

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A CORRELATION-BASED INVERSION METHOD FOR AEROSOL PROPERTY (CIMAP)
RETRIEVAL FROM AERONET MEASUREMENTS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

TAOZHONG HUANG

Norman, Oklahoma

2021

A CORRELATION-BASED INVERSION METHOD FOR AEROSOL PROPERTY (CIMAP)
RETRIEVAL FROM AERONET MEASUREMENTS

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Feng Xu, Chair

Dr. Jens Redemann

Dr. Guifu Zhang

ACKNOWLEDGMENTS

I would like to appreciate my advisor, Dr. Feng Xu, for his guidance and help throughout the two-year period of my graduate study. With his suggestions and support, I have much improved my technical skills and got more and more interest in atmospheric science and remote sensing theory. I would also like to thank Dr. Lan Gao, Dr. Connor Flynn, and David P. Nowicki, who provided useful suggestions and discussions through weekly meetings of the Atmospheric Radiation & Remote Sensing Group. It was a really great team.

I am grateful to Dr. Jens Redemann and Dr. Guifu Zhang for joining my graduate committee and proofreading my thesis. I am thankful to them for providing me with feedback and comments, which have helped further improve my work.

I would like to thank all of my family and my friends who have helped and encouraged me to overcome the difficulties. Their support brought me the courage to challenge every obscurity.

Again, I'm grateful to everyone above. Without the help from them, it is not possible to complete this graduate thesis of extending CIMAP for retrieving AERONET observations.

ABSTRACT

We developed a principal-component based inversion approach – A correlation-based Inversion Method for Aerosol Property (CIMAP), to retrieve aerosol properties from AERONET measurements of aerosol optical depths (AODs) and sky radiances. Capitalizing on the correlation in aerosol properties, CIMAP reduces the aerosol retrieval parameter space and saves retrieval time by more than 80% from adopting 7-8 principal components in the retrieval process, compared to retrievals in regular parameter space. As algorithm test, CIMAP is implemented to retrieve 60 sets of AERONET observations in several selected areas in southern California, southern Africa, and north China around Beijing, and the results are compared to the AERONET measurement and operational retrieval products. The mean absolute differences are found to be 0.004, 0.019, 0.032, 0.039, 0.003, and 0.044 μm for aerosol optical depth, single scattering albedo, Ångström exponent, real and imaginary parts of refractive index, and effective radii of whole size distribution, respectively. These differences are within or close to AERONET measurement or retrieval uncertainties.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	IV
ABSTRACT	V
1. INTRODUCTION	1
2. GENERAL STRUCTURE OF CIMAP ALGORITHM	3
2.1 PRINCIPAL COMPONENT ANALYSIS	4
2.2 AEROSOL OPTICAL PROPERTY COMPUTATION AND RADIATIVE TRANSFER MODELING	7
2.3 INVERSION	11
3. RETRIEVAL IMPLEMENTATION	12
3.1 PRE-ANALYSIS OF AERONET DATA TO DETERMINE PCS	13
3.2 DETERMINATION OF AN OPTIMAL NUMBER OF PCS	19
3.3 RETRIEVAL DEMONSTRATION	22
4. SUMMARY AND OUTLOOK	42
5. PUBLICATIONS	44
6. REFERENCE	45

1. Introduction

The radiative effects of aerosols and their interactions with clouds form the most significant uncertainty of radiative forcing in climate change studies. Many passive remote sensing techniques and instruments have been developed and provide irreplaceable observational constraints to assess aerosol-induced radiative forcing from regional to global scales. The essential products of many satellite-borne instruments, such as the Moderate Resolution Imaging Spectroradiometer (MODIS, [1][2]), Multi-angle imaging spectroradiometer (MISR, [3]), and the Polarization and Directionality of Earth's Reflectances (POLDER, [4]) instrument, include aerosol optical depth (AOD), single-scattering albedo (SSA), particle size distribution, and complex refractive index (RIR and RII) as a proxy of aerosol composition [5][6].

To validate the satellite-borne aerosol retrievals and trend analysis of aerosol variation, the ground-based Aerosol Robotic NETwork – (AERONET) has been developed as a global network of sun/sky radiometers that monitors AOD and aerosol properties [7]. The main instrument component of AERONET is a spectral radiometer that measures sun and sky radiances at a set of wavelengths from visible to the near-infrared spectral range. The direct sun measurements are made in a selection of spectral bands within 340 and 1020 nm that are sensitive to aerosol properties and have little contamination from gas absorption [7]. As AERONET's two sky observation functions, the almucantar and the principal plane measurements acquire sky radiances at a large scattering angular range at four wavelengths, including 440, 670, 870, and 1020nm. The angular measurements contain information about aerosol phase function, which further constrains the retrievals of aerosol loading, size distribution, refractive index, and non-sphericity. Since direct Sun radiance measurements provide information about aerosol optical depths, a combination of sun and sky radiance measurements brings up the remote sensing accuracy to a

high level, such as 0.01-0.015 for AOD direct measurements and ~ 0.03 for retrieved aerosol SSA for moderate and high aerosol loadings [8]-[10].

To fully capitalize on the information contained in AERONET observations, significant progress has been made on operational retrieval algorithm development toward retrieving aerosol loading, size distribution, refractive index, and non-spherical particle fraction from AERONET observations (e.g. [11][12]). With decades of effort, AERONET inversion algorithms have become highly mature, and their products are widely adopted for climate study and satellite remote sensing validation. By analyzing AERONET and other remote sensing aerosol products, significant correlations among aerosol size distribution, spectral refractive indices, and spherical particle fraction have been observed [13]. Indeed, aerosol variability is typically contributed by a finite number of prevailing aerosol types over a targeted domain. In other words, combinations of aerosol properties are unlikely to be arbitrary. Under such a context, aerosol properties are intrinsically correlated in contributing to the observed signals across multiple dimensions (angular, spectral, etc.). Mathematically, the spatial, temporal, and spectral correlations among aerosol physical quantities or quantities of other atmospheric constituents can be well captured by deploying the dominating principal components (PCs) [14]-[16]. These PCs not only preserve the correlations in original aerosol properties with high fidelity but also constitute a significantly reduced parameter space. Therefore, we propose to replace the direct retrieval of aerosol properties in regular parameter space by retrieving PCs. By doing so, retrieval efficiency and stability can be improved. By retrieving pixel-specific PC weights and spatiotemporally effective PC vectors, the correlation-based inversion method for aerosol properties (CIMAP) has been developed to invert observations of Airborne Multiangle Spectro Polarimetric Imager (AirMSPI [17]) [13]. Comparison of AOD and SSA derived from 27 data sets of AirMSPI and reference

data from AERONET indicates the mean absolute differences (MADs) of AOD and SSA to be ~ 0.029 and 0.038 , respectively [13]. Such accuracy is comparable to that achieved by the non-correlation based multi-pixel inversion approach [18], which solves for aerosol properties in regular parameter space. However, the CIMAP algorithm yields a substantial increase in retrieval efficiency as the parameter space for weights and elements of the first few dominating PCs are much smaller than that of regular aerosol parameters.

This graduate thesis applies CIMAP to retrieve ground-based AERONET observations. A simplification is made here by determining PC weights only while PC vectors are informed from the PC analysis of the aerosol climatology datasets over targeted areas. Our graduate thesis starts with a general structure of the algorithm in Section 2. Retrieval demonstration is given in Section 3. A short summary and outlook is given in Section 4.

2. General structure of CIMAP algorithm

The general structure of CIMAP is presented in table. 1. Aerosol properties ($\mathbf{a}_{\text{aerosol}}$), including spectral refractive index and volumetric size distribution from AERONET climatology over a targeted area, are used as a training data set for PC analysis. The training set generates a set of PC vectors that captures correlations in aerosol properties and the spatial and temporal mean vector of all aerosol fields (\mathbf{x}_{mean}). The dominating PCs in matrix \mathbf{v} and the mean \mathbf{x}_{mean} are then used to inform aerosol retrieval. The task of retrieval then becomes to determine an optimal set of PC weights ($\mathbf{w}_{\text{optimal}}$) that generate the aerosol parameters in logarithmic space via $\mathbf{x}_{\text{optimal}} = \mathbf{x}_{\text{mean}} + \mathbf{v} \times \mathbf{w}_{\text{optimal}}$. Further taking the exponential of $\mathbf{x}_{\text{optimal}}$ gives the vector $\mathbf{a}_{\text{aerosol,optimal}}$ that contains regular aerosol parameters in regular parameter space, namely $\mathbf{a}_{\text{aerosol,optimal}} = \exp(\mathbf{x}_{\text{optimal}})$, which in turn generates the modeled signals that fit the AERONET measurements of spectral AOD (t_1) and sky radiance (L_1) with the lowest fitting residue. A light scattering model and radiative transfer

(RT) model are adopted to model AODs and sky radiances with the input $\mathbf{a}_{\text{aerosol}}$. Following an iterative optimization process, the PC weights are adjusted until the optimal solution is achieved.

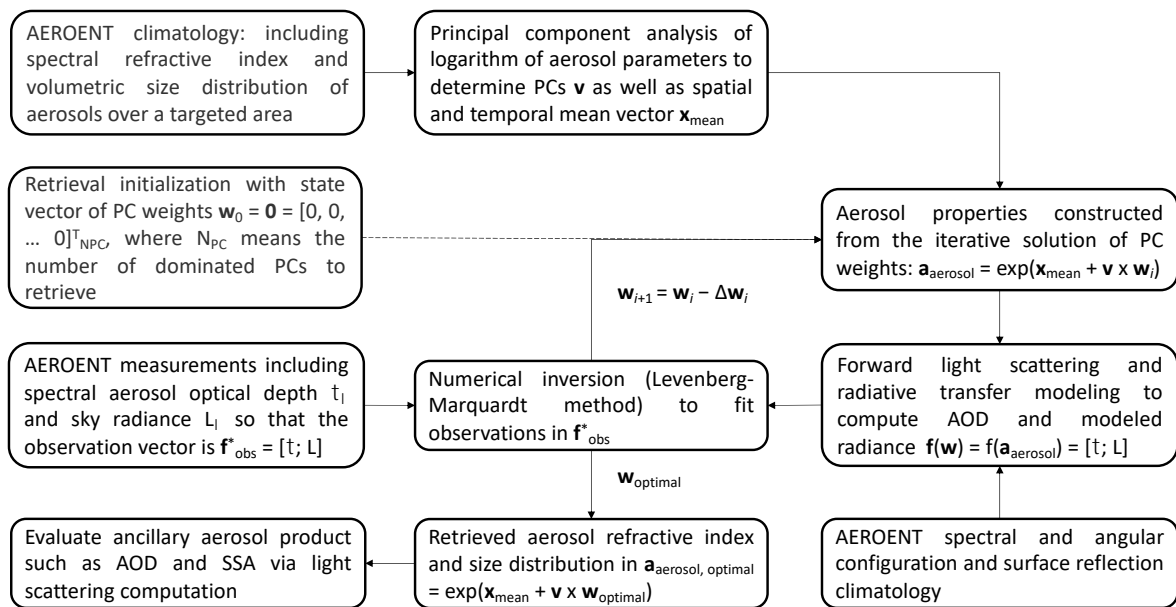


Figure. 1. General structure of the CIMAP algorithm for retrieving AERONET measurements.

In the following three sections, we will introduce several key components of CIMAP algorithm, including a PC analysis module (Section 2.1), a light scattering and radiative transfer computation module (Section 2.2), and an inversion module (section 2.3).

2.1 Principal component analysis

The correlation in aerosol properties can be captured by PC analysis of the spatially and temporally varying aerosol parameters ($\mathbf{a}_{\text{aerosol}}$). To demonstrate the effect, we perform PC analysis of AERONET aerosol climatology reported by its multiple sites in southern California. Figure 2a displays the geographical location of the sample area, which includes 21 AERONET sites. The aerosol climatology collected from these sites is utilized to generate a training dataset for PC analysis. The upper panel of Fig. 2b demonstrates the original aerosol products from the

training dataset, including volumetric size distribution resolved into 22 size bins, and refractive index at 440, 670, 870, and 1020 nm. Incorporating these aerosol products into the training dataset, we perform PC analysis of it in logarithmic space to leverage the difference of magnitudes in aerosol parameters. Then every single record of aerosol fields can be represented by PCs, namely:

$$\mathbf{x}_{\text{aerosol},i} = \log(\mathbf{a}_{\text{aerosol},i}) = \mathbf{x}_{\text{mean}} + \mathbf{v} \times \mathbf{w}_i \quad (1)$$

where $\mathbf{a}_{\text{aerosol},i}$ is the i^{th} record of the training dataset, \mathbf{x}_{mean} is a vector containing the spatial and temporal mean of aerosol parameters under PC analysis, \mathbf{v} denotes the PC matrix (each column is associated with a single PC vector), and \mathbf{w} is the matrix containing PC weights to multiply with PC vectors. Indeed, performing PC analysis in logarithmic space ensures non-negativity of the solution after reconstructing aerosol properties $\mathbf{a}_{\text{aerosol}}$ from taking the exponential of $\mathbf{x}_{\text{aerosol}}$ in Eq.(1). Moreover, the matrix \mathbf{v} in Eq. (1) can be truncated to contain only a desired number of PCs (N_{PC}) to reconstruct aerosol properties ($\mathbf{a}_{\text{aerosol}}$), namely

$$\mathbf{v} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \ \dots \ \mathbf{v}_{N_{\text{PC}}}], \quad (2)$$

where the k^{th} PC vector \mathbf{v}_k contains N_{aerosol} elements:

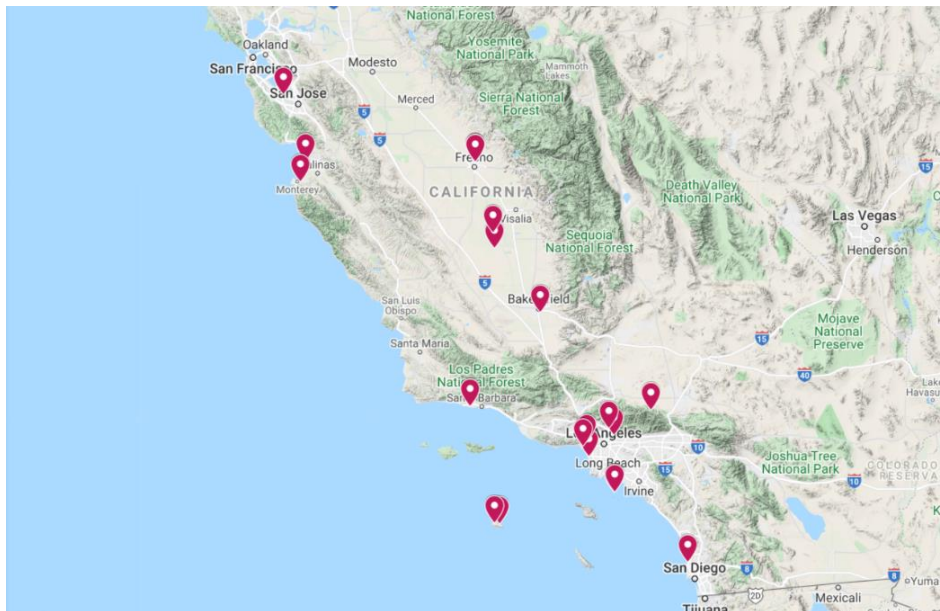
$$\mathbf{v}_k = [v_k(1); v_k(2); \dots ; v_k(N_{\text{aerosol}})]. \quad (3)$$

In the above equation, “;” means the elements in \mathbf{v}_k are vertically arranged. For a specific case of aerosol retrieval or specific aerosol record under PC analysis, the row vector consists of N_{PC} weights, namely

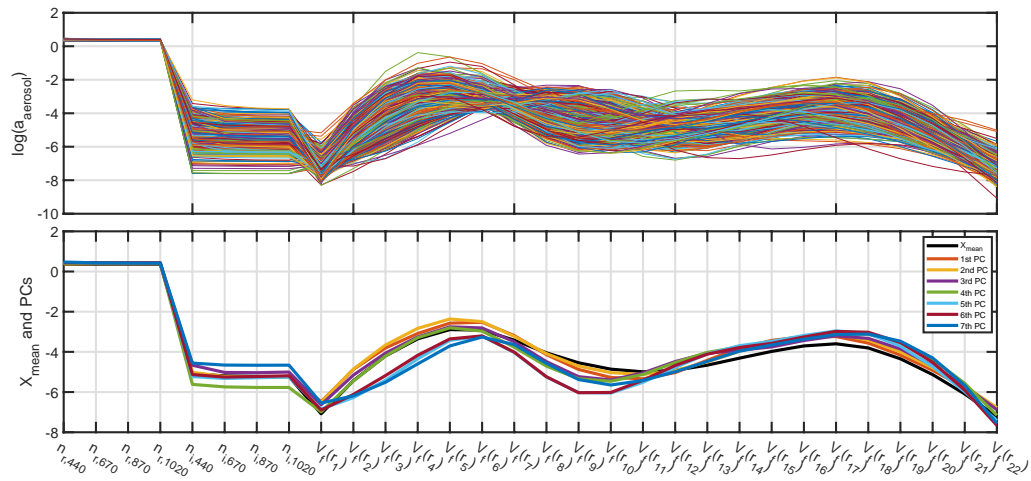
$$\mathbf{w}_s = [w_{s,1}, w_{s,2}, \dots, w_{s,N_{\text{PC}}}] \quad (4)$$

