describes the particle CRI. The CRI is usually unknown in experiments, and thus needs to be estimated like the other PMPs. The extinction coefficients are denoted as  $g = \alpha$ , and the backscatter coefficients are denoted as  $g = \beta$ . The PMPs can be easily estimated from the solution  $f$ .

On the basis of the average theorem, Eq. (1) can be rewritten as

$$f(\xi) \int_{r_{\min}}^{r_{\max}} K_g(\lambda, m, r) dr = g(\lambda), \quad (2)$$

with  $\xi \in [r_{\min}, r_{\max}]$ . From the physical point of view, we can treat this transformation as if a given PSD is replaced by the rectangle function, in other words, by the uniform law defined on the domain  $[r_{\min}, r_{\max}]$  and without changing the particle optical properties, i.e.,

$$f(r) = f(\xi) = \text{const}. \quad (3)$$

Depending on the definition of  $f$ , i.e., whether the PSD is represented in terms of number ( $n$ ), surface-area ( $s$ ), or volume ( $v$ ) concentration, we denote the kernel functions with an extra superscript:  $K_g^n$ ,  $K_g^s$ , and  $K_g^v$ , respectively. In the case of  $f(r) = s(r)$ , the kernel functions represent the optical cross-section per particle surface-area, i.e.,

$$K_g^s(\lambda, m, r) = Q_g(\lambda, m, r)/4. \quad (4)$$

It is a known fact that for the extinction efficiency [14], we find

$$Q_a \approx \text{const} \rightarrow \langle Q_a \rangle = 2.1 \quad \text{for } r \rightarrow \infty.$$

That relation works well starting for a particle radius larger than 1  $\mu\text{m}$ . If the particle radius is  $r \in [0.05; 0.5] \mu\text{m}$ , we obtain  $\langle Q_a \rangle \in [2; 3]$ . The symbol  $\langle \dots \rangle$  means that we are averaging over the respective radius interval.

We considered two asymptotic cases. These two cases describe small ( $r \rightarrow 0$ ) and big ( $r \rightarrow \infty$ ) particles, respectively. Moreover, we work with averaged, or to be more precise, with integral magnitudes over the size range which is larger than  $r - r_{\min} > 0.4 \mu\text{m}$  for both cases. From the physical point of view, we can treat both cases as limits for all possible cases, including particles in the radius interval from 0.5 to 1  $\mu\text{m}$ . In other words, the radius interval from 0.5 to 1  $\mu\text{m}$  is intermediate

for the two asymptotic cases. From the mathematical point of view, if we try to estimate  $\langle Q_\alpha \rangle$  over the range from 0.05 to 1  $\mu\text{m}$  (that is wider than 0.05–0.5  $\mu\text{m}$ ), we obtain the value of approximately  $\langle Q_\alpha \rangle \approx 2$ , which again belongs to the interval [2; 3].

In the case of using the approximation as written in Eq. (3), the surface-area concentration can be expressed as

$$s = \int_{r_{\min}}^{r_{\max}} s(r) dr \approx \int_{r_{\min}}^{r_{\max}} s(\xi) dr = s(\xi)(r_{\max} - r_{\min}). \quad (5)$$

The left hand side of Eq. (2) can be transformed on the basis of the average theorem into

$$\begin{aligned} f(\xi) \int_{r_{\min}}^{r_{\max}} K_\alpha(\lambda, m, r) dr &= \frac{s\langle Q_\alpha \rangle}{4(r_{\max} - r_{\min})} \int_{r_{\min}}^{r_{\max}} dr \\ &= \begin{cases} \frac{2+1}{4}s; & r > 1.0 \mu\text{m} \\ \frac{3}{4}s; & 0.05 \mu\text{m} \leq r \leq 0.5 \mu\text{m} \end{cases} \end{aligned} \quad (6)$$

The parameter  $s$  can then be directly estimated from particle extinction as

$$s = \begin{cases} \frac{4}{2.1}\alpha(\lambda) = 1.9\alpha(\lambda); & r > 1.0 \mu\text{m} \\ \frac{4}{3}\alpha(\lambda) = 1.3\alpha(\lambda); & 0.05 \mu\text{m} \leq r \leq 0.5 \mu\text{m} \end{cases} \quad (7)$$

Obviously Eq. (7) means that estimating the uncertainty of  $s$  is less than  $\sim 20\%$  if we consider the whole particle radius range, i.e.,  $r \in (0.05; \infty)$ . In fact, we have two limits, i.e., 1.3 (minimal) and 1.9 (maximal). The mean value between both limits is  $(1.9 + 1.3)/2 = 1.6$ . The uncertainty for the mean value of 1.6 is  $(1 - 1.3/1.6) \times 100\% \approx (1.9/1.6 - 1) \times 100\% \approx 20\%$ .

Let us consider the asymptotic relation with respect to volume concentration. Using the definition of effective radius

$$r_{\text{eff}} = \frac{3v}{s}, \quad (8)$$

and Eq. (7), the ratio of volume concentration to effective radius can be estimated as

$$\frac{v}{r_{\text{eff}}} = \begin{cases} 0.6\alpha(\lambda); & r > 1.0 \mu\text{m} \\ 0.4\alpha(\lambda); & 0.05 \mu\text{m} \leq r \leq 0.5 \mu\text{m} \end{cases} \quad (9)$$

Finally, we can find an asymptotic relation with respect to number concentration. Using the definitions of the geometrical standard deviation

$$\sigma^2 = \frac{1}{n} \int_{r_{\min}}^{r_{\max}} (r - r_{\text{mean}})^2 n(r) dr, \quad (10)$$

number concentration

$$n = \int_{r_{\min}}^{r_{\max}} n(r) dr, \quad (11)$$

mean radius

$$r_{\text{mean}} = \frac{1}{n} \int_{r_{\min}}^{r_{\max}} r n(r) dr, \quad (12)$$

and surface-area concentration

$$s = 4\pi \int_{r_{\min}}^{r_{\max}} r^2 n(r) dr, \quad (13)$$

we obtain

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \int_{r_{\min}}^{r_{\max}} r^2 n(r) dr - 2 \frac{1}{n} r_{\text{mean}} \int_{r_{\min}}^{r_{\max}} r n(r) dr \\ &\quad + \frac{1}{n} r_{\text{mean}}^2 \int_{r_{\min}}^{r_{\max}} n(r) dr = \frac{s}{4\pi n} - 2r_{\text{mean}}^2 + r_{\text{mean}}^2. \end{aligned} \quad (14)$$

From this equation, we find that

$$4\pi n(r_{\text{mean}}^2 + \sigma^2) \equiv s, \quad (15)$$

and in view of Eq. (7), we can derive the asymptotic relation for the product

$$n(r_{\text{mean}}^2 + \sigma^2) = \begin{cases} 0.15\alpha(\lambda); & r > 1.0 \mu\text{m} \\ 0.11\alpha(\lambda); & 0.05 \mu\text{m} \leq r \leq 0.5 \mu\text{m} \end{cases} \quad (16)$$

The asymptotic relations of Eqs. (7), (9), and (16) show that intensive parameters (IP), such as  $s$ , the ratio of parameters  $v/r_{\text{eff}}$ , and the product  $n(r_{\text{mean}}^2 + \sigma^2)$ , can be directly estimated from the lidar extinction measurements (at one wavelength) with an uncertainty that is less than 20%. This level of uncertainty of 20% allows us to round the coefficients to the tenths in Eqs. (7) and (9), and to the hundredths in Eq. (16).

For example, we find

$$n(r_{\text{mean}}^2 + \sigma^2) = 0.13\alpha(\lambda) \pm 20\%,$$

$$\text{where } 0.13 = (0.15 + 0.11)/2.$$

From the statistical point of view, it also means that the IPs  $p$  and the corresponding extinction coefficient  $\alpha(\lambda)$  are linearly correlated. We find the following regression equations

$$p = a_p \alpha(\lambda) + b_p, \quad (17a)$$

and

$$p = s, \quad v/r_{\text{eff}}, \quad n(r_{\text{mean}}^2 + \sigma^2). \quad (17b)$$

The regression coefficients (RC) in that case are

$$a_s \in [1.3; 1.9], \quad a_v \in [0.4; 0.6], \quad a_n \in [0.11; 0.15], \quad (18a)$$

$$b_p = 0, \quad p = s, \quad v/r_{\text{eff}}, \quad n(r_{\text{mean}}^2 + \sigma^2). \quad (18b)$$

In view of Eqs. (8) and (15), we note that  $a_v = a_s/3$ , and  $a_n = a_s/(4\pi)$ .