The lower panel of Fig. 2b demonstrates the mean of aerosol parameters (namely \mathbf{x}_{mean} , in the black curve) and the first seven (dominating) PCs (namely \mathbf{v}_{1-7} in different colors). Figure 2c demonstrates captured variance associated with 1-10 PCs. It is observed that 78% and 97% of

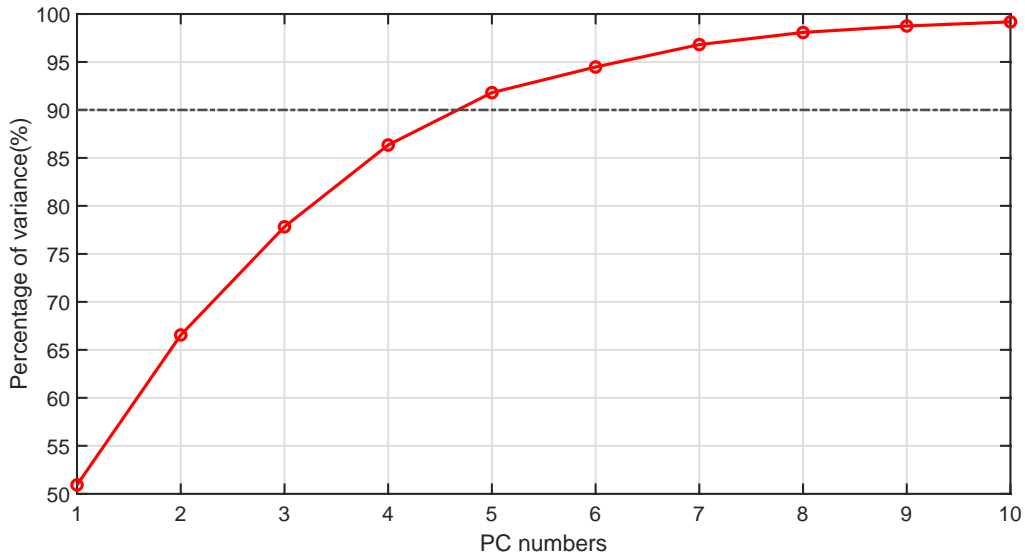
variance are captured by three and seven dominating PCs, respectively. Including more PCs will capture more variance in the original fields. However, due to the uncertainties of these retrieved aerosol properties which originate from measurement uncertainties, adding more PCs does not necessarily guarantee more fidelity to the true correlation in aerosol parameters. To be demonstrated later, we determine the optimal number of PCs based on the cost function which accounts for the measurement uncertainties and on a user specified threshold of percentage variance captured by PCs. Since most variance in the training set can be captured by a small number of dominating PCs, CIMAP is based on retrieval in PC space. In this study, PCs and the mean vector are pre-determined from target-specific training datasets. Retrieval is then implemented to resolve the weights of the few PCs that dominate the correlation.



(a)



(b)



(c)

Figure 2. a). Geographical area of AERONET sites whose aerosol climatology is used for sample PC analysis. Twenty-one sites in total cross 670 km in southern California and are labeled in crimson. The northernmost site is NASA Ames located at longitude 122.058°W and latitude 37.42°N, and the southernmost site is La Jolla located at longitude 117.25017°W and latitude 32.87063°N. b). Top panel: An example of AERONET data records in logarithmic scale under PC analysis. The data are reported by multiple AERONET sites in southern California. Every record includes real and imaginary parts of refractive index at four AERONET wavelengths 440, 670, 870, and 1020nm and volumetric size distribution discretized into 22 size bins. The bottom panel shows the spatial and temporal mean of the data record. c) Percentage of data variance captured by the first 1-10 PCs.

2.2 Aerosol optical property computation and radiative transfer modeling

AERONET provides Sun and sky radiance measurements. To fit the measurements, the radiances across relevant AERONET channels and angles are computed. To this purpose, light scattering computations of AOD, SSA, and phase matrix are performed, and the results are used as the input to radiative transfer modeling. As the input to light scattering computation, we made the same assumption as AERONET that aerosol refractive index is dependent on wavelength but independent of aerosol size [5][11]. In addition, we discretize the volumetric aerosol size distribution into a total of 22 size bins ($N_{\text{bin}} = 22$), with the lower size bound $r_{\text{min}} = 0.05 \mu\text{m}$ and upper bound $r_{\text{max}} = 15 \mu\text{m}$. The size grids are uniformly distributed in the logarithmic space so that $\Delta \ln r_{\text{grid}}$ is a constant. These assumptions are consistent with those made by the AERONET standard retrieval algorithm.

Calculations of aerosol scattering and extinction coefficients are informed by refractive index n and size distribution ($dV/d\ln r$) of aerosols. The former is a complex number consisting of a real part (n_r) and an imaginary part (n_i) so that

$$n(\lambda) = n_r(\lambda) + i \times n_i(\lambda). \quad (5)$$

Denoting the total column volumetric concentration by $C_{\text{v,column}}$ (m^3/m^2), aerosol volume concentration in i^{th} size bin ($V_f(r_i)$ $1 \leq i \leq N_{\text{bin}}$) is

$$V_f(r_i) = \frac{dV(r_i)}{d\ln r} = C_{\text{v,column}} f_i = C_{\text{v,column}} \frac{dv(r_i)}{d\ln r_i}. \quad (6)$$

where the normalized volumetric frequency of i^{th} size bin $f_i = \frac{dv(r_i)}{d\ln r_i}$. Assuming spherical aerosols,

the column extinction coefficient ($K_{\text{ext}}^{\text{aer}}(\lambda)$, m^{-1}) in the above equation is evaluated by

$$K_{\text{ext}}^{\text{aer}} = \frac{C_{\text{v,column}}}{DH} \int_{r_{\text{min}}}^{r_{\text{max}}} \frac{dv(r)}{d\ln r} \frac{3}{4\rho r^3} c_{\text{ext}}(n,r) d\ln r = \frac{C_{\text{v,column}}}{DH} \sum_{i=1}^{N_{\text{bin}}} f_i \int_{\ln r_i - \frac{\Delta \ln r}{2}}^{\ln r_i + \frac{\Delta \ln r}{2}} \frac{3}{4\rho r^3} c_{\text{ext}}(n,r) d\ln r, \quad (7)$$

where ΔH is the geometric thickness of the aerosol layer and c_{ext} is the aerosol extinction cross section (m^2) as a function of aerosol refractive index n and aerosol size r . For spherical aerosols, c_{ext} is calculated from using Mie theory [19].

In a similar way, the column scattering coefficient ($K_{\text{sca}}^{\text{aer}}(\lambda)$, m^{-1}) is evaluated by

$$K_{\text{sca}}^{\text{aer}}(l) = C_{\text{v,total}} \sum_{i=1}^{N_{\text{bin}}} f_i \int_{\ln r_i - \frac{\text{Dln}r}{2}}^{\ln r_i + \frac{\text{Dln}r}{2}} \frac{3}{4\rho r^3} c_{\text{sca}}(n, r, l) \text{dln}r, \quad (8)$$

The $(j, k)^{\text{th}}$ element of the four-by-four bulk phase matrix $\mathbf{P}^{\text{aer}}(\lambda)$ for the whole aerosol size distribution is evaluated by

$$P_{j,k}^{\text{aer}}(l; q) = \frac{\sum_{i=1}^{N_{\text{bin}}} f_i \int_{\ln r_i - \frac{\text{Dln}r}{2}}^{\ln r_i + \frac{\text{Dln}r}{2}} \frac{3}{4\rho r^3} c_{\text{sca}}(n, r, l; q) P_{j,k}(n, r, l; q) \text{dln}r}{\sum_{i=1}^{N_{\text{bin}}} f_i \int_{\ln r_i - \frac{\text{Dln}r}{2}}^{\ln r_i + \frac{\text{Dln}r}{2}} \frac{3}{4\rho r^3} c_{\text{sca}}(n, r, l; q) \text{dln}r}, \quad (9)$$

where $P_{j,k}$ in the integrand on the right-hand side of the equation is a single-particle phase matrix element.

The relationship between AOD ($\tau(\lambda)$) and volumetric extinction coefficient $K_{\text{ext}}^{\text{aer}}$ and column volume concentration (m^3/m^2) is straightforward,

$$\tau_{\text{ext}}^{\text{aer}}(\lambda) = C_{\text{v,column}} K_{\text{ext}}^{\text{aer}}(\lambda). \quad (10)$$

In combination with the scattering coefficient ($K_{\text{sca}}^{\text{aer}}$), the SSA is calculated by

$$\omega_0^{\text{aer}}(\lambda) = \frac{K_{\text{sca}}^{\text{aer}}(\lambda)}{K_{\text{ext}}^{\text{aer}}(\lambda)}. \quad (11)$$

Further on, the optical depth $\tau_{\text{ext}}^{\text{total}}(\lambda)$, single scattering albedo $\omega_0^{\text{total}}(\lambda)$ and phase matrix $P_{ij}^{\text{total}}(\theta; \lambda)$ of the atmospheric layer that contains both aerosols and air molecules are evaluated by,

$$\tau_{\text{ext}}^{\text{total}}(\lambda) = \tau_{\text{ext}}^{\text{aer}}(\lambda) + \tau_{\text{ext}}^{\text{mol}}(\lambda), \quad (12)$$

$$\omega_0^{\text{total}}(\lambda) = \frac{\tau_{\text{ext}}^{\text{aer}}(\lambda)\omega_0^{\text{aerosol}}(\lambda) + \tau_{\text{ext}}^{\text{mol}}(\lambda)\omega_0^{\text{mol}}(\lambda)}{\tau_{\text{ext}}^{\text{total}}(\lambda)}, \quad (13)$$

and

$$P_{j,k}^{\text{total}}(\lambda; \theta) = \frac{\tau_{\text{ext}}^{\text{aer}}(\lambda)\omega_0^{\text{aer}}(\lambda)P_{j,k}^{\text{aer}}(\lambda; \theta) + \tau_{\text{ext}}^{\text{mol}}(\lambda)\omega_0^{\text{mol}}(\lambda)P_{j,k}^{\text{mol}}(\lambda; \theta)}{\tau_{\text{ext}}^{\text{aer}}(\lambda)\omega_0^{\text{aer}}(\lambda) + \tau_{\text{ext}}^{\text{mol}}(\lambda)\omega_0^{\text{mol}}(\lambda)}, \quad (14)$$

respectively. In the absence of absorption, molecule single scattering albedo equals unity (namely $\omega_0^{\text{mol}}=1$).

Surface reflectance is another key input to radiative transfer modeling. CIMAP builds in a Ross-Li model for Bidirectional Reflectance Distribution Function (BRDF), which has the following form [20]:

$$\rho(\lambda; \omega_i, \omega_v) = f_{\text{iso}} + f_{\text{vol},\lambda} K_{\text{vol}}(\omega_i; \omega_v) + f_{\text{geo},\lambda} K_{\text{geo}}(\omega_i; \omega_v) \quad (15)$$

where f_{iso} , f_{vol} and f_{geo} are spectral weights of isotropic kernel, volumetric kernel (K_{vol}), and geometric kernel (K_{geo}) of surface reflection, respectively, ω_i represents the Sun incidence geometry composed of the incident zenith and azimuthal angles (θ_i, ϕ_i) and ω_v represents the view geometry composed of view zenith and azimuthal angles (θ_v, ϕ_v). We adopted RossThick kernel [21] for K_{vol} and Li-Sparse kernel [22] for K_{geo} . These two kernels are dependent on incidence or view geometries, but independent of wavelength.

The above optical properties of the atmospheric layer, including AOD, SSA, and phase matrix and surface reflection, are then used as input to the Markov chain radiative transfer (RT) model coupled with adding doubling method (MarCh-AD, [23][24]) to simulate the downward

diffuse radiances. The MarCh-AD RT model has been used for AirMSPI aerosol retrieval [24] and adapted to AERONET angular and spectral configurations in this work.

2.3 Inversion

A numerical inversion module is used to adjust aerosol parameters so that the modeled spectral sky radiances fit observations with the lowest residue. The retrieval parameters are contained in a state vector. As the kernel of our inversion module, we use the Levenberg-Marquardt (LM) algorithm [25][26]. The LM algorithm combines the strengths of the Gauss-Newton method and gradient descent method by incorporating a regularization term into the normal equation system. The optimal solution is approached in an iterative way, namely the increment of solution ($\Delta \mathbf{w}_q$) following q^{th} iteration is derived by solving the following linear system of equations in matrix form:

$$[\mathbf{J}^T \mathbf{W} \mathbf{J} + \lambda_q \text{diag}(\mathbf{J}^T \mathbf{W} \mathbf{J})] \Delta \mathbf{w}_q = \mathbf{J}^T \mathbf{W} [\mathbf{f}(\mathbf{w}_q) - \mathbf{f}_{\text{obs}}^*]. \quad (16)$$

where \mathbf{J} is the Jacobian matrix with its element evaluated by $J_{ij} = \partial f_i(\mathbf{x}_q) / \partial x_j$ and \mathbf{W} is a weighting matrix with its $(i, i)^{\text{th}}$ diagonal element evaluated by

$$W_{ii} = c \cdot \sigma_{ii}^{-2}. \quad (17)$$

In the above equation, c is a constant number equals to $N_\tau/2$ and $N_L/2$ for AERONET AOD and radiance signals, respectively, where N_τ and N_L are the total numbers of AOD and radiance signals, respectively. Assuming no correlations among errors in observational signals, the off-diagonal terms of the weighting matrix is zero. $\mathbf{f}_{\text{obs}}^*$ is the observational vector containing all measurements, including AODs and sky radiances. $\mathbf{f}(\mathbf{w}_q)$ is the model fit with q^{th} iterative solution of PC weights \mathbf{w}_q , which is associated with the aerosol solution $\mathbf{a}_{\text{aerosol},q} = \exp(\mathbf{x}_{\text{mean}} + \mathbf{v} \times \mathbf{w}_q)$. Note that in Eq.(16), the damping factor λ_q evolves during the iterative process. If the solution approaches the

solution in a fast manner, then λ_q can be adjusted to be small. In this case, the LM algorithm is close to the Gauss-Newton algorithm. But if convergence is slow, we increase the value of λ_q to reduce the impact of the term $\mathbf{J}^T \mathbf{W} \mathbf{J}$, so the solution is dominated by the term $\mathbf{J}^T \mathbf{W} [\mathbf{f}(\mathbf{w}_q) - \mathbf{f}_{\text{obs}}^*]$. This essentially means the LM algorithm is similar to the steepest descent method. Indeed, the matrix $\text{diag}(\mathbf{J}^T \mathbf{W} \mathbf{J})$ in Eq.(16) was suggested by Fletcher [27]. It scales each component of the gradient according to the curvature to ensure larger movement along with the directions where the gradient is smaller. The scaling operation avoids slow convergence in the direction of a small gradient. There are many discussions about the proper choice of the damping parameter λ_q . We followed the recommendation of Marquardt [26] by initializing the damping factor with a value λ_0 and a scale factor $\nu > 1$ and computing the residual cost function after an iteration. We then compare the cost functions of two successive iterations. If the cost function is reduced with the updated solution

$$\mathbf{w}_{q+1} = \mathbf{w}_q - \Delta \mathbf{w}_q, \quad (18)$$

then λ_0 is updated to be λ_0/ν for the next iteration; otherwise, λ_0 is updated to be $\lambda_0\nu$. In our current study, the scale factor ν is empirically determined as 10. As a measure of the degree of fitting, the Chi-square function form (c^2) of the cost function is defined with respect to the measurement uncertainties associated with the solution from q^{th} iteration:

$$\chi^2(q) = [\mathbf{f}(\mathbf{w}_q) - \mathbf{f}_{\text{obs}}^*]^T \mathbf{W} [\mathbf{f}(\mathbf{w}_q) - \mathbf{f}_{\text{obs}}^*]. \quad (19)$$

3. Retrieval implementation

The observational vector consists of spectral AODs and sky radiances measured by AERONET at multiple angles in almucantar scan configuration. In this case, $\mathbf{f}_{\text{obs}}^*$ is expressed as

$$\begin{aligned}
\mathbf{f}_{\text{obs}}^* = & \log [\tau(\lambda_1); L(\lambda_1; \theta_1; \phi_1); L(\lambda_1; \theta_2; \phi_2); \dots; L(\lambda_1; \theta_{N_\theta}; \phi_{N_\theta}); \dots \\
& \tau(\lambda_2); L(\lambda_2; \theta_1; \phi_1); L(\lambda_2; \theta_2; \phi_2); \dots; L(\lambda_2; \theta_{N_\theta}; \phi_{N_\theta}); \dots \\
& \tau(\lambda_3); L(\lambda_3; \theta_1; \phi_1); L(\lambda_3; \theta_2; \phi_2); \dots; L(\lambda_3; \theta_{N_\theta}; \phi_{N_\theta}); \dots \\
& \tau(\lambda_4); L(\lambda_4; \theta_1; \phi_1); L(\lambda_4; \theta_2; \phi_2); \dots; L(\lambda_4; \theta_{N_\theta}; \phi_{N_\theta})],
\end{aligned} \tag{20}$$

where in $\lambda_1 = 440\text{nm}$, $\lambda_2=670\text{nm}$, $\lambda_3=870\text{nm}$, and $\lambda_4=1020\text{nm}$, N_θ is the number of view angles at which sky radiances are measured. To evaluate the weighting matrix, the absolute uncertainty of AOD measurement is assumed to be 0.015 (adapted from an estimate of 0.01-0.02 in [9]), and the relative uncertainty of radiance measurement is assumed to be 5% (a conservative estimate in [5]). In the above equation, the radiance is converted into logarithmic space to leverage the differences in the magnitudes of signals. Note that though the CIMAP algorithm allows more complicated surface reflection to be considered via the Ross-Li BRDF model (Eq. (15)), its application to AERONET data retrieval assumes the surface to be Lambertian so that the spectral albedos reported in AERONET aerosol inversion products are used directly. Indeed, our test using the BRDF retrieved by GRASP algorithm [28] indicates minimal improvements for all cases under the current study.