The regression equation (17) can be extended, as it is generally accepted that there exists a correlation between the effective radius  $r_{\text{eff}}$  and the extinction Ångström exponent (EAE)  $\hat{a}_\alpha$ , which is

$$\hat{a}_\alpha = \ln[\alpha(\lambda_1)/\alpha(\lambda_2)]/\ln(\lambda_2/\lambda_1). \quad (19)$$

That means, the larger the EAE the smaller the particle effective radius. For small variations of EAE, we can apply the approximation

$$r_{\text{eff}} = a_r \hat{a}_\alpha + b_r, \quad (20)$$

and the RC  $a_r$  and  $b_r$  can be found from the statistical analysis of a synthetic data bank we used in our study, see Eqs. (33a)–(33c) and Fig. 1(a).

We obtain four regression equations, Eq. (17) to Eq. (20) for the six parameters  $r_{\text{mean}}$ ,  $r_{\text{eff}}$ ,  $\sigma$ ,  $n$ ,  $s$ , and  $v$ . All six parameters can be uniquely defined from the lidar data products considered in our study if six regression equations are available. Therefore, another two regression equations are needed in order to obtain a closed system. We find the relationship that links  $r_{\text{mean}}$ ,  $r_{\text{eff}}$ , and  $\sigma$ , and which describes particle radius.

Approximating the number PSD by the uniform law of Eq. (3), and taking into account the definitions of Eqs. (8), (10)–(13) and the definition of volume concentration

$$v = \frac{4\pi}{3} \int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r) dr, \quad (21)$$

we can show the following new relationships (A)–(E):

(A) The mean particle radius can be written as

$$\begin{aligned} r_{\text{mean}} &= \frac{1}{n} \int_{r_{\text{min}}}^{r_{\text{max}}} r n(r) dr \approx \frac{1}{n} \int_{r_{\text{min}}}^{r_{\text{max}}} r n(\xi) dr \\ &= \frac{1}{2} (r_{\text{max}} + r_{\text{min}}), \end{aligned} \quad (22)$$

(B) the standard deviation can be written as

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \int_{r_{\text{min}}}^{r_{\text{max}}} (r - r_{\text{mean}})^2 n(r) dr \approx \frac{1}{n} \int_{r_{\text{min}}}^{r_{\text{max}}} (r - r_{\text{mean}})^2 n(\xi) dr \\ &= \frac{1}{3} (r_{\text{max}}^2 - r_{\text{min}}^2), \end{aligned} \quad (23)$$

(C) the surface-area concentration can be written as

$$\begin{aligned} s &= 4\pi \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r) dr \approx 4\pi \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(\xi) dr \\ &= \frac{4\pi}{3} n(\xi) (r_{\text{max}}^3 - r_{\text{min}}^3), \end{aligned} \quad (24)$$

(D) the volume concentration can be written as

$$\begin{aligned} v &= \frac{4\pi}{3} \int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r) dr \approx \frac{4\pi}{3} \int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(\xi) dr \\ &= \frac{\pi}{3} n(\xi) (r_{\text{max}}^4 - r_{\text{min}}^4), \end{aligned} \quad (25)$$

(E) the effective radius can be written as

$$r_{\text{eff}} = \frac{3v}{s} \approx \frac{3r_{\text{max}}^4 - r_{\text{min}}^4}{4r_{\text{max}}^3 - r_{\text{min}}^3}. \quad (26)$$

We can also rewrite Eqs. (22) and (23) as

$$r_{\text{max}} - r_{\text{min}} \approx \frac{r_{\text{max}}^2 - r_{\text{min}}^2}{2r_{\text{mean}}}, \quad (27)$$

and

$$r_{\text{max}} r_{\text{min}} \approx r_{\text{mean}}^2 - 3\sigma^2, \quad (28)$$

respectively, and we obtain the following important relationships for effective radius and geometric standard deviation of the investigated PSDs:

$$r_{\text{eff}} \approx \frac{3\sigma^2 + r_{\text{mean}}^2}{\sigma^2 + r_{\text{mean}}^2} r_{\text{mean}}, \quad (29)$$

and

$$\sigma^2 + r_{\text{mean}}^2 \approx \frac{2r_{\text{mean}}^3}{3r_{\text{mean}} - r_{\text{eff}}} = \frac{2\theta^3 r_{\text{eff}}^2}{3\theta - 1}, \quad (30)$$

where  $\theta = r_{\text{mean}}/r_{\text{eff}}$ .

The corollaries useful for our study follow from these relationships:

(A) parameter  $\theta$  belongs to the interval  $(1/3; 1)$  in order to fulfill the inequalities

$$\begin{aligned} r_{\text{mean}}^2 + \sigma^2 &> 0 \\ r_{\text{mean}} &< r_{\text{eff}} \end{aligned} \quad (31a)$$

[see Eqs. (30) and (29), respectively],

(B)

$$\sigma^2 + r_{\text{mean}}^2 \geq \frac{1}{2} r_{\text{eff}}^2 \quad \text{for } \forall \theta \in (1/3; 1), \quad (31b)$$

(C)

$$\sigma^2 + r_{\text{mean}}^2 \leq r_{\text{eff}}^2 \quad \text{for } \forall \theta \in [\sim 11/30; 1), \quad (31c)$$

We stress that Eq. (22) works for atmospheric particle size distributions in special cases in which the PSD consists of one mode fraction (or elementary fraction). Any attempt to use Eq. (22) for multi-modal PSDs is useless because the total mean radius depends drastically on the intensity of each mode. Besides that, an in-depth analysis shows that Eq. (29) works for any PSD if  $r_{\text{eff}} < 3r_{\text{mean}}$ . However, there are mathematical laws describing the behavior of the PSD for which the last inequality is not valid. If the PSD has a logarithmic-normal shape, then it can be characterized even for  $r_{\text{eff}} > 5r_{\text{mean}}$  if  $\sigma_{\text{Gauss}} > 2$ .

Let us consider any log-normal law with large values of  $\sigma_{\text{Gauss}} \geq 2$  in terms of number and volume PSD. The maximal values of both PSDs result in totally different mode radii. For example, if the mode radius of a number PSD is about 0.2  $\mu\text{m}$ , then it is about 1  $\mu\text{m}$  for the corresponding volume PSDs at  $\sigma_{\text{Gauss}} = 2$ . In spite of the fact that the logarithmic-normal law describes a unimodal function, we cannot treat it as a one-modal PSD that describes either the fine or the coarse mode because of the rather large fraction of big particles with  $r > 1 \mu\text{m}$ . In this case, Eq. (29) underestimates the actual effective radius that is estimated with Eq. (8). We can use this particularity as a flag parameter, indicating that the investigated PSD is quite “wide” or, in other words, contains a significant fraction of particles in the coarse mode. Such a flag parameter is important for the automated, unsupervised inversion method that is being developed for the world-wide first airborne multiwavelength  $3\beta + 2\alpha$  system, i.e., NASA Langley’s high-spectral-resolution lidar (HSRL-2) [9].

### 3. NUMERICAL SIMULATION

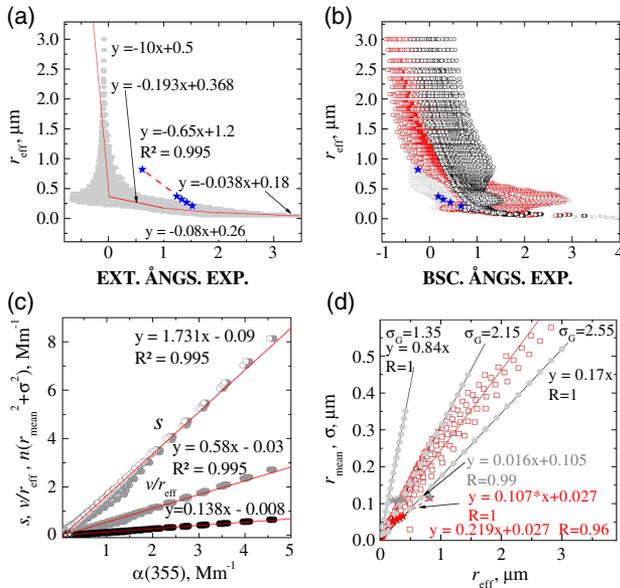
#### A. Synthetic Optical Data

We used SOD to identify the aforementioned correlation relationships. We generated the SOD for the HSRL-2 configuration ( $3\beta + 2\alpha$ ), i.e., for particle backscatter coefficients measured at 355, 532 and 1064 nm and extinction coefficients measured at 355 and 532 nm [4]. We took into account Eq. (1). We predefined values for  $f$  and  $m$ . We used the logarithmic-normal law for describing  $f$ . We selected a wide range

of values for the mean radius, i.e.,  $r_{\text{Gauss}} \in [15 \text{ nm}, 335 \text{ nm}]$  with a stepsize of 20 nm. With regard to the mode width, we used  $\sigma_{\text{Gauss}} \in [1.35, 2.55]$  with a stepsize of 0.1. We selected the following values for the real and imaginary parts:  $m_{\text{R}} \in [1.3, 1.7]$  and  $m_{\text{I}} \in [0.0, 0.05]$  with a stepsize of 0.025 and 0.005, respectively. With regard to the imaginary part, we used extra values close to 0:  $5e-4$ ,  $1e-3$ ,  $2e-3$ ,  $3e-3$ ,  $4e-3$ , and  $7.5e-3$ . We used all possible combinations of the four parameters ( $r_{\text{Gauss}}$ ,  $\sigma_{\text{Gauss}}$ ,  $m_{\text{R}}$ ,  $m_{\text{I}}$ ), and in that way we created 63,869 sets of  $3\beta + 2\alpha$  data for the SOD bank.

Figure 1 presents the statistical analysis of the SOD bank. We see that  $r_{\text{eff}}$  is nearly inversely proportional to  $\hat{a}_{\alpha}$  [Fig. 1(a)]. The range of EAE varies from  $-0.5$  to  $4$ . For example, if  $\hat{a}_{\alpha} > 2$ , which means that  $\delta r_{\text{eff}} < 0.05 \mu\text{m}$ , we find for the RC,  $a_r = -0.038$ , and  $b_r = 0.18$ . For  $\hat{a}_{\alpha} \in [1; 2]$ , the RCs are  $a_r = -0.08$ , and  $b_r = 0.26$ , and the threshold is  $\delta r_{\text{eff}} \leq 0.07 \mu\text{m}$ . The analysis of all entries of the SOD bank shows that these two parameters vary in the intervals:

$$a_r \in [-10; -0.038] \quad \text{and} \quad b_r \in [0.18; 0.5]. \quad (32)$$



**Fig. 1.** Statistics of the entries in the synthetic data bank we used in our study. (a)  $r_{\text{eff}}$  versus EAE. (b)  $r_{\text{eff}}$  versus BAE for the wavelength pair of 532 and 1064 nm. The red circles describe particles with strong light absorption and a high real part ( $m_{\text{R}} > 1.4$ ), the black circles describe particles of low light absorption and low real part ( $m_{\text{R}} \leq 1.4$ ), and the gray circles are the remaining cases and describe cases which (according to literature and our own experience) have not been observed in the atmosphere but are important in order to study the robustness of our data analysis scheme. (c)  $s$  (open circle),  $v/r_{\text{eff}}$  (gray circle), and  $(r_{\text{mean}}^2 + \sigma^2)n$  (black circle) versus  $\alpha(355)$ . (d)  $r_{\text{mean}}$  (gray circle) and  $\sigma$  (red square) versus  $r_{\text{eff}}$ . The solid and dotted lines describe the regression equations  $y = ax + b$  with the correlation coefficient  $R^2$ . The stars show the results for  $r_{\text{eff}}$  (blue stars in plot a and plot b),  $s$  (gray symbols in plot c),  $(r_{\text{mean}}^2 + \sigma^2)n$  (black symbols in plot c),  $\sigma$  (red symbols in plot d), and  $r_{\text{mean}}$  (gray symbols in plot d) for the bimodal PSDs. The volume concentration in the coarse mode with respect to the total volume concentration of the PSD is increasing from 0% to 90%. The effective radius of the coarse mode was kept constant in that case. Symbol  $\sigma_G$  denotes  $\sigma_{\text{Gauss}}$ .

In practice, the following relations are useful:

$$r_{\text{eff}} = \begin{cases} -0.038\hat{a}_{\alpha} + 0.18 \pm 0.05 \mu\text{m}; & 2 < \hat{a}_{\alpha} < 4 & (33a) \\ -0.082\hat{a}_{\alpha} + 0.26 \pm 0.07 \mu\text{m}; & 1 < \hat{a}_{\alpha} < 2 & (33b) \\ -0.193\hat{a}_{\alpha} + 0.37 \pm 0.08 \mu\text{m}; & 0.5 < \hat{a}_{\alpha} < 1 & (33c) \end{cases}$$

The linear regression is our idealization to constrain the solution space. We use the property that the effective radius is distributed near the trend (regression) line. Only the region that is enclosed by two lines  $\text{trend} \pm \delta r_{\text{eff}}$  (where  $\delta r_{\text{eff}}$  belongs to  $[0.05; 0.08]$ ) is useful for our constraints. We accept 0.5 as the lower boundary of the EAE, where it is still possible to reasonably limit the spread of the effective radius near the trend line with a threshold value  $\pm 0.08 \mu\text{m}$ .

It is important to investigate the statistics of the backscatter-related Ångström exponent (BAE)  $\hat{a}_{\beta}$  for the wavelength pair of 532 and 1064 nm. Figure 1(b) shows that any BAE value less than 3 results in a wide spread of the effective radius. Even for the interval (1; 2), the effective radius varies from  $\sim 0$  to  $0.5 \mu\text{m}$ .

Lidar measures the EAE and BAE simultaneously. Therefore, it is valuable to analyze the structure of the SOD bank with regard to these parameters. We consider three intervals for these two parameters, i.e., low  $\hat{a}_{\beta}$  and  $\hat{a}_{\alpha} < 0.5$ , moderately high  $\hat{a}_{\beta}$  and  $\hat{a}_{\alpha} = 0.5-1.5$ , and high  $\hat{a}_{\beta}$  and  $\hat{a}_{\alpha} > 1.5$ . The results of this analysis are presented in Table 1. This table shows the effective radius, the lidar ratio (LR) at 355 nm, and the CRI intervals within which the particle parameters vary for the different combinations of  $\hat{a}_{\beta}$  and  $\hat{a}_{\alpha}$  in their predefined intervals.

In particular, Table 1 shows that case #1, according to the intersection of the intervals  $\hat{a}_{\alpha} < 0.5$  and  $\hat{a}_{\beta} < 0.5$ , can be presented by particles with any LR from 2 to 1095 sr, any real part of the CRI from 1.3 to 1.7, any imaginary part of the CRI from 0 to 0.05, as well as a wide range of effective radii exceeding  $0.2 \mu\text{m}$ . With regard to case #9 ( $\hat{a}_{\alpha} > 1.5$  and  $\hat{a}_{\beta} > 1.5$ ), the particle parameters vary much less, i.e., LR and  $r_{\text{eff}}$  are less than 133 sr and  $0.12 \mu\text{m}$ , respectively.

Our SOD bank contains unrealistic values; for example, the LR can exceed 150 sr (cases #1–8), and it can be as high as 1095 sr (case #1). The effective radius reaches  $0.02 \mu\text{m}$  (cases #8, 9). Therefore, we constrained our SOD such that more “natural” aerosol properties are included, i.e.,

- effective radius  $r_{\text{eff}} \geq 0.05 \mu\text{m}$ ,
- high light-absorbing particles with real part  $m_{\text{R}} > 1.4$  and imaginary part  $m_{\text{I}} > 0.001$ ,
- low light-absorbing particles with real part  $m_{\text{R}} \leq 1.4$  and imaginary parts  $m_{\text{I}} \leq 0.015$ .