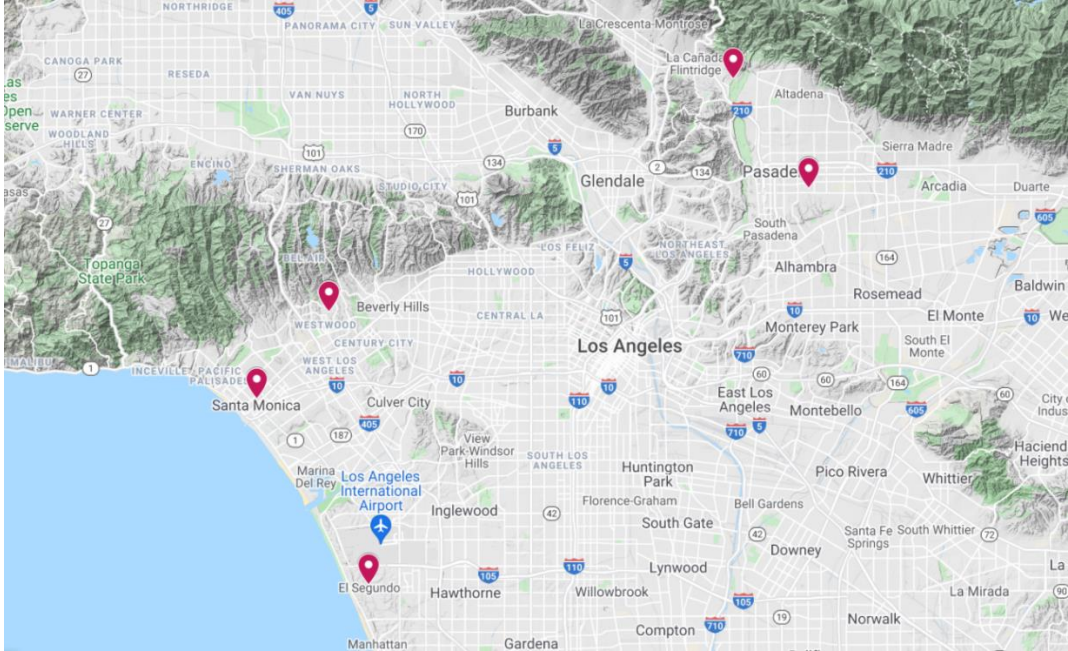
3.1 Pre-analysis of AERONET data to determine PCs

We test CIMAP retrieval of anthropogenic aerosols, biomass burning aerosols, and particulate air pollutants using a total of 60 sets of AERONET measurements in a southern California area (~150km domain), a southern Africa area (~1400km domain), and an area in north China around Beijing (~1500km domain). A total of 20 aerosol scenarios are selected for each area. These scenarios cover a diversity of aerosol loading, absorption, and size distribution. Figure 3 demonstrates the geographical locations of AERONET sites where the 60 measurements are

collected. The southern California area, southern Africa area, and north China area are displayed in Fig. 3a, 3b, and 3c, respectively. The coordinates of AERONET are labeled in crimson. AERONET site Caltech, UCLA, Santa Monica, El Segundo, and MISR-JPL are selected for southern California area; Namibe, Mwinilunga, Lubango, Windpoort, Tsumkwe, and Huambo are selected for southern Africa area; Beijing, XiangHe, AOE Baotou, Xinglong, Yulin, and Yufa PEK are chosen for north China area. Every region is selected separately to maintain the spatial correlations required by PC analysis. Note that the training dataset mentioned in Section 2.1 and displayed in Fig. 2 differs from the southern California dataset used for CIMAP retrieval demonstration. The dataset for CIMAP retrieval demonstration is a subset of the training dataset, which reduces the spatial climatology and targets a more precise area around Los Angeles (longitude 118.28°W and latitude 34.038°N). Tables 1-3 list the measurement information of the selected cases. For example, the second and third columns of these tables give the names and locations of selected AERONET sites in the three areas, respectively. The first step is to perform PC analysis of aerosol climatology (AERONET Version 2 products) over the targeted areas. All historical data, each having a full record of aerosol refractive index at the four AERONET wavelengths and volumetric size distribution discretized into 22 size grids, are analyzed to determine the PC vectors. For a given aerosol record $\mathbf{a}_{\text{aerosol}}$, it is expressed as

$$\mathbf{a}_{\text{aerosol}} = [n_r(\lambda_1), n_r(\lambda_2), n_r(\lambda_3), n_r(\lambda_4), n_i(\lambda_1), n_i(\lambda_2), n_i(\lambda_3), n_i(\lambda_4), \frac{dV(r_1)}{d\ln r_1}, \frac{dV(r_2)}{d\ln r_2}, \dots, \frac{dV(r_{22})}{d\ln r_{22}}]. \quad (21)$$

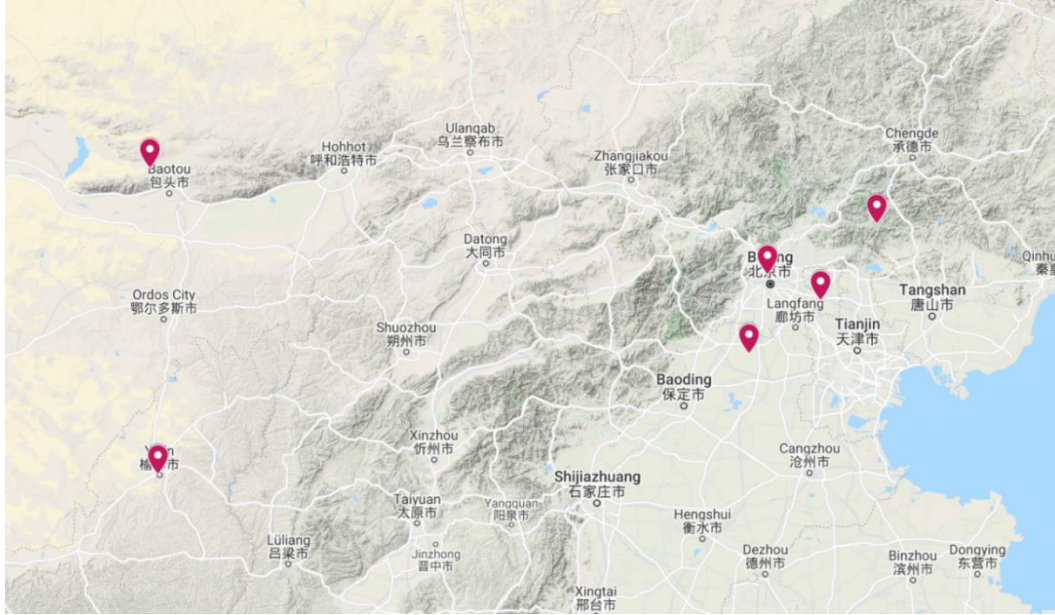
To simplify our algorithm, we excluded non-spherical particles [29] in our retrieval. Therefore, PC analysis of aerosol climatology was performed only over the cases with volumetric spherical particle fraction more than 95%.



(a)



(b)



(c)

Figure.3. a). Geographical locations of selected southern California area including AERONET site Caltech, UCLA, Santa Monica, El Segundo, and MISR-JPL (~150km domain); b). Geographical locations of selected southern Africa area including AERONET site Namibe, Mwinilunga, Lubango, Windpoort, Tsumkwe, and Huambo (~1400km domain); c). Geographical locations of selected north China area including AERONET site Beijing, XiangHe, AOE Baotou, Xinglong, Yulin, and Yufa PEK (~1400km domain).

Table 1. Information of AERONET measurements to test CIMAP retrieval (sites are in a southern California area).

Case#	AEROENT Site	Lon(°W), Lat(°N)	Sun angle (°)	Date	Measurement Time (UTC)	AOD (440nm)
1	Caltech-T1	118.13, 34.14	70.61	04:03:2011	00:09:08	0.480
2	Caltech-T2	118.13, 34.14	57.37	02:06:2013	22:08:50	0.469
3	Caltech-T3	118.13, 34.14	61.61	19:01:2011	22:05:06	0.561
4	Caltech-T4	118.13, 34.14	69.44	19:01:2011	23:05:05	0.460
5	UCLA-T1	118.45, 34.07	60.44	26:07:2006	00:28:05	0.428
6	UCLA-T2	118.45, 34.07	76.22	10:10:2006	00:15:16	0.403
7	UCLA-T3	118.45, 34.07	70.66	14:03:2000	00:21:30	0.416
8	UCLA-T4	118.45, 34.07	60.12	06:04:2004	23:49:37	0.660
9	UCLA-T5	118.45, 34.07	70.63	07:04:2004	00:41:09	0.472
10	UCLA-T6	118.45, 34.07	75.61	07:04:2004	01:05:19	0.525
11	Santa Monica-T1	118.47, 34.02	76.53	30:06:2013	01:54:00	0.428
12	Santa Monica-T2	118.47, 34.02	59.69	30:06:2013	15:25:16	0.407
13	Santa Monica-T3	118.47, 34.02	77.69	11:09:2013	01:03:49	0.487
14	Santa Monica-T4	118.47, 34.02	71.20	15:11:2013	23:00:01	0.505
15	El Segundo-T1	118.38, 33.91	57.05	11:09:2013	23:20:20	0.525
16	El Segundo-T2	118.38, 33.91	70.02	21:01:2015	16:56:05	0.427
17	El Segundo-T3	118.38, 33.91	67.22	21:01:2015	17:15:06	0.414
18	MISR-JPL-T1	118.17, 34.20	70.60	03:04:2002	00:36:26	0.480
19	MISR-JPL-T2	118.17, 34.20	60.06	03:04:2002	23:45:11	0.469
20	MISR-JPL-T3	118.17, 34.20	70.60	04:04:2002	00:37:12	0.561

Table 2. Same as Table 1 but for AERONET sites in a southern Africa area

Case#	AEROENT Site	Lon(°E), Lat(°S)	Sun angle (°)	Date	Measurement Time (UTC)	AOD (440nm)
1	Namibe-T1	12.178, 15.159	57.52	23:06:2016	08:18:52	0.461
2	Namibe-T2	12.178, 15.159	53.83	26:06:2016	08:39:51	0.460
3	Namibe-T3	12.178, 15.159	50.34	26:06:2016	13:26:04	0.429
4	Namibe-T4	12.178, 15.159	64.84	26:06:2016	14:47:59	0.409
5	Namibe-T5	12.178, 15.159	73.10	26:06:2016	15:28:53	0.442
6	Namibe-T6	12.178, 15.159	76.99	26:06:2016	15:47:33	0.455
7	Mwinilunga-T1	24.431, 11.740	54.26	10:08:2000	13:37:36	0.954
8	Mwinilunga-T2	24.431, 11.740	60.31	10:08:2000	14:05:24	0.938
9	Lubango-T1	13.445, 14.958	63.28	25:07:2016	14:48:49	0.452
10	Lubango-T2	13.445, 14.958	71.40	25:07:2016	15:27:23	0.437
11	Lubango-T3	13.445, 14.958	61.52	22:07:2017	14:38:27	0.403
12	Windpoort-T1	15.483, 19.366	76.61	27:06:2017	15:22:34	0.470
13	Tsumkwe-T1	20.442, 19.617	58.83	09:09:2017	14:12:47	0.445
14	Tsumkwe-T2	20.442, 19.617	63.97	09:09:2017	14:35:43	0.438
15	Tsumkwe-T3	20.442, 19.617	71.66	09:09:2017	15:09:29	0.461
16	Tsumkwe-T4	20.442, 19.617	56.05	26:09:2017	06:52:34	0.660
17	Tsumkwe-T5	20.442, 19.617	56.82	26:09:2017	14:09:58	0.482
18	Huambo-T1	15.705, 12.868	57.05	27:07:2018	14:15:39	0.408
19	Huambo-T2	15.705, 12.868	62.33	27:07:2018	14:41:05	0.450
20	Huambo-T3	15.705, 12.868	71.37	30:07:2018	15:24:10	0.548

Table 3. Same as Table 1 but for AERONET sites in a north China area around Beijing

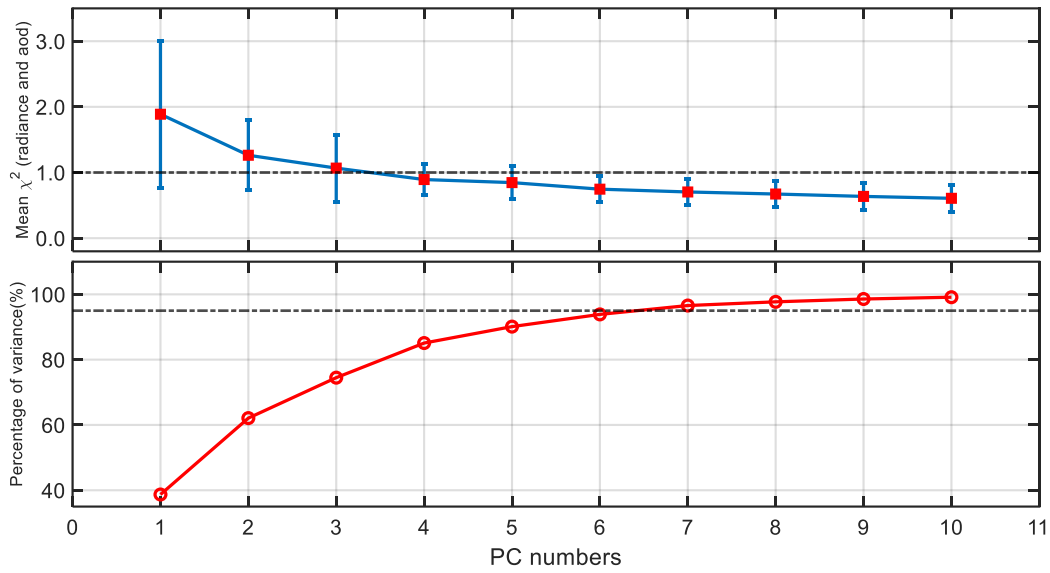
Case#	AERONET Site	Lon(°E), Lat(°N)	Sun angle (°)	Date	Measurement Time (UTC)	AOD (440nm)
1	Beijing-T1	116.381, 39.977	59.56	06:06:2002	23:38:53	1.973
2	Beijing-T2	116.381, 39.977	75.74	05:05:2004	09:51:14	0.542
3	Beijing-T3	116.381, 39.977	73.07	19:11:2008	07:02:14	0.422
4	Beijing-T4	116.381, 39.977	73.99	25:11:2010	07:03:23	0.441
5	Beijing-T5	116.381, 39.977	74.33	04:01:2011	01:27:34	0.629
6	XiangHe-T1	116.962, 39.754	59.64	10:09:2004	00:33:16	0.807
7	XiangHe-T2	116.962, 39.754	53.56	12:02:2007	04:28:50	1.339
8	XiangHe-T3	116.962, 39.754	51.41	01:10:2012	06:04:00	0.523
9	XiangHe-T4	116.962, 39.754	53.96	01:04:2013	07:18:22	0.613
10	AOE Baotou-T1	109.629, 40.852	54.93	25:10:2013	05:28:18	0.560
11	AOE Baotou-T2	109.629, 40.852	60.09	25:10:2013	06:28:13	0.478
12	Xinglong-T1	117.578, 40.396	53.56	14:05:2006	00:15:53	0.635
13	Xinglong-T2	117.578, 40.396	75.70	30:03:2008	09:12:42	0.404
14	Xinglong-T3	117.578, 40.396	57.90	05:02:2010	03:26:21	0.900
15	Yulin-T1	109.717, 38.283	75.75	27:05:2002	10:31:59	0.403
16	Yulin-T2	109.717, 38.283	53.67	06:07:2002	00:42:25	0.934
17	Yulin-T3	109.717, 38.283	59.67	19:08:2002	00:39:18	1.083
18	Yufa PEK-T1	116.184, 39.309	75.20	14:08:2006	22:50:18	0.568
19	Yufa PEK-T2	116.184, 39.309	60.53	16:08:2006	08:31:38	0.731
20	Yufa PEK-T3	116.184, 39.309	54.55	24:08:2006	07:49:25	1.610

3.2 Determination of the optimal number of PCs

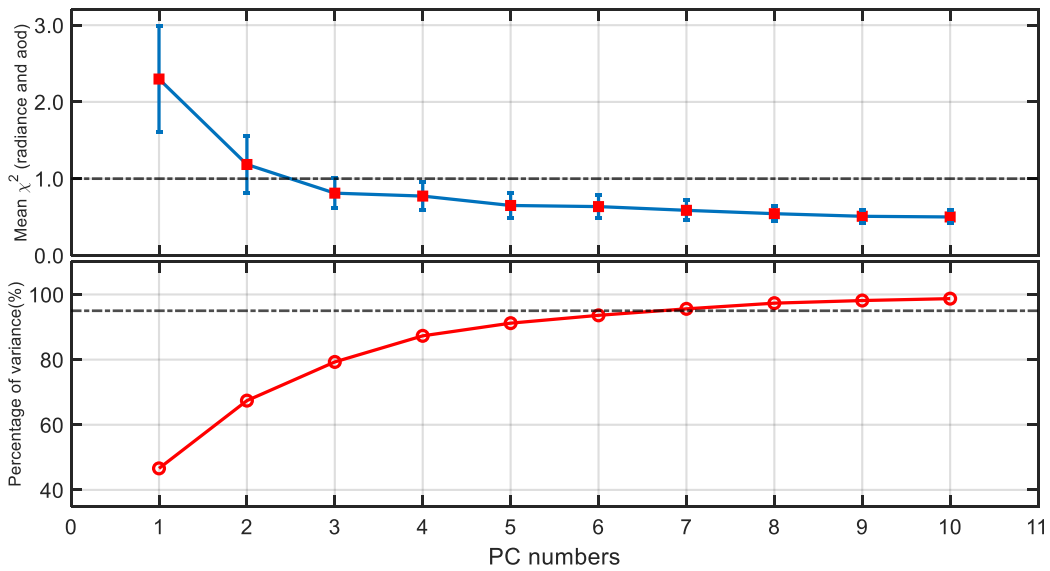
PCs are mutually orthogonal. Their number adopted by CIMAP is a critical configurational parameter. In the extreme case that the PC number equals to the number of aerosol parameters (namely $N_{PC} = N_{aerosol}$), the algorithm essentially degenerates to assuming no correlations in

aerosol properties. This is because an arbitrary set of regular aerosol parameters can be constructed from N_{aerosol} mutually orthogonal PCs, and the retrieval parameter space is the same as that in the traditional AERONET optimization. There is no gain of retrieval efficiency in this case. On the other hand, the imposed correlation constraint will be too strong if too few PCs are involved in retrieval. Consequently, retrieval is impacted by the errors in the pre-determined PC vectors from the training dataset for the targeted area. Therefore, the determination of an optimal number of PCs needs to leverage the capitalization of correlations in aerosol properties and maximization of inversion efficiency. To this purpose, we perform a test of CIMAP retrieval using different numbers of PCs and investigate the dependence of fitting residue (evaluated by Eq.(19)) on the numbers of PC. Taking CIMAP retrievals of 20 scenarios over an area in southern California as an example, the upper panel of Fig. 4a demonstrates the dependence of fitting residue on the number of PCs. The vertical bar indicates the standard deviation of the residue from the statistics of 20 randomly selected retrieval cases. With the addition of PCs into retrieval, a gradually descending trend of fitting residues can be observed. However, using a total of >6 PCs will bring the fitting errors to be within measurement uncertainties such that the fitting residue falls below the instrumental measurement uncertainties (namely $\chi^2 \leq 1$). Further requiring $>95\%$ aerosol variance over a targeted area to be captured by PCs, we adopt 7 PCs as the minimum to keep enough data variance, which is shown in the bottom panel in Fig. 4a. Hence, we adopt 7 PCs to demonstrate CIMAP retrievals of AERONET observations over southern California. In Fig. 4b and 4c, we demonstrate the minimum PCs which match each criterion for the rest two areas. In the upper panel of Fig. 4b and 4c, a similar PC analysis indicates that the first 4 and 8 PCs ensure $\chi^2 \leq 1$ in the southern Africa area and the north China area, respectively. A similar retrieval test indicates that the first 7 PCs capture $>95\%$ variance of aerosols in both southern Africa area and

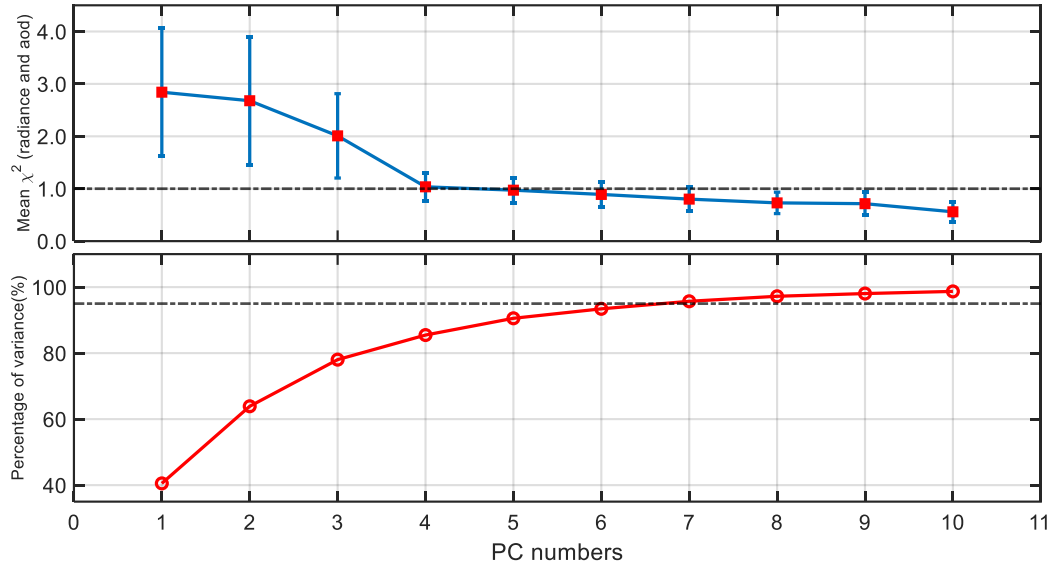
the north China area, respectively, which is shown in bottom panel of Fig. 4a and 4b. Therefore, we used the first 7 and 8 PCs, which are the larger PCs that satisfy both criteria, as the optimal PC number for CIMAP retrieval of southern Africa area and the north China area.



(a)



(b)



(c)

Figure. 4. Upper panels: Chi-square (χ^2) fitting errors as a function of numbers of PCs adopted in CIMAP retrieval. The square dots are the average of a total of 20 randomly selected retrievals of AERONET measurements in the a). southern California area; b). southern Africa area; c). northern China area. The vertical bar indicates the spread (1σ) of the fitting errors. Bottom panel: Percentage of variance captured by the specific PCs in a). southern California area; b). southern Africa area; c). northern China area. The black dash lines indicate the threshold of 95 % captured variances.

3.3 Retrieval demonstration

In this subsection, we start demonstrating CIMAP retrieval with a case study for each selected area. Then a systematic comparison of 20 case aerosol retrievals to AERONET reference data in each area will be made. Figure 5 demonstrates CIMAP retrievals of aerosol properties from AOD and sky radiance measurements acquired at the AERONET UCLA site with 118.45°W longitude and 34.07°N latitude. The UTC of data acquisition is 00:15:16 on Oct. 10 of 2006 (namely Case 6 in Table 1). In the upper two panels, CIMAP retrievals of real and imaginary parts of refractive index at the four AERONET wavelengths (440, 670, 870, and 1020nm) are compared to the AERONET reference products. In the bottom panel, the volumetric aerosol size distribution ($dV(r)/lnr$) is compared. AERONET and CIMAP retrieval results are plotted in black and red, respectively. The difference in real and imaginary parts of refractive index are found less than

0.02 and 0.015, respectively. While retrievals show a bi-modal distribution of aerosol volumetric size distributions, the locations of the mode radius of fine aerosols from the two retrievals differ slightly (by about $0.05\mu\text{m}$). For the coarse model aerosols, both retrievals indicate the maximum aerosol concentration to be around $3\mu\text{m}$.

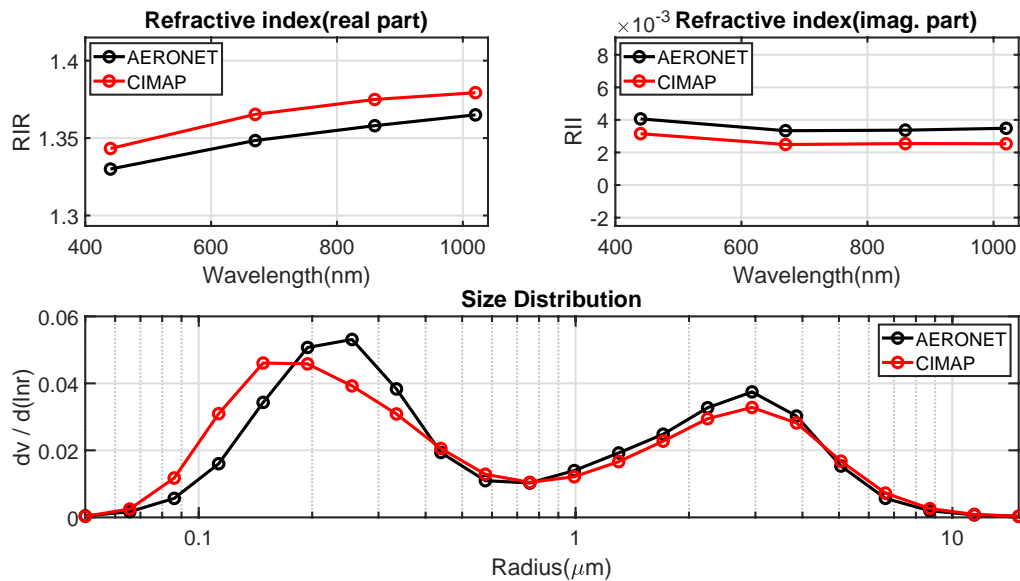


Figure. 5. The spectral aerosol refractive index and size distribution retrieved from AOD and sky radiance measurements (almucantar mode) acquired at the AERONET UCLA site. The UTC time of acquisition is 00:15:16 of October 10 of 2006 (namely Case 6 in Table 1). CIMAP results are plotted in red curve with empty circles. AERONET operational retrievals are plotted in black curve with empty circles. Upper left panel: real part of refractive index (RIR) at the four AERONET wavelengths 440, 670, 870, and 1020nm; Upper right panel: imaginary part of refractive index (RII); Bottom panel: volumetric aerosol size distribution ($dV(r)/d\ln r$).

The CIMAP retrievals of spectral AODs and SSAs are compared to AERONET reference data in Fig. 6. The left panel shows that CIMAP retrievals of spectral AODs match well with the AERONET measurement. The absolute differences are indeed within the measurement uncertainty of 0.015. The right panel shows a reasonable agreement of the SSA retrieval: the difference is within AERONET reported uncertainty 0.03 for SSA of non-absorbing aerosols when $\tau_{\text{ext},440\text{nm}}^{\text{aer}} > 0.2$ and 0.04 for absorbing aerosols when $\tau_{\text{ext},440\text{nm}}^{\text{aer}} \geq 0.5$ [8][9]. Using the retrieved aerosol abundance and properties, we calculate the modeled sky radiances and compare

them to the almucantar measurements in Fig. 7. Reasonable agreements are found across all wavelengths. The relative differences (not plotted here) are within measurement uncertainty (5%).

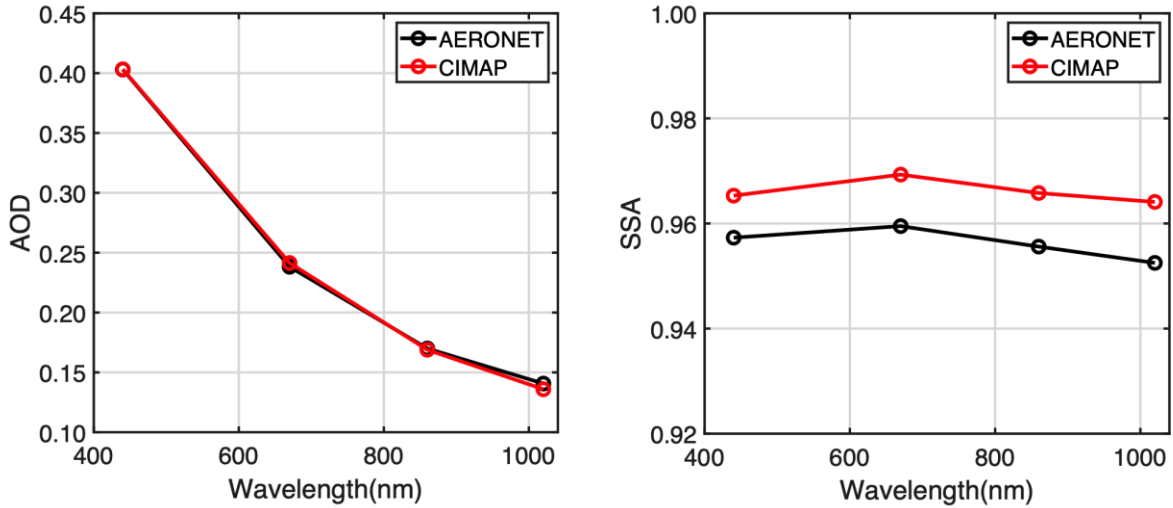


Figure 6. Left panel: comparison of CIMAP retrievals of AODs (red curve with empty circles) at the four AERONET wavelengths 440, 670, 870, and 1020 nm to those measured by AERONET (black curve with empty circles); Right panel: comparison of CIMAP retrievals of SSA to those from AERONET operational retrieval.

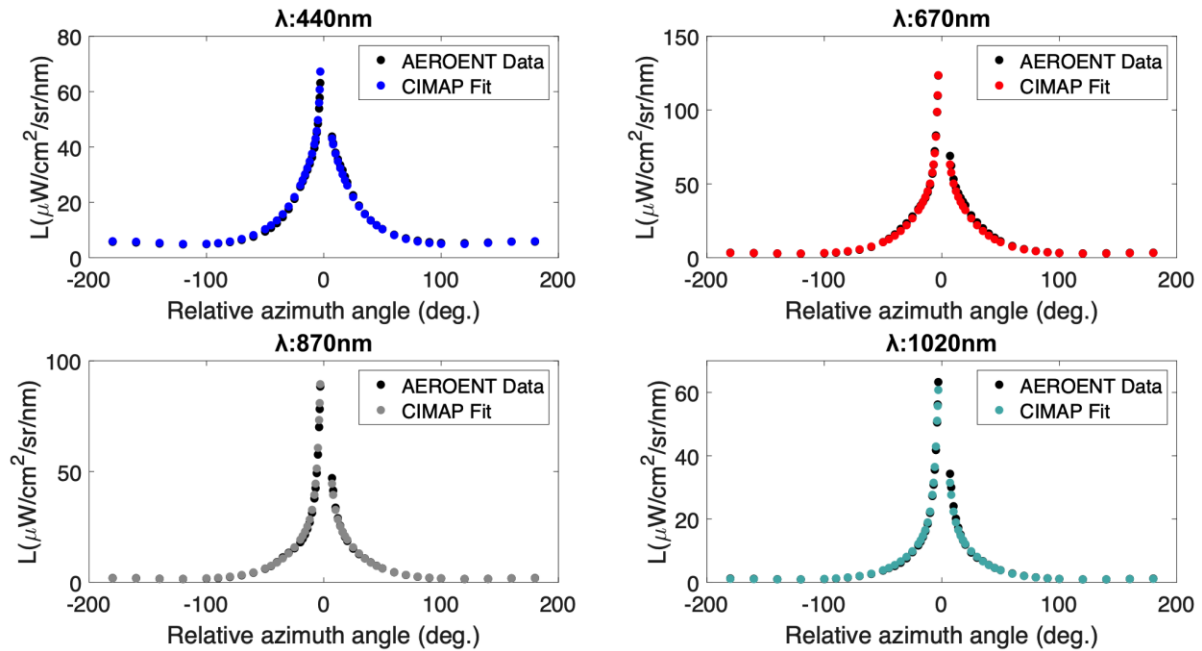
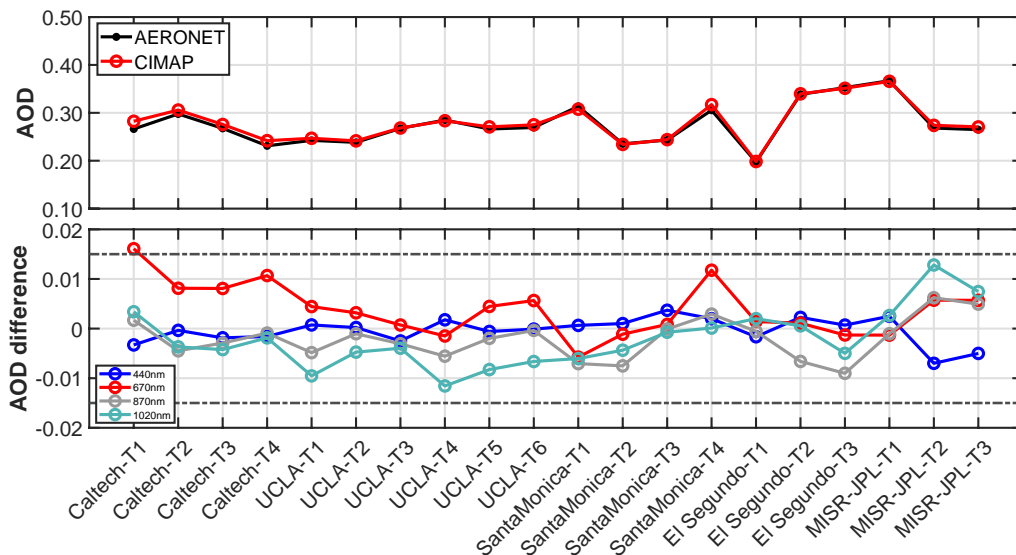
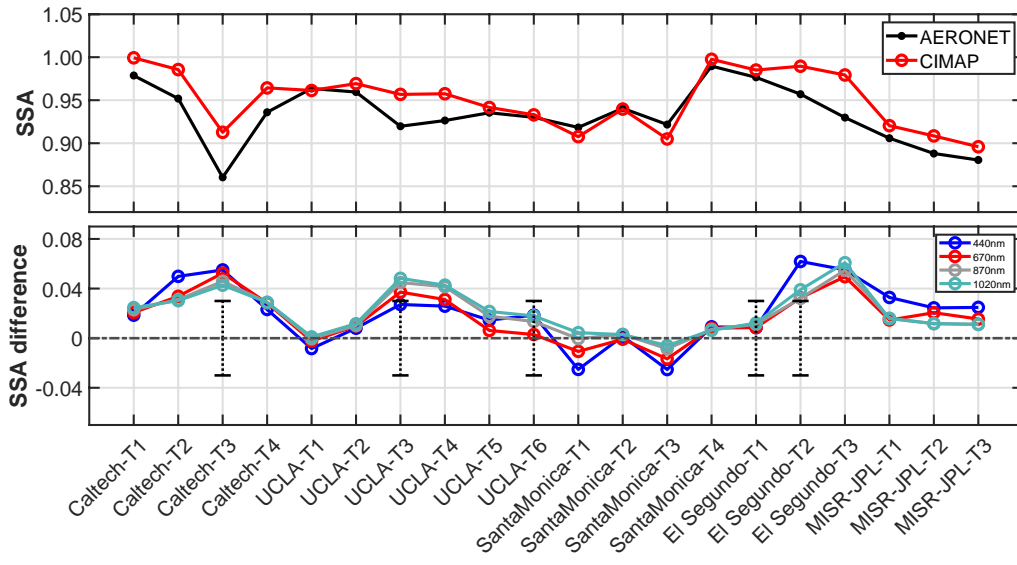


Figure 7. Comparison of modeled sky radiances to the measurements by the almucantar. The modeled radiances are calculated from the retrieved solution. The comparison of radiances in upper left, upper right, bottom left and bottom right panels are for AERONET wavelength 440, 670, 870, and 1020 nm, respectively.