Table 2 shows the CRI values used in the constrained SOD bank in more detail. The red and black circles in Fig. 1(b) indicate high and low light-absorbing particles, respectively. All other cases in Fig. 1(b) are shown by gray circles and are likely not to occur under realistic atmospheric conditions. In comparison to the “complete” SOD, the “constrained” SOD is about half the size and includes 35,404 data sets.

Two cases in Table 1 deserve special attention. First, in case #3 in our “constrained” SOD bank, there is no monomodal PSD that can produce the high EAE ( $\hat{a}_{\alpha} > 1.5$ ) and the low BAE ( $\hat{a}_{\beta} < 0.5$ ) simultaneously. As we will discuss below, it is possible for bimodal PSDs. Second, case #2 contains just

**Table 1. Structure of the SOD Bank in Dependence of the EAE and BAE Intervals for the Wavelength Pairs (355, 532) nm and (532, 1064) nm, Respectively<sup>a</sup>**

		EAE		
		$\dot{a}_\alpha < 0.5$	$\dot{a}_\alpha = 0.5\text{--}1.5$	$\dot{a}_\alpha > 1.5$
BAE	$\dot{a}_\beta < 0.5$	case #1	case #2	case #3
		$r_{\text{eff}} > 0.2(0.4) \mu\text{m}$	$r_{\text{eff}} = 0.14 - 0.4(0.14\text{--}0.22) \mu\text{m}$	$r_{\text{eff}} = 0.14\text{--}0.19(-) \mu\text{m}$
		1–1095 (4–269) sr	62–720 (62–155) sr	147–277 (-)sr
		$m_1 = 0.0\text{--}0.05$	$m_1 = 0.02\text{--}0.05 (0.035\text{--}0.05)$	$m_1 = 0.015\text{--}0.05(-)$
		$m_R = 1.3\text{--}1.7$	$m_R = 1.3\text{--}1.7 (1.525\text{--}1.7)$	$m_R = 1.3 - 1.5(-)$
	$\dot{a}_\beta = 0.5\text{--}1.5$	case #4	case #5	case #6
		$r_{\text{eff}} > 0.18 \mu\text{m}$	$r_{\text{eff}} = 0.08\text{--}0.4 \mu\text{m}$	$r_{\text{eff}} = 0.08\text{--}0.22 \mu\text{m}$
		1–400 sr (4–216)	18–522 (18–219) sr	39–278 (39–154) sr
		$m_1 = 0.0\text{--}0.05$	$m_1 = 0.0\text{--}0.05$	$m_1 = 0.0\text{--}0.05$
	$m_R = 1.3\text{--}1.7$	$m_R = 1.3\text{--}1.7$	$m_R = 1.3\text{--}1.7$	
$\dot{a}_\beta > 1.5$	case #7	case #8	case #9	
	$r_{\text{eff}} = 0.18\text{--}0.55 \mu\text{m}$	$r_{\text{eff}} < 0.3 \mu\text{m}$	$r_{\text{eff}} < 0.12 \mu\text{m}$	
	2–67 (4–67) sr	9–204 (12–72) sr	9–133 (21–109) sr	
	$m_1 = 0.0\text{--}0.05$	$m_1 = 0.0\text{--}0.05 (0\text{--}0.045)$	$m_1 = 0.0\text{--}0.05$	
	$m_R = 1.4\text{--}1.7$	$m_R = 1.3\text{--}1.7 (1.4\text{--}1.7)$	$m_R = 1.3\text{--}1.7$	

<sup>a</sup>The values in brackets were obtained for the constrained SOD bank and with the CRIs shown in Table 2.

0.2% of all “constrained” data sets in our SOD. All particles in this case are small ( $r_{\text{eff}} \leq 0.22 \mu\text{m}$ ) and highly light-absorbing ( $m_1 \geq 0.035$ ).

To the best of our knowledge, we cannot find literature that shows one-modal atmospheric PSDs for which the BAE is less than 0.5 and  $r_{\text{eff}} < 0.5 \mu\text{m}$  simultaneously. The region between 0.25 and 0.7 for the BAE and  $r_{\text{eff}} \sim 0.25 \mu\text{m}$  is characterized by comparably high values of the imaginary part of the CRI, i.e.,  $m_1 \geq 0.035$ .

We find a linear correlation with a correlation coefficient of  $R^2 = 0.995$  for  $s$  versus  $\alpha(355)$  [see Fig. 1(c)]. If we use all values of our SOD bank, we find for the RC that  $a_s = 1.73$  and  $b_s = -0.09 \approx 0$ . These numbers agree with our theoretical estimate, see Eq. (18a).

A more thorough analysis of the SOD bank shows the following variations

$$a_s \in [1.33; 1.78], \quad b_s \in [-0.09; 0.03]. \quad (34)$$

These values depend, for example, on  $\sigma_{\text{Gauss}}$ . The strong correlation guarantees that the surface-area concentration can

**Table 2. Real and Imaginary Parts of the CRI used in the Constrained SOD Bank<sup>a</sup>**

		$m_R$				
		1.3–1.35	1.35–1.4	1.4–1.45	1.45–1.5	1.5–1.7
$m_1$	0.03–0.05	gray	gray	gray	gray	red
	0.015–0.03	gray	gray	gray	red	red
	0.01–0.015	gray	black	red	red	red
	0.005–0.01	gray	black	red	red	red
	0.001–0.005	black	black	red	red	gray
	0–0.001	black	black	gray	gray	gray

<sup>a</sup>The red cells describe particles with strong light-absorption and a comparably high real part ( $m_R > 1.4$ ). The black cells describe particles of low light absorption and a comparably low real part ( $m_R \leq 1.4$ ).

be directly estimated from the extinction with an uncertainty that is less than the measurement error.

Furthermore, Fig. 1(c) shows that  $v/r_{\text{eff}}$  and  $\alpha$  as well as  $n(r_{\text{mean}}^2 + \sigma^2)$  and  $\alpha$  are linearly correlated, with a correlation coefficient of  $R^2 = 0.995$ . We find the following values of the RCs for the interval  $\sigma_{\text{Gauss}} \in [1.35; 2.55]$ :

$$a_v \in [0.4; 0.6], \quad b_v \in [-0.03; 0.01], \quad (35)$$

$$a_n \in [0.11; 0.14], \quad b_n \in [-0.05; 0.002]. \quad (36)$$

These numbers fulfill the conditions in Eq. (18).

We stress the fact that the linear correlations of the IPs of  $s$  and  $v/r_{\text{eff}}$  and  $n(r_{\text{mean}}^2 + \sigma^2)$  versus  $\alpha$ , and the RCs given by Eq. (18) hold true for any distribution law that describes the PSDs, particularly if the PSDs are mono- and bimodal. However, the RCs defined by Eqs. (33a)–(33c) are not valid if we consider correlations of effective radius versus EAE and BAE for arbitrarily shaped PSDs.

To illustrate, for example, the contribution of coarse-mode particles in these correlation relations, we calculated the OD and the IPs of bimodal PSDs. The mean radii of the fine mode and the coarse mode of these bimodal PSDs were  $r_{\text{Gauss}} = 0.1$  and  $1 \mu\text{m}$ , respectively. The mode width was  $\sigma_{\text{Gauss}} = 1.5$ , and the CRI was  $m = 1.45 - i0.005$  for both modes in that study. We used these parameters to generate five PSDs by considering different fractions of the coarse-mode number concentration (respectively, volume concentration). We used 0.05% (33%), 0.10% (50%), 0.15% (60%), 0.20% (67%), and 1.0% (91%) for the number (volume) concentration of the coarse mode. The results are shown as stars in Fig. 1.

We see that even relatively large EAEs close to 1 do not only describe small particles with effective radii of less than  $0.25 \mu\text{m}$ , see Fig. 1(a) (stars). Indeed, in the case of monomodal PSDs, the effective radius does not exceed  $0.25 \mu\text{m}$  at  $\dot{a}_\alpha > 1$ , whereas for bimodal PSDs an EAE equal to one leads to  $r_{\text{eff}} > 0.5 \mu\text{m}$ . The reason for this effect is that the contribution of coarse-mode particles to the total particle volume concentration of

the bimodal PSD is much more than the contribution of coarse-mode particles to the total extinction coefficients at 355 and 532 nm. In other words, the sensitivity of the effective radius to the fraction of coarse-mode particles is higher than the sensitivity of the extinction coefficient to that same fraction of coarse mode particles.