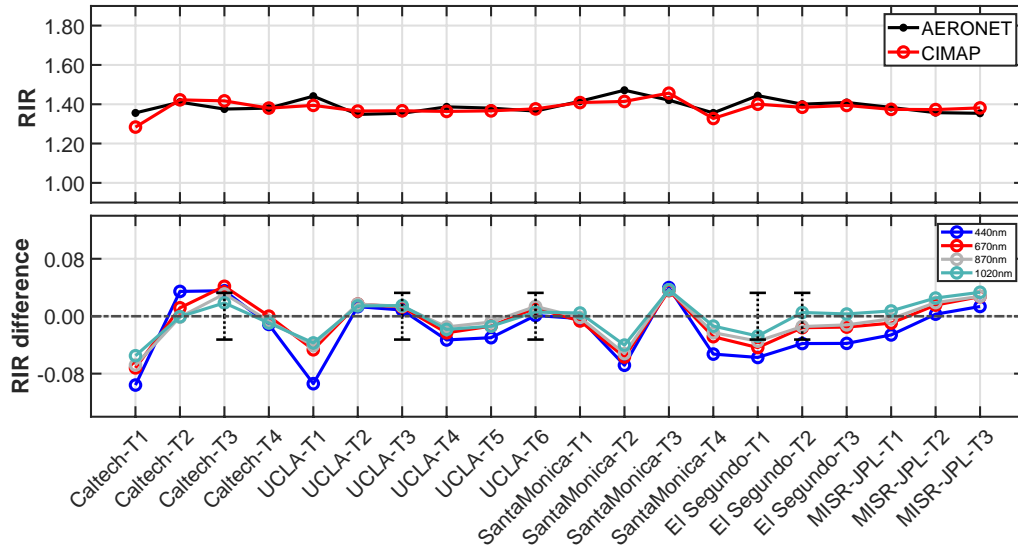
Following the above case study, we demonstrate 20 randomly selected aerosol retrievals from AERONET observations in southern California and compare the results to AERONET reference data. The upper panel of Fig. 8a illustrates the measured and retrieved AOD at 670 nm. The lower panel demonstrates the difference of AOD at the four AERONET wavelengths, which are mostly within the ± 0.015 measurement uncertainty indicated by the two dashed lines. Figure 8b presents the comparison of retrievals of SSA. AERONET retrieval uncertainties of SSA (0.0325) were plotted in bars. They are averaged from the uncertainties estimated for weakly absorbing aerosols (0.025) and absorbing aerosols (0.04) in [8] when AOD at 440nm is larger than 0.5. The differences in CIMAP and AERONET retrievals of SSA are within or close to the estimated AERONET SSA retrieval uncertainties for most retrievals. Comparisons of the real and imaginary parts of aerosol refractive index are demonstrated in Fig. 8c-d, respectively. Comparison to estimated AERONET retrieval uncertainties plotted in bars indicates CIMAP's reasonably high qualities of refractive index retrieval. Though not demonstrated here, CIMAP itself is also subjected to retrieval uncertainties. This also explains part of the difference observed in CIMAP and AERONET retrievals as well.



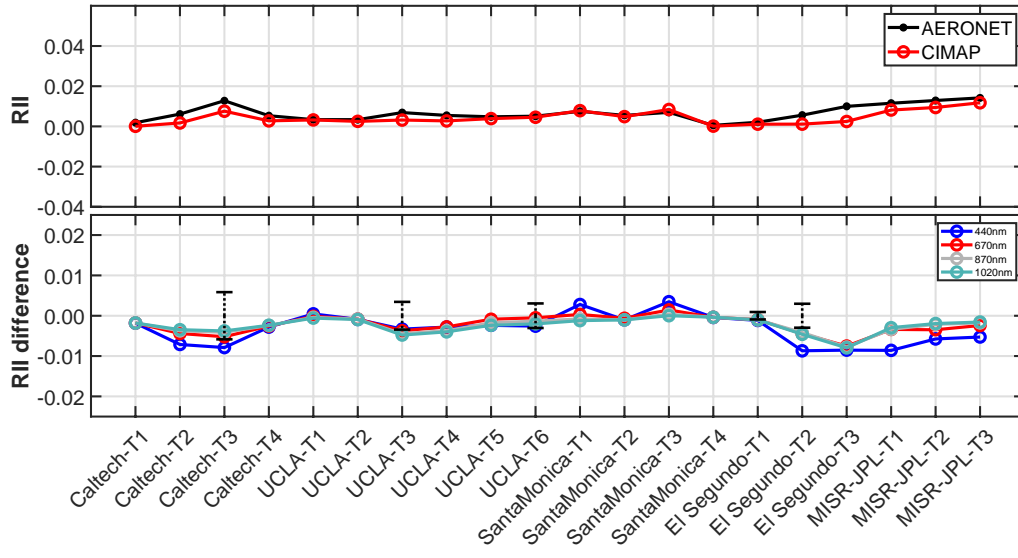
(a)



(b)



(c)



(d)

Figure 8. a) Comparison of CIMAP retrievals of AODs with AERONET measurements (black) from 20 retrieval cases; b) Comparison of CIMAP retrievals of SSA with AERONET reference data; c) Comparison of refractive index (real part, RIR); d) Comparison of refractive index (imaginary part, RII). The labels on the x-axis contain the names of AERONET sites located in southern California and the series number of measurements. In the subfigures, the upper panels demonstrate a direct comparison of AOD, SSA, RIR, and RII values at 670 nm, while lower panels demonstrate the difference for all AERONET wavelengths. The difference is calculated as $X_{\text{CIMAP}} - X_{\text{AERONET}}$, where $X = \text{AOD, SSA, RIR, or RII}$. The AERONET measurement or retrieval uncertainties are denoted by the dashed lines or bars in the lower panels.

To assess the retrieved spectral dependence of AOD on wavelength, we calculate the Ångström exponent (AE) from the retrieved AODs at 440 and 870nm by $AE = -\log(\tau_{\text{ext},440\text{nm}}^{\text{aer}}/\tau_{\text{ext},870\text{nm}}^{\text{aer}})/\log(440/870)$ and compare the results to the AERONET reference data. Figure 9 demonstrates the AE for all cases. Most AEs vary between 1 to 2. Indeed, aerosols can be classified according to their AE values [30]. For the AERONET sites located near the ocean, such as site El Segundo, smaller AEs are found, which is mainly contributed by coarse mode sea-salt particles. For the urban areas, such as AERONET's Caltech site, relative larger AEs are retrieved due to the increased contribution of fine mode aerosols. For all cases, consistencies are found between AEs from CIMAP and AERONET operational retrievals. The MAD is 0.03.

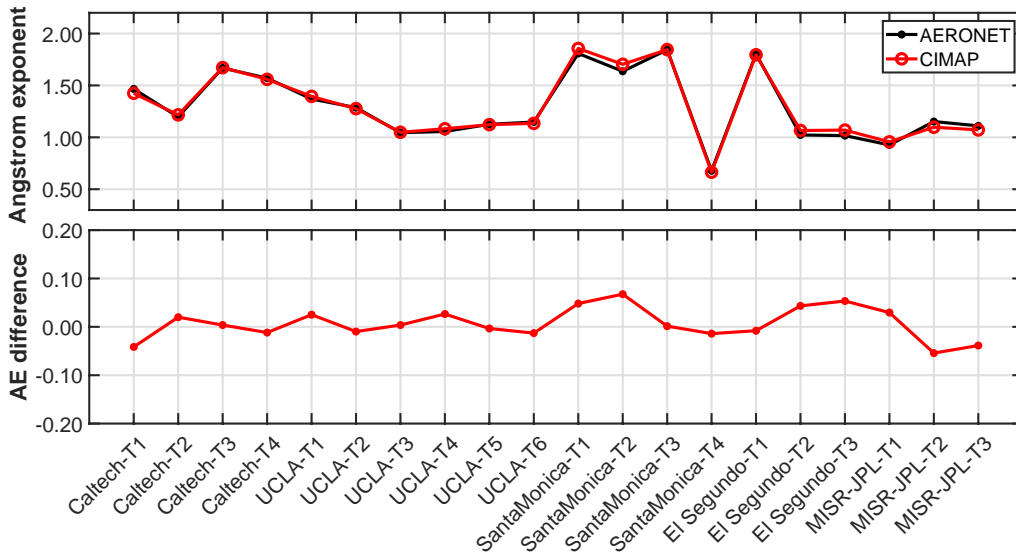


Figure 9. Comparison of Ångström exponents derived from CIMAP and AERONET retrievals of AODs at 440nm and 870nm.

To measure the quality of the retrieved aerosol size, we further calculate the effective radii of fine mode aerosols ($r_{\text{eff},\text{fine}}$), coarse mode aerosols ($r_{\text{eff},\text{coarse}}$), and the whole size distribution ($r_{\text{eff},\text{total}}$) and compare them to AERONET reference data. The equation for effective radius calculation is:

$$r_{\text{eff}} = \frac{\int_{r_{\min}}^{r_{\max}} \frac{dV(r)}{d \ln r} d \ln r}{\int_{r_{\min}}^{r_{\max}} \frac{1}{r} \frac{dV(r)}{d \ln r} d \ln r}, \quad (22)$$

where the quadrature bounds, r_{\min} and r_{\max} , depend on the particle size mode. Fine and coarse mode aerosols are distinguished by the relative minimum of size distribution frequency within the range of 0.439 to 0.992 μm of size interval [31]. Figure 10 demonstrates the effective radii from AERONET and CIMAP retrievals. For effective radii of fine ($r_{\text{eff},\text{fine}}$) and coarse mode ($r_{\text{eff},\text{coarse}}$) aerosols, the comparison demonstrates a high quality of agreement. The differences are generally less than 0.05 and 0.5 μm for fine and coarse mode aerosols, respectively. While the mode-specific effective radius from CIMAP and operational retrievals are quite consistent with each other, the

difference in the effective radius of the whole size distribution ($r_{\text{eff,total}}$) is remarkable for the 4th case of Santa Monica (labeled as “Santa Monica-T4” on the x-axis of the plot). Though the size distribution of fine and coarse mode aerosols from the two retrievals are similar to each other, CIMAP’s retrieval of fine mode aerosol fraction, in this case, is lower than that of AERONET operational retrieval. This causes the difference in the effective radius of whole size distribution, while the differences in the effective radii of fine and coarse mode aerosols alone are not obvious.

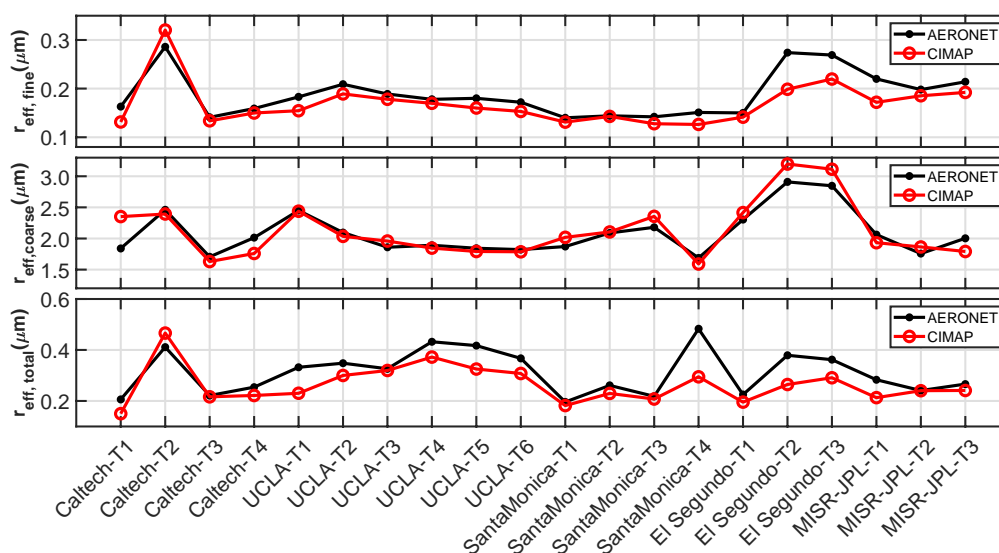
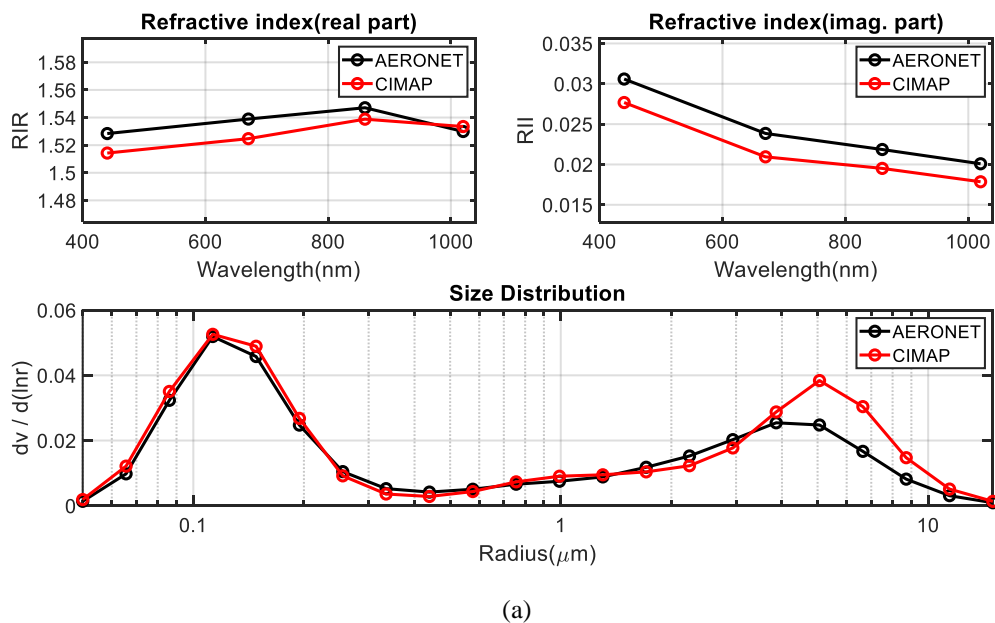


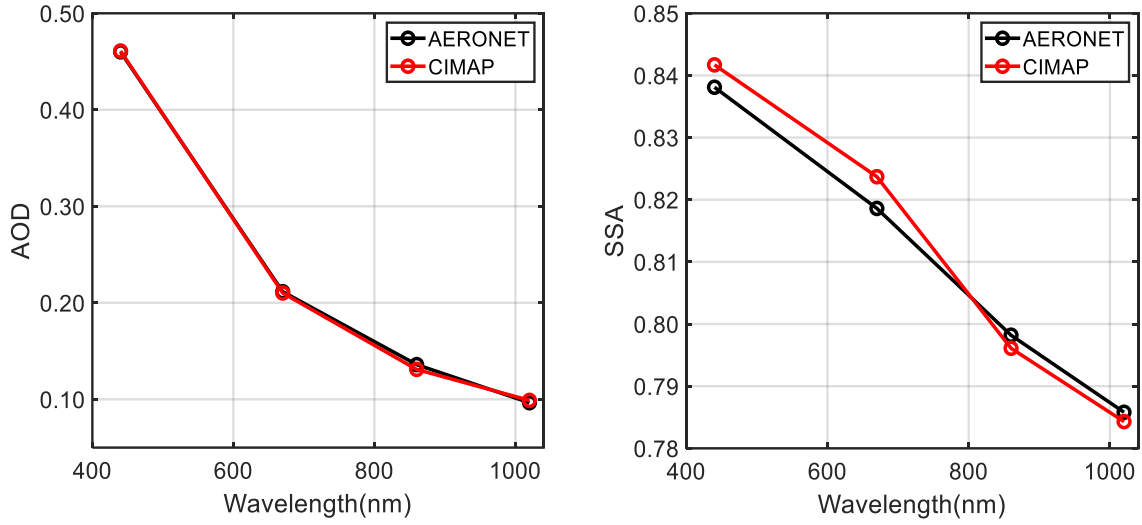
Figure. 10. Effective radii of fine mode and coarse mode aerosols and the whole size distribution as retrieved from CIMAP algorithm (red) and from AERONET operational algorithm (black).