As a result, the RC  $a_r$  decreases from values of  $(-0.08) - (-0.19)$  to as low as  $(-0.652)$  at  $\hat{a}_\alpha > 0.5$ . The parameter  $b_r$  increases from the values of  $0.26-0.37$  to  $1.2$ . We investigated other bimodal PSDs with coarse modes, which are described by  $r_{\text{Gauss}} > 1 \mu\text{m}$  and  $\sigma_{\text{Gauss}} > 1.5$ . The slope of the regression line depends on the coarse-mode effective radius (in this case  $\hat{a}_\alpha \approx 0$ ) and the fine-mode effective radius, i.e.,  $r_{\text{eff}} < 0.3 \mu\text{m}$  (in this case  $\hat{a}_\alpha > 0.5$ ).

Figure 1(c) shows that the star symbols accumulate at the origin of the coordinate axis. The effective radius of the coarse mode of the bimodal PSD is about  $1.5 \mu\text{m}$ . The maximal effective radius in our SOD is  $3 \mu\text{m}$ . The number concentration is about  $1 \text{ cm}^{-3}$  for both cases. As a result the surface-area concentration in the first case is much less than  $s$  in the second case. The extinction coefficients in both cases are defined by just particle size, which is larger for the case with  $r_{\text{eff}} = 3 \mu\text{m}$ .

Another important conclusion can be made if we look at the stars shown in Fig. 1(b). They represent bimodal PSDs with an EAE of  $\hat{a}_\alpha = 0.5-1.5$  [see Fig. 1(a)]. If we plot all monomodal PSDs with  $\hat{a}_\alpha > 1$  in Figs. 1(a) and 1(b), we would see that there is not a single point of the BAE that is less than  $0.5$  (not shown here). This is case #2, shown in Table 1. However, the position of the stars in these plots occurs quite often in practice, i.e., the BAE is less than  $0.5$ , and, simultaneously, the EAE is larger than  $1$ . This set of numbers for the BAE and EAE is only possible if the PSD contains coarse-mode particles. As we mentioned, the only exception to this rule is that the PSD contains strong light-absorbing particles for which  $m_1 \geq 0.035$ .

The next point we need to discuss in this section concerns the regression equation [Eq. (17)] at  $p = n(r_{\text{mean}}^2 + \sigma^2)$ . The disadvantage of this regression relation is that it allows for a high degree of freedom for each of the individual parameters  $n$ ,  $r_{\text{mean}}$ , and  $\sigma^2$ . There are a lot of possible combinations of these parameters that can lead to the same result for  $n(r_{\text{mean}}^2 + \sigma^2)$ . We want to constrain, for example,  $(r_{\text{mean}}^2 + \sigma^2)$  in order to find a more stable (accurate) solution for  $n$ . In that regard, we can use the inequalities described by Eqs. (31b) and (31c). However, these two inequalities work only for “narrow” PSDs and cannot be used for a bimodal PSD. In section 2, we defined “narrow” PSDs in the sense that Eqs. (29) and (8) deliver close estimations of an effective radius. In terms of logarithmic-normal functions, that means  $\sigma_{\text{Gauss}} < 2$ . The analysis of our SOD bank shows that all PSDs with  $\sigma_{\text{Gauss}} \geq 2$  (apart from PSDs with  $r_{\text{Gauss}} < 0.05 \mu\text{m}$  at  $\sigma_{\text{Gauss}} \approx 2$ ) produce EAEs that are less than  $1$ . The values of  $\sigma_{\text{Gauss}} < 2$  and  $\hat{a}_\alpha > 1$  could be treated as definitions of what we denote as “narrow” PSDs or as fine-mode fraction of the PSD. However, logarithmic-normal PSDs with  $\sigma_{\text{Gauss}} < 2$  and  $r_{\text{Gauss}} > 0.2 \mu\text{m}$  (which is “narrow”) can produce EAE values close to  $0$ . It means that an EAE above one cannot be used as a “universal definition” for the “narrowness” or monomodality of a PSD.

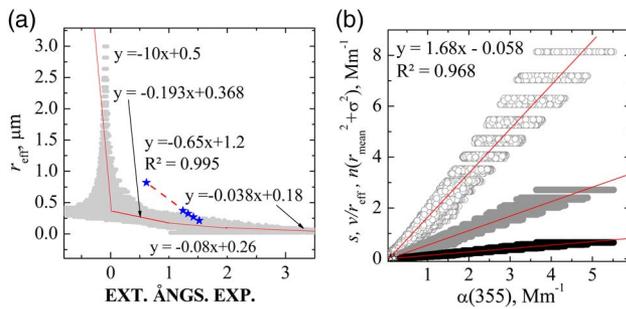
We are also interested in a flag parameter that indicates the presence of coarse-mode particles in a PSD. It is a fact that  $\langle Q_\alpha \rangle \rightarrow \sim 2$  for  $r \rightarrow \infty$ . That means that the extinction values at any wavelength converge to each other for big particles with a radius larger than  $1 \mu\text{m}$  (they present coarse-mode particles). Therefore, the “best” flag parameter we can currently find on the basis of our SOD is the ratio  $d = \alpha(532)/\alpha(355)$ . This parameter belongs to the interval  $[0.19; 1.32]$  of our SOD bank. However, it narrows to  $d \in (1; 1.07)$ , i.e.,  $\hat{a}_\alpha \in (-0.17; 0)$  in terms of the EAE, if we take into account all the cases for which  $r_{\text{eff}} > 1.1 \mu\text{m}$ . We will exploit this property of the coarse mode more intensively in the present study.

We analyzed if  $r_{\text{mean}}$  and  $\sigma$  are correlated with  $r_{\text{eff}}$ . In view of Eqs. (29) and (30), we may assume that an increase (decrease) of  $r_{\text{mean}}$  and/or  $\sigma$  leads to an increase (decrease) of  $r_{\text{eff}}$ . From a physical point of view, it is a quite natural assumption that this effect should happen. The statistical analysis of our data bank confirms this assumption.

Figure 1(d) shows that  $\sigma$  and  $r_{\text{eff}}$  are correlated with the values of  $a_\sigma = 0.22$ ,  $b_\sigma = 0.027$ , and  $R^2 = 0.96$  for monomodal PSDs. The mean and effective radii are correlated with  $R^2 = 1$  in that case, but  $\sigma_G$  is at a fixed value. Depending on  $\sigma_G$ , the RC of the mean radius  $a_m$  changes from  $0.17$  to  $0.84$ , and  $b_m$  is equal to zero. A similarly consistent behavior can be found for bimodal PSDs, see Fig. 1(d) (stars). However, in that case, the “slope” (RC) can vary significantly. The important point is that the values of the RCs, i.e.,  $a_m$  and  $a_\sigma$  converge to zero if the investigated PSD is “wide” (we define “wide” if  $\sigma_{\text{Gauss}} > 2$ ) or contains particles in the coarse mode. We can use this fact again as a flag parameter in the automated inversion software, i.e., we have the option to identify if our optical input data are affected by a bimodal PSD. We stress this point, as in general, coarse-mode particles most likely consist of irregularly shaped dust particles for which we do not have a light-scattering model that allows us to reliably reproduce measured particle backscatter coefficients.

The second part of our series of papers (part 2, simulation) is focused on error analysis. However, we can make a preliminary assessment of how measurement errors may influence the uncertainty of the PMPs. If we accept that the PMPs and the OD (obtained from the measurements) are linearly interdependent, we can substitute the extinction coefficient into Eq. (17) and the EAE into Eqs. (33a)–(33c) together with their uncertainties. It is clear that we obtain the respective PMPs with the same uncertainty as the measurement error, apart from the mathematical error, which is  $20\%$  in the case of  $s$ .

Figures 2(a) and 2(b) show the same set of parameters that are shown in Figs. 1(a) and 1(c), but the EAE and extinction values are distorted by up to  $20\%$ . For example,  $s$  and  $\alpha(355)$  are correlated but with a lower correlation coefficient of  $R^2 = 0.97$ . The  $s$  uncertainty/spread is wider than the one for the error free case, see Fig. 1(c), but Eq. (17) does not permit for the strong outliers we deal with in practice in some cases. An effective radius versus a distorted EAE is almost the same as in the case of an error-free EAE. At least there is no change of the  $r_{\text{eff}}$  uncertainties that we use in Eqs. (33a)–(33c).



**Fig. 2.** Same as Fig. 1(a) and 1(c), but (a) EAE and (b) extinction coefficients are distorted by up to 20% in the synthetic data bank.

## B. Correlation Relationships between CRI and Lidar Data Products

We investigated our SOD bank with regard to the question if the CRI is somehow correlated with lidar data products. We use the results shown in Figs. 1(a) and 1(b) for that purpose. The figures show the statistics for effective radius versus EAE and BAE, respectively. As we discussed before, for a fixed BAE, a wider spread of effective radii is possible in comparison to the spread that can be found for a fixed EAE. This feature can be explained by the strong dependence of the BAE on both the effective radius and the CRI, whereas the EAE only weakly depends on the CRI.

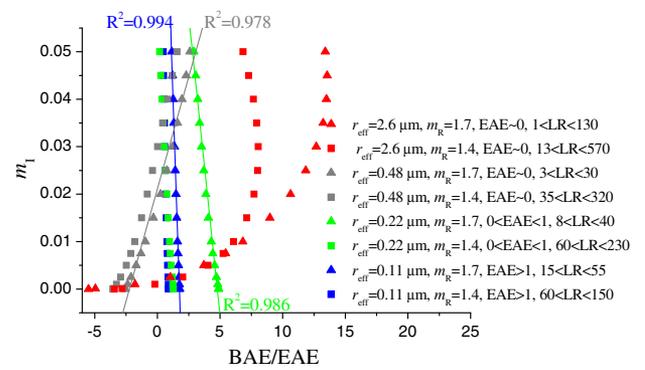
Let us consider the ratio of BAE to EAE (RBE) at the wavelength pair of 355 and 532 nm (not shown here since its behavior is very similar to the behavior of the BAE at the wavelength pair of 532 and 1064 nm) for an arbitrary value  $\tilde{a}_\alpha$ . That  $\tilde{a}_\alpha$  value corresponds to a certain set of effective radii [see Fig. 1(a)]. Apparently we can find the set of BAE values corresponding to that set of effective radii. If we fix the effective-radius value, we can expect that the RBE will depend only on the CRI.

We can simplify the statistical analysis of CRI versus RBE for a fixed effective radius if we also consider a fixed value of the real part of the CRI. The result of this approach is shown in Fig. 3. It shows the SOD bank statistics of the imaginary part of the CRI versus the RBE for a few fixed values of effective radii and real parts of the CRI. We fix the effective radius

- at  $r_{\text{eff}} = 0.11 \mu\text{m}$  which describes small particles of the fine-mode fraction and an EAE above one (green symbols),
- at  $r_{\text{eff}} = 0.22 \mu\text{m}$  for which the EAE is between zero and one, and which also describes particles of the fine-mode fraction (green symbols),
- at  $r_{\text{eff}} = 0.48 \mu\text{m}$  for which the EAE is close to 0, and which describes submicron particles (gray symbols),
- at  $r_{\text{eff}} = 2.6 \mu\text{m}$  for which the EAE is close to 0, and which describes particles of the coarse mode (red symbols).

In addition we consider real parts of the CRI at 1.4 (square) and 1.7 (triangle) for each particle size.

We see that the imaginary part and the RBE are linearly correlated with a correlation coefficient  $R^2$  that varies between 0.978 for submicron particles and 0.994 for the smallest particles (gray, green, and blue symbols). We stress the fact that a similar correlation exists between the real part of the CRI and the RBE if we set the imaginary part to a fixed value instead of



**Fig. 3.** Statistics of the SOD bank: correlation between the imaginary part of the CRI and the ratio of BAE to EAE. The triangles and squares describe the real part of 1.7 and 1.4, respectively. Blue, green, gray, and red colors describe small particles with  $r_{\text{eff}} = 0.1 \mu\text{m}$  and  $\tilde{a}_\alpha > 1$ , small particles with  $r_{\text{eff}} = 0.22 \mu\text{m}$  and  $\tilde{a}_\alpha > 1$ , submicron particles with  $r_{\text{eff}} = 0.48 \mu\text{m}$  and  $\tilde{a}_\alpha \approx 0$ , and big particles with  $r_{\text{eff}} = 2.6 \mu\text{m}$  and  $\tilde{a}_\alpha \approx 0$ , respectively.

setting the real part for “natural” aerosol particles (not shown here) to a fixed value. With regard to big particles in the coarse mode, the correlation between the CRI and the RBE is more complicated (red symbols). Furthermore in this case an ambiguous relationship appears for  $m_i$  versus RBE, for example, at RBE close to 7.5 and  $m_R = 1.4$ .

We need to understand if there is any practical value in this statistical analysis. Indeed, we allow for the fact that effective radius and real part (or imaginary part) are known simultaneously. However, we have other pieces of information which are presented by dimensionless parameters, i.e., we make use of relations among two of these parameters; such dimensionless parameters do not depend on particle concentration.

It is well known that the LR  $\alpha(\lambda)/\beta(\lambda)$  strongly depends on the CRI. We can use this property in our statistical analysis. For that reason, Fig. 3 also shows the LR intervals at  $\lambda = 355 \text{ nm}$  in the legend for each effective radius and each real part of the CRI. For example, we find the maximum LR of 55 sr at  $r_{\text{eff}} = 0.11 \mu\text{m}$  and  $m_R = 1.7$ , whereas the minimal LR is 60 sr at  $r_{\text{eff}} = 0.11 \mu\text{m}$  and  $m_R = 1.4$  (blue symbols). It means that we do not need to know the real part of the CRI if the LR is available in our analysis. Furthermore, as we discussed in the context of big particles, the interdependence  $m_i$  versus RBE is ambiguous for some RBE ranges. However, if we take into account the LR for that respective range, we find that the LR extremely changes. For example, the LR changes from 250 to 400 sr at a RBE close to 7.5 for  $m_R = 1.4$ .

On the basis of the statistical analysis of our SOD bank, we can postulate the following. Let us choose the following set of parameters of a monomodal PSD:

- effective radius,
- LR at 355 and 532 nm,
- BAE and EAE at the wavelength pairs of 355 and 532 nm.

The effective radius, the LR at 355 and 532 nm, and the BAE and EAE at the wavelength pair of 355 and 532 nm uniquely define the CRI of aerosols described by monomodal

PSDs. In other words, there is only one value of CRI which can produce that set of parameters.

That conclusion has important practical value. In fact, state-of-the-art Raman lidar and HSRL can provide  $3\beta + 2\alpha$  OD sets, and several algorithms of PMP retrievals provide us with an effective radius, see Ref. [1–5]. Taking into account such products as LR, BAE, and EAE measured with lidar and an effective radius retrieved with these available algorithms, we can put these values into our SOD bank and find the respective CRI. Of course, we deal with measurements. Therefore, retrieval errors need to be considered, but we can obtain an estimate of the uncertainty of the CRI estimation with the help of such simulation studies. If we consider different errors of the ODs and effective radii in our SOD bank, it allows us to make such estimations of the CRI.

We considered the correlation characteristics for our SOD computed on the basis of Mie theory that can be used only in the case of spherical particles. The question is if we can find modified correlation equations in the case of aspherical particles and if the postulate used in our present study can still be applied.

It is known that extinction coefficients weakly depend on particle shape. Therefore, Eqs. (7), (9), (16), and (33a)–(33c) can be applied in the general case of particles of any shape. At the same time, the postulate takes into account the LR and BAE. These parameters strongly depend on backscatter coefficients that are sensitive to particle shape. First of all, we can exclude in practical application the analysis of cases for which non-spherical particles are present. The most straightforward situation is that we only use data from instruments that measure the (linear) particle depolarization ratio at least at one wavelength. Modern aerosol lidars usually measure this parameter at 532 nm and/or 355 and/or 1064 nm. We can also set a threshold value of the linear particle depolarization ratio to a maximum value below which we still accept the OD for data analysis. Such threshold values may be feasible, as it is not only coarse-mode dust particles that may cause depolarization ratios larger than 0. Fine-mode biomass burning particles and fine-mode dust particles may also be responsible for small, but non-negligible depolarization ratios.