Following the above demonstration of CIMAP retrievals of anthropogenic aerosols in southern California, we then implement CIMAP to retrieve 20 sets of AERONET measurements of AODs and radiances at six sites located in a southern Africa area (see Table 2) where carbon aerosols prevail and another 20 sets of measurements at six sites in an area of north China around Beijing (see Table 3) where particulate air pollutants prevail. Same as southern California area, case studies of southern Africa area and north China area will be displayed at first. Figure 11 demonstrates the CIMAP retrieval of RIR, RII, particle size distribution, AOD, SSA, and

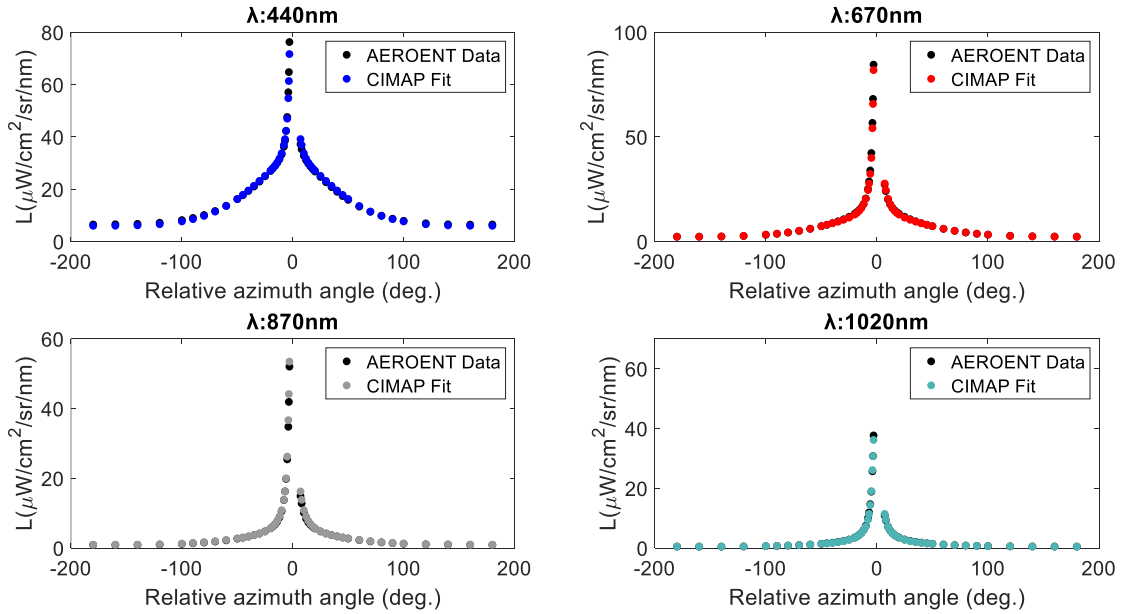
almucantar sky radiance acquired at AERONET Namibe site at 12.18°E longitude and 15.16°S latitude on June. 26 2016 (namely Case 2 in Table 2). The UTC for case studies of the Namibe site is 08:39:51.

In Fig.11a, the upper two plots show the CIMAP retrieval results of complex refractive index, where differences of RIR and RII are compared less than 0.013 and 0.0029, respectively. The lower plot shows the volumetric particle size distribution. While the location of coarse mode particle size distribution is around 5.06 μm , a difference of 0.013 of maximum aerosol concentration is found between AERONET measurement and CIMAP retrieval. The maximum aerosol concentration of fine mode particle size distribution is found around 0.011 μm for both AERONET measurement and CIMAP retrieval. In Fig. 11b, CIMAP retrieval of AODs and SSAs are performed, where a high level of agreement of AODs is observed and differences of SSAs are less than 0.05. For the case study of Namibe site, modeled sky radiance has reasonable fitting agreement compared to AERONET almucantars measurement, which is demonstrated in Fig. 11c.





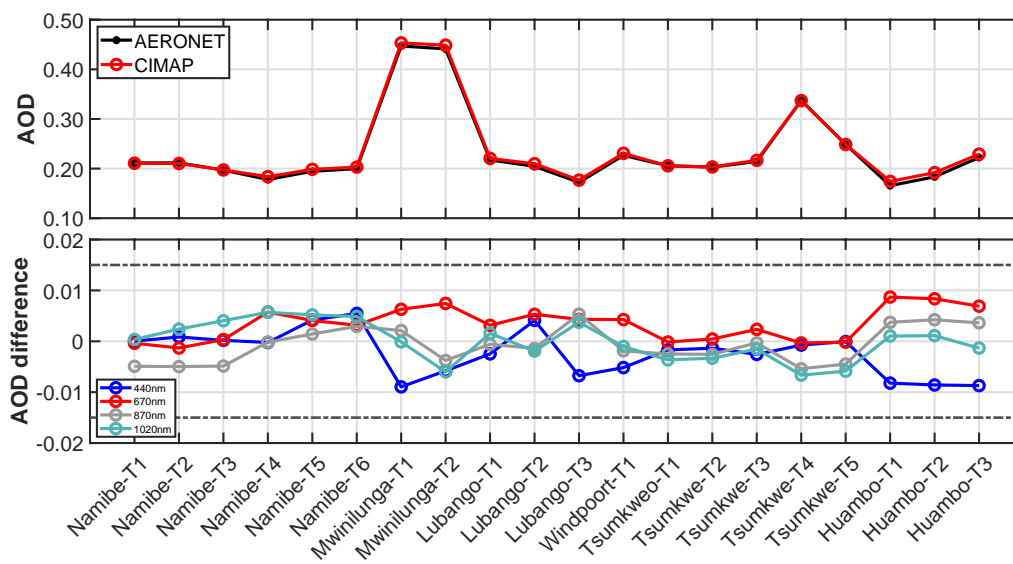
(b)



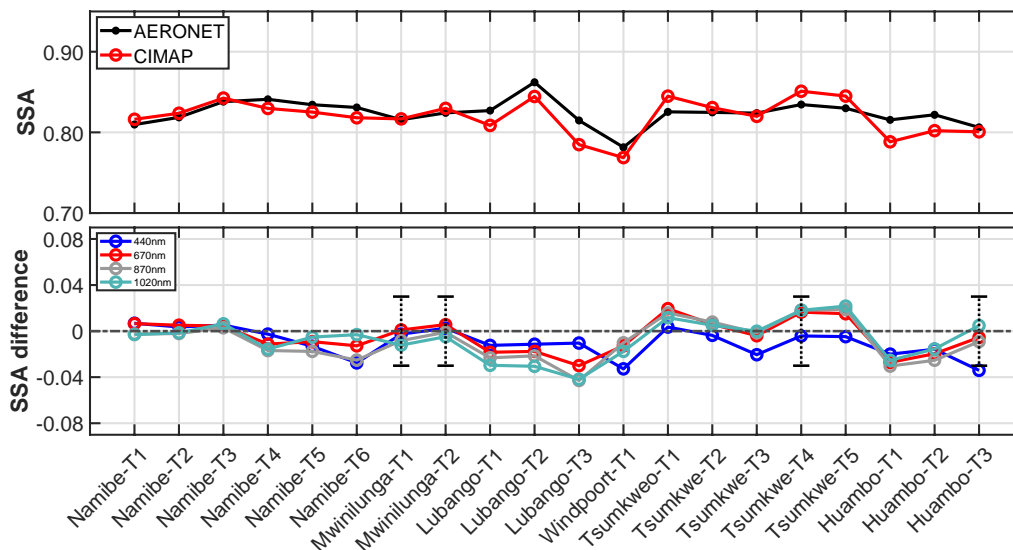
(c)

Figure. 11. Same as Fig. 5-7, but the case study is selected from AERONET Namibe site at 12.18°E longitude and 15.16°S latitude on 08:39:51, June. 26 2016 (namely Case 2 in Table 2). a). CIMAP retrieval results of spectral aerosol refractive index and size distribution acquired at the AERONET Namibe site. CIMAP results are plotted in red curve with empty circles. Upper left panel: real part of refractive index (RIR) at the four AERONET wavelengths 440, 670, 870, and 1020nm; Upper right panel: imaginary part of refractive index (RII); Bottom panel: volumetric aerosol size distribution ($dV(r)/d\ln r$). b). Comparisons of CIMAP retrievals of AODs (left panel) and SSAs (right panel) at the four AERONET wavelengths 440, 670, 870, and 1020 nm to those measured by AERONET (black curve with empty circles). c). Comparison of modeled sky radiances to the measurements by the almucantar in the upper left, upper right, bottom left, and bottom right panels are for AERONET wavelength 440, 670, 870, and 1020 nm, respectively.

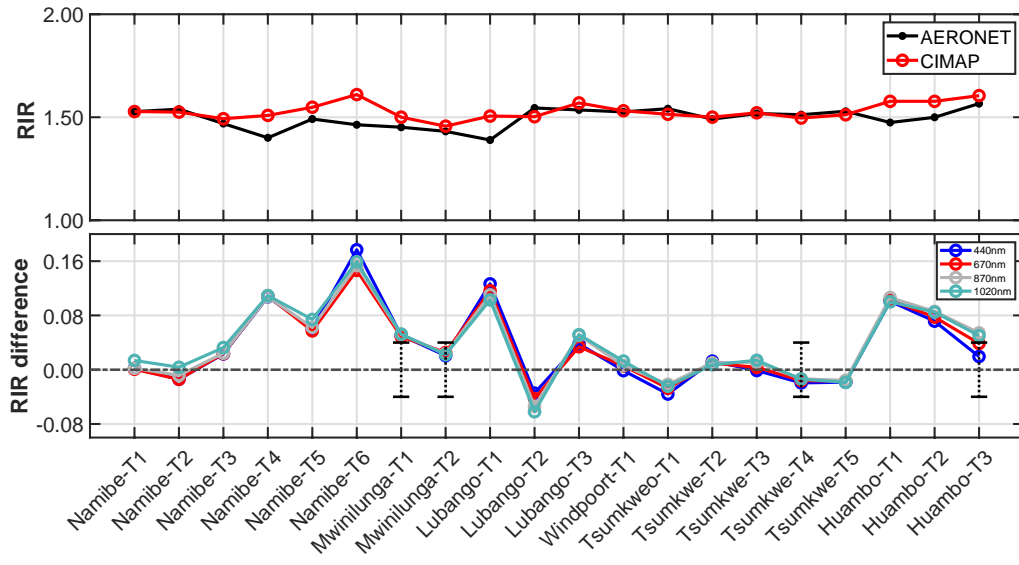
Same as Fig. 8-10, figure 12a-12f perform CIMAP retrievals of AODs, SSAs, RIRs, RIIs, AEs, bi-mode and total effective radius, respectively. AERONET measurements and retrievals are set to be references. An uncertainty of ± 0.015 is labeled for AODs. For cases with AODs at 440nm larger than 0.5, uncertainties of ± 0.03 , ± 0.04 , $\pm 30\%$ are set for SSA, RIR, and RII in Fig. 12 b-d, respectively. Since CIMAP retrievals of most cases are within measurement uncertainties, a high level of agreement is received from the comparisons of southern Africa area.



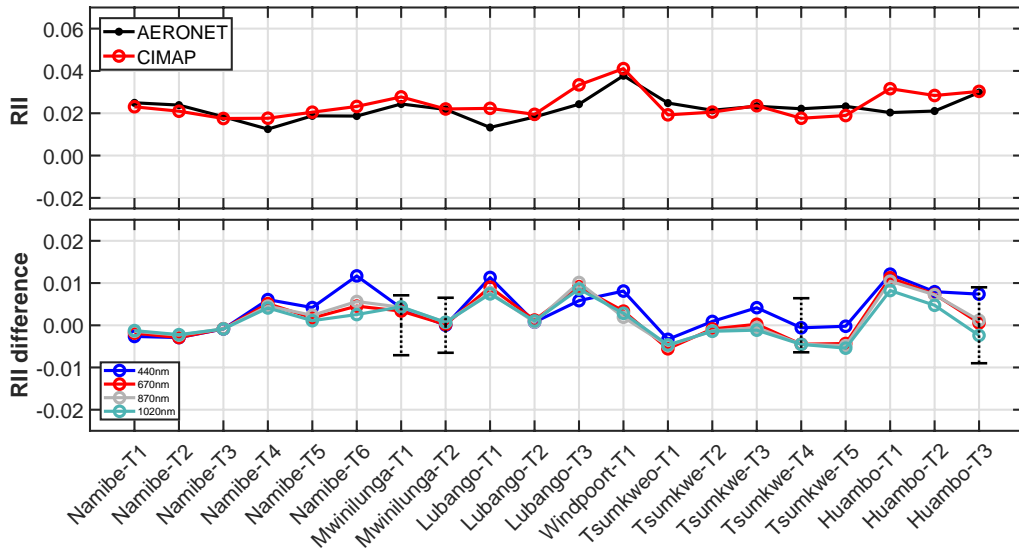
(a)



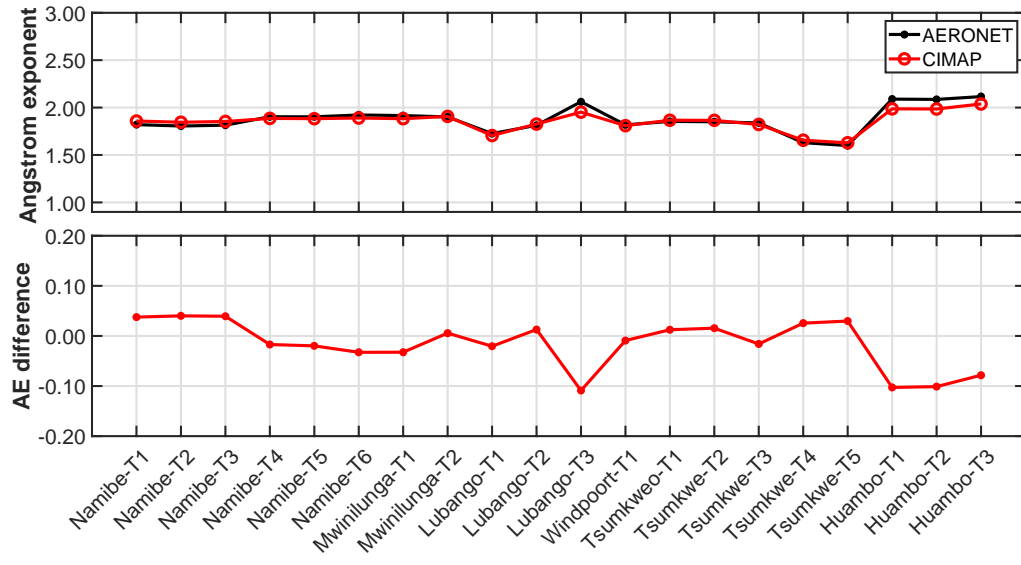
(b)



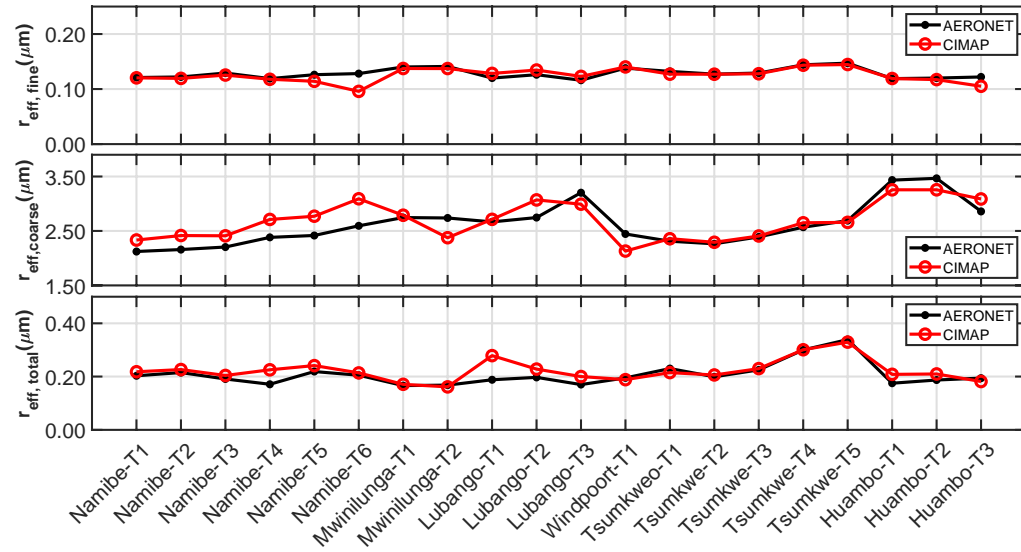
(c)



(d)



(e)