In our future studies, we will compute our SOD on the basis of a particle spheroid model [15] in order to investigate if the postulate can be applied in that special case. In the framework of this investigation, it will be also important to understand the significance of using depolarization information for retrievals of PMPs. However, such a study is way beyond the scope of our current work because there are too many uncertainties involved with regard to the light-scattering model that we want to use [15]. This model is simply not validated for lidar applications and/or may not even be able to reproduce certain lidar-measured parameters like the depolarization ratio and the backscatter coefficients.

#### 4. VALIDATION

In the past years, only a few experiments with simultaneous lidar and *in situ* measurements were carried out. Results of the PMP retrievals from OD taken with lidar and *in situ* instruments are presented in Refs. [2,12,16]. In the following, we attempt to validate our correlation relationships by taking into

account the statistics we collected during these experiments. Retrievals of microphysical properties as well as statistical results for big experiments like Two-Column Aerosol Project (TCAP) have been obtained in the past several years and data have in part already been published [9,17]. That means that there exists a rather robust basis for carrying out validation/comparison studies in future studies.

In Ref. [12], we investigated in detail the PMP retrievals from the SAMUM data. The statistics of all retrieval results were collected in Fig. 15 from Ref. [12]. Figure 15(b) in that publication shows that the surface-area concentration and the extinction coefficient at 355 nm are almost linearly correlated with  $a_s \approx 1.55$ ,  $b_s \approx 0$ , and  $R^2 = 0.98$ . Now, we can explain this result in view of Eqs. (17) and (18). Besides this explanation, we showed in Fig. 15(a) of Ref. [12] that effective radius and EAE are approximately linearly correlated with  $a_r \approx -0.8$ ,  $b_r \approx 1.1$ , and  $R^2 = 0.85$ . As we discussed before, such a low value of  $a_r$  indicates that the underlying PSD is bimodal and contains a significant contribution of coarse-mode particles. Therefore, we cannot investigate the CRI with the help of our SOD bank at the moment.

In Ref. [16], we investigated a measurement case of biomass burning particles (measurement from 22 July 2004). In the present contribution, we collect the statistics of the PMP retrieval results.

Figure 4 shows the results for (a)  $r_{\text{eff}}$  versus EAE,  $s$  (open circles),  $\nu/r_{\text{eff}}$  (gray bullets), and (b)  $(r_{\text{mean}}^2 + \sigma^2)n$  (black bullets) versus  $\alpha(355)$ . Also shown in Fig. 4(c) are the correlations of  $r_{\text{mean}}$  (gray bullets) and  $\sigma$  (black bullets) versus  $r_{\text{eff}}$ . We see that all IPs and data points are linearly correlated with high correlation coefficients of  $R^2 > 0.97$  [Figs. 4(a) and 4(b)]. The lowest correlation coefficient  $R^2 = 0.91$  exists between  $r_{\text{mean}}$  and  $r_{\text{eff}}$  [Fig. 4(c)].

We need to stress the fact that the RCs (see regression equations in Fig. 4) of all IPs belong to intervals we found theoretically as well as from the analysis of our SOD bank. Relatively small values of  $a_r = -0.46$  and  $a_\sigma = 0.19$  indicate that the profile contains aerosol particles that can be described by a bimodal PSD. At the same time, there are layers where values of EAE and BAE are close to 1 which points to a monomodal PSD (see Fig. 3(a) in Ref. [16]).

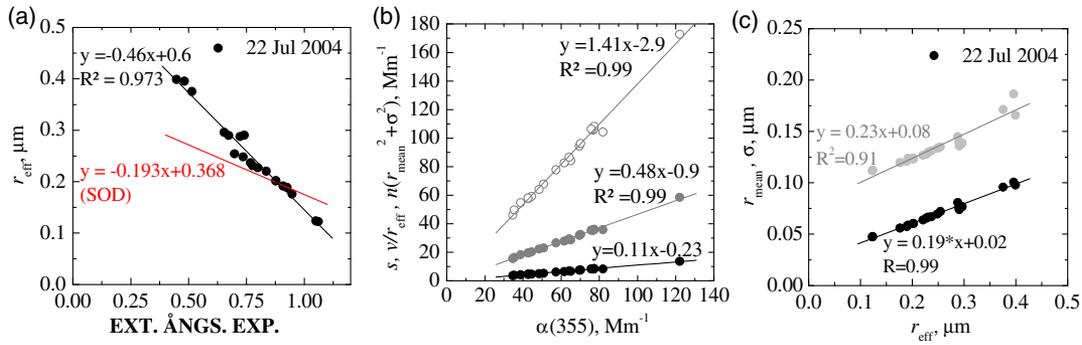
We look in more detail at the particle layer in 3.8 km height. In that height (see Fig. 2 in [16]), we find the following values:

- $\hat{a}_\alpha = 0.88$  and  $\hat{a}_\beta = 1.55$  for the wavelength pair  $\lambda = 355$  and 532 nm,
- LRs are 41 and 54 sr at  $\lambda = 355$  and 532 nm, respectively,
- $r_{\text{eff}} = 0.2 \mu\text{m}$ .

In the following, we list the values from our SOD bank that are closest to the EAEs, BAEs, and LRs measured with lidar, and the effective radius that has been retrieved with our two-dimensional regularization approach [16]:

- $\hat{a}_\alpha = 0.84$  and  $\hat{a}_\beta = 1.52$  for the wavelength pair  $\lambda = 355$  and 532 nm,
- LRs are 41 and 54 sr at  $\lambda = 355$  and 532, respectively,
- $r_{\text{eff}} = 0.21 \mu\text{m}$ .

This result means that the average discrepancy between the data from our SOD bank (look-up table) and the respective data from the experiment, defined as



**Fig. 4.** Statistics of the experimental data taken on 22 July 2004 [15]. (a)  $r_{\text{eff}}$  versus EAE together with linear fit (black line) and regression line obtained from the SOD bank for  $\hat{\alpha}_\alpha = 0.5\text{--}1$  (red line). (b)  $s$  (open circles),  $v/r_{\text{eff}}$  (gray bullets), and  $(r_{\text{mean}}^2 + \sigma^2)n$  (black bullets) versus  $\alpha(355)$  together with the linear regression lines. (c)  $r_{\text{mean}}$  (gray bullets) and  $\sigma$  (black bullets) versus  $r_{\text{eff}}$  together with the linear regression lines.

$$\rho_{\text{LUT}} = \frac{1}{5} \sum_p \frac{|p_{\text{LUT}} - p_{\text{experiment}}|}{p_{\text{LUT}}},$$

$$p = \hat{\alpha}_\omega \hat{\alpha}_{\beta(355), \beta(532), \alpha(355)/\beta(355), \alpha(532)/\beta(532)} r_{\text{eff}} \quad (37)$$

is approximately  $\rho_{\text{LUT}, \text{min}} \approx 2\%$ . According to this set of parameters, we find for the CRI from our SOD bank the value  $m = 1.475 - i0.001$ . This value is close to the value  $m = 1.52 - i0.01$  that was retrieved with our regularization approach, and which is within the uncertainties of the real ( $\pm 0.08$ ) and imaginary ( $\pm 0.01$ ) parts that we obtain from our regularization method.

With the help of our SOD bank, we can explain the reasons why the regularization approach leads to such large CRI uncertainties. For example, the next two sets of entries in the SOD bank with discrepancies [Eq. (37)] closest to  $\rho_{\text{LUT}, \text{min}}$  ( $\rho_{\text{LUT}} \approx 3.5$  and  $3.6\%$ ) give the CRIs  $m = 1.475 - i0.002$ , and  $m = 1.5 - i0.075$ . In fact, the SOD bank contains another 30 sets ( $\rho_{\text{LUT}} < 10\%$ ) of solutions for which the variation of the effective radius is  $\delta r_{\text{eff}} = 0.21 \mu\text{m} \pm 8\%$ , the variation of the real part of the CRI is  $\delta m_{\text{R}} = 1.5 \pm 0.05$ , and the variation of the imaginary part is  $\delta m_{\text{I}} = 0.001\text{--}0.02$ .

The effective radius was measured *in situ*. A value of  $0.16 \mu\text{m}$  has been reported [16]. This value is within the retrieval uncertainty of the effective radius derived with our regularization approach ( $\pm 0.1 \mu\text{m}$ ). The SOD bank relates the value  $m = 1.6 - i0.025$  to  $r_{\text{eff}} = 0.16 \mu\text{m}$ . This result explains the higher CRI measured *in situ* ( $m = 1.55 - i0.02$ ). We see that the SOD bank shows the following property: the higher the real part, the larger the imaginary part. This pattern agrees, for example, with previous research [18] and results shown in Ref. [19].

Extra information, for instance BAE  $\hat{\alpha}_{\beta(532;1064)}$  at the wavelength pair  $\lambda = 532$  and  $1064$  nm, would be useful in order to decrease the retrieval uncertainty. The measured value of the BAE for that pair of wavelengths is  $1.17$ . It means that this value describes particles of case #5 in Table 1, and the particles in this case can be described by a monomodal PSD. However, if we take into account the discrepancy of Eq. (37), including  $\hat{\alpha}_{\beta(532;1064)}$ , the structure of the solution space is not changed in the vicinity of  $\rho_{\text{LUT}, \text{min}}$ . In general, in view of the retrieval

uncertainty, we find good agreement between the results obtained from the experiment and from our correlation analysis.

We need to stress that the retrieval results in the experiments described in Refs. [2, 16, 12] were obtained by experienced staff doing manual (hand operated) data analysis, and it took significant time to make the in-depth analysis of the solution space. In contrast, if we use the correlation relationships *a priori*, as for example in the automated unsupervised inversion algorithm [9], we can considerably speed up the analysis of the solution space, and the results are more accurate. Part 2, “Improved identification of the solution space of aerosol microphysical properties derived from the inversion of profiles of lidar optical data, part 2: simulations with synthetic optical data”, of our series of articles is devoted to this question.

## 5. CONCLUSION

We present correlation relationships between particle bulk parameters and measured optical information. We find that the linear correlation between particle surface-area concentration and the extinction coefficient at  $355$  nm is approximately  $R^2 \approx 1$ . The IPs  $v/r_{\text{eff}}$  and  $n(r_{\text{mean}}^2 + \sigma^2)$  are linearly correlated with the extinction coefficient  $\alpha(355)$  and with the same correlation coefficient.

With regard to particles of the fine-mode fraction of the PSD, the linear correlation holds true between the imaginary part of the CRI and the ratio of the BAE to the extinction-related Ångström exponent for fixed values of the effective radius and the real part of the CRI. We find that for particles of any mode fraction of the PSD, the CRI is uniquely defined by the EAE and BAE, the LRs at two wavelengths, and the effective radius of the investigated PSD.

The relationships we find for particle bulk parameters were validated with an SOD bank that contains  $63,869$  sets of  $3\beta + 2\alpha$  data that we generated. We also collected statistics from experiments in which lidar and *in situ* measurements were carried out simultaneously. We find agreement between our theoretical and numerical simulations and the experimental results.

These newly found relationships could be incorporated into one of the next versions of the automated algorithm [9] in order to make the analysis of the solution space faster and more accurate.

**Funding.** NASA Langley Research Center; University of Hertfordshire (UH); Science Systems and Applications.

## REFERENCES

1. D. Müller, U. Wandinger, and A. Ansmann, "Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory," *Appl. Opt.* **38**, 2346–2357 (1999).
2. I. Veselovskii, A. Kolgotin, V. Griaiznov, D. Müller, U. Wandinger, and D. Whiteman, "Inversion with regularization for the retrieval of tropospheric aerosol parameters from multiwavelength lidar sounding," *Appl. Opt.* **41**, 3685–3699 (2002).
3. C. Böckmann, I. Miranova, D. Müller, L. Scheidenbach, and R. Nessler, "Microphysical aerosol parameters from multiwavelength lidar," *J. Opt. Soc. Am. A* **22**, 518–528 (2005).
4. E. Chemyakin, D. Müller, S. Burton, A. Kolgotin, C. Hostetler, and R. Ferrare, "Arrange and average algorithm for the retrieval of aerosols parameters from multiwavelength HSRL/Raman lidar data," *Appl. Opt.* **53**, 7252–7266 (2014).
5. D. P. Donovan and A. I. Carswell, "Principal component analysis applied to multiwavelength lidar aerosol backscatter and extinction measurements," *Appl. Opt.* **36**, 9406–9424 (1997).
6. A. N. Tikhonov and V. Y. Arsenin, eds., *Solution of Ill-Posed Problems* (Wiley, 1977).
7. A. Kolgotin and D. Müller, "Theory of inversion with two-dimensional regularization: profiles of microphysical particle properties derived from multiwavelength lidar measurements," *Appl. Opt.* **47**, 4472–4490 (2008).
8. A. Ångström, "The parameters of atmospheric turbidity," *Tellus* **16**, 64–75 (1964).
9. D. Müller, C. A. Hostetler, R. A. Ferrare, S. P. Burton, E. Chemyakin, A. Kolgotin, J. W. Hair, A. L. Cook, D. B. Harper, R. R. Rogers, R. W. Hare, C. S. Cleckner, M. D. Obland, J. Tomlinson, L. K. Berg, and B. Schmid, "Airborne multiwavelength high spectral resolution lidar (HSRL-2) observations during TCAP 2012: vertical profiles of optical and microphysical properties of a smoke/urban haze plume over the northeastern coast of the US," *Atmos. Meas. Tech.* **7**, 3487–3496 (2014).
10. N. T. O'Neill, O. Dubovik, and T. F. Eck, "Modified Ångström exponent for the characterization of submicrometer aerosols," *Appl. Opt.* **40**, 2368–2375 (2001).
11. N. T. O'Neill, T. F. Eck, B. Holben, A. Smirnov, and O. Dubovik, "Bimodal size distribution influences on the variation of Ångström derivatives in spectral and optical depth space," *J. Geophys. Res.* **106**, 9787–9806 (2001).
12. D. Müller, I. Veselovskii, A. Kolgotin, M. Tesche, A. Ansmann, and O. Dubovik, "Vertical profiles of pure dust (SAMUM 2006) and mixed-smoke-dust plumes (SAMUM 2008) inferred from inversion of multiwavelength Raman lidar data and comparison to AERONET retrievals," *Appl. Opt.* **52**, 3178–3202 (2013).
13. Z. Han, F. Moshary, Y. Wu, and B. Gross, "Use of lidar derived optical extinction and backscattering coefficients near cloud base to explore aerosol-cloud interactions," in *Proceedings of ILRC 27*, New York, USA, July 2015 (2015), paper PS-B4-17.
14. C. F. Bohren and D. R. Huffman, eds., *Absorption and Scattering of Light by Small Particles* (Wiley, 1983).
15. O. Dubovik, A. Sinyuk, T. Lapyonok, B. N. Holben, M. Mishenko, P. Yang, T. F. Eck, H. Volten, O. Muñoz, B. Veihelmann, W. J. van der Zande, J.-F. Leon, M. Sorokin, and I. Slutsker, "Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust," *J. Geophys. Res.* **111**, D11208 (2006).
16. D. Müller, A. Kolgotin, I. Mattis, A. Petzold, and A. Stohl, "Vertical profiles of microphysical particle properties derived from inversion with two-dimensional regularization of multiwavelength Raman lidar data: experiment," *Appl. Opt.* **50**, 2069–2079 (2011).
17. E. Kassianov, L. K. Berg, M. Pekour, J. Barnard, D. Chand, C. Flynn, M. Ovchinnikov, A. Sedlacek, B. Schmid, J. Shilling, J. Tomlinson, and J. Fast, "Airborne aerosol in situ measurements during TCAP: a closure study of total scattering," *Atmosphere* **6**, 1069–1101 (2015).
18. A. Kolgotin, D. Müller, E. Chemyakin, and A. Romanov, "Perspectives of the explicit retrieval of the complex refractive index of aerosols from optical data taken with lidar," in *Proceedings of ILRC 27*, New York, USA, July 2015 (2015), paper PS-B5-17.
19. D. Müller, U. Wandinger, D. Althausen, and M. Fiebig, "Comprehensive particle characterization from three-wavelength Raman-lidar observations: case study," *Appl. Opt.* **40**, 4863–4869 (2001).