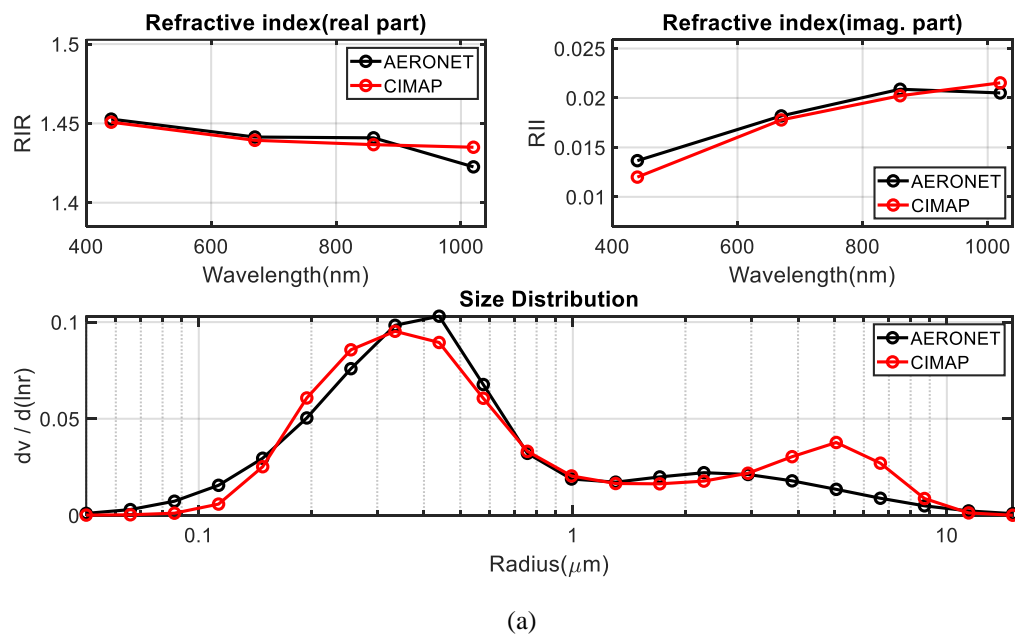


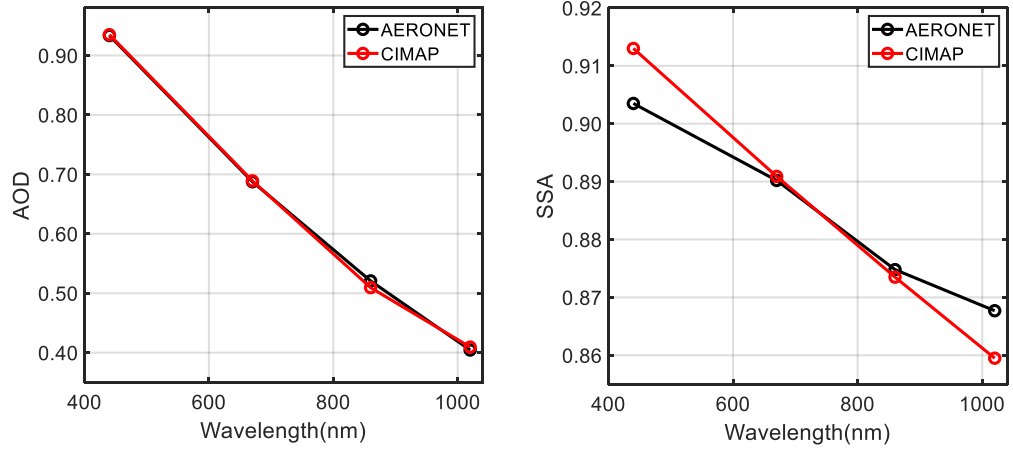
(f)

Figure 12. Same as Fig. 8-10, but the retrievals are performed for 20 sets of AERONET measurements in Southern Africa. The AERONET retrieval uncertainty (bars) for SSA and refractive index refer to the values reported for biomass burning aerosols when AOD at 440nm is larger than 0.5 [8].

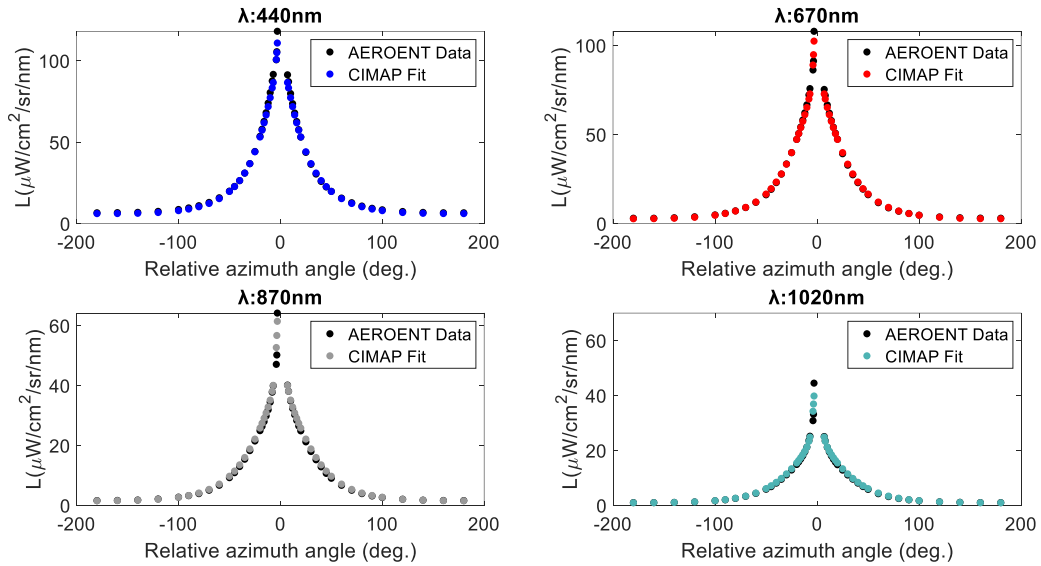
The following figures perform the case study and overall demonstration of 20 selected cases in Beijing area. AERONET climatology from Yulin site is retrieval for case study (namely Case 16 in Table 3). The detailed date of this case is 00:42:25, July. 6 2002. Comparison of CIMAP retrieval with AERONET references are plotted in Fig. 13. In Fig. 13a, comparisons of complex

refractive index and volumetric particle size distribution are made. Differences of RIR and RII are found less than 0.0123 and 0.0017 in the upper two plots, respectively. Particle size distribution are compared in the bottom panel, where the maximum particle concentrations differ about $0.1\mu\text{m}$ and $3\mu\text{m}$ for fine and coarse mode size distribution, respectively. In Fig. 13b, retrievals of AODs and SSAs are plotted in the left and the right panel. The maximum differences of AODs and SSAs are about 0.02 and 0.01, respectively. Modeled almucantar radiances generated by retrieved aerosol properties are shown in Fig. 13c. Differences in almucantar radiances at four AERONET wavelengths of 440, 670, 870, and 1020nm are not obvious, which is within 5% of measurement uncertainties.





(b)

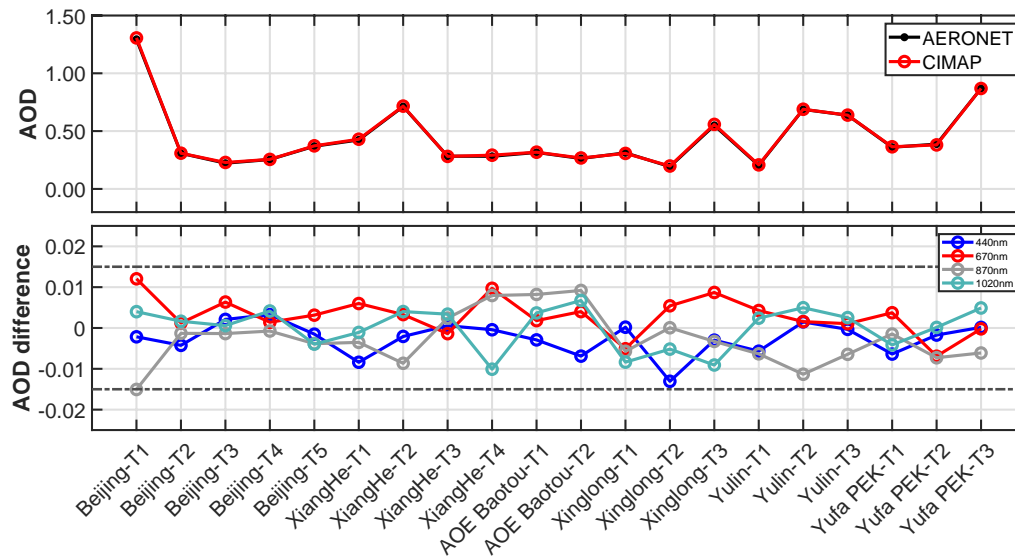


(c)

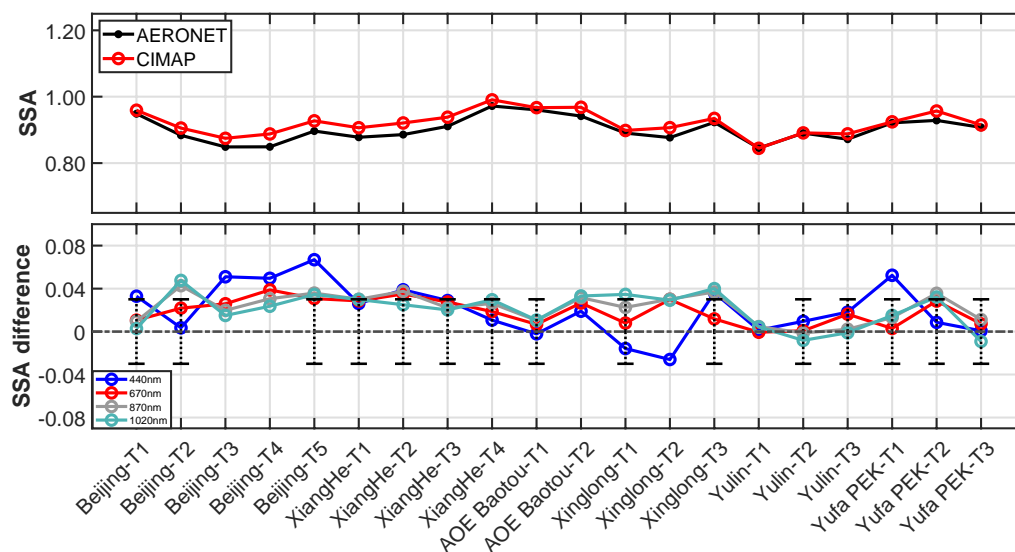
Figure 13. Same as Fig. 5-7 and Fig. 11, but the case study is selected from AERONET Yulin site at 109.72°E longitude and 38.28°N latitude on 00:42:25, July. 6. 2002 (namely Case 16 in Table 3). a). CIMAP retrieval results of spectral aerosol refractive index and size distribution acquired at the AERONET Yulin site. CIMAP results are plotted in red curve with empty circles. Upper left panel: real part of refractive index (RIR) at the four AERONET wavelengths 440, 670, 870, and 1020nm; Upper right panel: imaginary part of refractive index (RII); Bottom panel: volumetric aerosol size distribution ($dV(r)/d\ln r$). b). Comparisons of CIMAP retrievals of AODs (left panel) and SSAs (right panel) at the four AERONET wavelengths 440, 670, 870, and 1020 nm to those measured by AERONET (black curve with empty circles). c). Comparison of modeled sky radiances to the measurements by the almucantar in the upper left, upper right, bottom left, and bottom right panels are for AERONET wavelength 440, 670, 870, and 1020 nm, respectively.

Figure 14 displays CIMAP retrievals of the overall Beijing area. Same as Fig. 8-10 and Fig. 12, AODs, SSAs, RIRs, RIIs, AEs, bi-mode and total effective radius are demonstrated in Fig. 13

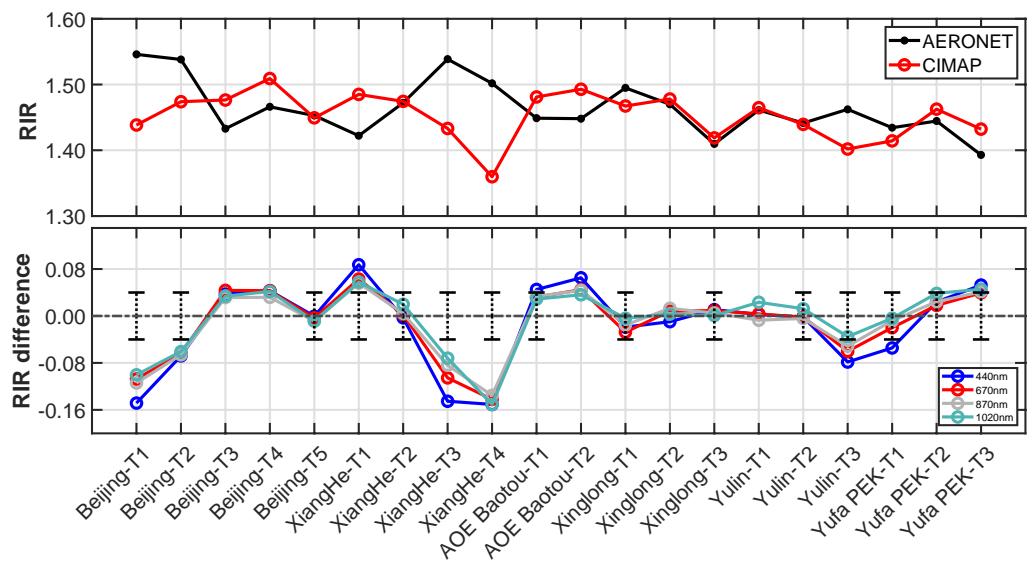
a-f, respectively. AERONET references are plotted with every parameter for comparison. The uncertainty set for AODs is ± 0.015 in Fig. 14a. Same as uncertainties labeled in the comparison of southern Africa area, in Fig. 14 b-d, ± 0.03 , ± 0.04 , and $\pm 50\%$ are set for SSA, RIR, and RII, respectively. The overall retrieval results in this area are also performed in high quality.



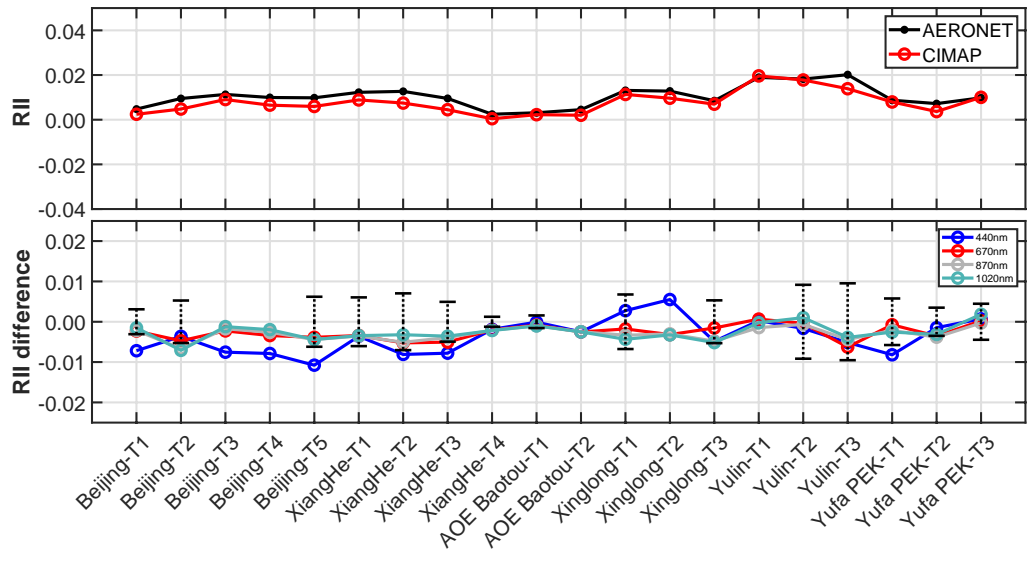
(a)



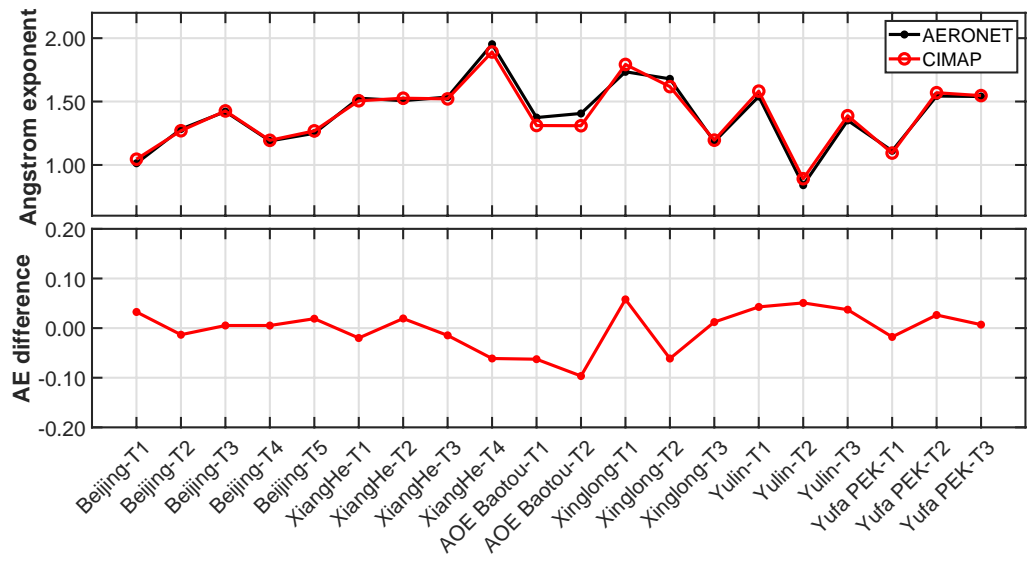
(b)



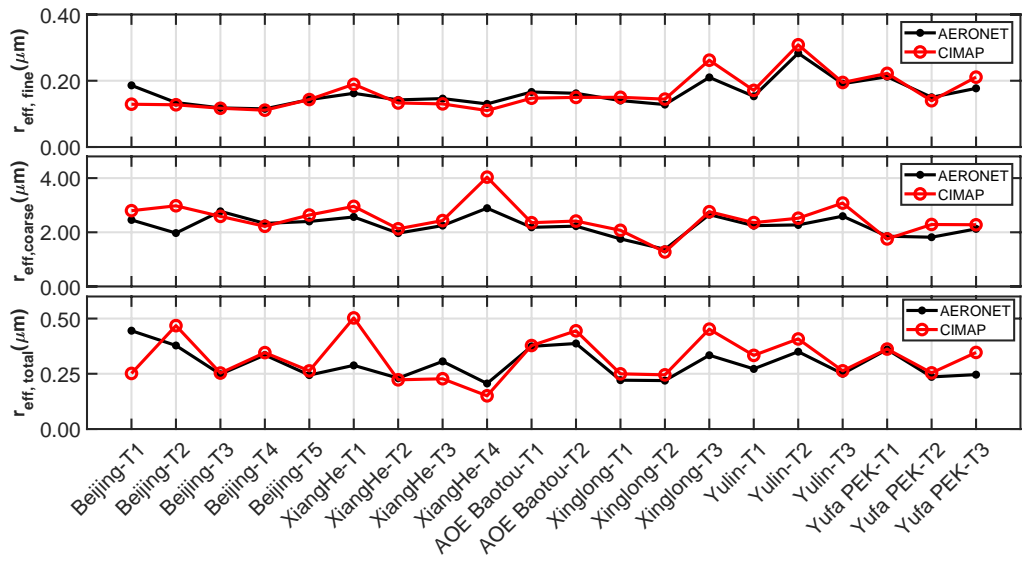
(c)



(d)



(e)



(f)

Figure. 14. Same as Fig. 8-10 and Fig. 12, but the retrievals are performed for 20 sets of AERONET measurements in north China. The AERONET retrieval uncertainty (bars) for SSA and refractive index refer to the values reported for absorbing aerosols (dust) when AOD at 440nm is larger than 0.5 [8].

Overall, the comparison of CIMAP and AERONET retrievals for all three areas indicates similar retrieval accuracies of the two algorithms. Table 4 summarizes the MADs of CIMAP retrievals and AERONET reference data of AOD, SSA, Ångström exponent, real and imaginary

parts of refractive index, and effective radii of whole size distribution. The MADs are found to be 0.004, 0.019, 0.032, 0.039, 0.003 and 0.044 μm for these quantities, respectively. These differences are within or close to AERONET measurement or retrieval uncertainties.

Table 4. Mean absolute differences of CIMAP and AERONET retrieval aerosol quantities.

Aerosol quantities	Southern California	Southern Africa	North China	All-case & all-site mean
AOD (440 nm)	0.002	0.004	0.003	
AOD (670 nm)	0.005	0.004	0.004	0.004 (incl. wavelength mean)
AOD (870 nm)	0.004	0.003	0.006	
AOD (1020 nm)	0.005	0.003	0.004	
SSA (440 nm)	0.026	0.012	0.025	
SSA (670 nm)	0.020	0.012	0.019	0.019 (incl. wavelength mean)
SSA (870 nm)	0.021	0.015	0.023	
SSA (1020 nm)	0.022	0.014	0.022	
Refr. index (real part, 440nm)	0.035	0.047	0.052	
Refr. index (real part, 670nm)	0.025	0.046	0.042	0.039 (incl. wavelength mean)
Refr. index (real part, 870nm)	0.022	0.049	0.039	
Refr. index (real part, 1020nm)	0.019	0.051	0.039	
Refr. index (imag. part, 440nm)	0.004	0.005	0.005	
Refr. index (imag. part, 670nm)	0.002	0.004	0.003	0.003 (incl. wavelength mean)
Refr. index (imag. part, 870nm)	0.002	0.004	0.003	
Refr. index (imag. part, 1020nm)	0.002	0.004	0.003	
Ångström exponent	0.026	0.038	0.033	0.032
Effective radius (fine mode, μm)	0.054	0.020	0.058	0.015
Effective radius (coarse mode, μm)	0.023	0.006	0.018	0.216
Effective radius (all size, μm)	0.138	0.199	0.311	0.044

Comparing to traditional direct retrieval of 30 aerosol parameters in regular parameter space, CIMAP capitalizes on the aerosol correlation and retrieves 7-8 PC weights only. Therefore,

it improves the retrieval efficiency. To get insight into CIMAP's gain of retrieval efficiency, we implemented an extra retrieval using the maximum number (30) of PCs. Considering the mutual orthogonality of PC vectors, 30 PC based retrieval essentially means no correlation is invoked so that the retrieval is considered as a proxy of traditional retrieval in regular parameter space. We then calculate the ratio (R) of retrieval time cost (t) from $R = t_{nPC} / t_{30PC}$, where $n_{PC} = 7, 7, \text{ and } 8$ for CIMAP retrievals of measurements over southern California, southern Africa and north China, respectively. Figure 15 demonstrates the "R" values for all 60 cases. We found the mean time cost of CIMAP retrieval to be 13%, 20% and 21% of the non-correlation-based retrievals for the three areas, respectively. This means an average time saving of ~82%. However, attention should be paid to that the operational AERONET further imposes extra smoothness constraints on the variations of aerosol refractive index as a function of wavelength and on the variations of aerosol volume concentrations across 22 size bins [11]. Same smoothness constraints have been formulated into the correlation-based retrieval and demonstrated using airborne observations [13]. In the next stage, we will test the effect of these extra constraints in enhancing the accuracy and efficiency of ground based AERONET retrievals.

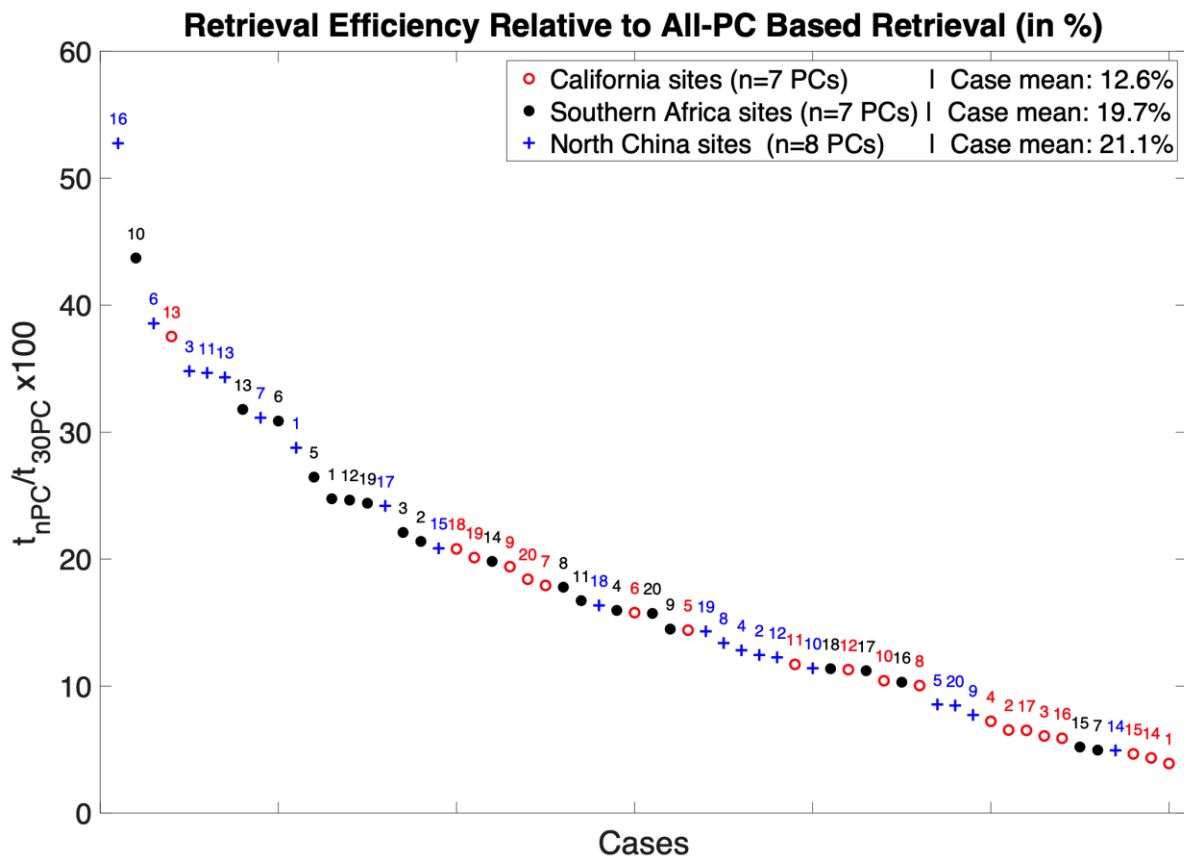


Figure 15. The ratio of retrieval time cost from using 7 or 8 PCs against using 30 PCs. Red, black and blue markers and text correspond to the selected retrievals of AERONET measurements in southern California, southern Africa, and north China, respectively. The numbers above markers are retrieval case numbers in Tables 1-3.

4. Summary and outlook

By capturing the correlation in aerosol properties, a principal-component based inversion approach – CIMAP is developed to retrieve aerosol properties, including refractive index and volumetric size distribution from AERONET measurements of AODs and sky radiances. With the adoption of 7-8 PCs, CIMAP retrievals of AERONET measurements in several selected areas in southern California, southern Africa, and north China show that it significantly reduces the retrieval parameter space and reduces the retrieval time by >80%. The mean absolute differences between CIMAP retrievals and AERONET standard retrieval products are 0.004, 0.019, 0.032, 0.039, 0.003, and 0.044 μm for AOD, SSA, Ångström exponent, real and imaginary parts of

refractive index and effective radius, respectively. The differences in CIMAP and AERONET retrievals are generally within or close to AERONET measurement or retrieval uncertainties. CIMAP's retrievals of ground-based measurements by AERONET in this work and airborne observations by AirMSPI in an earlier work [13] form a stepstone toward aerosol properties from satellite-borne measurements, which involves the inversion of a large number of imaging pixels. In this case, both retrieval accuracy and efficiency are important concerns. In addition to the retrieval efficiency gain by reducing retrieval parameter space, CIMAP also offers a higher degree of flexibility to interface with diverse sources of a priori information about aerosols (e.g., climatology, transport models, reanalysis, etc.) as long as these sources can inform the retrieval with a proper initialization of PCs. Under such a context, CIMAP is expected to mitigate the ill-posedness which arises from retrieving a large set of aerosol parameters (such as aerosol species' abundance and properties) in regular parameter space when observations alone contain insufficient information.

5. Publications

[1]. T. Huang, F. Xu, L. Gao, C. Flynn and O. Dubovik, “A correlation-based inversion method for inversion of aerosol properties (CIMAP) from AERONET measurements,” *J. Quant. Spectrosc. Radiat. Transfer*. Accepted, 2021.

[2]. F. Xu, T. Huang, D.J. Diner, L. Gao, C. Flynn, J. Redemann, and O. Dubovik, “A correlation-based inversion method for aerosol property retrieval from AirMSPI and AERONET measurements,” *The 19th Electromagnetic and Light Scattering Conference*, online, 12-16 July, 2021.

6. REFERENCE

- [1] Salomonson VV, Barnes WL, Maymon PW, Montgomery HE, Ostrow H. MODIS: advanced facility instrument for studies of the Earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*. 1989;27:145-53.
- [2] King MD, Kaufman YJ, Menzel WP, Tanre D. Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS). *IEEE Transactions on Geoscience and Remote Sensing*. 1992;30:2-27.
- [3] Diner DJ, Beckert JC, Reilly TH, Bruegge CJ, Conel JE, Kahn RA, et al. Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*. 1998;36:1072-1087.
- [4] Deschamps P, Breon F, Leroy M, Podaire A, Bricaud A, Buriez J, et al. The POLDER mission: instrument characteristics and scientific objectives. *IEEE Transactions on Geoscience and Remote Sensing*. 1994;32:598-615.
- [5] Holben B, Eck TF, Slutsker I, Tanré D, Buis JP, Setzer A, et al. AERONET-a federated instrument network and data archive for aerosol Characterization. *Remote Sensing of Environment*. 1998;66:1-16.
- [6] Vermote E, Vermeulen A. Atmospheric correction algorithm: spectral reflectances (MOD09). ATBD version. 1999;4:1-107.
- [7] Holben B, Tanre D, Smirnov A, Eck T, Slutsker I, Abuhassan N, et al. An emerging ground-based aerosol climatology: Aerosol Optical Depth from AERONET. *Journal of Geophysical Research: Atmospheres*. 2001;106.
- [8] Dubovik O, Smirnov A, Holben BN, King MD, Kaufman YJ, Eck TF, et al. Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network

- (AERONET) Sun and sky radiance measurements. *Journal of Geophysical Research: Atmospheres*. 2000;105:9791-806.
- [9] Holben B, Eck T, Slutsker I, Smirnov A, Sinyuk A, Schafer J, et al. Aeronet's Version 2.0 quality assurance criteria: SPIE; 2006.
- [10] Eck TF, Holben BN, Reid JS, Dubovik O, Smirnov A, O'Neill NT, et al. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *Journal of Geophysical Research: Atmospheres*. 1999;104:31333-49.
- [11] Dubovik O, King MD. A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. *Journal of Geophysical Research: Atmospheres*. 2000;105:20673-96.
- [12] O'Neill N, Royer A. Extraction of bimodal aerosol-size distribution radii from spectral and angular slope (Angstrom) coefficients. *Applied Optics* 1993;32:1642-5.
- [13] Xu F, Diner DJ, Dubovik O, Schechner Y. A Correlated Multi-Pixel Inversion Approach for Aerosol Remote Sensing. *Remote Sensing*. 2019;11:746.
- [14] Kopparla P, Natraj V, Limpasuvan D, Spurr R, Crisp D, Shia R-L, et al. PCA-based radiative transfer: Improvements to aerosol scheme, vertical layering and spectral binning. *Journal of Quantitative Spectroscopy and Radiative Transfer*. 2017;198:104-11.
- [15] Liu X, Smith WL, Zhou DK, Larar A. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Applied Optics*. 2006;45:201-9.
- [16] Natraj V, Jiang X, Shia R-l, Huang X, Margolis JS, Yung YL. Application of principal component analysis to high spectral resolution radiative transfer: A case study of the band. *Journal of Quantitative Spectroscopy and Radiative Transfer*. 2005;95:539-56.
- [17] Diner DJ, Xu F, Garay MJ, Martonchik JV, Rheingans BE, Geier S, et al. The Airborne

- Multiangle SpectroPolarimetric Imager (AirMSPI): a new tool for aerosol and cloud remote sensing. *Atmospheric Measurement Techniques*. 2013;6: 2007-2025.
- [18] Dubovik O, Herman M, Holdak A, Lapyonok T, Tanré D, Deuzé JL, et al.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, *Atmospheric Measurement Techniques*. 2011;4:975–1018.
- [19] van de Hulst HC. *Light Scattering by Small Particles* (Dover, New York, 1981).
- [20] Wanner W, Li X, Strahler AH. On the derivation of kernels for kernel-driven models of bidirectional reflectance, *Journal of Geophysical Research: Atmospheres*. 1995, 100: 21077.
- [21] Ross J. *The radiation regime and architecture of plant stands*. The Hague: Dr W. Junk Publishers.; 1981.
- [22] Li X, Strahler AH. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*. 1992;30:276.
- [23] Xu, F., Dubovik, O., Zhai, P.-W., Diner, D. J., Kalashnikova, O. V., Seidel, F. C., Litvinov, P., Bovchaliuk, A., Garay, M. J., van Harten, G., and Davis, A. B.: Joint retrieval of aerosol and water-leaving radiance from multispectral, multiangular and polarimetric measurements over ocean, *Atmospheric Measurement Techniques*. 2016;9:2877-2907.
- [24] Xu F, van Harten G, Diner DJ, Kalashnikova OV, Seidel FC, Bruegge CJ, et al. Coupled retrieval of aerosol properties and land surface reflection using the Airborne Multiangle SpectroPolarimetric Imager. *Journal of Geophysical Research: Atmospheres*. 2017;122:7004-7026.
- [25] Levenberg K. A METHOD FOR THE SOLUTION OF CERTAIN NON-LINEAR

- PROBLEMS IN LEAST SQUARES. Quarterly of Applied Mathematics. 1944;2:164-8.
- [26] Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. Journal of the Society for Industrial and Applied Mathematics. 1963;11:431-41.
- [27] Fletcher R. A modified Marquardt subroutine for non-linear least squares. 1971.
- [28] Dubovik O, Lapyonok T, Litvinov P, Herman M, Fuertes D, Ducos F, et al. GRASP: a versatile algorithm for characterizing the atmosphere. SPIE Newsroom. 2014.
- [29] Dubovik O, Sinyuk A, Lapyonok T, Holben BN, Mishchenko M, Yang P, et al. Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, Journal of Geophysical Research: Atmospheres. 2006;111:D11208.
- [30] O'Neill NT, Dubovik O, Eck TF. Modified Ångström exponent for the characterization of submicrometer aerosols. Applied Optics 2001;40:2368-75.
- [31] AERONET Inversion Products, http://aeronet.gsfc.nasa.gov/new_web/Documents/Inversion_products_V2.pdf